3. Whole exome sequencing reveals rapid acquisition of driver mutations and branching evolution in a case of NPM1 positive CMML transforming to AML

3.1 Introduction

Chronic myelomonocytic leukaemia (CMML) is a clonal disorder characterised by the accumulation of monocytes in the peripheral blood together with abnormal myeloid differentiation, which is either dysplastic or proliferative or both. CMML develops due to the stepwise accumulation of genetic mutations in haematopoietic stem cells (HSC). Although several recurrent mutations have been described, none are specific to CMML (Itzykson and Solary, 2013). Cancers are thought to evolve with a complex branching clonal architecture, however the evidence so far in CMML is that the majority of mutations accumulate in a linear manner, with limited branching through loss of heterozygosity (LOH) (Itzykson et al., 2013b). CMML progresses to secondary AML in 20-30% of cases. This is thought to be driven by a clone acquiring a novel fitness conferring mutation, although the expansion of a clone with high fitness in the absence of new genetic lesions has not been excluded (Itzykson and Solary, 2013).

Nucleophosmin (*NPM1*) mutations are described in around 30% of cases of AML, making it one of the commonest driver lesions in this disease (Falini et al., 2005; TCGA_Research_Network, 2013). Mutations in *NPM1* are thought to be an early although not necessarily initiating event (Shlush et al., 2014) and define a large subgroup of AML with distinct clinical, pathological and molecular characteristics (Swerdlow, 2008). These mutations result in cytoplasmic dislocation of the *NPM1* protein and are mutually exclusive to the fusion genes, which are presumed initiating lesions in several other subtypes of AML (TCGA_Research_Network, 2013).

NPM1 is frequently overexpressed in solid malignancies and is involved in chromosomal translocations in various haematological and solid tumours, however *NPM1* mutations are generally considered specific to AML (Falini et al., 2011; Grisendi

et al., 2006). Therapy related AMLs and AMLs secondary to myeloproliferative neoplasms (MPN) or myelodysplasia (MDS) are sometimes found to have cytoplasmic *NPM1*, but *NPM1* mutations are much more common in de novo disease (Falini et al., 2005; Fernandez-Mercado et al., 2012; Gale et al., 2008; Schnittger et al., 2011). Also, *NPM1* mutations have been reported in a small proportion of cases of MDS(Bains et al., 2011; Falini et al., 2011; Zhang et al., 2007). However, *NPM1* mutant AML may have dysplastic features (Falini et al., 2011) and as the distinction between MDS and AML is based on a 20% threshold of detectable blasts, the diagnosis is subject to sampling and inter-observer variation. It is therefore unclear if these cases represent true MDS or early evolving AML. Similarly there are occasional reports of *NPM1* mutant CMML (Bains et al., 2011; Caudill et al., 2006; Courville et al., 2013; Itzykson et al., 2013a), but these cases generally progress rapidly to AML (Bains et al., 2011; Caudill et al., 2006; Courville et al., 2013). Similarly, some experts question whether these cases are CMML or AML which has been detected in an early, sub-clinical phase, accompanied by marked monocytic differentiation (Falini et al., 2011).

In the acute leukaemia clinic at Addenbrooke's Hospital we were treating one such patient, who was initially diagnosed with CMML, but progressed to clinically overt AML within three months. Routine diagnostic tests performed at the hospital at the time of AML presentation detected both *NPM1* and *FLT3-ITD* mutations, but on retrospective assessment of the CMML sample only the *NPM1* mutation was identified. In order to understand the nature of the CMML to AML progression in this uncommon situation where the *NPM1* mutation was detectable prior to the onset of overt AML, I studied the paired CMML and AML diagnostic samples using deep sequencing.

3.2 Clinical Case

A 50y.o. woman presented with an eight week history of non-specific symptoms. She was anaemic (haemoglobin 7.5g/dL) and had a peripheral blood monocytosis (1.74 x 10^9/L) with normal neutrophil count (5.48 x 10^9/L) and a total white cell count (WCC) of 8.7 x 10^9/L. Her bone marrow was hypercellular, with myeloid to erythroid ratio of >10:1 and dysplastic changes, but <5% blasts. There were no high risk features on FISH analysis and the cytogenetic study was normal. However 83 days later she was admitted to hospital with fevers, abdominal symptoms and a WCC of

209x10^9/L (82% blasts). The diagnosis of AML was confirmed on bone marrow examination and molecular testing identified *FLT3-ITD* and *NPM1* mutations. She was treated according to the AML17 trial protocol with cytarabine, daunorubicin and etoposide (ADE) chemotherapy and initially went into complete remission with no detectable *FLT3-ITD* or *NPM1* mutation by PCR. However, following the first consolidation cycle of ADE chemotherapy she relapsed with 38% blasts on her bone marrow. She responded to salvage treatment with FLAG-IDA (fludarabine, cytarabine, idarubicin and G-CSF) followed by high dose cytarabine and was in complete remission with incomplete peripheral recovery of blood counts at the time of allogeneic transplant, 300 days after her initial presentation with CMML. Unfortunately three months after transplant she died from relapsed disease. We studied samples taken at different time points during her clinical course. The timing of the analysed samples along with her peripheral blood WCC, haemoglobin and clinical course are shown in figure 3.1.
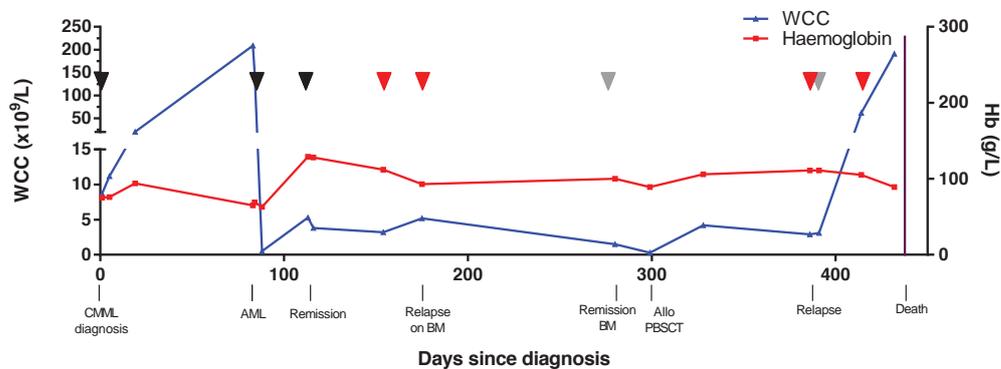

Figure 3.1: Disease timecourse. Major clinical events are noted. Black arrowheads indicate BM samples used in both exome sequencing and PCR validation. Those indicated by red and grey arrowheads were used for PCR and MiSeq only. Red = blood, grey = BM

3.3 Results

Exome sequencing was performed on whole bone marrow samples taken at diagnosis of both CMML and AML and during the first complete remission, using Agilent SureSelect Human Exon 50Mb Kit baits and the Illumina HiSeq2000 sequencing platform. This generated between 84 and 90 million, 75bp, paired end reads in all three samples. In each sample 88% of all reads were mapped to the genome covering over 98% of the targeted regions. The median coverage was 100 fold across the exome with 80% 40 fold or higher coverage in all three samples.

After comparison to the remission sample and standard filtering, ten insertions and deletions were identified on Pindel analysis in the AML samples and eight in the CMML (appendix 3A). These included a four nucleotide TCTG insertion at 5:170837547 in both samples, consistent with a type A *NPM1* mutation. The only other shared aberration called by Pindel was a complex abnormality involving 2127 bases within CDC27 on chromosome 17, of uncertain significance.

Caveman analysis for single nucleotide variants (SNV) identified 43 mutant calls in the AML sample and 62 in the CMML sample, of which 23 were common to both (table 3.1) (appendix 3B). As each read of an Illumina sequencing run derives from a single molecule of genomic DNA, the proportion of independent sequencing reads reporting a variant allele can be used to estimate the proportion of cells in a DNA sample carrying that mutation(Campbell et al., 2008). Exome sequencing of the CMML sample revealed three mutations which are likely to have a driver role in leukaemogenesis, with an allelic frequency suggesting they were present in the dominant clone; *DNMT3A*, *TET2* and *NPM1* (figure 3.2).

In addition to these three driver lesions, mutations in *PTPN11* and *SMC3*, which are both genes that are recurrently mutated in AML, were also found in the CMML sample. However, these mutations had a low variant allele frequency (VAF) in the CMML sample, suggesting they were occurring in minor sub-clones (figure 3.2). Sequencing of the AML sample demonstrated an expansion of the *SMC3* clone, but a reduction in *PTPN11* mutant reads. In fact the *PTPN11* mutation was detected in such small number in the AML sample that it did not pass filtering to be included on the output list on Caveman analysis, but mutant reads were evident on review of the raw sequencing data mpileup (Niccolo Bolli). The contrasting pattern in VAF of these two

mutations on serial sampling suggests they may not be co-occurring within a single tumour sub-clone although with this level of coverage they were still grouped together on the Dirichilet analysis.

Gene	CHR	Position	cDNA	Protein	Type	Allele		Depth			% Mutant in		
						WT	MT	Normal	CMML	AML	Normal	CMML	AML
PGLYRP4	1	153317825	c.173G>A	p.R58H	Misssense	C	T	88	100	97	0	42	45.36
DNMT3A	2	25457243	c.2644C>T	p.R882C	Misssense	G	A	90	70	73	0	42.86	47.95
ARL6	3	97503812	c.268A>G	p.I90V	Misssense	A	G	93	87	115	0	29.89	43.48
MFI2	3	196746550	c.835G>A	p.V279I	Misssense	C	T	198	210	230	0.51	32.38	40.87
TET2	4	106196940	c.5273C>G	p.S1758*	Nonsense	C	G	164	140	167	0	42.86	46.11
ACCN5	4	156775338	c.476C>T	p.A159V	Misssense	G	A	137	116	172	1.46	34.48	50
ACSL6	5	131326623	c.308G>A	p.G103D	Missense	C	T	69	60	73	0	43.33	47.95
HLA-A	6	29911272	c.571T>G	p.W191G	Missense	T	G	48	60	67	2.08	11.67	7.46
CLCN1	7	143029920	c.1355C>T	p.P452L	Missense	C	T	229	188	197	0	35.11	38.58
ZMAT4	8	40532213	c.577+10G>T	p.?	Splice	C	A	101	81	107	0.99	45.68	43.93
AQP7	9	33395107	c.113T>C	p.L38P	Misssense	A	G	106	129	133	2.83	6.98	9.02
ZCCHC7	9	37356935	c.1302G>T	p.W434C	Misssense	G	T	106	98	84	1.89	26.53	50
EPB41L4B	9	112003846	Non-coding	r.2171u>c	3'UTR	A	G	35	35	53	0	45.71	45.28
SMC3	10	112360841	c.2597T>C	p.L866P	Misssense	T	C	500	500	500	0	4.2	14
ENSG00000196779	11	57876606	c.528C>G	p.F176L	Misssense	G	C	108	135	128	0	42.96	52.34
STRCP1	15	43892822	c.4903G>T	p.V1635F	Misssense	C	A	72	118	78	2.78	8.47	8.97
KRT14	17	39742894	c.193C>T	p.L65L	Silent	G	A	16	28	27	0	28.57	33.33
KRT14	17	39742898	c.189C>T	p.C63C	Silent	G	A	15	26	27	0	26.92	37.04
DIRAS1	19	2717179	Non-coding	r.784a>c	3'UTR	T	G	54	70	41	3.7	10	21.95
PSG8	19	43258444	Non-coding	r.1284g>a	3' UTR	C	T	391	378	361	0.26	36.51	40.17
ZNF41	X	47308797	c.372C>T	p.P124P	Silent	G	A	72	72	80	0	30.56	40
ZXDB	X	57618870	c.389G>A	p.G130D	Misssense	G	A	17	25	36	0	20	19.44
THOC2	X	122754807	c.4226G>A	p.R1409H	Misssense	C	T	399	374	440	0	36.9	42.5

Table 3.1: SNV shared by both the CMML and AML samples as detected by Caveman Analysis. CHR = Chromosome, WT = wildtype and MT mutant allele

There were several recurrent leukaemia associated mutations that were identified in the AML sample and were not detected at the CMML stage on standard Caveman and Pindel analysis. The dinucleotide insertion in *CEBPA* is one such example. Although this was evident in occasional reads in the CMML sample when the data was reviewed in mpileup, this mutation represented a much higher proportion of reads in the AML sample. It is possible the acquisition of a bi-allelic *CEBPA* mutation within a sub-clone contributed to this increased VAF, however the degree of change indicates there was expansion of this *CEBPA* containing clone. Analysis of the AML sample also revealed an additional *NRAS* mutation, which was not detected at the CMML stage. The allelic frequency indicates this mutation was present within a small AML sub-clone.
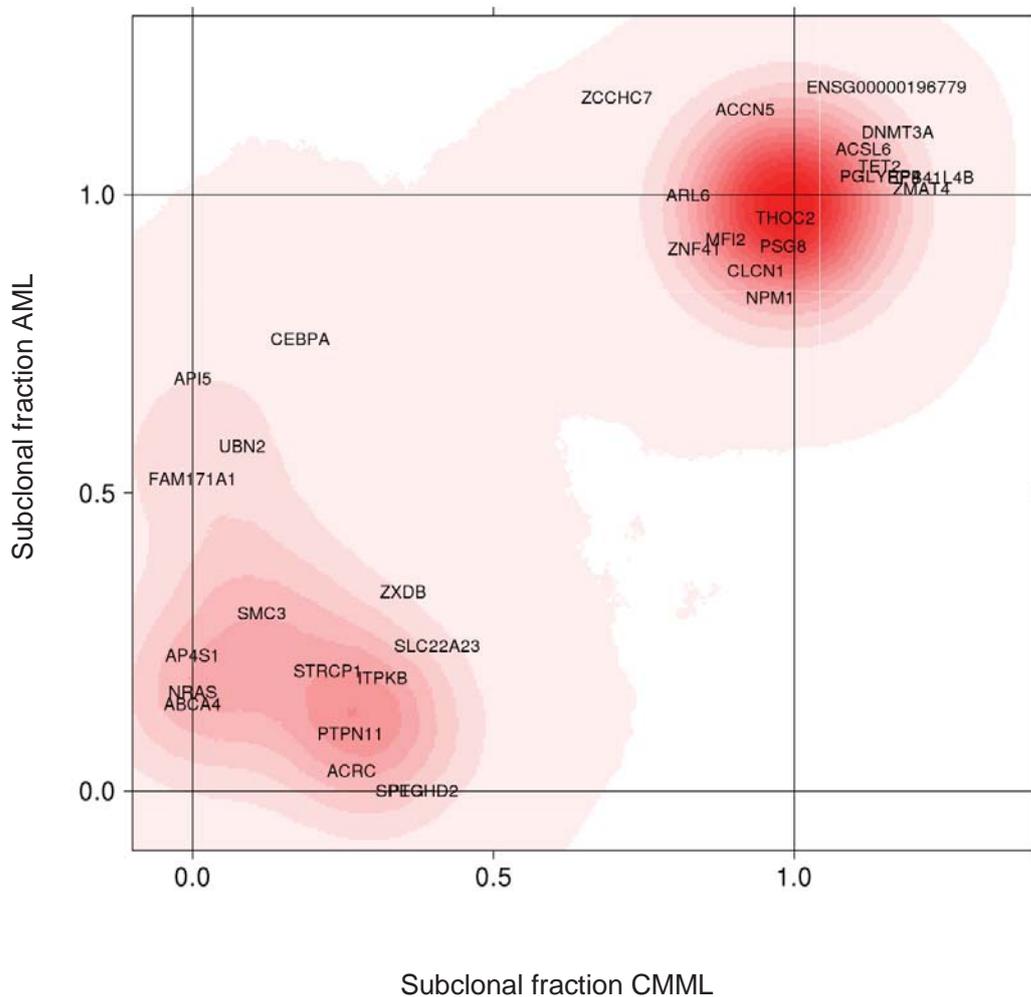
FIGURE 3.2: Two dimensional density plot showing the fraction of tumour cells carrying the mutations indicated, and their clustering. Increasing intensity of red indicates high posterior probability of a cluster. The *FLT3-ITD* mutation is not represented as it was not detected by Pindel or Caveman analysis. Manual annotation (Niccolo Bolli) was used to determine the subclonal fractions for *CEBPA* and *NPM1*.

Interestingly, the *FLT3-ITD* mutation which was picked up in the AML sample by PCR and agarose gel analysis in the diagnostic laboratory (figure 3.3) was not detected using exome sequencing and Pindel analysis. This is a recurrent problem with this specific type of mutation (Dr Eli Papaemmanuil, personal communication). Sanger sequencing was performed on DNA from the mutant band and identified the *FLT3-ITD* sequence as an 81bp duplication of 13:28608238-28608319. However, re-

analysis of the exome data performed by Dr Eli Papaemmanuil, specifically screening for this *FLT3-ITD* sequence, failed to detect mutant reads.

Copy number analysis (ASCAT) (Van Loo et al., 2010) was also performed based on the exome sequencing data. This showed sub-clonal copy-neutral loss of heterozygosity (LOH) in chromosome 1p and 13 in the AML sample, which would be consistent with acquired uniparental disomy in the *NRAS* and *FLT3* loci (figure 3.4).
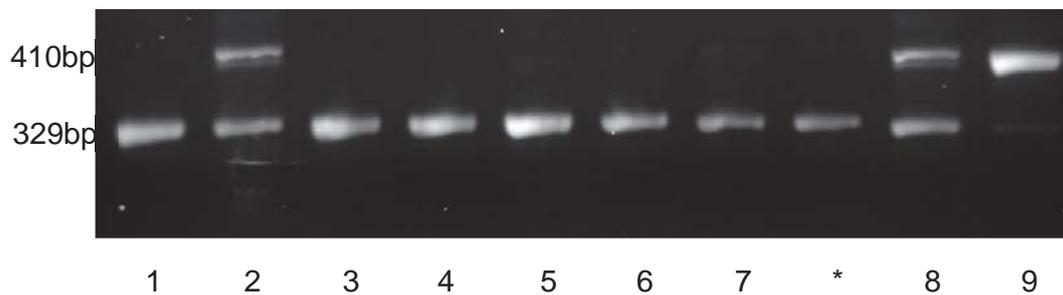

FIGURE 3.3: PCR and gel electrophoresis for the FLT3-ITD mutation performed at the diagnostic laboratory (Anthony Bench). Samples 1-9 correspond to the samples used for MiSeq analysis.
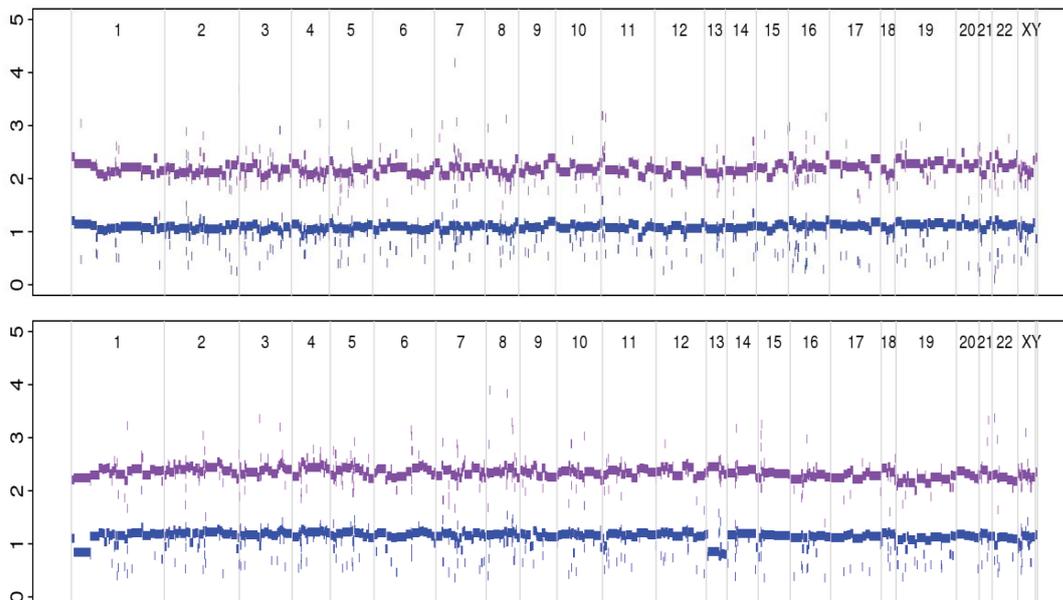

FIGURE 3.4: Copy number changes between CMML (top) and AML (bottom)
The CMML had normal karyotype but there was sub-clonal copy neutral LOH in chromosome 1p and 13 in the AML sample.

The eight presumed driver mutations that have been described already (figure 3.2) along with thirteen probable passenger mutations that clustered with them were validated using read counting of Illumina sequencing (MiSeq) of non-allele specific PCR products. This analysis was performed on nine blood and bone marrow samples that were collected through the clinical course (figure 3.1) including the remission and diagnostic samples used for exome sequencing, and a control DNA from a person with no diagnosis of haematological malignancy. After excluding samples with <1000 reads for a given gene, the remaining gene and sample combinations gave an average of 65 000 fold coverage at the mutant loci (median 51023). As the *ZXDB* locus amplified poorly and failed in the majority of samples, it was excluded from further analysis (figure 3.5).
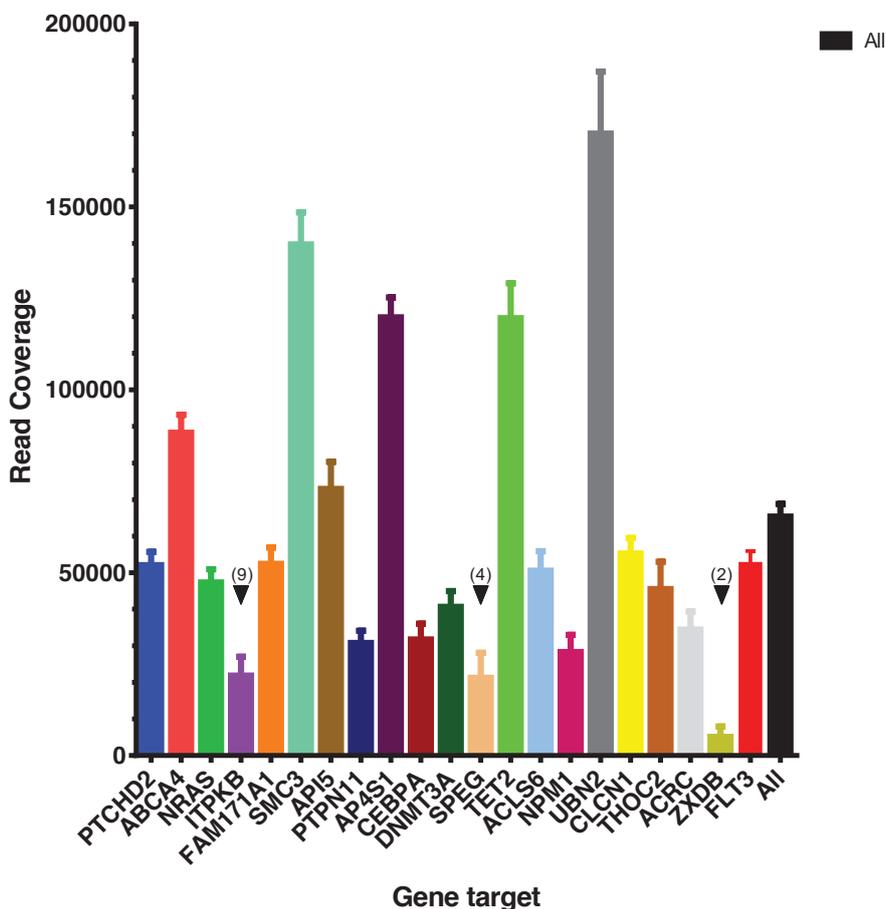

Figure 3.5: Coverage for the 10 samples across 21 target genes. The mean and standard error of the mean are shown for each site. Where a sample gave <1000 reads at a site this was excluded from the analysis. When less than10 samples were included the number of samples analysed are given in parenthesis.

Using these deep sequencing results we were able to track the allelic frequency through the serial samples. On Dirichlet analysis twelve clusters were identified of which six contained only a single mutation (figure 3.6). The findings for *CEBPA* in the CMML sample differed to the exome sequencing where only occasional *CEBPA* mutant reads were detected. Notably, the VAF of the *CEBPA* mutation rose above 50% in the last relapse sample. The VAF agreed closely with the exome data for all three samples studied and for all genes except *CEBPA* thus confirming the quantitative nature of the PCR method.

There were five mutations, *TET2*, *THOC2*, *DNMT3A*, *NPM1* and *CEBPA*, which had allelic frequencies of over 30% in both the CMML and AML samples and rose to a similar level in the last relapse sample, although these were clustered into five separate groups. The *TET2*, *THOC2* and *DNMT3A* mutations were detectable above baseline at low level in the day 175 peripheral blood sample; the time of relapse 1 (sample 5), whereas the *CEBPA* and *NPM1* mutations were not (table 3.2). *SMC3* and *UBN2*, which were grouped together on Dirichilet analysis, were also increased at relapse 1 but unlike *TET2*, *DNMT3A* and *THOC2* these were not detected in the later relapse samples. Both of these mutations were detected at low level in the CMML sample and were present in around 30% of cells in the initial AML sample (sample 2). The allelic frequencies of *TET2*, *DNMT3A, THOC2*, *SMC3* and *UBN2* in the relapse 1 sample (sample 5) were low (around 1.5% in most) and comparable to the level seen in the initial remission sample used for exome sequencing (sample 3) for *TET2* and *DNMT3A*. However, this represented a 3-12 fold increase and an absolute rise in mutant read number of several hundred compared to the mean coverage in the deep remission samples (sample 4 and 6) and control sample for each of these genes. Such changes were not seen in sample 5 for *CEBPA*, *NPM1*, *NRAS* or *FLT3*.

Sample	TET2 Number	TET2 Proportion (%)	THOC2 Number	THOC2 Proportion (%)	DNMT3A Number	DNMT3A Proportion (%)	SMC3 Number	SMC3 Proportion (%)	UBN2 Number	UBN2 Proportion (%)	CEBPA Number	CEBPA Proportion (%)	NPM1 Number	NPM1 Proportion (%)	NRAS Number	NRAS Proportion (%)	FLT3 Number	FLT3 Proportion (%)
1	64014	40.35	12352	40.93	27901	44.37	4148	4.11	8013	4.36	8466	34.16	17210	43.11	848	1.55	12	0.07
2	53654	46.58	21338	43.98	16895	50.57	22440	14.39	28614	13.77	10313	35.50	5635	44.42	5325	12.13	5400	23.81
3	1278	1.19	329	0.31	684	2.11	290	0.27	366	0.17	2	0.01	0	0.00	73	0.20	10	0.03
4	611	0.69	105	0.33	53	0.24	298	0.24	224	0.19	2	0.01	72	0.31	113	0.20	7	0.04
5	2717	1.79	2163	4.63	987	1.79	2351	1.51	2440	1.49	16	0.03	0	0.00	69	0.16	3	0.01
6	250	0.28	127	0.51	65	0.25	265	0.20	285	0.11	3	0.02	-	0.00	51	0.16	6	0.03
7	5041	4.06	4715	8.01	1817	4.69	284	0.21	95	0.07	1819	6.31	1045	3.78	63	0.19	1008	4.40
8	32897	19.93	10229	26.06	8710	21.46	410	0.24	88	0.05	10062	24.50	3126	16.79	86	0.15	8148	24.25
9	53843	47.10	20911	47.02	17428	45.31	401	0.21	217	0.14	20764	56.57	9832	46.09	111	0.23	16832	86.75
Control	473	0.57	397	1.58	123	0.22	256	0.20	52	0.06	11	0.02	0	0.00	123	0.19	10	0.03

Table 3.2: The absolute number and proportion of reads assigned to the variant allele for nine of the mutations. The first relapse sample (sample 5) is shaded.

Although *NPM1* and *CEBPA* were clustered separately on the Dirichilet analysis, they were both detected in the CMML, AML and late relapse samples but not in the initial relapse (sample 5). Cluster 9, which contained *CLCN1* and *ACLS6* had a similar pattern of occurrence to *NPM1* and *CEBPA*, but were present at roughly half the allelic frequency. *FLT3-ITD* was also clustered separately on Dirichilet analysis. It had an allelic frequency of 24% in the AML diagnosis sample, was not detected in the initial relapse and had an allelic frequency of 87% in the final sample, which suggests the majority of cells had bi-allelic mutations at that time. *NRAS* clustered with *ABCA4* and *AP4S1* but is in a separate clone to the other driver mutations. It was present in the initial AML sample, but not in either of the relapses.
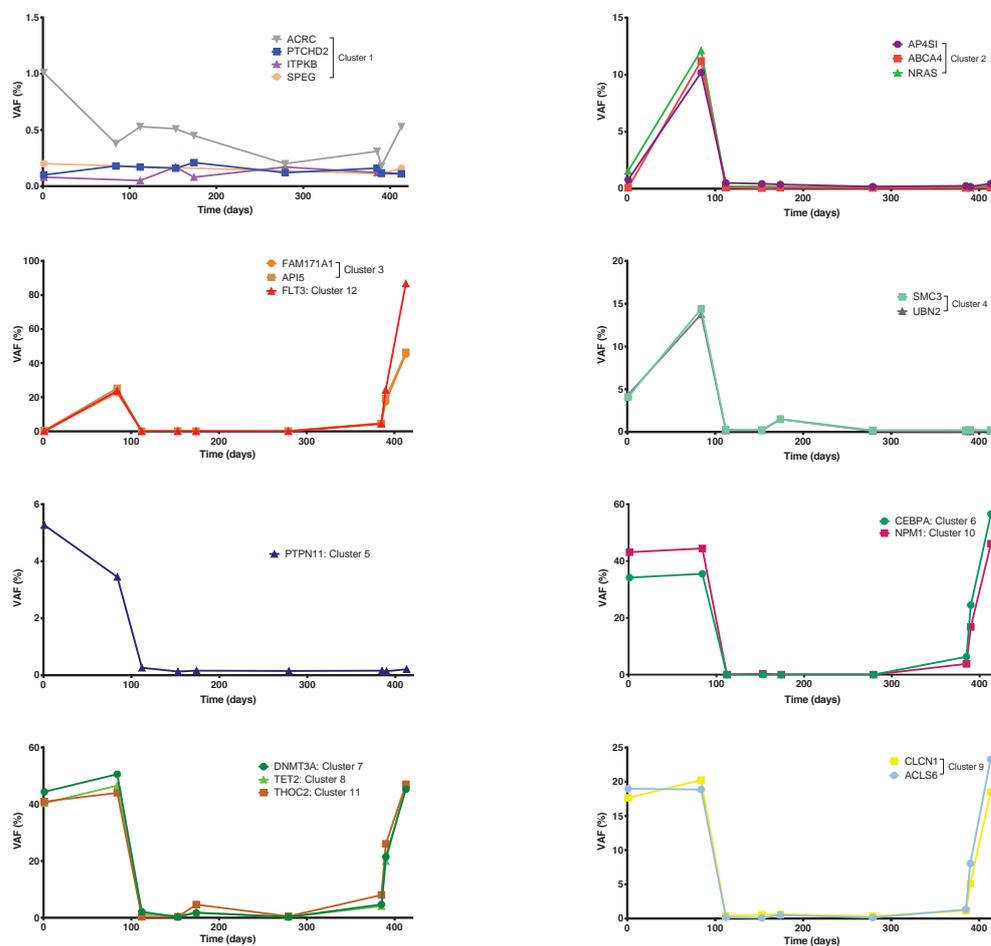

Figure 3.6: Variant allele frequencies (VAF) for the various mutations over time. The cluster to which the mutation was assigned on the Dirichilet analysis is also shown.

3.4 Discussion

This case is striking for the sheer number of AML associated mutations which were identified in the CMML genome. The detection of a combination of *TET2*, *DNMT3A* and *NPM1* mutations in the predominant clone at the time of CMML was surprising as one might have expected these three powerful mutations to cause full blown AML. It has been suggested that as few as two driver mutations may be sufficient to generate leukaemia (Welch et al., 2012). However, it also appears that rather than occurring due to the simple expansion of this CMML clone, the clinical progression to AML was driven by the acquisition of new mutations. The *FLT3-ITD* containing clone, or a sub-clone that evolved from it, also led to the fatal relapse following bone marrow transplantation. The evolution of pre-leukaemia clones to AML is thought to be a stochastic process. However, in this case with multiple pre-existing mutations the independent acquisition of *FLT3-ITD* and *NRAS* in separate clones in such a short time suggests that the evolution to AML was almost inevitable, akin to a deterministic process.

The combination of whole exome sequencing to identify disease specific mutations, with a targeted gene re-sequencing approach across multiple serial samples has proven effective in deciphering the branching clonal evolution of this individual leukaemia. In AML disease relapse typically arises from a pre-existing clone (Ding et al., 2012). In this study we used a targeted sequencing approach on the relapse samples and therefore we could not identify new mutations specific to the relapse. However, the pattern of allele frequencies for the various mutations during the disease course provided evidence for which mutations were co-occurring within a single disease clone and the order of acquisition of mutations (figure 3.7).
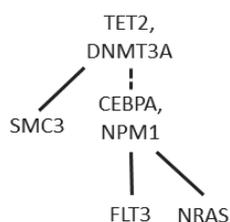

Figure 3.7: Phylogram showing the order of acquisition of the driver mutations. *TET2*, *DNMT3A*, *CEBPA* and *NPM1* are all in the major clone. The distinction between *TET2/DNMT3A* and *CEBPA/NPM1* is based on their presence in the early relapse. The differing patterns of recurrence of *SMC3*, *FLT3* and *NRAS* indicate these were likely to be in separate sub-clones.

Non-allele specific PCR and deep sequencing is a powerful approach for analysis of specific mutations at high read depth. Although whole genome and whole exome sequencing approaches have provided unique insights into the heterogeneity of mutations and sub-clonal architecture of diseases such as AML, these approaches are limited in their ability to detect mutations occurring in minor sub-clones with VAFs of less than 10%(TCGA_Research_Network, 2013). Although the proportion of mutant reads detected in the initial relapse sample from this patient were in the order of 1.5-4.6%, this represented a clear increase above baseline at these sites (table 3.2).

The pattern of mutations detected in the first relapse sample suggests that the *TET2* and *DNMT3A* mutations were acquired before the *NPM1* and *CEBPA* mutations in clonal evolution. This concurs with recent evidence that the former are early events (Busque et al., 2012; Jan et al., 2012; Shlush et al., 2014). Nevertheless, the mutant VAFs were closely concordant at CMML and AML diagnosis and in the second relapse. The first relapse sample was a peripheral blood sample and the detected mutations were at low VAF despite the fact there were 38% blasts reported on a bone marrow taken at the time. Unfortunately we were unable to obtain any DNA from the marrow sample. One potential criticism of the targeted re-sequencing approach is that PCR may introduce bias to the VAF and it is possible the *NPM1* and *CEBPA* mutations were only absent due to reduced sensitivity for detection compared to the other mutations, hence the dotted line in figure 3.7. PCR bias is unlikely to be a problem for SNVs, and the mutated base was positioned in the middle of the PCR product where possible. Although PCR bias is potentially more of an issue with insertions and deletions both the *NPM1* mutation (a tetra-nucleotide repeat) and the *CEBPA* insertion (two base pairs) are very short. It is unlikely that such small changes introduce a significant bias to the PCR reaction and with such deep coverage we would expect to find some evidence of the mutation at the initial relapse. There were over 50 000 MiSeq reads for each of these genes at this time-point, so for a heterozygous variant in 3% of cells there should be 750 mutant reads. Even if the PCR favoured the wild-type allele 10:1, 75 mutant reads would be detectable with this depth of read coverage. Furthermore, both mutations were clearly evident in the seventh sample, with VAF of 3.7% and 6.3% respectively. These results are in line with the VAF of the *TET2* (4.1%) and *DNMT3A* (4.6%)

mutations in this blood sample, when all of these mutations are present in the relapsing clone.

The data suggests that the two AML relapses occurred from different sub-clones, the first of which contained *SMC3* but not the second. The absence of *SMC3* in relapse 2 indicates this mutation was in a branching sub-clone from the one that progressed causing fatal disease, as was the *NRAS* mutation seen at diagnosis (sample 2). The final relapse sample contained *TET2*, *THOC2*, *DNMT3A*, *NPM1*, *CEBPA*, *FAM171A1*, *API5* and *FLT3-ITD* mutations at a VAF that implies they were present in close to 100% of nucleated cells in the peripheral blood.

Several of the mutations that were targeted for re-sequencing were thought to be passenger mutations. These were selected as they appeared to track with driver mutations based on their VAF in the exome sequencing. Including these mutations improved the confidence with which we could track sub-clones, nevertheless some of these, such as THOC2 were grouped separately in the MiSeq analysis. Most of the mutations with a sub-clonal fraction of close to one on the exome sequencing were clustered separately in the Dirichilet analysis of the MiSeq data. This is likely to be a consequence of the deeper read coverage and does not imply these are in separate clones.

The *FLT3-ITD* was not detected on exome sequencing, but on PCR re-sequencing had an allelic frequency of 24% in the AML diagnosis sample. The copy number analysis on the AML exome sample also suggests LOH at the *FLT3* locus. The *FLT3-ITD* sequence could not be detected in the exome sequencing output despite specifically searching for this and other *FLT3-ITD* mutations in the raw data. A possible explanation is that the 81bp duplication affected the 'pull down' of the mutant allele during hybridisation to the RNA baits, therefore selecting against the mutant allele in the exome sequencing protocol. In the MiSeq analysis it is also possible that the PCR performed ahead of re-sequencing introduced an amplification bias. However, given the larger size of the mutant PCR product any such bias is more likely to favour the wild-type allele and the mutant allele frequency of 87% in the final relapse sample suggests that this mutation is not strongly selected against in the MiSeq library preparation. Furthermore, the MiSeq data correlates well with the PCR gel electrophoresis which was performed by the diagnostic laboratory.

DNMT3A and *NPM1* mutations are among the commonest mutations in de novo AML and frequently co-occur with each other and with *FLT3* mutations within the same tumour(TCGA_Research_Network, 2013). In fact this combination of mutations was reported in the first published case of AML analysed by whole genome sequencing(Ley et al., 2010; Ley et al., 2008). The *NPM1* mutation identified in this case is the commonest, type A mutation which affects the critical 288 and 290 tryptophan residues disrupting the nucleolar localisation signal (Falini et al., 2005) as well as introducing a new nuclear export signal(2005; Falini et al., 2006). *DNMT3A* encodes one of a group of DNA methyltransferases, which catalyse the addition of a methyl group to cytosine residues of CpG dinucleotides. Increased methylation of CpG islands is typically associated with reduced expression of downstream genes(Ley et al., 2010). $DNMT3A^{R882C}$ is a missense mutation commonly found in AML and previously described in CMML-derived AML(Jankowska et al., 2011; Ley et al., 2010).

TET2 mutations also have an effect on epigenetic regulation. *TET2* catalyses the conversion of 5-methylcytosine to 5-hydroxymethylcytosine (5hmC) and therefore plays are role in DNA demethylation (Figueroa et al., 2010; Ko et al., 2010). *TET2* mutant samples display low levels of 5hmC (Ko et al., 2010). *TET2* mutations are one of the commonest lesions in CMML (Jankowska et al., 2011; Meggendorfer et al., 2012), are reported in 7-23% of de novo AML cases and seem to be more prevalent in elderly patients (Gaidzik et al., 2012; Metzeler et al., 2011; TCGA_Research_Network, 2013). The particular nonsense *TET2* mutation found in this case was in exon 11, and to our knowledge it has not been previously reported. However, exon 11 encodes the alpha ketoglutarate binding domain and is one of the most frequently mutated *TET2* exons in AML(Gaidzik et al., 2012). As opposed to *TET2* and *IDH1/2* mutations which are mutually exclusive in AML(Figueroa et al., 2010), *TET2* and DNMT3A are known to co-occur(TCGA_Research_Network, 2013). In fact the combination of *NPM1*, *DNMT3A*, *TET2* and *FLT3* mutations was reported in one case in the recent study of 200 AML patients using exome and whole genome sequencing(TCGA_Research_Network, 2013).

Only about 2-4% of *NPM1* mutated AML cases also carry a *CEBPA* mutation and over 90% of these are single *CEBPA* mutations (Dufour et al., 2010; Green et al., 2010; Taskesen et al., 2011; TCGA_Research_Network, 2013). In contrast across all non-

acute promyelocytic leukaemia (APL) cases of AML around 7% of patients are found to have *CEBPA* mutations of which over half are double mutations (Dufour et al., 2010; Green et al., 2010). It appears that the good prognostic impact of *CEBPA* mutations are limited to this double mutant group, which has distinct molecular characteristics (Green et al., 2010; Taskesen et al., 2011; Wouters et al., 2009). Surprisingly, in this patient with a *NPM1* mutation the VAF of the CEBPA mutation is 57% in the final relapse sample. This suggests either our patient has acquired a bi-allelic *CEBPA* mutation or lost her wildtype allele or that there is a bias for this variant in the MiSeq PCR library preparation.

The mutations in *NRAS* and *SMC3* found in this patient are probable driver lesions, even though they were not found in the major disease clone at AML diagnosis or in the final relapse. The $NRAS^{G12D}$ mutation is frequently described in human cancers including AML and has been shown to co-operate with other mutations in mice to induce AML(Li et al., 2011; Ward et al., 2012). Cohesin complex genes, including *SMC3* are mutated in 6-13% of cytogenetically normal AML(Kon et al., 2013; TCGA_Research_Network, 2013) and in 10% of cases of CMML. In AML they frequently co-exist with *NPM1* mutations (Ding et al., 2012; TCGA_Research_Network, 2013; Thol et al., 2013) and across all myeloid malignancies they are commonly found in association with mutations in *TET2*, *ASXL1* and *EZH2* (Kon et al., 2013). The cohesin complex mutations result in reduced amounts of chromatin bound cohesin components and are thought to have global effects on gene expression (Kon et al., 2013). The missense mutation of $SMC3^{L866P}$ found in this case is not reported in the COSMIC database. However, to date no particular mutation hotspot has been identified in *SMC3* in AML or other myeloid malignancies and the majority of the reported mutations are of the missense type as found in this case (Cosmic Database).

Although mutations in *PTPN11* which encodes the protein tyrosine phosphatase SHP-2 are widely reported in AML and other myeloid malignancies (Loh et al., 2004; Tartaglia et al., 2003) the 'driver' credentials of the particular mutation found in this case are less clear. The $PTPN11^{N308D}$ mutation was detected in the CMML sample with a VAF around 5%, had a lower allelic ratio in the AML sample and was not detectable above baseline in the remaining samples. The missense mutation $PTPN11^{N308D}$ is described as a germ-line mutation in Noonan Syndrome, but has not

been associated with AML(Tartaglia et al., 2003). Phosphatase assays using wild type and mutant SHP-2 proteins have shown the phosphatase activity of the *N308D* mutant is less than JMML associated SHP-2 mutants but greater than the wild-type protein. In *in vitro* assays, proliferation in cells transiently expressing the *N308D* mutation was also intermediate between normal cells and those with a JMML associated mutation. The prevalence of *PTPN11* mutations is lower in adult than in childhood AML (Tartaglia and Gelb, 2005) and whether this particular somatic mutation is a true driver in the context of adult AML is uncertain.

Regardless of the veracity of the *PTPN11* mutation as a driver, this case clearly demonstrates the molecular complexity of AML and shows a branching clonal evolution. Although the dominant clone at the time of CMML diagnosis already contained co-occurring *TET2*, *DNMT3A* and *NPM1* mutations, progression to frank AML was associated with the clear acquisition of new driver mutations. Three distinct sub-clones carrying unique driver mutations were detected in the AML sample. The initial relapse appears to have developed from a clone carrying *SMC3* along with the *TET2* and *DNMT3A* mutations. The ultimate relapse leading to death was from the re-emergence of a clone containing at least five driver mutations; *FLT3-ITD*, *CEBPA*, *NPM1*, *TET2* and *DNMT3A* but not *SMC3* or *NRAS*. This case highlights that clonal evolution is a dynamic process and forces us to question if the progression to AML is inevitable in the setting of a CMML clone with a high mutational burden.