Title: Intraclonal heterogeneity and distinct molecular mechanisms characterize the development of t(4;14) and t(11;14) myeloma.

Short title: Exome sequencing of t(4;14) and t(11;14) myeloma

Authors: Brian A. Walker[1*], Christopher P. Wardell[1*], Lorenzo Melchor[1], Sanna Hulkki[1], Nicola E. Potter[1], David C. Johnson[1], Kerry Fenwick[2], Iwanka Kozarewa[2], David Gonzalez[1], Christopher J. Lord[2], Alan Ashworth[2], Faith E. Davies[1] and Gareth J. Morgan[1]

Affiliations:

[1]Haemato-Oncology Research Unit, Division of Molecular Pathology, The Institute of Cancer Research, London, UK; [2]The Breakthrough Breast Cancer Centre, Division of Breast Cancer Research, The Institute of Cancer Research, London, UK.

* B.A.W and C.P.W. contributed equally to this work.

Corresponding Author: Professor Gareth Morgan, Haemato-Oncology Research Unit, Division of Molecular Pathology, The Institute of Cancer Research, 15 Cotswold Road, Sutton, UK, SM2 5NG. Email: gareth.morgan@icr.ac.uk Telephone: +44 2087224130 Fax: +44 2087224432

Scientific Category: Lymphoid Neoplasia

Abstract word count:	**199**
Text word count:	**5000**
Number of Figures:	**5**
Number of Tables:	**5**
Number of References:	**37**

Abstract

We have used whole exome sequencing to compare a group of presentation t(4;14) with t(11;14) cases of myeloma in order to define the mutational landscape. Each case was characterized by a median of 24.5 exonic non-synonymous SNVs and there was a consistently higher number of mutations in the t(4;14) group, but this did not reach statistical significance. We show that the transition/transversion rate in the two subgroups is similar suggesting that there was no specific mechanism leading to mutation differentiating the two groups. Only 3% of mutations were seen in both groups and recurrently mutated genes include *NRAS*, *KRAS*, *BRAF* and *DIS3* as well as *DNAH5*, a member of the axonemal dynein family. The pattern of mutation in each group was distinct with the t(4;14) group being characterized by deregulation of chromatin organization, actin filament and microfilament movement. Recurrent RAS pathway mutations identified subclonal heterogeneity at a mutational level in both groups with mutations being present as either dominant or minor subclones. The presence of subclonal diversity was confirmed at a single cell level using other tumor acquired mutations. These results are consistent with a distinct molecular pathogenesis underlying each subgroup and have important impacts on targeted treatment strategies.

Introduction

The t(4;14) is seen in 11% of presenting myeloma and defines a unique subtype associated with a poor clinical outcome.[1] In contrast to other molecular subtypes it is seen at lower frequencies in MGUS suggesting that the genes deregulated by the t(4;14) rapidly push plasma cells to a more aggressive phenotype.[2,3] While *FGFR3* was initially suggested to be the critical driver gene in t(4;14) myeloma it is lost by deletion of the der(14) chromosome in 25-30% of cases, focusing attention on the other deregulated gene *MMSET*.[4,5] *MMSET* is a member of a family of genes which have histone methyltransferase (HMTase) activity. Other members of the family are known oncogenes relevant to the pathogenesis of acute leukemia providing strong support for the central importance of this gene in myeloma.[6] Pathologically, MMSET has been shown to disrupt the distribution of di-methylation marks on histone H3 throughout the genome and are relevant to disease progression.[7-9] A specific DNA methylation pattern associated with t(4;14) samples is also thought to be pathologically relevant, though the relationship of these patterns to changes in histone methylation marks is uncertain.[10]

The t(4;14) alone is insufficient to lead directly to myeloma, being seen in MGUS and smoldering myeloma.[2,3] Therefore, it is likely there are collaborating mutations at other sites throughout the genome and that the spectrum of these will be unique to this subtype of disease. In this respect the changes in histone and DNA methylation seen in this subset may affect both the spectrum and number of mutations. It has also been suggested that MMSET functions in DNA repair.[8] Consequently, based on these arguments we hypothesized that there may be a unique mutational spectrum associated with t(4;14) myeloma.

Until recently the optimum way of identifying collaborating mutations has been to study regions of copy number abnormality (CNA) within which there are relevant mutational events.[5,11,12] Comprehensive mapping analyses show that CNA are common, the most frequent being deletions of 1p, 6q, 8p, 12p, 13q, 14q, 16q, and 17p.[5,13] Mutational analysis of genes within regions of homozygous deletion has led to the discovery of key genes involved in the pathogenesis of myeloma including *FAM46C* on 1p, *BIRC2*/*BIRC3* on 11q, *RB1* on 13q, *TRAF3* on 14q, *CYLD* on 16q and *TP53* on 17p.[5,14-19] Despite the success of this approach we have not, as yet, identified a specific pattern associated with the t(4;14) cases. More recently whole genome- and whole exome-based sequencing approaches have been developed which provide a means of global mutational screening that have given us the ability to define mutations outside of regions of CNA. In addition to identifying the range of mutations present within a tumor, this technology can provide semi-quantitative analysis of the size of the clonal population carrying the abnormality.[20] This technical advantage allows

for the characterization of the tumor substructure, whereas previous tools such as gene mapping arrays have only analyzed the predominant clone.

The first results generated from the application of next generation sequencing approaches were carried out on a mix of presentation and relapse samples and confirmed previous hypotheses suggesting that there are critical pathways that are deregulated leading to the malignant transformation of plasma cells.[14] This study did not focus on the t(4;14) subgroup where we understand in more detail the basis of its molecular etiology and as such the changes associated with this group remain unknown. We have integrated genome-wide SNP-based mapping array data with whole exome sequencing derived from a series of presenting cases of t(4;14) myeloma and compared the mutations present with a group of cases defined by the t(11;14). The t(11;14) is a good comparator group as it also has a clear pathogenesis which is defined by the deregulation of Cyclin D1. The aim of this work was to inform whether the collaborating mutational events in these different etiological settings are the same or different as well as how these events are acquired leading to tumor evolution and treatment resistance.

Materials and methods

Exome Sequencing

Whole exome sequencing (WES) was performed on 22 presentation myeloma samples entered in the MRC Myeloma IX trial consisting of 10 t(4;14) samples and 12 t(11;14) samples. All patients consented to storage and use of tissue in research studies and the study was approved by local and regional ethics committees. The MRC Myeloma IX trial is registered under ISRCTN68454111. CD138-positive bone marrow plasma cells were selected to a purity >95%, as determined by cytospin, using magnetic assisted cell sorting (Miltenyi Biotech). Tumor DNA and RNA were extracted using the AllPrep kit (Qiagen) and the DNA concentration assessed using Pico-green (Invitrogen). Non-tumor DNA was extracted from a white cell pellet from peripheral blood samples using the Flexigene kit (Qiagen).

Libraries were prepared from tumor and non-tumor DNA from the same patient and sequenced to identify acquired single nucleotide variations (SNVs) and indels in the tumor sample. 50 ng of genomic DNA was used to capture the exome using the SureSelect Human All Exon 50Mb target enrichment set (Agilent). We have previously validated this approach and shown it to have parity with approaches using larger starting amounts of DNA.[21] Briefly, 50 ng of DNA was fragmented on a Covaris E-series sonicator to a mean length of 200-400

bp, as assessed on a High Sensitivity DNA Bio-analyzer chip (Agilent). Fragmented DNA underwent end repair and A-tailing before ligation of Illumina adaptor sequences. The prep underwent subsequent PCR amplification for 8 cycles and 250 ng of amplified product was hybridized overnight to the SureSelect Human All Exon baits (Agilent). After hybridization and washing the captured DNA underwent a further 8 cycles of PCR. Purified libraries were sequenced on a GAIIx sequencer (Illumina) giving 76 bp paired-end reads. Samples were sequenced to a median depth after deduplication of 61x, with 99% >1x and 85% >20x exomic coverage. Following base calling and quality control metrics the raw fastq reads were aligned to the reference human genome.

Exome Analysis

The Genome Analysis Tool Kit was used to call single nucleotide variants (SNVs). These variant calls were recalibrated and soft filters applied to remove potential false-positives using dbSNP 132, HapMap 3.3 and the thousand genomes project as truth sets. Variants that occurred in both the normal and tumor samples were filtered out and the tumor-specific variants were annotated using Annovar.[22] Tumor acquired indels and deleted exons were also called. Full details of data analysis are in the Supplementary Data. As well as the identification of commonly affected genes, functional annotation enrichment analysis was used to identify commonly affected pathways.

Mutation Validation

Confirmation of whole exome sequenced variants was performed on whole genome amplified DNA. 50 ng of genomic DNA was amplified in a linear fashion using the Repli-g kit (Qiagen) to generate approximately 40 µg of amplified DNA. Samples included the 22 which had undergone whole exome sequencing as well as an additional 127 samples from the MRC Myeloma IX study. Individual PCR primer sequences and conditions can be found in the Supplementary data. Mutations in *NRAS*, *KRAS* and *BRAF* were confirmed on an individual sample basis.

Copy Number and Gene Expression Analysis

GeneChip Mapping 500K Array sets (Affymetrix) were performed as previously described.[5,15,23] The SNP genotypes and inferred copy number were obtained using GTYPE and dChip. Data have been deposited into GEO under accession number GSE21349.[5] U133 Plus 2.0 gene expression arrays (Affymetrix) had previously been performed using total RNA from these samples. In total 260 samples were available to determine the proportion of samples expressing that gene. Genes were placed in bins according to the proportion of samples expressing them and the number of variants counted for each bin. Genes defined

as having spiked expression in myeloma were defined according to the criteria outlined in a prior publication.[24]

Single Cell Sorting and Genotyping

Single cells from a patient were sorted on a FACSaria (Becton Dickenson, Oxford, UK) using a panel of antibodies known to discriminate abnormal from normal plasma cells (Flores-Montero et al., submitted): anti-CD45-PB (Dako, UK), anti-CD38-FITC (Caltag MedSystems, UK), anti-CD138-PO (Caltag MedSystems, UK), anti-CD56-PE (Caltag MedSystems, UK), anti-CD19-PE-Cy7 Beckman Coulter, UK), anti-IGK-APC (Dako, UK), and anti-Lambda-APC-Cy7 (Becton Dickson, UK). Abnormal plasma cells were identified as $CD38^+$, $CD138^+$, $CD45^-$, $CD56^-$, $CD19^-$ and IGK^+. 270 single cells were sorted on three 96-well plates (90 cells/plate). In parallel, lymphocytes from a healthy donor patient were also sorted to serve as wild type control for further genotyping analysis (5 cells/plate). Finally, 1 well per plate was left empty to serve as no template control for the analysis.

This novel approach for single cell multiplex real-time quantitative PCR (Q-PCR) analysis was followed according to N.E.P *et al*. (in preparation). Briefly, single cells were sorted directly into lysis buffer. Specific (DNA) target amplification (STA) was performed before Q-PCR. This multiplex STA reaction involves the simultaneous amplification of target regions of interest using custom designed Taqman assays. Genotyping assays specific for the mutations of interest were custom designed according to manufacturer's guidelines. The STA product was diluted prior to Q-PCR interrogation for mutations using the 96.96 dynamic microfluidic array and the BioMark HD (Fluidigm, UK) as recommended by the manufacturer. Genotyping analysis was performed by Genotyping SNP Software (Fluidigm, UK). Hierarchical clustering was performed using Pearson correlation and average linkage on the Rock platform.[25]

Results

Background rates of mutation in the two molecular subgroups

After accounting for the variants in the non-tumor DNA there were on average 74 tumor-specific SNVs per sample (range 54-99), including all captured sequences comprising exonic and surrounding intronic regions. There was a median of 36 exonic SNVs per sample (range 27-64) and of these 24.5 were non-synonymous (NS) (range 15-41). While there was not a marked difference in the frequency of SNVs between the t(11;14) and t(4;14) groups, the number of SNVs was consistently higher in the t(4;14) samples, **Table 1 and Supplementary Figure 3**. The composition of somatic variations across both groups shows a predominance of C•G→T•A transitions (37%) and T•A→C•G transitions (20%) , **Figure 1A**. A high proportion of C•G→T•A transitions has been previously noted,[14] and is due to spontaneous deamination of methyl-C residues to uracil, which is subsequently replaced by thymine.

While we have previously shown that t(4;14) samples have more CpG methylation than other cytogenetic groups, in this analysis we found no significant difference in the rates of transition mutations between t(4;14) and t(11;14) samples suggesting that this is not responsible for the pattern of mutations seen. There were a total of 39 acquired indels over the 22 samples (mean=1.77; range 0-4). Of these, all were exonic and 28 (71.8%) introduced a frameshift and 11 (28.2%) were in-frame with the coding sequence. All 39 indels affected unique genes, **Supplementary Table 4**.

As intragenic deletions have been reported in myeloma for key genes such as *KDM6A* (*UTX*) we investigated the prevalence of these. This analysis identified 276 regions of deletion in the tumor samples excluding physiologic immunoglobulin gene deletions. There was a median of 4 exonic deletions per sample, but one sample had an abnormally high number of deletions (n=181) due to a large homozygous deletion on chromosome 17 at the keratin cluster, which was also seen by mapping array.

There were significantly more mutations within genes which were not commonly expressed (p=0.022, Chi-squared) and there was an increasing mutation rate as the proportion of samples that did not express the gene increased (**Supplementary Figure 4)**. However, for the subset of genes expressed in >75% of myeloma samples an elevated mutation rate was seen. Genes in this subset include known oncogenes (*KRAS*, *NRAS*, *MYC*) and tumor suppressor genes (*CYLD*, *FAM46C*). We also examined genes with spiked expression in myeloma and we found that *PTPRG, FGFR3, ICAM1, LGMN, BMP2K, BCL6, MYC* and *ROBO1* were mutated.

In order to gain mechanistic insights into the relationship of copy number abnormalities and mutations we integrated the mutation data with genome mapping data. There were 102 distinct regions of copy number loss (totaling 3,678 Mb) and 91 regions of copy number gain (4,296 Mb) containing 33 and 48 NS acquired SNVs, respectively. In regions of normal copy number there were 517 mutations over 51,070 Mb of DNA. Per Mb of DNA the mutation rate is 1/111.4 Mb in regions of loss, 1/110.3 Mb in regions of normal copy number and 1/89.5 Mb in regions of gain. Taking into account the number of alleles available for analysis the results are consistent with a mutation rate in regions of loss of heterozygosity of one SNV per 111 Mb compared to 1 per 220 Mb in regions of normal copy number and 1 per 268 Mb in regions of gain, consistent with an increased mutation rate in regions of loss (**Table 2**). There were 33 non-synonymous acquired mutations in regions of loss of heterozygosity (LOH) within the 22 samples. These were found in the common regions of LOH, namely 8p, 13q, 14q, 16q, 17p, 22 and X in females. There was only one recurrently

mutated gene within a region of LOH and this was *DIS3*, on 13q. Other non-recurrently mutated genes of interest in regions of LOH include *MYC*, *CYLD* and *FLT3* (**Table 3**).

Almost all t(4;14) cases have del(13) and using acquired SNVs on chromosome 13 as a tool we have investigated the dynamics of the development of del(13) and the mutation of genes carried on it, **Figure 2 and Supplementary Data**. Mutational frequency was determined by mutant sequence reads and copy number by a combination of SNP array copy number and loss of heterozygosity. We show that in the majority of instances del(13) occurs first and is followed by the acquisition of a mutation, but when mutations occur first these are more likely to be non-synonymous. Samples in which uniparental disomy (UPD) and acquired mutations occur together can be used to validate this finding. Three samples had UPD and nonsynonymous mutations. The first is on 1q, which has LOH and is also trisomic, which has two mutations at frequencies of 14% (*FCRL5*) and 27% (*IL6R*). The second on chromosome 7 has two mutations present in 49% (*BUD31*) and 38% (*CTTNBP2*) of reads. The third is on chromosome 8 and has 43% mutated reads (*CSMD3*). These data suggest that deletion and subsequent duplication, to create UPD, occurred before mutation, which is the order determined by our analysis of chromosome 13.

DIS3 is located on chromosome 13 and is recurrently mutated in myeloma (10% reported previously and 18% in this set). We have identified 4 mutations in this gene, one of which has been previously described (R750K) suggesting a gain in function of this gene. Interestingly, to date, mutation of *DIS3* has only been seen in cases with either a t(4;14) or t(11;14). In samples with deletion of the whole chromosome mutation of *DIS3* was seen as both a subclonal minority (29%) and as a majority population (69%, 81%), **Table 4**. In contrast, in the sample with an isolated interstitial deletion of *RB1* the *DIS3* mutation was present at 92%. It seems that *DIS3* mutation always occurs in parallel with deletion of the *RB1* region (13q14) raising the possibility that these are collaborating events. We also noted that of the 4 samples with *DIS3* mutations in a previous dataset[14] two samples had del(13q) and the remaining two samples a t(4;14), which is highly associated with del(13q). Thus, it is possible that mutation and selection of *DIS3*, as a driver mutation, is dependent on deletion of 13q14.

Recurrently mutated genes and families

We identified 28 genes with recurrent mutations, non-synonymous SNVs and indels (**Figure 3** and **Supplementary Table 1**). The most frequently mutated genes were *KRAS* (n=7), *DIS3* (n=4), *NRAS* (n=4) and *DNAH5* (n=3). In total, 5 genes were recurrently mutated in the t(4;14) samples (*KRAS*, *ABCA13*, *DIS3*, *DNAH5* and *PRKD2)* and 11 genes in the t(11;14)

samples (*KRAS*, *NRAS*, *ACTG1*, *DIS3*, *EFNB2*, *F8*, *LRR1Q1*, *PCLO*, *PLD1*, *SEC31A* and *SSPO*). The genes mutated in the two cytogenetic groups were distinct and out of the 284 genes mutated in t(4;14) and 305 genes in t(11;14) samples only 18 (3.0%) genes were found mutated in both cytogenetic groups (**Figure 1B**). The majority of mutations occurred exclusively in either subgroup and we looked at the pathways deregulated in each of the subgroups to determine whether they defined different pathobiological routes to myeloma. Gene Ontology analysis of genes disrupted in the t(4;14) group suggests an enrichment for genes involved in cytoskeleton organization, microtubule movement and actin filament based processes as well as chromatin organization, **Table 5**. The t(11;14) group showed an enrichment for protein phosphorylation, phosphate metabolism and Ras signaling.

Mutated genes driving the different pattern in the t(4;14) group included chromatin modulating and cellular motility genes. Interstitial loss of exons of the histone demethylase *KDM6A* (*UTX*) have been described in myeloma and we identified a mutated case in this analysis. MMSET is an HMTase gene but in an extended analysis we did not identify an inverse relationship between MMSET and interstitial deletions of *KDM6A*. Other genes involved in DNA or histone modifications were also mutated and include *ACTL7B*, *ASXL1*, *CHAF1B*, *ESCO1*, *KDM2A*, *KDM4B*, *KDM6A*, *MLL*, *MLL2*, *MLL5*, *MYST1*, *PCGF5, SETD1B, SETD6* and *TOX*. *DNAH5*, which encodes an axonemal dynein heavy chain, was mutated in 3 out of the 22 cases and was also mutated in 2 cases in a previous report, accounting for 8% of myeloma samples sequenced to date. Several other genes encoding dynein associated proteins were also mutated including *DNAH2*, *DNAH3*, *DNAH10*, *DNAH11*, *DYNC2H1*, *TXNDC2*, *CCDC154*, *WWC1*, *TRPS1*, *RPRG* and *ODF1*.

In the t(11;14) group Ras signaling seemed prevalent including *NRAS*, *KRAS* and *BRAF* mutations which were validated by orthogonal technologies in an extended dataset confirming the prevalence at 22%, 18% and 4%, respectively (**Supplementary Figure 5**). We show that the RAS pathway mutations (*KRAS*, *NRAS* and *BRAF*) are mutually exclusive (**Figure 3**) and that deregulation of the pathway at any point is sufficient to lead to a clonal growth advantage. Interestingly, the mutations were not associated with any specific prognosis. In addition to these RAS/MAPK pathway variants other RAS-related genes were found to have acquired mutations or deletions including *ARHGEF10*, *ARHGEF6*, *ARL2BP*, *MAP3K4*, *MAPKBP1*, *RALGAPA2*, *RANGAP1*, *RASA4*, *RASGRF2,* and *SPRED2* both in our dataset and the Broad dataset. These findings reflect the importance of deregulation of this pathway in presenting myeloma linking to the phenotypic function deregulated by these events in myeloma.

Both molecular subgroups showed deregulation of plasma cell biology and cell cycle related genes. Genes involved in B cell biology which were identified as being mutated include *PRDM1*, *PAX2*, *PAX7*, *IL6R*, *IRF4*, *IL10RA*, *BCL6*, *MAX*, *MYC*, *CYLD*, *MALT1*, *BMP2K*, *CARD6*, *TNIP1*, *IRF5*, *MKI67* and *LTB.* Of these only *PRDM1, MKI67* and *LTB* carried recurrent mutations. Several cell cycle genes have mutations including *ATM*, *ATR*, *CCAR1*, *CDC14B*, *PTPLAD1*, *PTPN21*, *PTPRU*, *TP53*, *TACC2*, *FAT1* and *FAT4*.

Intraclonal heterogeneity at the level of mutation

Deregulation of the RAS/MAPK pathway is clearly a central feature of both groups. Importantly, despite their driver status, clones with a RAS pathway mutation were not always present in the dominant clone. The subclonal populations can be visualized by calculating the percentage of mutant reads for all acquired mutations within sequence of depth >30x in a sample and adjusting for copy number and cell purity to generate a frequency of mutated cells for each mutation. These data can be used to generate kernel density plots, with peaks showing the dominant clone and any subclones that may be present. In doing this we see that in all samples with a RAS pathway mutation there is at least one subclonal population, and frequently more than one (**Figure 4**).

In the case of *NRAS*, in the 4 samples with mutations, mutated reads were seen at frequencies varying from 16-48%, equating to 32-96% of the tumor population (**Table 4**). Examples of the two extremes are provided by samples 983 and 1890. Sample 983 has a mutation in *NRAS* (c.C181A) which is present in 47% of the sequencing reads equating to 94% of the clonal cells. In contrast, sample 1890 has a mutation (c.A182G) in 17% of the sequencing reads, consistent with 34% of the clone. Similarly in the case of *KRAS*, we found mutated reads in 7 samples at a frequency between 10-36% of all reads or 20-72% of the tumor cell population (**Table 4, Figure 4**). Such intraclonal heterogeneity was also seen in *KRAS* where 2 samples had 2 acquired mutations each. In the first (sample 3) both SNVs were present in minor clones (10% and 24% of reads) yet both were activating *KRAS* mutations (c.G34C p.Q61R, and c.A182G p.G12R). Thus, it is likely that these 2 mutations were present in different cell populations and had evolved independently. Conversely, in the second example (sample 879 which harbors both c.A199C p.M67L, and c.A183C p.Q61H) the two mutations were present at similar frequencies, 70%, and are present in the same cells (both mutations are present in the same sequencing reads). However, M67L is not a known activating mutation and, therefore, it is likely that the M67L mutation preceded or occurred concurrently with Q61H mutation and its clonal expansion is due to the activating mutation. *BRAF* was mutated in 2 samples and was also present as a major and minor clone (92% and 36% of cells, 46% and 12% of reads). The sample with only 12% of mutated reads was trisomic for chromosome 7, the location of *BRAF*, consistent with the mutation

occurring after the chromosome became trisomic. Previous data on *BRAF* in myeloma reported 2.4% of samples with a V600E mutation and an additional 1.8% with a K601N.[14] In our dataset we identified 4% with the V600E mutation and no samples with the K601N mutation.

In order to provide further information on how the subclonal architecture at a mutational level develops we studied a case of t(11;14), for which we had exome sequencing data and single cells stored. Exome sequencing identified 38 acquired non-synonymous mutations. Upon integration with copy number data from a SNP 6.0 array (Affymetrix) the frequency of each SNV in the cell population was calculated and a mutation kernel plot produced (**Figure 5A**). This identified three potential populations which could be summarized through analysis of four mutations, *ATM* (estimated at 95% of population), *FSIP2* (57%), *CLTC* (24%) and *GLMN* (33%). These four mutations were picked for confirmation of clonal frequency by single cell analysis. Genotyping for all four variants was performed on 270 single cells and showed the presence of the *ATM* mutation in 97% of single cells, *FSIP2* in 59%, *CLTC* in 30% and *GLMN* in 27% (**Figure 5B**). This confirmed that sequencing allelic reads can be used to accurately estimate the frequency of the mutation in the cell population. In addition to this, the single cell analysis also indicated that all the cells contained the *ATM* mutation and there were three subpopulations, one of which only contained the *ATM* mutations, another with and *ATM* and *FSIP2* mutation and a third with *ATM*, *CLTC* and *GLMN* mutations (**Figure 5C and 5D**).

Discussion

All myeloma passes through a MGUS phase yet there are well described differences in the pathogenesis and clinical outcome of the t(4;14) and t(11;14) subgroups of myeloma.[26,27] It has been estimated that the transition from MGUS to myeloma takes a median of 25 years[28] during which time mutations may accumulate.[10] The frequency at which the t(4;14) translocation occurs in MGUS patients is markedly lower than that seen in myeloma patients (4% versus 11%), whereas the t(11;14) occurs in relatively equal proportions (10-14%).[2,3] This difference in frequency reflects the unstable phenotype induced by the t(4;14) which decreases the time to develop myeloma and could affect mutational load at presentation. We show that t(4;14) samples at presentation have an increased mutational load compared to t(11;14) samples (median 27 vs. 23.5 nonsynonymous mutations, respectively) and with increased sample sizes this trend may become significant. We can conclude that the number and spectrum of mutations necessary to be a presenting case of myeloma is similar in the two molecular subgroups. As it is likely that cases with t(4;14) develop myeloma over a shorter time period than those with t(11;14) we conclude that the genome of the t(4;14) cases is more unstable.

Additionally, transition and tranversion rates between the two groups were equivalent. Previously we have shown that the t(4;14) subgroup has increased DNA hypermethylation compared to other subgroups.[10] It may be expected, therefore, that there would be an increase in transversions in the t(4;14) group, due to the deamination of 5-methylcytosine to uracil, which is replaced by thymine. However, this was not the case but previous studies had identified an increased mutation rate at CpG islands[14] and it may be that the exome capture, which has a bias against first exons and CpG rich sequences, has not detected this.

We incorporated our data with data from the other myeloma genome sequencing dataset that together identified 228 recurrently mutated genes in 60 samples. The most frequently mutated genes were *KRAS*, *NRAS*, *DIS3*, *FAM46C*, *LRP1B*, *DNAH5* and *WHSC1*. There were some differences in the genes identified between the datasets, but this be accounted for by the techniques used (whole genome versus whole exome; *WHSC1, LRP1B*) or the cytogenetic groups studied (random sampling versus translocations; *DIS3*) or treated and non-treated patients (*FAM46C*). In order to address the passenger/driver status of the genes identified we utilized the COSMIC database to identify recurrently mutated genes in other cancer cell types. Of the 571 genes found mutated in our own dataset 63 genes have not been identified as mutated in any cancer type. 21 genes were identified as being mutated in hematological cancers and of these 6 genes were recurrently mutated in our dataset (*NRAS*, *KRAS*, *BRAF*, *PRDM1*, *SEC31A*, *PDE4DIP*) (**Supplementary Table 3**).

One of the novel findings of the first myeloma genome publication was the identification of *DIS3* mutations in 10% of myeloma samples. We have confirmed the presence of *DIS3* mutations in myeloma, and found that the proportion of affected samples was higher (18%). This difference is due to the samples analyzed, as *DIS3* mutations have been found exclusively in t(4;14) or t(11;14) samples, of which our dataset is composed. In addition, we found that all samples with *DIS3* mutations also have deletion of 13q14, the region surrounding *RB1*, and so deletion of 13q14 and mutation of *DIS3* may be collaborating lesions. The gene of interest at 13q14 has been debated for many years and may be *RB1* or one of the microRNAs implicated in chronic lymphocytic leukemia (miR-15a/16).[29,30] It remains unclear which region(s) of chromosome 13 are important but the discovery of yet another locus will confound the situation further and only through analyses of large datasets, comprising copy number and mutational data, will the true nature of these abnormalities be discovered. *DIS3* mutations have also been found in relapse acute myeloid leukemia (AML) which is also known to have deletion or uniparental disomy of chromosome 13.[31,32]

When we compare the intersection of mutations in the two groups we see that only 3% are shared. The majority of mutations in each group are distinct suggesting that the pathological contribution of the genes mutated in each group is different. We show that genes expressed at >75% have a greater rate of mutation which could explain some of the differences between the subgroups. Gene Ontology analysis of genes disrupted in the t(4;14) group suggests an enrichment for genes involved in cytoskeleton organization, microtubule movement and actin filament based processes as well as chromatin organization, **Table 5**. The t(11;14) group showed an enrichment for protein phosphorylation and Ras signaling. This observation is consistent with the idea that collaborating mutations are likely to interact with the initiating events in each molecular subtypes of myeloma to hijack the normal physiological role of plasma cells, contributing to their malignant transformation. RAS pathway mutations have been shown to occur in myeloma[33] and here we show convincingly that RAS/MAPK pathway mutations tend to be mutually exclusive, consistent with the lack of a clonal survival advantage for cells acquiring additional mutations within a pathway already deregulated.

We demonstrate that driver mutations are present as minor subclones in half of the samples. One sample was identified which contained two activating *KRAS* mutations present in different subclonal populations, comprising 20% and 48% of cells. This suggests that subclones are continually at risk of developing driver mutations that can confer a growth and survival advantage leading to clonal dominance over time. Single cell analysis confirmed the presence of subclonal substructure at a mutational level. In one example we showed that all cells contained an *ATM* mutation, but two subpopulations existed with mutually exclusive mutations in *FSIP2* and *GLMN/CLTC*. The mutant cell frequency plot for this sample shows 3 distinct peaks (at 0.3, 0.5 and 1.0) corresponding to the 3 main clones in this sample (mutated at *ATM*, *ATM+FSIP2*, or *ATM+GLMN/CLTC*). This type of subclonal structure has been described for other hematological malignancies such as AML, acute lymphoblastic leukemia, breast cancer and prostate cancer using FISH, single cell sequencing or ultra-deep sequencing.[32,34-37] The realization that tumor biology consists of multiple independent subclones has important ramifications for tumor progression and personalized treatment strategies.

We provide evidence that there is a distinct mutational landscape in the t(4;14) and t(11;14) groups of myeloma. Following initial immortalization by the acquisition of a translocation each group follows a distinct pathway to becoming a presenting case of myeloma. Copy number abnormalities may occur relatively early on in the pathogenesis of myeloma and regions of LOH undergo mutation more frequently. Mutations in regions of LOH tend to be

nonsynonymous, pathologically relevant variants which contribute to clonal progression. The rate of mutation seems to be higher in the t(4;14) group which may lead to their more rapid progression to myeloma. The net mutational load, however, that is required to develop presenting myeloma is consistent in the two groups. We demonstrate intraclonal heterogeneity and the continued acquisition of mutations within subclones which leads to clonal outgrowth and disease progression. The clonal architecture described here, with multiple subclones vying for dominance, is reminiscent of the Darwinian natural selection process with a branching, non-linear accumulation of mutations in malignant plasma cells.

Acknowledgements

We acknowledge grant support from Myeloma UK and the National Institute of Health Biomedical Research Centre at the Royal Marsden Hospital as well as Children with cancer UK. We gratefully acknowledge the expertise of Ian Titley for assistance in flow cytometry sorting of single cells and Ricardo Morilla for assistance in setting up the myeloma flow panel.

Author Contributions

B.A.W. designed and performed research, analyzed data and wrote the paper.
C.P.W. performed computational research, analyzed data and wrote the paper.
L.M. performed research and analyzed data.
N.P. developed single cell protocols.
S.H. performed research.
D.G. designed research.
D.C.J. performed research.
I.K. designed and performed research.
C.J.L. provided resources.
A.A. provided resources.
F.E.D. designed research.
G.J.M. designed research, analyzed data and wrote the paper.

Conflict of Interest Disclosures

The authors have no conflicts of interest to disclose.

References

1. Boyd KD, Ross FM, Chiecchio L, 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(2):349-355.
2. Lopez-Corral L, Gutierrez NC, Vidriales MB, et al. The progression from MGUS to smoldering myeloma and eventually to multiple myeloma involves a clonal expansion of genetically abnormal plasma cells. *Clin Cancer Res*. 2011;17(7):1692-1700.
3. Ross FM, Chiecchio L, Dagrada G, 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(7):1221-1225.
4. Keats JJ, Reiman T, Maxwell CA, et al. In multiple myeloma, t(4;14)(p16;q32) is an adverse prognostic factor irrespective of FGFR3 expression. *Blood*. 2003;101(4):1520-1529.
5. Walker BA, Leone PE, Chiecchio L, et al. A compendium of myeloma-associated chromosomal copy number abnormalities and their prognostic value. *Blood*. 2010;116(15):e56-65.
6. Eguchi M, Eguchi-Ishimae M, Greaves M. Molecular pathogenesis of MLL-associated leukemias. *Int J Hematol*. 2005;82(1):9-20.
7. Kuo AJ, Cheung P, Chen K, et al. NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. *Mol Cell*. 2011;44(4):609-620.
8. Pei H, Zhang L, Luo K, et al. MMSET regulates histone H4K20 methylation and 53BP1 accumulation at DNA damage sites. *Nature*. 2011;470(7332):124-128.
9. Marango J, Shimoyama M, Nishio H, et al. The MMSET protein is a histone methyltransferase with characteristics of a transcriptional corepressor. *Blood*. 2008;111(6):3145-3154.
10. Walker BA, Wardell CP, Chiecchio L, et al. Aberrant global methylation patterns affect the molecular pathogenesis and prognosis of multiple myeloma. *Blood*. 2011;117(2):553-562.
11. Carrasco DR, Tonon G, Huang Y, et al. High-resolution genomic profiles define distinct clinico-pathogenetic subgroups of multiple myeloma patients. *Cancer Cell*. 2006;9(4):313-325.
12. Walker BA, Morgan GJ. Use of single nucleotide polymorphism-based mapping arrays to detect copy number changes and loss of heterozygosity in multiple myeloma. *Clin Lymphoma Myeloma*. 2006;7(3):186-191.
13. Walker BA, Leone PE, Jenner MW, et al. Integration of global SNP-based mapping and expression arrays reveals key regions, mechanisms and genes important in the pathogenesis of multiple myeloma. *Blood*. 2006;108(5):1733-1743.
14. Chapman MA, Lawrence MS, Keats JJ, et al. Initial genome sequencing and analysis of multiple myeloma. *Nature*. 2011;471(7339):467-472.
15. Jenner MW, Leone PE, Walker BA, et al. Gene mapping and expression analysis of 16q loss of heterozygosity identifies WWOX and CYLD as being important in determining clinical outcome in multiple myeloma. *Blood*. 2007;110(9):3291-3300.
16. Leone PE, Walker BA, Jenner MW, et al. Deletions of CDKN2C in multiple myeloma: biological and clinical implications. *Clin Cancer Res*. 2008;14(19):6033-6041.
17. Boyd KD, Ross FM, Walker BA, et al. Mapping of chromosome 1p deletions in myeloma identifies FAM46C at 1p12 and CDKN2C at 1p32.3 as being genes in regions associated with adverse survival. *Clin Cancer Res*. 2011;17(24):7776-7784.
18. Annunziata CM, Davis RE, Demchenko Y, et al. Frequent engagement of the classical and alternative NF-kappaB pathways by diverse genetic abnormalities in multiple myeloma. *Cancer Cell*. 2007;12(2):115-130.
19. Keats JJ, Fonseca R, Chesi M, et al. Promiscuous mutations activate the noncanonical NF-kappaB pathway in multiple myeloma. *Cancer Cell*. 2007;12(2):131-144.
20. Campbell PJ, Pleasance ED, Stephens PJ, et al. Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. *Proc Natl Acad Sci U S A*. 2008;105(35):13081-13086.

21. Kozarewa I, Rosa-Rosa JM, Wardell CP, et al. A modified method for whole exome resequencing from minimal amounts of starting DNA. *PLoS ONE*. 2012;7(3):e32617.
22. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
23. Dickens NJ, Walker BA, Leone PE, et al. Homozygous deletion mapping in myeloma samples identifies genes and an expression signature relevant to pathogenesis and outcome. *Clin Cancer Res*. 2010;16(6):1856-1864.
24. Kassambara A, Hose D, Moreaux J, et al. Genes with a spike expression are clustered in chromosome (sub)bands and spike (sub)bands have a powerful prognostic value in patients with multiple myeloma. *Haematologica*. 2011.
25. Sims D, Bursteinas B, Gao Q, et al. ROCK: a breast cancer functional genomics resource. *Breast Cancer Res Treat*. 2010;124(2):567-572.
26. Weiss BM, Abadie J, Verma P, Howard RS, Kuehl WM. A monoclonal gammopathy precedes multiple myeloma in most patients. *Blood*. 2009;113(22):5418-5422.
27. Bergsagel PL, Kuehl WM. Molecular pathogenesis and a consequent classification of multiple myeloma. *J Clin Oncol*. 2005;23(26):6333-6338.
28. Blade J, Dimopoulos M, Rosinol L, Rajkumar SV, Kyle RA. Smoldering (asymptomatic) multiple myeloma: current diagnostic criteria, new predictors of outcome, and follow-up recommendations. *J Clin Oncol*. 2010;28(4):690-697.
29. Calin GA, Ferracin M, Cimmino A, et al. A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N Engl J Med*. 2005;353(17):1793-1801.
30. Davies FE, Dring AM, Li C, et al. Insights into the multistep transformation of MGUS to myeloma using microarray expression analysis. *Blood*. 2003;102(13):4504-4511.
31. Raghavan M, Smith LL, Lillington DM, et al. Segmental uniparental disomy is a commonly acquired genetic event in relapsed acute myeloid leukemia. *Blood*. 2008;112(3):814-821.
32. Ding L, Ley TJ, Larson DE, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature*. 2012;481(7382):506-510.
33. Corradini P, Ladetto M, Voena C, et al. Mutational activation of N- and K-ras oncogenes in plasma cell dyscrasias. *Blood*. 1993;81(10):2708-2713.
34. Anderson K, Lutz C, van Delft FW, et al. Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature*. 2011;469(7330):356-361.
35. Mullighan CG, Phillips LA, Su X, et al. Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. *Science*. 2008;322(5906):1377-1380.
36. Greaves M, Maley CC. Clonal evolution in cancer. *Nature*. 2012;481(7381):306-313.
37. Navin N, Kendall J, Troge J, et al. Tumour evolution inferred by single-cell sequencing. *Nature*. 2011;472(7341):90-94.

Table 1. Comparison of mutation rates between the t(11;14) and t(4;14) subgroups

	t(4;14) (n=10)	t(11;14) (n=12)	Total (n=22)
SNVs			
Median Acquired Exonic SNVs	**38**	**33.5**	**36**
(Range)	**(28-63)**	**(26-50)**	**(26-63)**
Median Acquired Non-synonymous Exonic SNVs	**27**	**23.5**	**24.5**
Acquired NS:S	**2.51**	**2.59**	**2.55**
Indels			
Median acquired exonic indels	**1.5**	**2**	**2**
(Range)	**(0-4)**	**(0-4)**	**(0-4)**
Median acquired frameshift indels	**1**	**1.5**	**1**
Median acquired nonframeshift indels	**0**	**0**	**0**
Exonic Homozygous Deletions			
Average acquired exonic deletions	**3**	**3.5**	**3**
(Range)	**(1-19)**	**(0-181)**	**(0-181)**

Table 2. Mutations in regions of copy number change

	Gain	Loss	Normal
Number of NS mutations	**48**	**33**	**517**
Number of regions	**91**	**102**	**NA**
Total size of regions (Mb)	**4296**	**3678**	**57010**
Mb per mutation	**89.5**	**111.4**	**110.3**

Table 3: List of mutated genes in regions of loss of heterozygosity

Gene	Sample	Chr.	Base position	Amino acid change
ZNF474	91603	5	121488236	p.H184R
PHAX	91603	5	125960427	p.N359S
FUT9	91603	6	96651572	p.E181X
SLC18A1	90738	8	20038466	p.T4P
MYC	91890	8	128752881	p.L348V
FLT3	91577	13	28592689	p.V819A
COG3	91827	13	46057379	p.D244E
DIS3	90468	13	73333955	p.K922T
DIS3	90405	13	73345947	p.Y501D
DIS3	90738	13	73347941	p.T344P
UGGT2	90357	13	96511903	p.N1256T
STK24	90482	13	99109461	p.S407F
CHD8	90245	14	21862547	p.R1830C
MAX	90245	14	65560493	p.R26H
TCL1A	90142	14	96180394	p.C4S
CYLD	90273	16	50783949	p.K114X
PKD1L2	90273	16	81209247	p.R849H
MYH4	90482	17	10354723	p.E1262G
KRTAP16-1	90405	17	39464278	p.T410P
ACTG1	90983	17	79478992	p.E100D
ZCCHC2	90142	18	60242631	p.R1106K
CRYBB2	91577	22	25620907	p.N26I
SEC14L3	91603	22	30857373	p.D335E
FAM47B	90738	X	34962196	p.K416N
USP11	90226	X	47092384	p.V24A
MORC4	90226	X	106200218	p.E468X
IL13RA2	90738	X	114251797	p.Y12X
MAP7D3	90738	X	135318488	fs

Table 4: Frequency of clonal subpopulations carrying a mutation

Sample	Translocation group	Gene	SNV	Copy Number	Frequency of SNV	Estimated % of tumor population*	Major/minor clone component
142	t(4;14)	*KRAS*	c.A182G	2	36%	72%	Major
738	t(4;14)	*KRAS*	c.G436C	2	36%	72%	Major
879	t(11;14)	*KRAS*	c.A199C	2	35%	70%	Major
879	t(11;14)	*KRAS*	c.A183C	2	34%	68%	Major
245	t(4;14)	*KRAS*	c.G38A	2	33%	66%	Major
3	t(11;14)	*KRAS*	c.A182G	2	10%	20%	Minor
3	t(11;14)	*KRAS*	c.G34C	2	24%	48%	Minor
827	t(11;14)	*KRAS*	c.A182G	2	14%	28%	Minor
1112	t(11;14)	*KRAS*	c.G38A	2	14%	28%	Minor
983	t(11;14)	*NRAS*	c.C181A	2	47%	94%	Major
203	t(11;14)	*NRAS*	c.C181A	2	48%	96%	Major
482	t(11;14)	*NRAS*	c.A182G	2	16%	32%	Minor
1890	t(4;14)	*NRAS*	c.A182G	2	17%	34%	Minor
468	t(11;14)	*BRAF*	c.T1799A	2	46%	92%	Major
405	t(4;14)	*BRAF*	c.T1799A	3	12%	36%	Minor
983	t(11;14)	*DIS3*	c.G2249A	2	46%	92%	Major
738	t(4;14)	*DIS3*	c.A1030C	1	81%	81%	Major
468	t(11;14)	*DIS3*	c.A2765C	1	69%	69%	Major
405	t(4;14)	*DIS3*	c.T1501G	1	29%	29%	Minor

*calculated using copy number and SNV frequency

Table 5: Pathway analysis for genes mutated in t(4;14),top, and t(11;14), bottom.

Term	Count	%	P-Value
GO:0007010 cytoskeleton organization	17	8.6	$1.92x10^{-5}$
GO:0030036 actin cytoskeleton organization	11	5. 6	$1.70 x10^{-4}$
GO:0030029 actin filament-based process	11	5.6	$2.85 x10^{-4}$
GO:0006325 chromatin organization	13	6.6	$8.10 x10^{-4}$
GO:0007017 microtubule-based process	10	5.1	0.001
GO:0051276 chromosome organization	13	6.6	0.006
GO:0016568 chromatin modification	9	4.6	0.009

Term	Count	%	P-Value
GO:0006468 protein amino acid phosphorylation	17	9.3	0.001
GO:0016044 membrane organization	12	6.6	0.002
GO:0006796 phosphate metabolic process	21	11.5	0.003
GO:0006793 phosphorus metabolic process	21	11.5	0.003
GO:0007265 Ras protein signal transduction	6	3.3	0.005

Figure Legends

Figure 1. A, Transition/Transversion rates in sample groups. B, Intersection of t(4;14) and t(11;14) genes with NS SNVs and indels.

Figure 2: The dynamic relationship of mutation to loss of heterozygosity on chromosome 13. A cell with 2 copies of a given chromosome (A) can either gain a mutation (star) in one copy (B) and subsequently delete the unaltered copy giving rise to a mixed population of cells (C), which then through selection results in a clonal population of cells with a mutation and deletion (D). Alternatively, the original cell (A) could delete one copy of the chromosome (E) and subsequently a sub-population mutates the remaining allele (F) which is the selected for giving rise to a clonal population with both deletion and mutation (D). Samples are shown next to the cell genotype, as defined by sequencing and copy number data, along with the mutated gene on chromosome 13. Synonymous and non-synonymous (underlined) SNVs shown. *= interstitial deletion of *RB1* region detected with mutation in a region of normal copy number.

Figure 3: Association of abnormalities and mutations. Cytogenetic data repeated at the top of each column.

Figure 4: Gaussian kernel density plots indicating the frequency of cells carrying all acquired mutations. Frequency is calculated by adjusting mutant allele burden by copy number of the loci mutated. Samples with mutated *KRAS* (A), *NRAS* (B) or *BRAF* (C) indicates that driver mutations may be in major (>0.5) or minor (<0.5) clone populations. Frequency of the mutated RAS pathway gene is indicated by a dotted line.

Figure 5: Clonal heterogeneity determined by single cell analysis. A, Mutant allele frequencies of 47 exonic synonymous and nonsynonymous variants identified through exome sequencing. The frequencies of four variants are shown. B, The proportion of cells identified with a mutated genotype for four genes as determined by exome sequencing allelic frequencies and single cell genotyping. C, Hierarchical clustering of the genotypes generated from the analysis of 270 single cells. Red, mutated; Black, reference. Individual populations are identified (1-4). D, Determined cell lineages showing the proportion of tumor population and genotypes for the four genes in the identified populations.

FIGURE 1

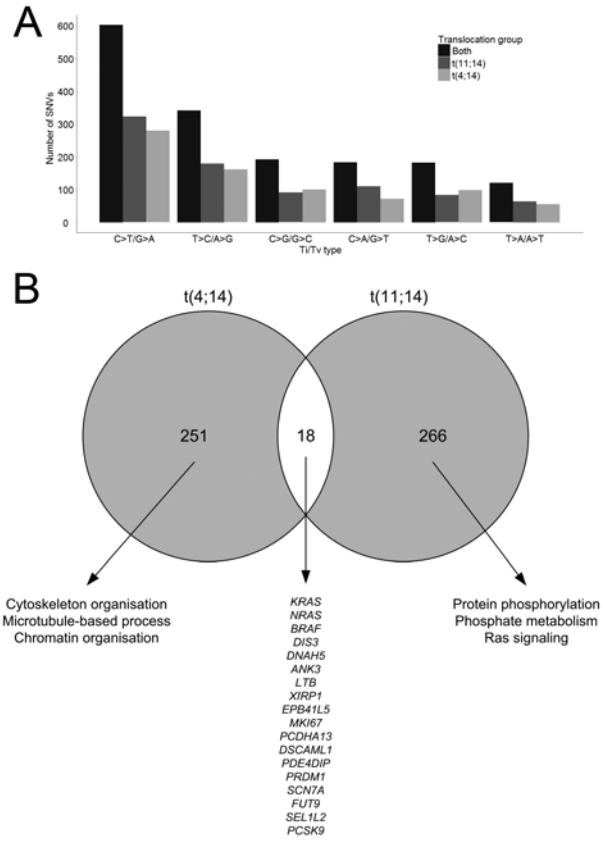

A
Translocation group
Both
t(11;14)
t(4;14)
Number of SNVs
600
500
400
300
200
100
0
C>T/G>A
T>C/A>G
C>G/G>C
C>A/G>T
T>G/A>C
T>A/A>T
Ti/Tv type
B
t(4;14)
t(11;14)
251
18
266
Cytoskeleton organisation
Microtubule-based process
Chromatin organisation
KRAS
NRAS
BRAF
DIS3
DNAH5
ANK3
LTB
XIRP1
EPB41L5
MKI67
PCDHA13
DSCAML1
PDE4DIP
PRDM1
SCN7A
FUT9
SEL1L2
PCSK9
Protein phosphorylation
Phosphate metabolism
Ras signaling

FIGURE 2

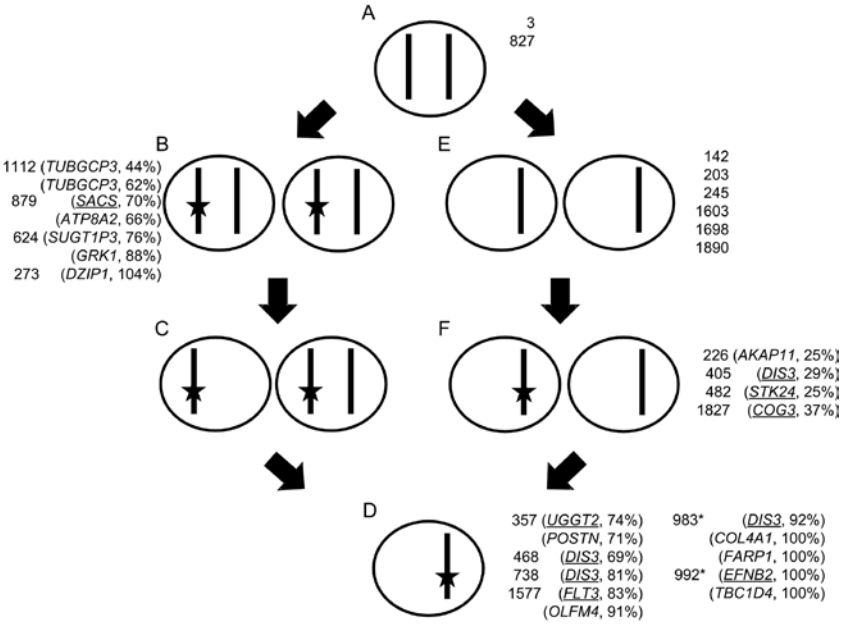

A
3
827
B
1112 (TUBGCP3, 44%)
(TUBGCP3, 62%)
879 (SACS, 70%)
(ATP8A2, 66%)
624 (SUGT1P3, 76%)
(GRK1, 88%)
273 (DZIP1, 104%)
E
142
203
245
1603
1698
1890
C
F
226 (AKAP11, 25%)
405 (DIS3, 29%)
482 (STK24, 25%)
1827 (COG3, 37%)
D
357 (UGGT2, 74%)
(POSTN, 71%)
468 (DIS3, 69%)
738 (DIS3, 81%)
1577 (FLT3, 83%)
(OLFM4, 91%)
983* (DIS3, 92%)
(COL4A1, 100%)
(FARP1, 100%)
992* (EFNB2, 100%)
(TBC1D4, 100%)

FIGURE 3

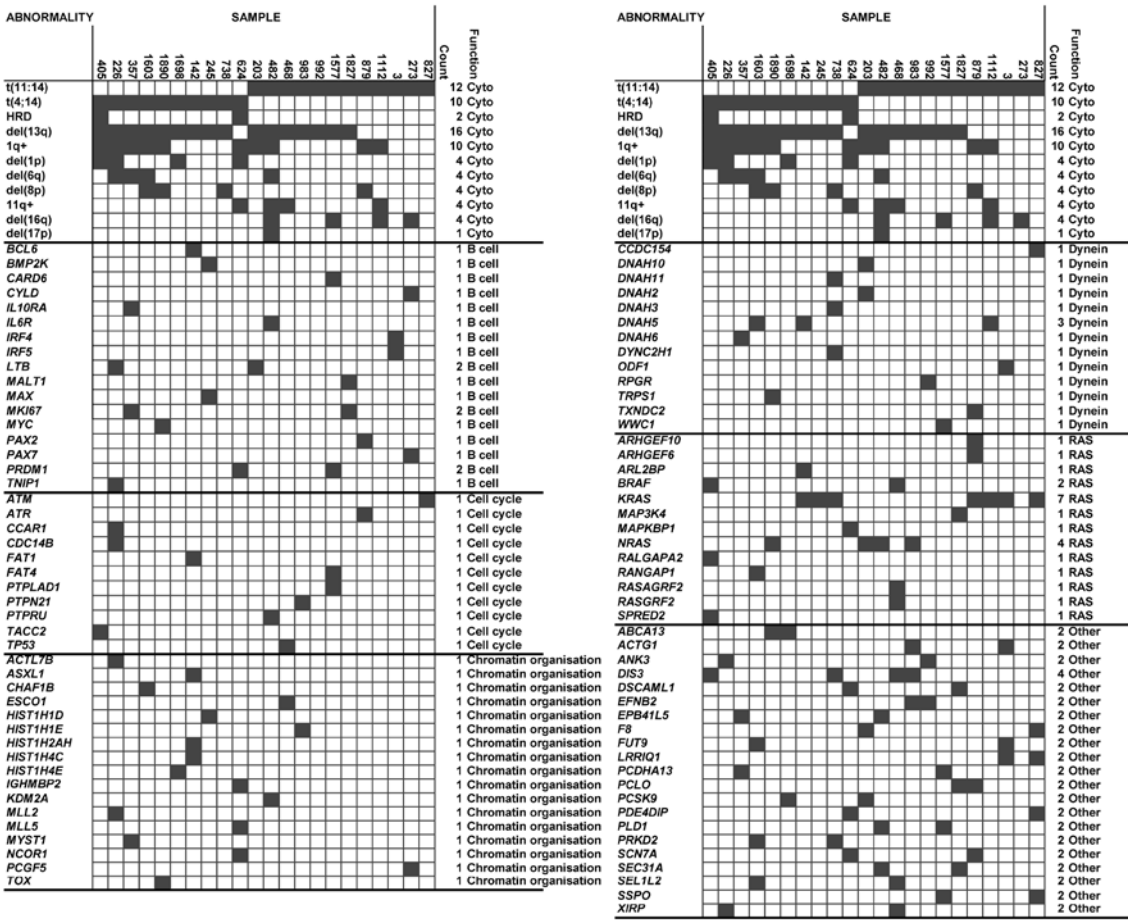

ABNORMALITY
SAMPLE
405 226 357 1603 1890 1698 142 245 738 624 203 482 468 983 992 1577 1827 879 1112 3 273 827
Count
Function
t(11:14) 12 Cyto
t(4;14) 10 Cyto
HRD 2 Cyto
del(13q) 16 Cyto
1q+ 10 Cyto
del(1p) 4 Cyto
del(6q) 4 Cyto
del(8p) 4 Cyto
11q+ 4 Cyto
del(16q) 4 Cyto
del(17p) 1 Cyto
BCL6 1 B cell
BMP2K 1 B cell
CARD6 1 B cell
CYLD 1 B cell
IL10RA 1 B cell
IL6R 1 B cell
IRF4 1 B cell
IRF5 1 B cell
LTB 2 B cell
MALT1 1 B cell
MAX 1 B cell
MKI67 2 B cell
MYC 1 B cell
PAX2 1 B cell
PAX7 1 B cell
PRDM1 2 B cell
TNIP1 1 B cell
ATM 1 Cell cycle
ATR 1 Cell cycle
CCAR1 1 Cell cycle
CDC14B 1 Cell cycle
FAT1 1 Cell cycle
FAT4 1 Cell cycle
PTPLAD1 1 Cell cycle
PTPN21 1 Cell cycle
PTPRU 1 Cell cycle
TACC2 1 Cell cycle
TP53 1 Cell cycle
ACTL7B 1 Chromatin organisation
ASXL1 1 Chromatin organisation
CHAF1B 1 Chromatin organisation
ESCO1 1 Chromatin organisation
HIST1H1D 1 Chromatin organisation
HIST1H1E 1 Chromatin organisation
HIST1H2AH 1 Chromatin organisation
HIST1H4C 1 Chromatin organisation
HIST1H4E 1 Chromatin organisation
IGHMBP2 1 Chromatin organisation
KDM2A 1 Chromatin organisation
MLL2 1 Chromatin organisation
MLL5 1 Chromatin organisation
MYST1 1 Chromatin organisation
NCOR1 1 Chromatin organisation
PCGF5 1 Chromatin organisation
TOX 1 Chromatin organisation
ABNORMALITY
SAMPLE
405 226 357 1603 1890 1698 142 245 738 624 203 482 468 983 992 1577 1827 879 1112 3 273 827
Count
Function
t(11:14) 12 Cyto
t(4;14) 10 Cyto
HRD 2 Cyto
del(13q) 16 Cyto
1q+ 10 Cyto
del(1p) 4 Cyto
del(6q) 4 Cyto
del(8p) 4 Cyto
11q+ 4 Cyto
del(16q) 4 Cyto
del(17p) 1 Cyto
CCDC154 1 Dynein
DNAH10 1 Dynein
DNAH11 1 Dynein
DNAH2 1 Dynein
DNAH3 1 Dynein
DNAH5 3 Dynein
DNAH6 1 Dynein
DYNC2H1 1 Dynein
ODF1 1 Dynein
RPGR 1 Dynein
TRPS1 1 Dynein
TXNDC2 1 Dynein
WWC1 1 Dynein
ARHGEF10 1 RAS
ARHGEF6 1 RAS
ARL2BP 1 RAS
BRAF 2 RAS
KRAS 7 RAS
MAP3K4 1 RAS
MAPKBP1 1 RAS
NRAS 4 RAS
RALGAPA2 1 RAS
RANGAP1 1 RAS
RASAGRF2 1 RAS
RASGRF2 1 RAS
SPRED2 1 RAS
ABCA13 2 Other
ACTG1 2 Other
ANK3 2 Other
DIS3 4 Other
DSCAML1 2 Other
EFNB2 2 Other
EPB41L5 2 Other
F8 2 Other
FUT9 2 Other
LRRIQ1 2 Other
PCDHA13 2 Other
PCLO 2 Other
PCSK9 2 Other
PDE4DIP 2 Other
PLD1 2 Other
PRKD2 2 Other
SCN7A 2 Other
SEC31A 2 Other
SEL1L2 2 Other
SSPO 2 Other
XIRP 2 Other

FIGURE 4

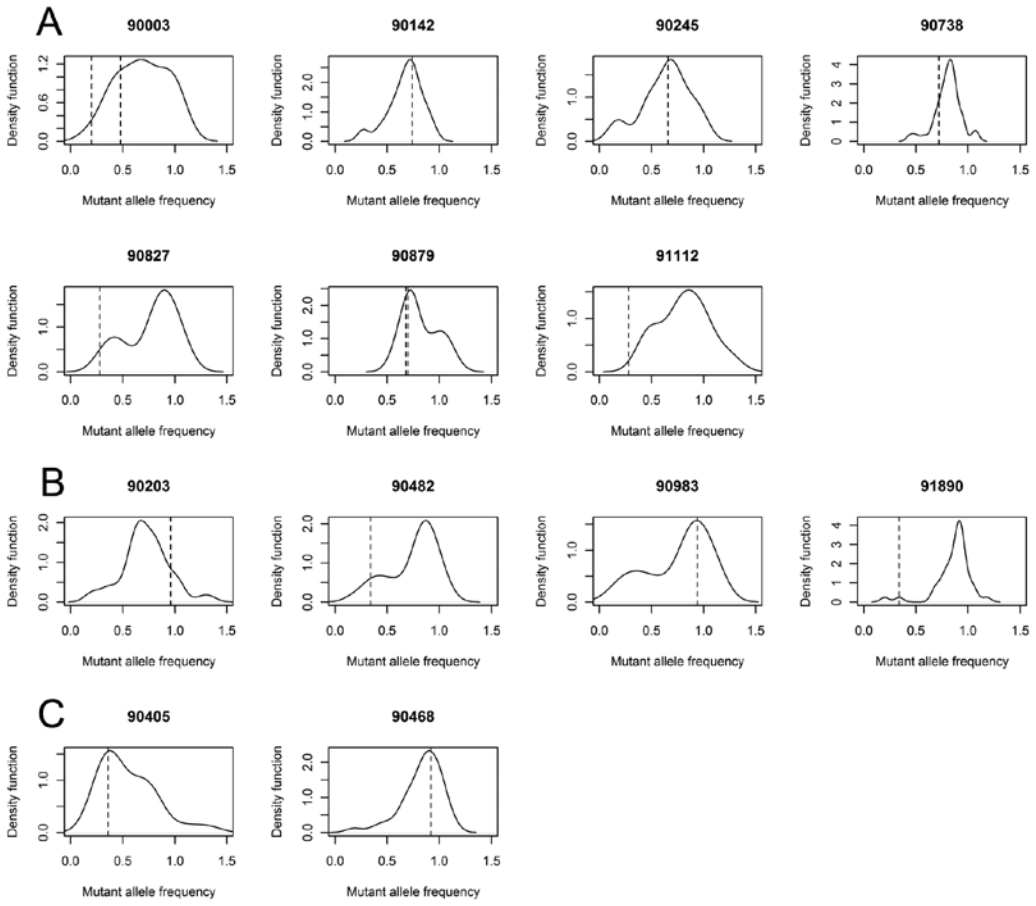
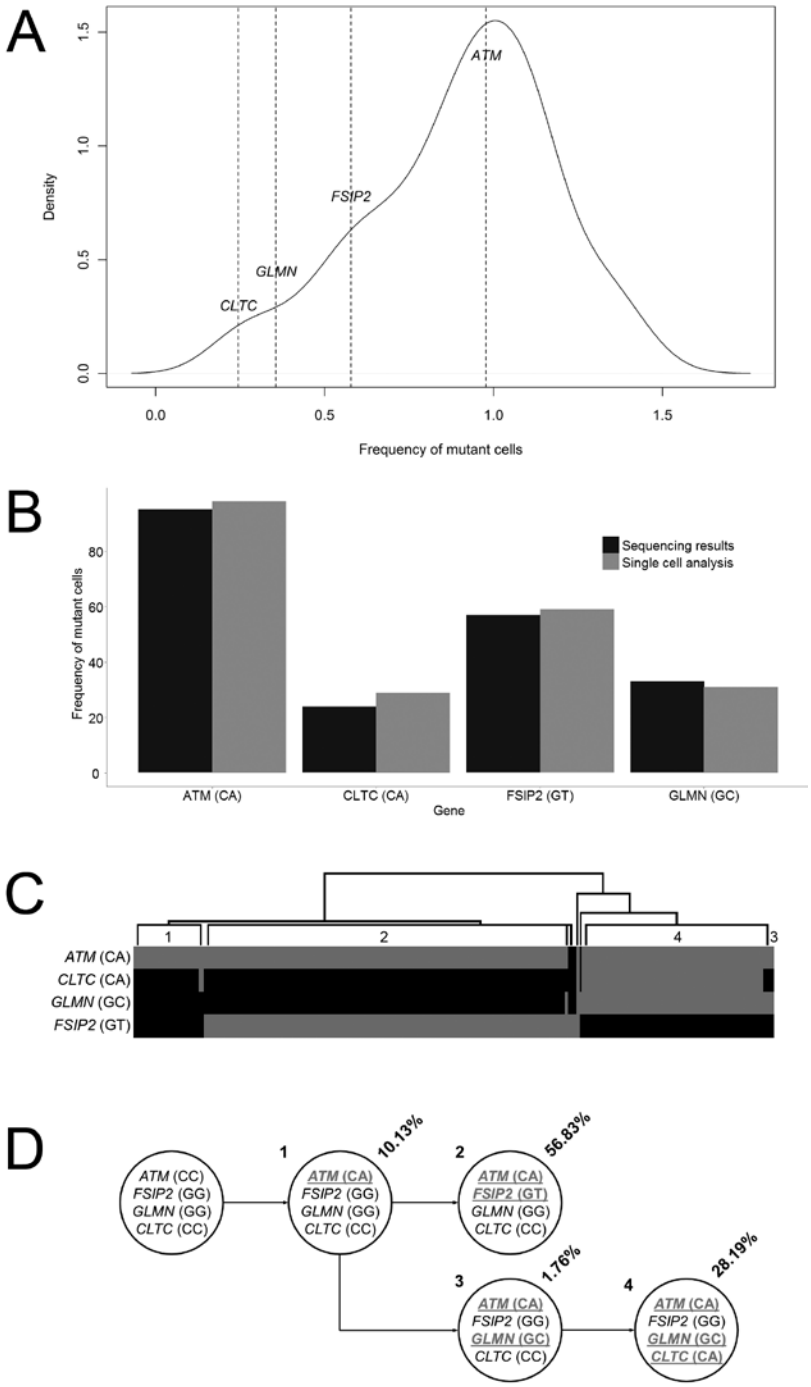

A
90003
90142
90245
90738
90827
90879
91112
B
90203
90482
90983
91890
C
90405
90468
Density function
Mutant allele frequency

FIGURE 5


A
Density
Frequency of mutant cells
ATM
FSIP2
GLMN
CLTC
B
Frequency of mutant cells
Sequencing results
Single cell analysis
ATM (CA)
CLTC (CA)
FSIP2 (GT)
GLMN (GC)
Gene
C
ATM (CA)
CLTC (CA)
GLMN (GC)
FSIP2 (GT)
D
ATM (CC) FSIP2 (GG) GLMN (GG) CLTC (CC)
ATM (CA) FSIP2 (GG) GLMN (GG) CLTC (CC)
10.13%
ATM (CA) FSIP2 (GT) GLMN (GG) CLTC (CC)
56.83%
ATM (CA) FSIP2 (GG) GLMN (GC) CLTC (CC)
1.76%
ATM (CA) FSIP2 (GG) GLMN (GC) CLTC (CA)
28.19%