AML1-ETO controls Genes involved in Apoptosis, Haemopoietic Differentiation and Histone Modifications

KIAA0125 is the most statistically significant gene and is down regulated although its function is unknown. It resides in the middle of the immunoglobulin heavy chain gene region. *C9orf89* is the second most statistically significant gene and is down regulated in this series. The gene contains a CARD domain that inhibits BCL10-mediated activation of NF-kappa B (Woo et al., 2004). The down regulation of this gene would lead to more BCL-10 related apoptosis. *SerpinB9*, an anti apoptosis gene and involved in the regulation of granzyme B, is up regulated. *GABARAPL1*, shown to increase apoptosis, has decreased expression. *GFI1B* is a transcriptional repressor that is essential for erythroid cells and megakaryocytes, but is also expressed in hematopoietic stem cells and early myeloid progenitors. It is usually up-regulated in AML but is down regulated in this series suggesting that t(8;21) does not necessarily need to work through this pathway to establish leukaemia (Vassen et al., 2009). *NFE2,* down regulated in t(8;21) patients, encodes for a transcription factor with an important role in erythroid and megakaryocyte differentiation being regulated by the ligase Itch (Lee et al., 2008). TBL1X is part of the SMRT complex and also binds Wnt target genes (Li and Wang, 2008). *TRERF1* encodes a zinc-finger transcriptional regulating protein that interacts with CBP/p300. Both these genes are up regulated in t(8;21) and could play a key role in the regulation of histone acetylation. Histone transcripts, as they lack poly-A tails, are not usually well represented on standard array platforms but *HIST1H2BM* is up regulated using the exon array platform thus highlighting the extra information that exon arrays provide.

UHRF1 & CD82 may be Key Genes Regulated by AML1-ETO

UHRF1 is up regulated in many carcinomas including breast and bladder cancer and its expression is regulated by E2F-1. Although initially discovered through binding topoisomerase gene promoter it appears to have a greater affinity for methylated CpG islands of various promoters. It is able to recruit HDAC1, suggesting that it might suppress expression of tumour suppressor genes through histone deacetylation (Unoki et al., 2004). Furthermore, it is also able to recruit DNMT1 to hemimethylated sites, allowing DNA methylation. Thus, UHRF1 appears to be a key molecule regulating cross-talk between DNA methylation and histone modifications (Hashimoto et al., 2009) (Figure 24).

CD82 encodes for a tetraspanin and was identified as a tumour metastasis suppressor gene in many types of cancer. Tetraspanins regulate migration, adhesion and proliferation at the cellular level. *CD82* is often down regulated in many human malignancies but its expression has been noted to be up regulated in leukaemia cells. Interestingly it is down regulated in t(8;21) samples. Furthermore, CD82 appears to work through Rho GTPases (Liu and Zhang, 2006). *CD82* can be activated by p53 and its expression and that of p53 are strongly correlated (Smith and Theodorescu, 2009) (Figure 25).
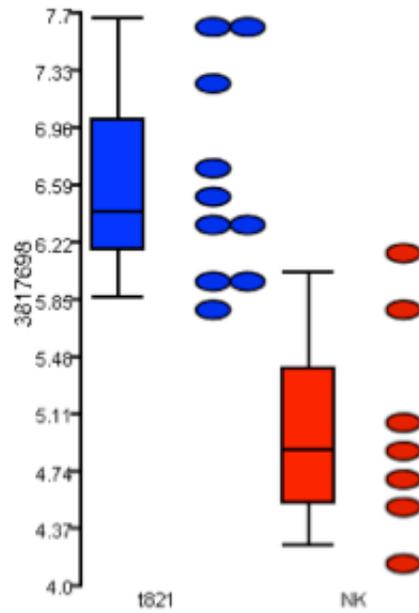
Figure 24 Box plot of expression signals for *UHRF* from t(8;21) patients in blue and NK patients in red.
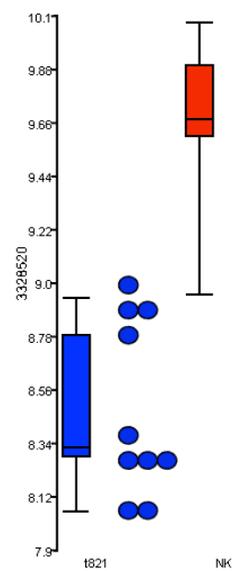

Figure 25 Box plot of expression signals for *CD82* from t(8;21) patients in blue and NK patients in red.

Further Comparisons & Validation of Exon Array Data

The comparison of the ABI and Human Exon expression arrays had 58% similarity in the gene lists generated. Some of these differences, as discussed, maybe related to the experimental design whilst others to non-overlapping annotation files. Nevertheless, the high concordance rate between the two lists appears to justify the use of the exon array for use in gene expression profiling.

To further validate the exon array platform, gene lists generated on the exon array platform were compared to additional gene lists derived from similar experiments conducted on standard expression arrays in both primary samples and cell lines. Data sets from primary AML samples collected from this laboratory using the ABI system, Kasumi cell line data from Illumina (Kasumi I), Affymetrix U133 plus 2.0 array (Kasumi A) and Exon array (Kasumi E) platforms and a data set from stem cells from Gene Expression Omnibus (Krejci) were used for comparisons with the exon array data (Table 14). This analysis described in detail below validates the platform but also suggests that investigation using the Kasumi cell line does not result in robust data.

Table 14 Summary of the comparison of the main features from gene expression experiments performed: Expression data derived from the ABI platform (ABI); published Krejci reference (Krejci); Kasumi cell line performed on Affymetrix platform (KASUMI A); Kasumi cell line performed on Illumina platform (KASUMI I) and Kasumi cell line on Exon array platform (KASUMI E) respectively. (see text below for further details)

	ABI	KREJCI	KASUMI A	KASUMI I	KASUMI E
SAMPLE	Patients	Stem Cells	Cell lines	Cell lines	Cell lines
NUMBERS	9 vs 7	9 vs 3	1 vs 1	unknown	1 vs 1
PLATFORM	ABI	Affy 3'	Affy 3'	Illumina	Exon
ANALYSIS	Partek	Partek	Partek	unknown	Partek

Using the same ABI data set and the same analysis methodology as described above a further analysis was performed to more closely mimic the exon array experiment. The previously identified 9 t(8;21) primary samples were compared to 7 M2 normal karyotype primary samples from this data set. This generated only 46 differentially expressed probesets of which 46% (21) were up regulated and 54% (25) were down regulated. Only 33 probesets were annotated making meaningful comparison difficult. However, comparison with the exon data (223 genes) showed 19 (58%) concordant genes. Of specific interest, a total of 12 HOX genes were present in the exon array data set of which 4 HOX genes were also present in the ABI gene list. A further 2 HOX genes were present in the ABI data set alone.

Data sets from the gene expression publication by Krejci (Krejci et al., 2008) were obtained from the Gene Expression Omnibus. In this study an AML1-ETO vector was introduced into stem cells and the expression profiles compared to control stem cells with mock transfections. A total of 9 sets of CD34+ cells treated with AML1-ETO vector and 3 controls were obtained. The .cel files were imported into the Partek software and gene lists were constructed using a FDR of $p < 0.05$ and fold change +2 to -2. 5638 probesets passed the above criteria. Of these 2047 (36%) were up-regulated and 3591 (64%) were down regulated. Of the 5638 probesets, 4532 had annotations. Comparison with exon data (229 probesets and 223 genes) showed 120 probesets representing 90 (40%) genes were concordant.

Expression profiling of Kasumi cell line correlates poorly with primary samples

Kasumi Cell line data was available form three sources; previously published gene expression profiling data from our laboratory using the Illumina platform (Dunne et al., 2006); previously performed standard 3' Affymetrix profiling from this laboratory by J. Dunne (unpublished) and finally Kasumi cell line experiments performed by the author on the exon array platform. In all these experiments Kasumi cells had the AML1-ETO gene knocked down with si-RNA and were compared to Kasumi cell with mock transfection (Table 15). [All knockdown experiments were performed by J. Dunne]

The published gene list generated from Illumina platform data was used for comparison (Dunne et al., 2006). Data from the Illumina platform generated 294

probesets of which 133 were up regulated and 161 down regulated. Of the 294 probesets 223 genes were annotated. Comparison with the patient exon data showed that 29 (12%) genes were present in both lists.

Gene lists from the Affymetrix data were generated using Partek with parameters of p<0.05 and FC +/-2. This experiment generated a 125 probesets of which 52 (42%) were up regulated and 73 (58%) down regulated. Of the 125 probesets, 122 had annotations. Comparison to the patient exon data showed that 14 (11%) genes were present in both data sets.

Data from the exon arrays for the Kasumi experiments using a FC criteria of +/- 2 and p <0.05 revealed 103 probesets differentially expressed with 67 down regulated and 36 up-regulated. Comparison of this Kasumi exon data to the patient exon data demonstrated only one common gene.

The expression analysis from Kasumi cell line studies appears to be poorly reproducible. They do not appear to correlate well with primary samples results. However the number of replicates performed was limited and care needs to be taken with this interpretation.

Table 15 Summary of Comparisons between Exon Array and other experiments. Experiments: Exon1 – exon array with Fold Change +/-1; Exon 2 exon array with Fold Change+/-2; Kas A – Kasumi cell line using Affymetrix platform; Kas I- Kasumi cell line using Illumina platform; Kas E- Kasumi cell line using exon array platform.

	PROBESETS (PS)	UP REGULATED PS	DOWN REGULATED PS	GENES	CONCORD -ANCE
Exon 1	448	234	214	436	N/A
Exon 2	229	113	116	223	N/A
ABI	46	21	25	33	59%
Krejci	5638	2047	3591	4532	40%
Kas A	125	52	73	122	11%
Kas I	294	133	161	223	12%
Kas E	103	36	67	101	1%

Experimental Validation

RQ-PCR correlates with expression results

RQ-PCR on 8 genes was performed to experimentally validate the array profiles (See Chapter 2 p63). RNA from all primary t(8;21) and four NK samples were tested. (Figure 26)
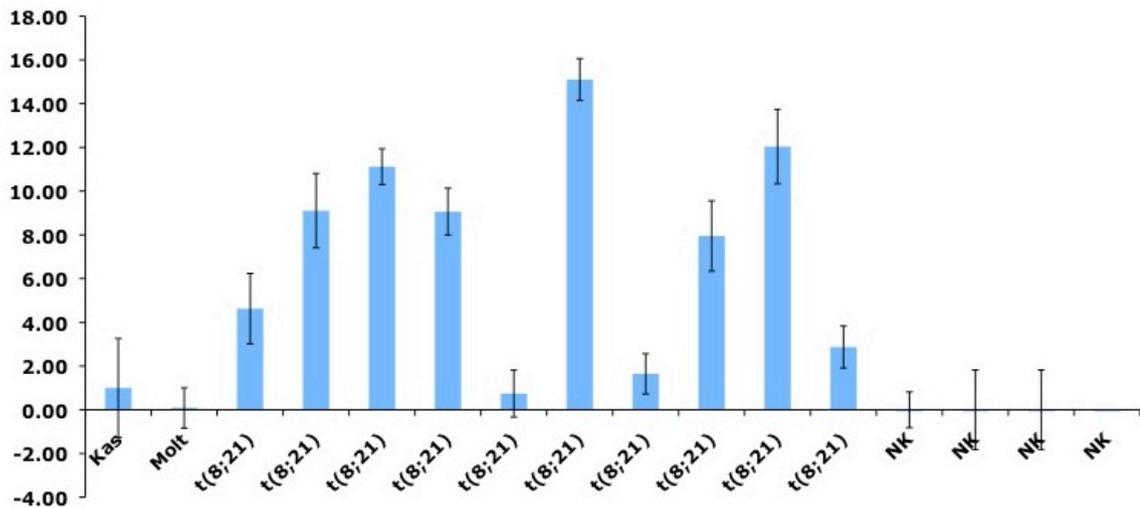
Figure 26 A representative sample of RQ PCR. Amplification of the BRN3a gene. The Kasumi cell line and Molt4 cell lines are used as positive and negative controls. Ten t(8;21) and four normal karyotpe (NK) patients have been tested. Tested samples have been plotted on the x-axis. The delta delta Ct relative to 18S relative to Kasumi is plotted on the y-axis. Experiments were tested in triplicate. Error bars are shown. Further details on methodology as described as in chapter 2 p63.

Values were compared to expression data values (Chapter 2 p63). Seven of the eight genes had a statistically significant correlation ($p<0.05$) and validated our experiment (Table 16).

Table 16 Correlation of RQ with expression for 8 genes.

Gene	Fold change	p-value	Correlation (r)	p-value
BRN3a	+17.1	6.23×10^{-07}	0.98	<0.0001
ROBO1	+15.0	1.37×10^{-08}	0.96	<0.0001
FGFR1	+2.96	1.46×10^{-08}	0.77	0.0027
GNAI1	+3.5	3.71×10^{-06}	0.83	0.0027
STK32B	+3.4	3.14×10^{-06}	0.87	0.0009
SIPA1L2	+7.3	2.24×10^{-07}	0.83	0.0008
GFI1B	-3.3	4.59×10^{-06}	0.54	0.0882
PRKCD	-2.8	1.13×10^{-05}	0.76	0.0041

Gene Ontology Functional Analysis

For a global assessment of the generated gene list, gene ontology functional analysis, using three sources of commercial software, GeneGo Inc., Ingenuity Pathway Analysis (IPA) from Ingenuity Systems[TM] and Gene Ontology (GO) was performed by collaborator P Chakraverty.

Using the IPA software, 448 genes derived from the exon array were interrogated for enrichment for processes. The top process was "cellular growth and proliferation" which validated this type of approach to interrogate our gene list (Table 17).

Table 17 Top process generated by IPA showing the network members and p value.

Process	p-value	Network members
Cellular Growth and Proliferation	1.32E-05	*TPM1, TRH, RECK, MBP, TBXA2R, ELA2, MEIS1, PALM, HOXA9, DEF6, VEGFA, ELF4, TGFB1, CAV1, ITGB4, TSPAN32, TNFRSF21, LST1, FGFR1, MPO, GFI1B, RUNX1T1, HSPG2 (includes EG:3339), NFE2, PRKCD, RCBTB1, CAT, VAV1, ITGAX*

Further detailed functional data mining was performed using the GeneGo software. The same gene list consisting of 448 genes was tested for enrichment of canonical pathways (Table 18) and processes (Table 19) using the GeneGo metacore tool.

Name	pValue	Network objects	Network Members
Development_Delta- and kappa-type opioid receptors signaling via beta-arrestin	3.99E-05	6/19	GNAI1, GNB5, GNG7, GRK6, PRKCD, DNM1
Transport_Alpha-2 adrenergic receptor regulation of ion channels	4.76E-05	7/28	GNAI1, GNB5, GNG7, PLCB2, PLCB3, ADRA2C, PRKCD
Development_EDG1 signaling pathway	3.42E-04	6/27	GNAI1, GNB5, GNG7, ADCY7, ADCY9, VAV1, PLCB2, PLCB3
Protein folding_Membrane trafficking and signal transduction of G-alpha (i) heterotrimeric G-protein	4.73E-04	5/19	GNAI1, GNB5, GNG7, PLCB2, PLCB3
Cell adhesion_Chemokines and adhesion	5.03E-04	11/93	GNAI1, GNB5, GNG7, CAV2, ITGB4, COL4A5, VEGFA, PLAU, CAV1, VCL, CTNNB1, VAV1, TRIO
Development_Alpha-2 adrenergic receptor activation of ERK	5.18E-04	7/40	GNAI1, GNB5, GNG7, PLCB2, PLCB3, ADRA2C, DAGLB
Development_Endothelin-1/EDNRA signaling	6.06E-04	7/41	GNAI1, ADCY7, ADCY9 PLCB3, MAP2K6, CTNNB1, CCND1
Cytoskeleton remodeling_Cytoskeleton remodeling	6.62E-04	11/96	PLAU, COL4A5, VEGFA, TGFB1, ITGB4, CAV1, VAV1, TRIO, CTNNB1, VCL, EIF4A1, SMAD3
Development_Mu-type opioid receptor signaling via Beta-arrestin	7.80E-04	5/21	GNAI1, GNB5, GNG7, DNM1, GRk6, GRK5
G-protein signaling_EDG5 signaling	7.80E-04	5/21	GNAI1, GNB5, GNG7, ADCY7, ADCY9, PLCB3
Chemotaxis_Lipoxin inhibitory action on fMLP-induced neutrophil chemotaxis	1.26E-03	6/34	GNAI1, GNB5, GNG7, LTB4R, VAV1, MAP2K6, COL4A5
Cell adhesion_Plasmin signaling	1.26E-03	6/34	PLAU, VEGFA, TGFB1, FGFR1, MAP2K6, COL4A5
Signal transduction_cAMP signaling	1.26E-03	6/34	GNAI1, GNB5, GNG7, ADCY7, ADCY9, PRKCD
Cytoskeleton remodeling_TGF, WNT and cytoskeletal remodeling	1.64E-03	11/107	VEGFA, PLAU, COL4A5, TGFB1, LRP5, CAV1, SMAD3, VAV1, CTNNB1, VCL, CCND1

Table 18 Enriched Pathways using GeneGO

G-Protein coupled receptors & cell adhesion pathways are enriched

Table 14 suggests there are two groups of molecules: those related to pathways through the use of G-protein coupled receptors and G-protein intermediaries including via beta-arrestins, EDG and MAPK pathways; and those involved in cell adhesion and cytoskeleton remodelling, including through WNT and TGFb pathways

Beta-arrestin-1 and beta-arrestin-2 act as scaffolds linking GPCR to a variety of signalling cascades. It was initially shown the beta2 adrenegic receptor induced activation of MAPK's through an Src/beta-arrestin complex. Subsequently other pathways requiring beta-arrestins for activation were identified and included the small GTPase RhoA, Ral GDP dissociation factor and the actin filament severing protein cofilin (Defea, 2008). EDG3, EDG5 and EDNRA are all G-protein coupled receptors (GPCR). EDG3 and EDG5 can regulate the activity of the Rho family GTPase Rac (Takuwa, 2002).

TGFb and the WNT pathways have been established to play a role in t(8;21) leukaemia; AML1 interacts with SMAD proteins and TGFb and AML1-ETO interacts with plakoglobin inducing the Wnt pathways (Pardali et al., 2000) (Muller-Tidow et al., 2004). The canonical Wnt pathways control the stability of b-catenin. Recent evidence shows evidence of Rho GTPase mediating in both the canonical and non-canonical Wnt pathways (Schlessinger et al., 2009). Similarly, non-Smad pathways, including the Rho-like GTPase signalling pathways, have also been implicated in TGFb signalling (Zhang, 2009).

Signal transduction & cell adhesion processes are enriched

The top two enriched processes identified by GeneGo are "cell adhesion" which is up regulated and "signal transduction" which is down regulated (Table 19). Network members involved in cell adhesion are displayed in Figure 27. (drawn by P Chakreverty using the GeneGo software)

Table 19 Top 2 processes using GeneGo and their network members

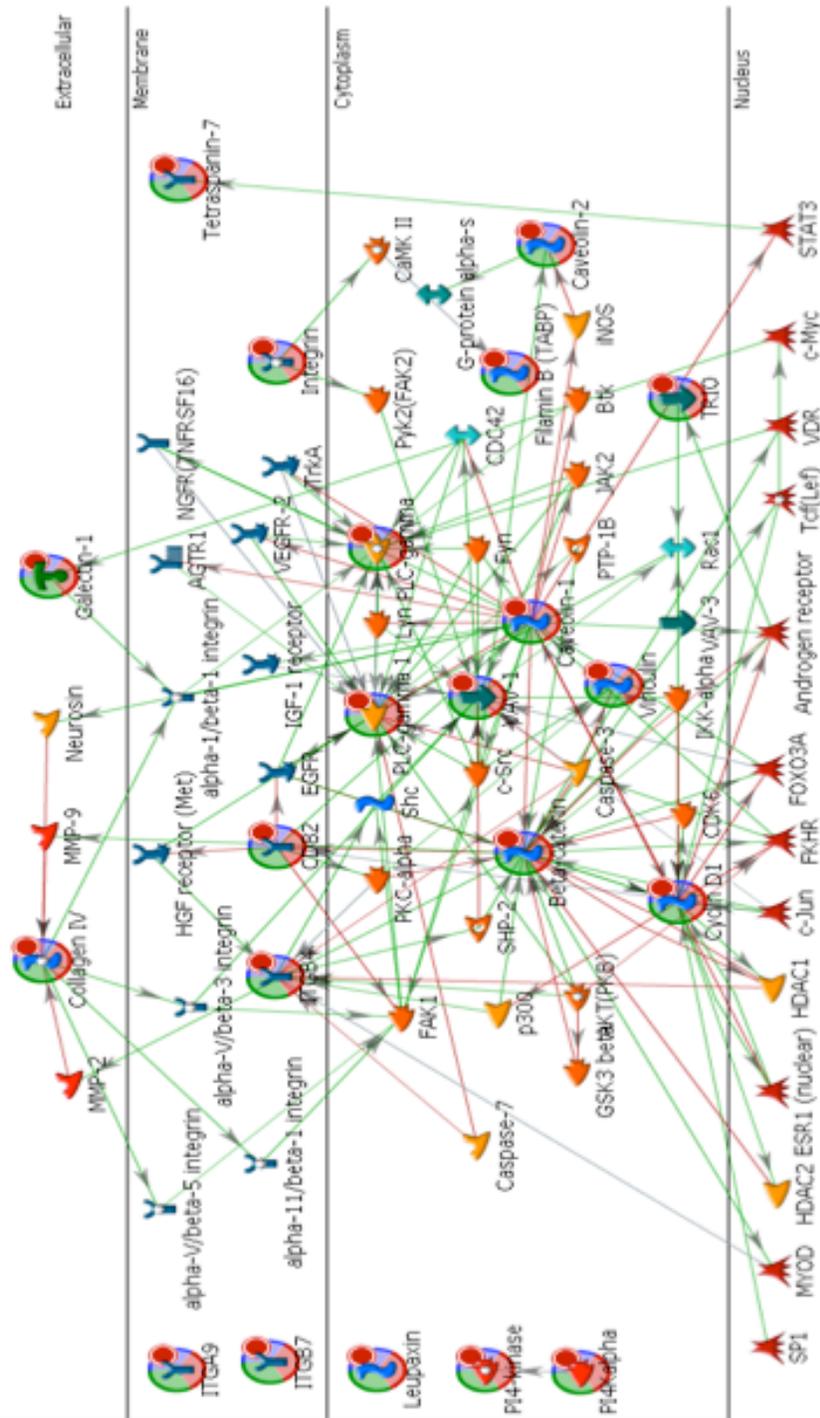| Cell Adhesion Molecules, integrin mediated cell-matrix adhesion (p-value 3.15E-06; network objects 21/210)	Signal Transduction (TGF beta & activin signalling) (p-value 6.31E-05; network objects 15/146)
Beta-catenin	C/EBPalpha
CD82	Cyclin D1
CAV-1	DOK1
CAV-2	Integrin
Collagen IV	MEK6
Cyclin D1	Mixl1
Filamin B (TABP)	PKC
Galectin-1	PLC-2
ITGA9	PLC-beta2
ITGB4	PLC-beta3
ITGB7	SMAD3
Leupaxin	Sno-N
PI4-kinase	TGF-beta
PI4Kalpha	VAV-1
PLC-gamma	KIT
PLC gamma1	
TRIO	
Tetraspanin	
VAV-1	
Integrin	

Figure 27 Interactions for molecules related to cell adhesion. Nodes in list are circled

G-Protein signal transduction pathways are enriched

To confirm the findings using the GeneGo software, a further functional data mining analysis was performed using the Gene Ontology (GO) software. The GO software is a multi-level hierarchy categorising functions into biological function, molecular process and cellular location. There was enrichment of genes involved in regulation of signal transduction. The top two enriched subgroups were noted to involve genes related to GTP protein signal transduction pathways thus, confirming the analysis using the GeneGO metacore tool. The genes involved in these subgroups and their interactions are demonstrated in Figure 28 (drawn by P. Chakraverty using the GeneGo software). These findings confirm the implications that G-protein coupled receptors play a key role t(8;21) leukaemogenesis. This observation may have implications in finding novel treatment targets and strategies.
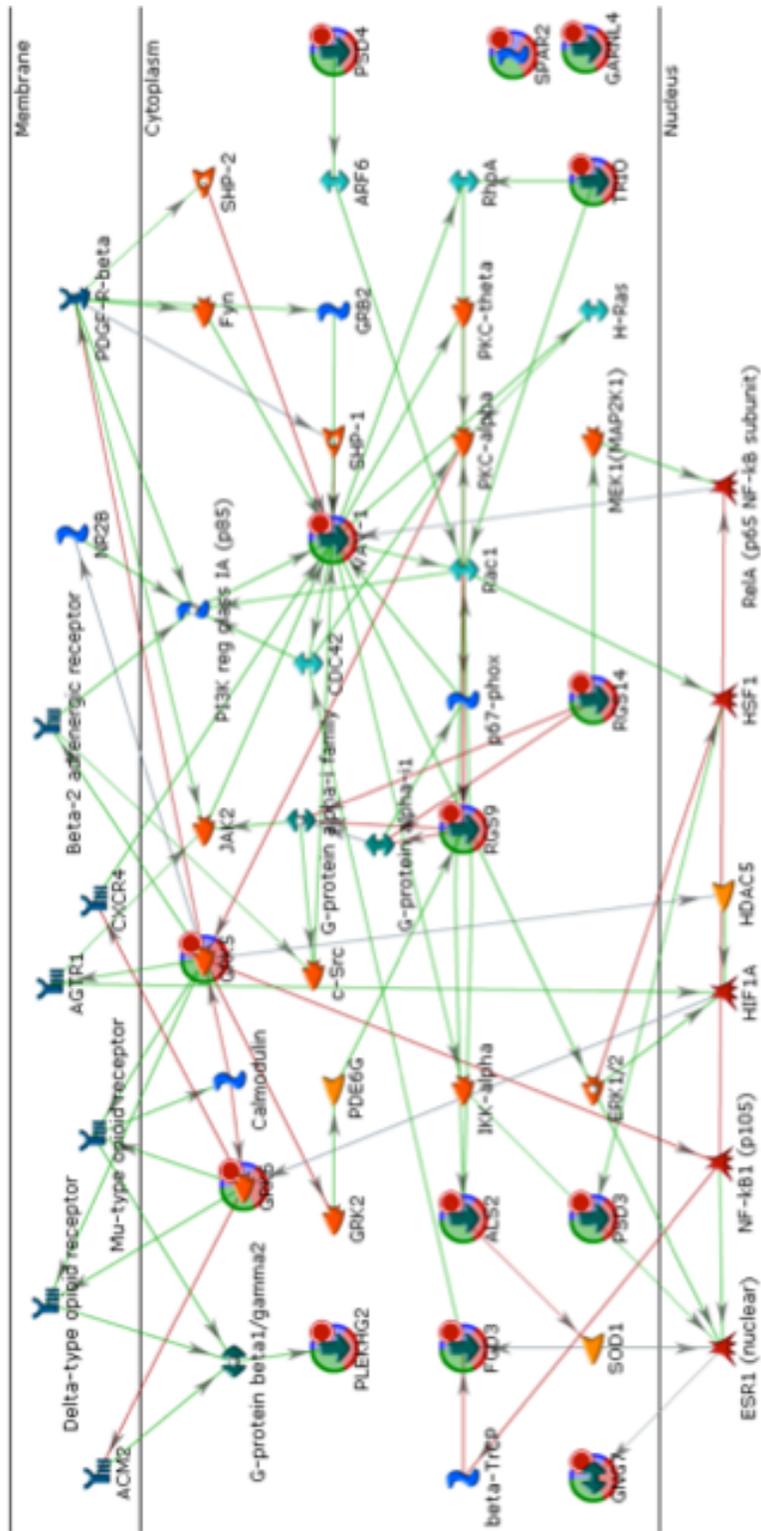
Figure 28 Interactions of molecules associated with G-proteins. Nodes in list are circled

Further Exon Array Analysis

In contrast to standard expression arrays, exon arrays in addition to global gene expression profiling can be analysed in a number of alternative ways. The design advantage enables interrogation for alternative exons and discovery of novel exons, some of which may represent miRs.

microRNA Discovery

The exon array has 190 probes to detect microRNAs as part of the extended metaprobeset file. On standard expression profiling using the extended and full metaprobeset files no microRNAs from the 190 interrogated on the array were noted to have differential expression. However, using these metaprobeset files, containing hypothetical exons, a number of probesets were detected which are differentially expressed but not part of any known annotated genes. These probesets may potentially represent important undiscovered genes and were therefore investigated as recently annotated microRNAs.

At the gene level using the full and extended metaprobeset files no "un-annotated" probesets corresponding to recently annotated microRNAs were discovered. However, at the exon level and using parameters of FDR $p<0.05$ and FC +/-1 a list of differentially expressed probesets was generated and inspection showed five probesets that are differentially expressed and part of microRNA genes. The five probesets identified are part of miR-155hg (BIC), miR-29c, miR22, miR-194-2 and

miR-223. The BIC gene is up regulated in this dataset while the other four are all down regulated.

Previous reports have shown silencing of miR-223 by AML1-ETO (Fazi et al., 2007). More recently both miR-29c and miR-223 were found to be down regulated in CLL and associated with a poor prognosis (Stamatopoulos et al., 2009). miR-29c targets include DNMT, which potentially can revert abnormal DNA methylation. miR-155 is up-regulated in a host of cancers including lymphomas and in bone marrow blasts of primary AML samples with FAB M4 and M5 subtypes. Furthermore, over expression of miR-155 in murine models resulted in a myeloproliferative disorder with preleukaemic aspects (O'Connell et al., 2008). Potential targets include CSF1R, Pu1 and CEBPA. As AML1-ETO also has direct interactions with these molecules it suggests a complex level of regulation.

These reports are consistent with our array data validating our approach. However, it is noted that although there is differential expression at the individual probeset level, at gene/transcript level the differences are not statistically significant (Table 20).

Table 20 Summary of details of probesets which show a greater than twofold differential expression and are part of known microRNAs

Probeset	Transcript ID	Gene	p-value	Fold Change
2453234	2453230	miR-29c	0.020	-2.85
3377262	3377258	miR-194-2	0.038	-2.93
3740673	3740664	miR-22	0.005	-2.37
3916516	3916508	miR-155hg	0.015	+2.42
3979749	3979745	miR-223	0.126	-2.04

Analysis by RQ PCR for miR-155, miR-29c and miR-223 on a series of primary samples has been previously performed in this laboratory and were used for validation (Dixon-McIver). Therefore, RQ values for 10 t(8;21) FAB M2 and 8 NK FAB M2 samples were selected and compared. There was no significant difference between the two groups and therefore did not validate the changes in microRNA expression between the two groups detected on the array.

Alternative Splicing

One of the aims for using the exon array platform was to discover genes whose splicing is regulated by AML1-ETO and to discover if the function of these genes is switched by alternative splicing. The hypothesis would be to discover genes involved in myeloid differentiation whose splicing is regulated by AML1-ETO and thus give a plausible rationale for the development of leukaemia.

Partek uses an alternative transcript algorithm from the exon level data to generate lists of differentially expressed probesets. Lists generated also contained a p-value score for the splicing index. Visual inspection of 500 genes with the lowest splicing index p-value was performed to detect genes whose exons were differentially expressed. Strikingly, two genes were noted to have a true differential exon usage unlike the other 498 genes, which only showed a degree of differential gene expression, but not differential individual exon usage. (Figures 29 & 30).

An RQ PCR strategy was used to attempt to confirm these changes. Primers on either side of the potentially alternatively spliced exon were designed. Primers between exons 6 & 7, 7 & 8 and 6 & 8 for the DHRS2 gene and between exons 18 & 19 and 18 & 20 were devised. cDNA from four t(8;21) and four NK primary samples were tested. The above observations of alternative splicing could not be confirmed by this method.
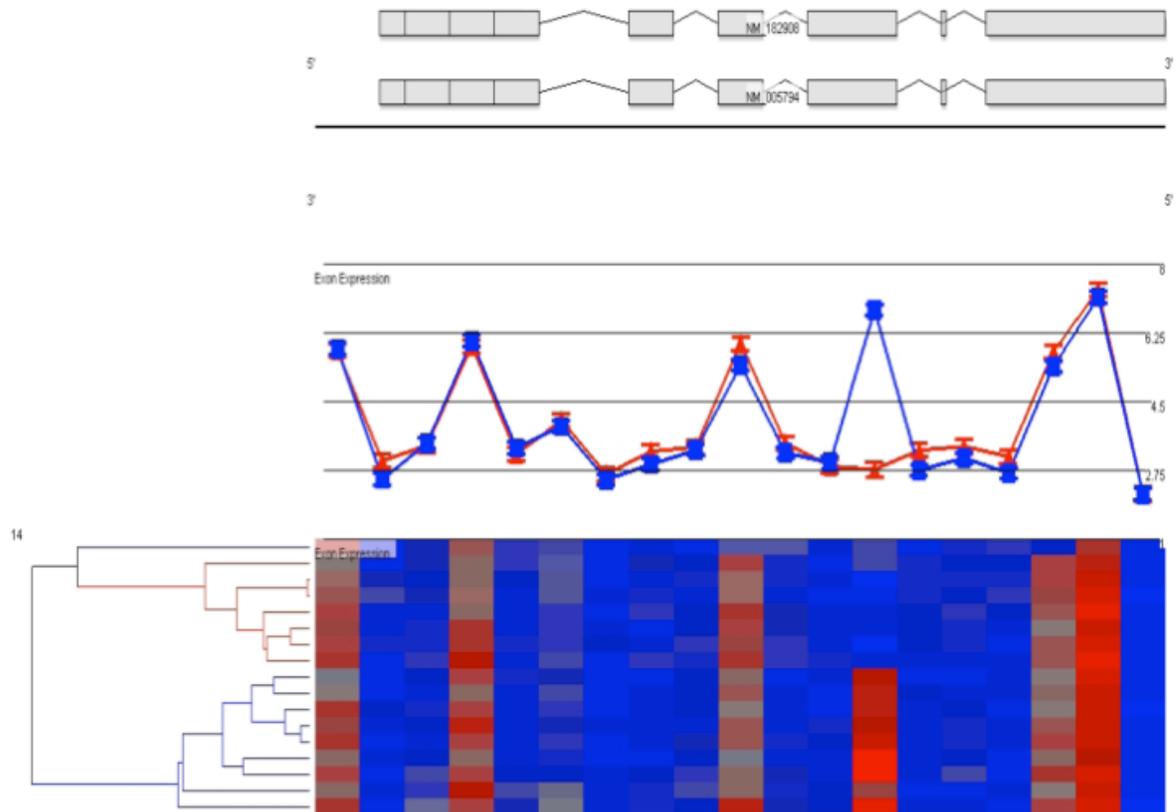
Figure 29 Representation of expression values for *DHRS2* gene. The gene annotations with filled boxes representing exons are at the top of figure. The expression values representing each probeset are depicted as a dot plot in the middle portion of the figure. Blue represents patients with t(8;21) and red with NK. The bottom of the figure is a heat map. On the left of the map are the individual patients with blue representing t(8;21) and red NK. The heat map for each of the samples for each probeset is represented with blue representing low expression and red high expression. One pair of probesets corresponding to the seventh exon from the 5' end shows differential expression between the two subsets of patient groups.

DHRS2 (Hep27) is a reductase involved in inactivation of carbonyls. It is expressed in hepatocellular carcinoma cells (Pellegrini et al., 2002) and monocytic derived dendritic cells (Heinz et al., 2002).
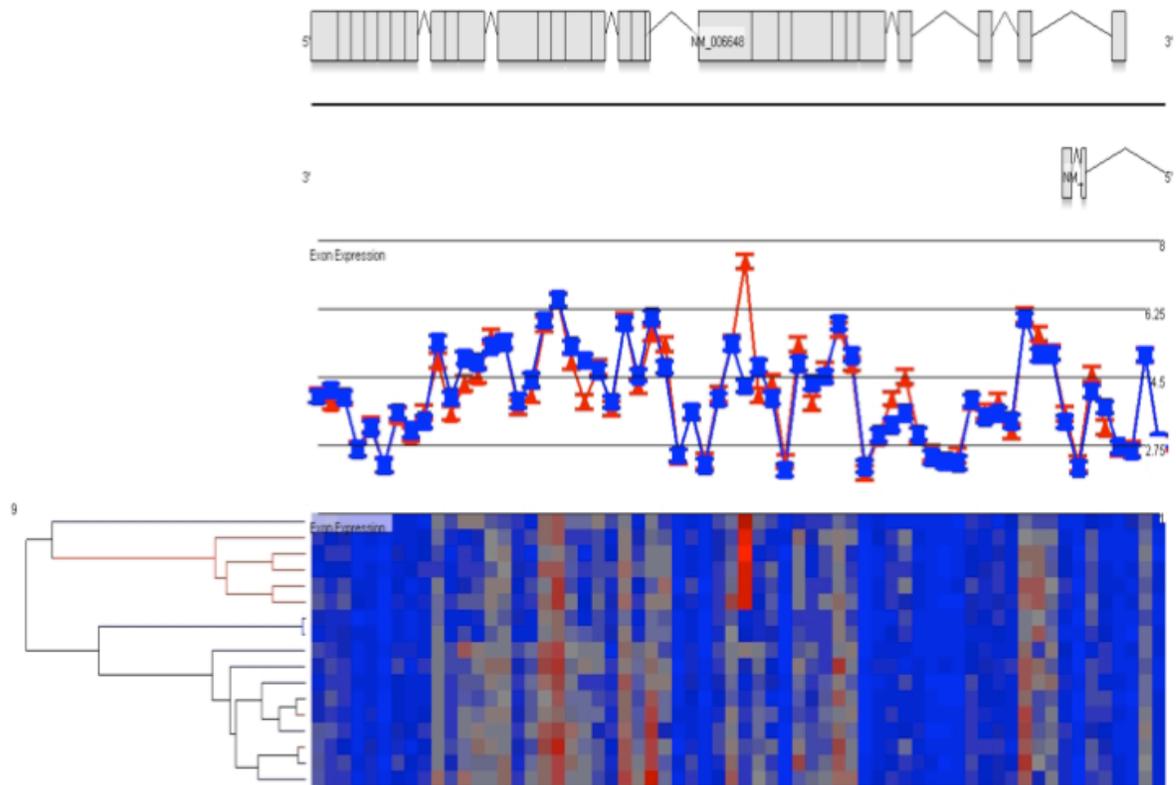
Figure 30 Representation of expression values for *WNK2* gene. The format is as described in figure 29. One of the probesets representing an exon near the middle of the gene is clearly demonstrated to show differential expression between the two subsets of patients.

WNK2 (with no K = lysine) is a lysine deficient protein kinase 2. It has been found in brain, liver, heart and gut and is involved in colon cancer and glial tumours (Moniz et al., 2007). Interestingly, WNK2 appears to modulate ERK1/2 activity through Rho GTPase intermediaries (Moniz et al., 2008). In glial tumours WNK2 undergoes marked DNA methylation and subsequent silencing, suggesting that this gene acts to suppress cell growth (Hong et al., 2007).

Discussion

Exon arrays can be used to profile individual exons for each gene. Studying the *AML1-ETO* gene specifically, it is noted that the ETO is transcriptionally silent in haemopoietic cells. Furthermore, the genomic breakpoint on *ETO* varies between patient samples. The ETO breakpoint has previously been reported to occur in three breakpoint regions (BCRs) in intron 1b (Zhang et al., 2002). A further breakpoint region has been identified in intron 1a (Buchholz et al., 2000). These reports appear consistent with the identification of at least 4 BCRs in our dataset. More samples may identify further breakpoint regions. Similarly, exon arrays may be able to identify genomic breakpoint locations in other translocations.

The exon array also detects a whole range of signals from putative exons, suggesting that novel exons may be identified using this platform. We were unable to detect these by RT-PCR methods, despite attempts to optimise the conditions. As described in the Materials and Methods the Affymetrix Human Exon 1.0 ST array uses a whole transcript assay. RNA is used as starting material and ribosomal RNA is removed. However, there is no attempt made to remove pre-messenger RNA in the RNA processing procedure. Therefore signal values generated may result from the intronic sequences of the unspliced pre-messenger RNA rather than the true exons from messenger RNA. In this case, the signal values associated with probesets against the putative exons may simply represent intronic signals explaining the negative result in these experiments.

Exon arrays were also used for global gene expression profiling. We confirm findings of previous expression data that AML1-ETO not only causes transcriptional repression but also activation. The most significant genes generated on these arrays appear to be similar to previous array data from t(8;21) primary samples validating the use of this platform for expression analysis.

Detailed comparison of exon array with ABI platform derived gene lists using primary samples, showed a degree of concordance. However, discrepancies were noted. The experiments that were compared used different controls. The exon array experiment used, the normal karyotype AML but with the same phenotype, in order to assess for specific AML1-ETO differences. The gene lists derived from the ABI platform compared t(8;21) to all other subtypes of AML. The resulting discordant genes may be as a result of this difference in controls. Using this control attempts to avoid detecting genes associated with differentiation stage of leukaemogenesis and highlight genes, which may play a key role in the early steps of leukaemia. Thus, genes that demonstrated differential expression on the exon array but not the ABI platform, may be more specific to the mechanisms of t(8;21) induced leukaemogenesis.

However, this supposition assumes that the platforms used for expression profiling are analogous and the only difference in the results is due to the experimental variation of the controls used. However, this notion may not be valid for a number of reasons. Platforms differ in their sets of genes arrayed, have their own bias in the

design of their probes and often have incorrect or incomplete annotation files. Experiments performed by different people can also affect results, with the most variable part of the exon array protocol being the ribosomal removal step. Using a whole transcript assay results in a different pool of amplified mRNA, with the 3' arrays omitting genes, such as histone genes, which lack a poly-A tail. Normalisation protocols, such as the use of RMA or PLIER can also result in differences. Furthermore, Partek uses different summarisation methods for expression data and exon data.

Despite these highlighted differences, the exon array platform has been validated for gene expression analysis in previous reports (Bemmo et al., 2008). An established dataset was compared on three platforms: Illumina, Affymetrix Hu133 plus 2.0 and Human Exon ST1.0. Generally all three platforms coincided well with over 90% concordance although the comparisons between the Illumina and Hu133 platforms were more robust then with the Exon platform.

Thus, comparison of gene lists between different experiments on different platforms is a valid method of highlighting differential gene expression. Specific genes have been highlighted which may play a key role in the oncogenic mechanisms of t(8;21) AML.

Comparing gene expression data from different platforms it was noted that data from Kasumi cell lines was poorly reproducible. In addition, there was poor concordance

between gene lists derived between primary samples and Kasumi cell lines on both the exon array platform and other platforms. Some of this variance is due to lack of replicates, which would give a greater confidence to the results and indeed previous work has suggested that many cell lines are faithful model systems (Rucker et al., 2006). However, primary samples are often quite biologically different to manipulated cell lines or stem cells. For example the Kasumi cell line has *p53* and *Kit* mutations. It is recognised that some genes, such as *POU4F1*, which are highly over expressed in primary samples, are minimally altered in the Kasumi cell line (Dunne et al., 2006). Subsequently, Kasumi cell line expression data from this study should be interpreted with a degree of caution. As Kasumi cell line investigation was not the focus of this study, replication of samples and further analysis was not performed to confirm these findings.

One of the main observations using the functional mining tools was the enrichment for G-protein coupled receptors and G protein related pathways in the t(8;21) set of patients, suggesting that expression changes in GPCR play an important role in AML1-ETO leukaemogenesis. Malignant cells can hijack the normal physiological functions of G-proteins and commonly stimulate G-proteins to be over expressed. Furthermore, the other main enriched pathway was that related to cytoskeletal remodelling through TGF beta and WNT. G-proteins are intermediaries in a number of pathways including those involving TGFb and WNT. These observations provide further insights into the mechanisms that the t(8;21) utilises leading to cancer.

Exploring these pathways may lead to development of specific drug targets, such as GPCR targeting agents, for the treatment of t(8;21) leukaemia.

Interestingly, it is noted that "cell adhesion" was the top process using the GeneGo metacore tool and was up regulated. The clinical observation that t(8;21) is associated with the development of granulocytic sarcoma (i.e. extramedullary AML) may be relevant in this context.

In addition to expression profiling, exon arrays are able to provide further information. Using extended and full metaprobeset lists in the analysis at exon or gene level, a number of probesets are identified which are differentially expressed but have no known annotations. These regions may be important and require further exploration. Some of these genes may code for microRNAs. Our findings suggested that at least at the probeset level there are important changes in expression of some microRNAs some of which have been implicated in AML. Although we observed differential expression at the exon level, the gene expression of these microRNAs on the array was not shown to be significantly different. Similarly, we were unable to detect significant gene expression changes of some these microRNAs by RQ analysis. The exon array although containing probesets to interrogate microRNAs has not been validated for microRNA analysis. The exon array does not attempt to remove pre-messenger RNA (mainly nuclear) and therefore the expression values might represent pre-microRNAs rather than the mature forms. Furthermore, for such small molecules the IVT amplification method used in the exon array protocol may

not be suitable. Some of the putative exons suggested by the array may not represent true exons and therefore mask true gene expression changes, which may have a significant effect for small genes. However, as microRNAs are relatively new and only recently being confirmed, mainly by computational models, the use of exon arrays using the full probeset list appears a justifiable way to attempt to identify them. Other annotations currently not classified as microRNAs may eventually result in being designated and thus exon arrays may give valuable information on microRNA expression as annotations become exhaustive.

Using the alternative splice algorithm in Partek the exon array data was used to detect differentially expressed exons between the groups. However, the use of exon arrays to detect differentially spliced exons, unlike their use for expression analysis, has been problematic. While many algorithms have been used to detect differentially spliced exons, no specific algorithm has been demonstrated to be robust and accurate. One of the main drawbacks is the lack of a gold standard to compare against a particular analysis model. The arrival of next-generation sequence technologies such as RNA-Seq that give further information about the transcriptome, may overcome this limitation and provide an accurate gold standard as alternative transcripts are discovered and analysed.

Alternative splicing analysis algorithms have to overcome a number of issues. Comparing probeset intensities across samples does not take into account the expression of the transcript across the sample; therefore probeset measurements

would reflect gene expression rather than differential exon usage. Use of algorithms such as the Splicing Index (SI) suggested by Affymetrix or Partek ANOVA method used in this analysis have been used to address this issue with both methods having the same limitations. The SI is calculated by dividing probe set intensity by the metaprobeset intensity i.e. exon-expression / gene-expression. However, using the SI many artefacts and false positives are generated. In these calculations the assumption is made that all probesets across a gene have a uniform response to changes in gene levels, which is often false. Probesets particularly affected by this are probesets at 3' and 5' ends, probesets that have low levels or levels similar to background, such as probesets identifying skipped exons (i.e. alternatively spliced but not differentially spliced exons) and probesets within highly expressed genes so that the detection range is saturated. The basis for the intensity variables at the 3' and 5' ends (termed edge bias effect) has been hypothesised. (Bemmo et al., 2008). The lower readings at the 3' end are thought to be simply due to reduced priming potential as it is at the end of the template. This results in values for these probesets being nearer background and having a high SI. Another consequence of this effect is that smaller genes appear to have lower expression than longer genes. The rationale for the low levels seen at the 5' end is believed to be due to the high GC content of the probesets, related to CpG islands. In addition, it has been demonstrated that genes showing strong changes in gene expression have an increased splice index (Gaidatzis et al., 2009). Whilst the gene expression levels appear to be robust the identification of alternative splicing is a technical artefact and probably reflects the original design of the arrays. Exon measurements are obtained

from the readings of only 4 probes, which are quite often overlapping whilst gene expression values are obtained from all probesets on a gene and quite often for many genes the number of probes used to detect measurements is vast. Another potential pitfall for detecting alternative splicing is that some probes may contain polymorphisms and when comparing different patient samples may lead to errors. Finally, the platform used does not detect alternative splice mRNA variants but only differences in exon usage so limiting the scope of the analysis available.

As a result of these limitations, all algorithms produce many false positives and therefore it is important to visually inspect highlighted genes to detect true alternative splicing events. We identified two genes, WNK2 and DHRS, to have a high splicing index, which on visual inspection were confirmed as having differential exon expression. The assumption is that AML1-ETO controls the splicing of these genes resulting in variable exon usage. The alternative transcripts of these genes may play a significant role in leukaemogenesis. However, although these genes were implicated it was unable to be experimentally validated using RQ techniques. This negative result may have arisen due to limitations of the RQ-PCR technique. The expression of the relevant exons, although different in the patient groups, was low and RQ PCR was perhaps not sensitive enough to detect these changes in expression between the groups at such low levels. The use of mRNA-seq may allow a more comprehensive and robust analysis of alternatively spliced genes involved in this particular disease and validate some these findings.

In summary, exon arrays are more complex to perform and analyse. The use of more patients may allow even more robust interpretation. It is noted that samples for analysis were chosen based on number of vials available and generally samples which, were collected by leucopheresis or had a high white count had more vials. These samples, associated with a high WCC, maybe associated with poorer prognosis and introduce bias. We were careful to avoid the bias of batch effect (experimental variation on a single processing batch) by ensuring that our arrays were performed in small, varied cohorts.

Using this platform for gene expression a robust dataset for differentially expressed genes specific to AML1-ETO has been produced. Both novel differentially expressed genes and key functional pathways and processes that are involved through AML1-ETO have been identified. Furthermore, exon arrays were able to identify alternative splicing of certain key molecules under the control of AML1-ETO.

Chapter 4 Results

Alternative Transcripts of AML1-ETO

Introduction

Alternative transcripts are likely to play a key role in increasing the diversity of the human genome. Recent developments in identifying alternative transcripts in the fusion gene *AML1-ETO* have highlighted the importance of this phenomenon.

An initial observation noted that a truncated AML1-ETO protein of 552 amino acids, produced as a result of the loss of the C-terminal part of the AML1-ETO protein, led to the promotion of leukaemia in mice (Yan et al., 2004). A subsequent report identified an alternative exon of ETO, exon 9a, whose utilisation resulted in alternative transcript of AML1-ETO. This results in an early stop codon at the end of ETO exon 8 and a truncated AML1-ETO product of 575 aa lacking the NHR3 and NHR4 domains. This product was found co-expressed with full length AML1-ETO in t(8;21) patients. Remarkably, in mouse models the expression of *AML1-ETO 9a* transcript leads to the development of a rapidly developing leukaemia. Furthermore, co-expression of this mutant with the full length *AML1-ETO* in mice resulted in an earlier onset of leukaemia (Yan et al., 2006). These observations suggest that firstly loss of repressor domains may confer leukaemogenic properties to AML1-ETO and secondly that alternative spliced transcripts may co-operate with the full-length protein in promoting leukaemogenesis.

Utilisation of yet another alternative exon, exon 11a, gives rise to an alternative truncated transcript that is also co-expressed with full length *AML1-ETO* transcript.

157

The resulting product of 688aa lacking the NHR4 domain, which is required for recruiting the co-repressors NCoR/SMRT, lost repressor activity. However, co-expression with the full length *AML1-ETO* restored this transcriptional repressor activity. It was shown that the variants formed multimers unlike wild type AML1-ETO that formed dimers. The interaction between the wild-type ETO and the variant were increased compared to homo-dimerisation of wild type (Kozu et al., 2005). The variant forms produced by dysregulated splicing may become the core of the ETO repressor complexes and stimulate further oligomerisation. Oligomerisation of transcriptional factors imposes altered interaction with transcriptional co-regulators and leads to the oncogenic conversion of transcription factors in leukaemia.

Other spliced forms of AML1-ETO have also been described (Figure 31). A further 3 transcripts have been identified which contain introns 2, 4 and 5 of ETO and result in in-frame stop codons occurring after exons 2, 4 and 5 respectively (Era et al., 1995, Kozu et al., 1993, Nisson et al., 1992). The later 2 variants have been shown to co-exist with full length AML1-ETO. In one sample, a variant with a 50-nucleotide deletion of exon 2 has been recognised as the only transcript. This caused a disruption of the reading frame and a premature stop codon resulting in only 31aa of ETO remaining (Lasa et al., 2002). This transcript lead to the characteristic t(8;21) with an M2 phenotype suggesting that very little of the ETO transcript is required for it to be leukaemogenic. AML1-ETO transcripts with insertions between exon 5 of AML1 and exon 2 of ETO have also been detected and result in out of frame transcript causing a stop codon and a truncated AML1-ETO with only the RHD

domain of AML1 translated (Tighe and Calabi, 1994, van de Locht et al., 1994). The

roles of all these transcripts in leukaemogenesis have not been investigated.
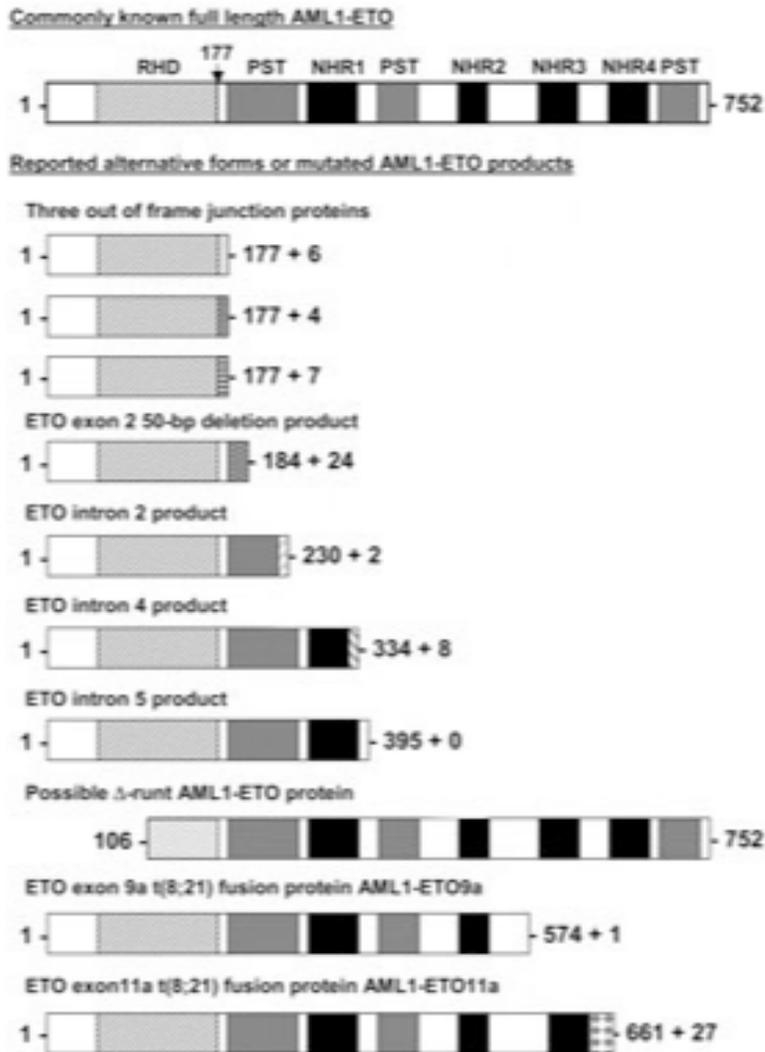
Figure 31 Alternative transcripts of AML1-ETO (Peterson et al., 2007a). See text for full details.

In addition to these isolated observations a further report identified a remarkable heterogeneity of AML1-ETO transcripts in 13 paediatric t(8;21) AML patients as a result of alternative transcripts and deletions (LaFiura et al., 2008). This suggests the fusion gene produces a whole range of proteins.

Further understanding of the roles of alternative splice variants has focussed on understanding the various roles of the NHR domains.

The NHR2 and NHR4 domains, containing the sites to bind co-repressors have been the most investigated. It was demonstrated that double deletion of NHR2 and NHR4 domains resulted in loss of the activity of the fusion protein in inhibiting differentiation and proliferation as well as promoting self-renewal of primary cells. However, the individual deletions of NHR2 and NHR4 domains did not cause this effect (Hug et al., 2002). In contrast, by introducing mutations specifically in the oligomerisation sequence of the NHR2 domain, AML1-ETO's ability to sustain clonogenic activity of stem cells in vitro was impaired. Disrupting oligomerisation eliminated AML1-ETO's ability to interact with other ETO molecules. However, oligomerisation was not required for interaction with co-repressors. Oligomerisation forming tetramers have the potential to increase recruitment by x4-fold repressor molecules (Liu et al., 2006). A polypeptide against the NHR2 domain disrupted high molecular weight complex formation, restored the expression of AML1-ETO target genes and reversed block in differentiation (Wichmann et al., 2007). Furthermore, it was shown that a heterologous dimerisation motif could rescue NHR2 deletions

suggesting that NHR2 plays a key role in homodimerisation and heterodimerization by bringing together NHR4 polypeptides (Zhang et al., 2001). Also contrary to the reports by Hug, it was shown that certain key mutations introduced in the NHR4 resulted in the development of a rapidly growing leukaemia (Ahn et al., 2008).

The NHR1 has been shown to interact with the activation domains of E-proteins, which are involved in cell cycle regulation (Zhang et al., 2004). Consequently, focussing further investigations on the leukaemic potential of this domain becomes an attractive strategy.

In this chapter the leukaemic potential of the NHR1 of ETO is further investigated. I report on the discovery of a novel exon, ETO exon 6a, whose utilisation can result in at least two further *AML1-ETO* transcripts. One of these transcripts contains the NHR1 and lacks the remaining three NHRs. I have assessed this transcript using a retroviral transduction assay and demonstrated that the NHR1 is not leukaemogenic on its own but may negatively influence the effect of the full-length transcript.

Results

In chapter 3, an attempt to detect novel exons of ETO using a PCR based strategy was performed. As a positive control for these experiments, RT-PCR to amplify the previously described exon 9a transcript was performed, using the primers for ETO exon 9a and AML1 exon 4, for three primary samples with t(8;21) (see Chapter 2 p65). Bands at the appropriate size (1699bp) in two patients were detected (Figure 32).
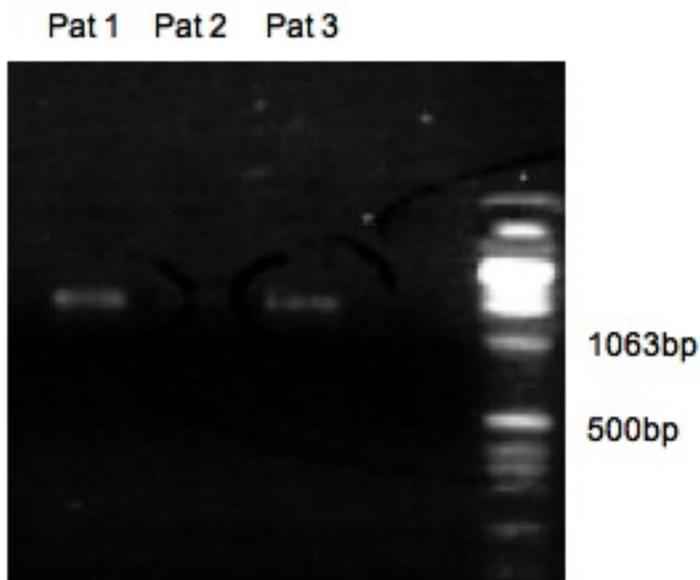
Figure 32 Agarose Gel of PCR experiments of three samples from patients with AML t(8;21). Primers to ETO exon 9a and AML1 exon 4 were used, detecting a product of 1699bp corresponding to the 9a transcript. Bands are detected in patients 1 and 3 but not 2.

The PCR products obtained from these two patients were cloned and sequenced to confirm the presence of AML1-ETO exon 9a transcript. Product A retained the expected sequence and is schematically represented in figure 33. Product B, in contrast, despite being an almost identical size, actually proved to be a novel transcript, which is represented in figure 34. The novel transcript has ETO exon 3 spliced out and an unidentified sequence inserted between ETO exons 5 and 6, giving a product of approximately similar size.

Figure 33 Schematic representation of product A derived from the sequence analysis. AML-ETO exons are shown with AML1 on the left and ETO on the right. The thick vertical bar represents the fusion site. The arrows represent primers to AML1 and ETO exon 9a. The forward sequence is initiated by the AML1 exon 4 primer and continues without interruption through AML exon 5, ETO exons 2, 3, 4 and part of 5. The reverse sequence is initiated by ETO exon 9a primer and continues through exons 8, 7, 6 and part of 5. The two sequences overlap in ETO exon 5. This confirms that product A is the AML-ETO exon 9a product.

Figure 34 Schematic representation of product B derived from the sequence analysis. Layout is as figure 33. This shows that the ETO exon three is spliced out but a novel sequence is identified between the ETO exons 5 and 6.

A blat search (http://genome.ucsc.edu/cgi-bin/hgBlat) confirmed that the unidentified sequence is part of the ETO gene and corresponds to part of the extended probeset on the Affymetrix Human Exon ST1.0 array between exon 5 and 6 (Figure 35). This new exon is referred to as ETO exon 6a.
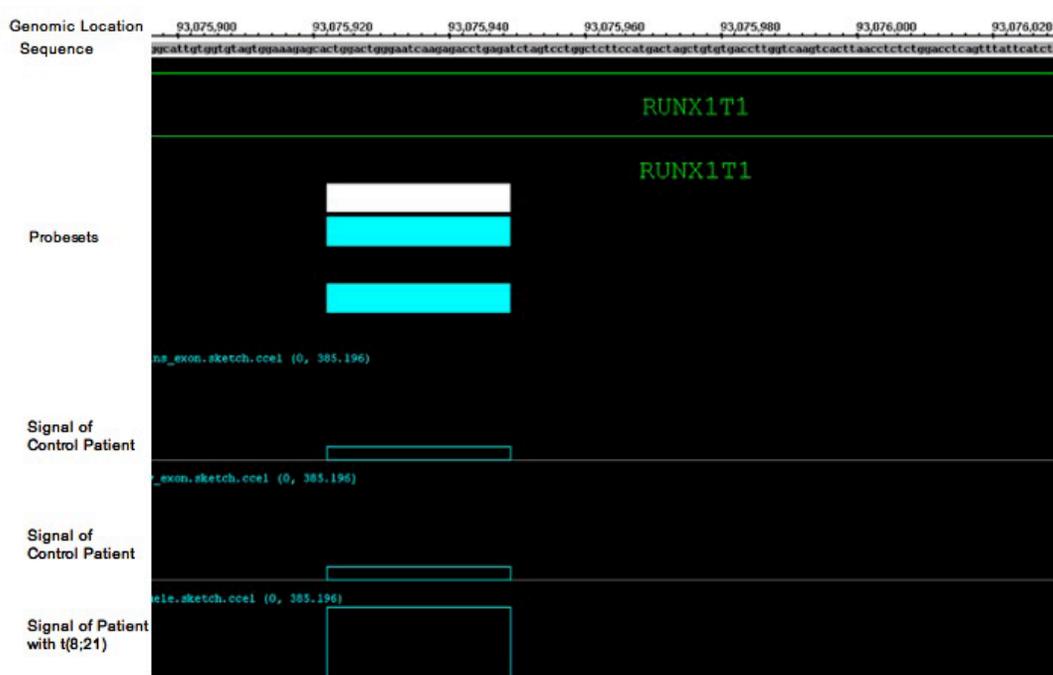
Figure 35 The reverse strand of ETO showing most of the sequence of exon 6a, illustrated on the IGB. The genomic location and base sequence of most of the exon 6a gene is shown at the top of the figure. The Affymetrix probeset for the 6a exon is represented by the blue and white bars. Signal estimates are low in the two controls and high in the t(8;21) patient.

To confirm the presence of the ETO exon 6a, RT-PCR on primary samples and the Kasumi cell line were performed, using the AML1 exon 4 and newly designed ETO exon 6a primers, specifically to amplify AML1-ETO 6a transcripts (Figure 36)
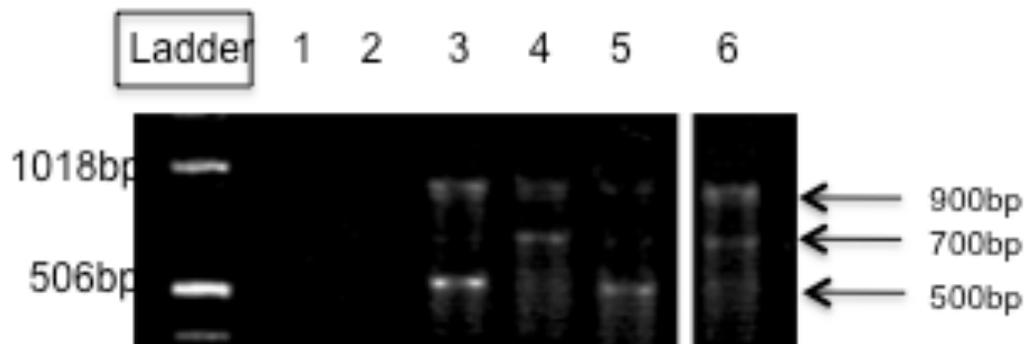
Figure 36 Agarose electrophoresis of RT-PCR products derived from amplifications using primers to *AML1* exon 4 and *ETO* exon 6a on the following templates: negative controls (lanes 1 and 2), three different t(8;21) patient samples (lanes 3 to 5) and Kasumi cells (lane 6). Products of the expected 971 bp are seen. In addition three other fainter bands are seen in these samples – one just below the main band at 900bp, one at approximately 700bp and one at 500bp

These products were cloned and sequenced to confirm the exon 6a transcripts. A

transcript with exon three spliced out was confirmed on a patient sample (Figure 37).

However, a different transcript incorporating the new exon but retaining exon 3 was

detected using the Kasumi cell line (Figure 38).
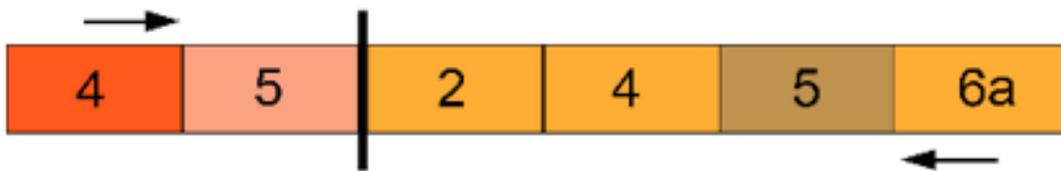
Figure 37 Schematic representation of a product derived from the sequence analysis using AML1 and ETO exon 6a primers in a second patient. The layout is as figure 33. This shows a transcript with ETO exon 3 spliced out and the new exon 6a retained.
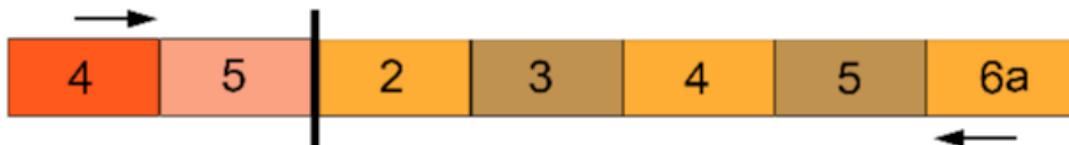
Figure 38 Schematic representation of a product derived from the sequence analysis using AML1 and ETO6a primers in the Kasumi cell line. The layout is as figure 33. This shows a transcript with both ETO exon 3 and the new exon 6a retained.

ETO exon 6a has 137 base pairs and is located at 93075890-930760270 (version hg18 Mar 2006). It is flanked by base pairs AG and GT, which are consensus motifs characteristic of splice donor and acceptor sites in eukaryotic genes (Shapiro and Senapathy, 1987). The full ETO exon 6a sequence is as follows:

ATGAATAAACTGAGGTCCAGAGAGGTTAAGTGACTTGACCAAGGTCACACAGC
TAGTCATGGAAGAGCCAGGACTAGATCTCAGGTCTCTTGATTCCCAGTCCAGT
GCTCTTTCCACTACACCACAATGCCAGTAAG.

The incorporation of exon 6a was found in at least two AML1-ETO transcripts potentially giving rise to two truncated forms of the AML1-ETO protein. In the shorter transcript (AML1-ETO 6a-sh), the splicing out of ETO exon 3 and retention of ETO exon 6a leads to disruption of the reading frame and the introduction of a premature stop codon at the end of exon 2. In the longer transcript (AML1-ETO 6a), both exon 3 and exon 6a from the primary transcript are retained, resulting in the introduction of a premature stop codon at the end of ETO exon 5. The protein generated by this transcript is predicted to be 395aa, retaining the NHR1 domain but lacking the NHR 2, 3 and 4 domains. AML1-ETO transcripts predicted to yield identical 395aa sequences have been previously described, although in these reports the termination of the transcripts were as a result of the introduction of other intronic sequences and not the novel exon described here (Era et al., 1995, Kozu et al., 1993).

Next, the presence of the exon 6a-containing transcripts in further t(8;21) patient samples was confirmed by RQ-PCR. A total of ten t(8;21) patients were identified and RNA obtained as previously described in Chapter 2 p63 (Table 21).

Table 21 Patient samples used in this study. Age in years, CR- Complete remission, TRM- transplant related mortality, O/S – Overall survival, DFS – Disease free survival

AGE	SEX	BLAST %	RESPONSE TO 1ST LINE RX	O/S (YEARS)	DFS (YEARS)	KARYOTYPE
24	F	98	Fail/ Resistant disease	0.49	0	46,XX,t(1;6)(p36;p23),t(8;21)(q22;q22)/46,idem,der(22)t(1;22)(q23;p11.2)/47, idem,+8
38	F	70	CR	15.79	0.82	46,XX,t(8;21)(q22;q22)
51	M	51	CR	1.00	0.65	46,XY,del(7)(q32q36), t(8;21)(q22;q22)
68	M	50	CR	5.63	5.6	45,X,-Y,t(8;21)(q22;q22)
20	M	60	CR	0.65	0.7	45,X,-Y,t(8;21),add(14q32)
27	F	88	CR	11.06	11.1	46,XX,t(8;21)(q22;q22)
49	F	NA	CR	5.07	5.1	46,XY,t(8;21)(q22;q22)
67	M	30	TRM	0.16	0	46,XY,t(8;21)(q22;q22)
34	M	NA	CR	7.09	7.1	45,X,-Y,t(8;21)(q22;q22)
18	M	50	CR	14.81	11.6	45,X,-Y,t(8;21)(q22;q22)

RQ primers were designed to amplify across the ETO exon 6a boundaries. Transcripts containing the ETO exon 6a were identified in all patient samples. The expression level of each isoform differs between patients, for example patient 5 has high levels and patient 6 low levels (Figure 39).
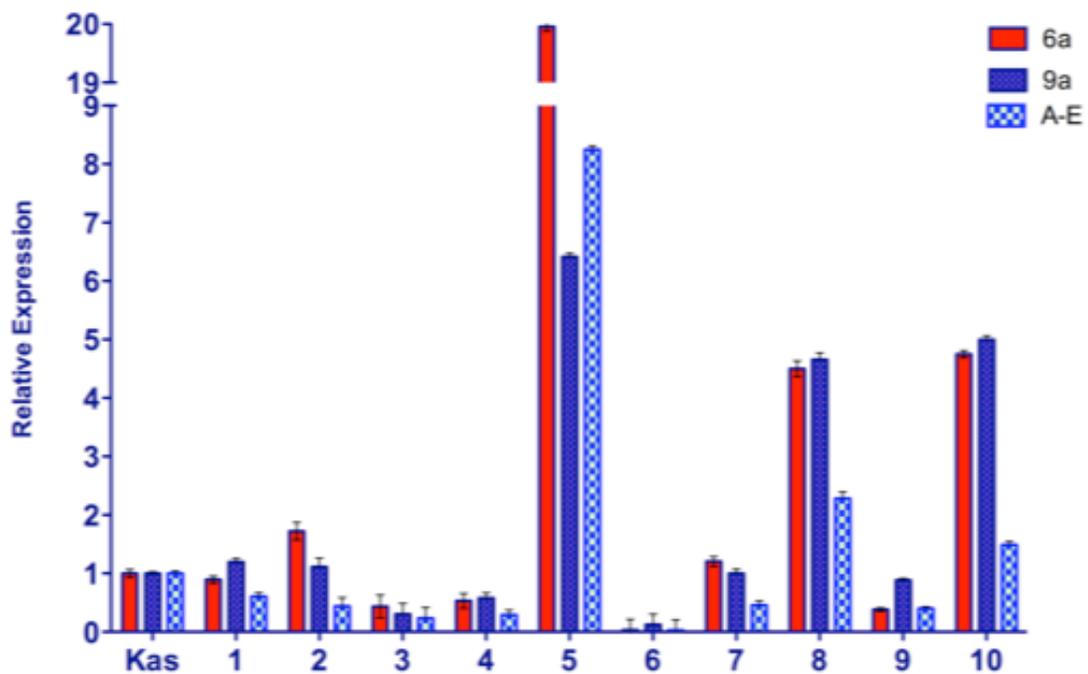
Figure 39 Bar charts showing the ΔΔCt (threshold cycle) results of real time PCR assays for ten AML patients and Kasumi cell line. Results, from probes designed to detect *AML1-ETO* 6a, 9a and full-length transcripts (A-E), relative to 18S ribosomal RNA normalized for Kasumi are shown. (Details Chapter 2 p66)

Comparisons of transcript levels within each patient demonstrate that 6a transcripts are generally but not always less abundant than the 9a transcript and consistently five- to ten-fold less abundant than the full length AML1-ETO transcript (Figure 40).
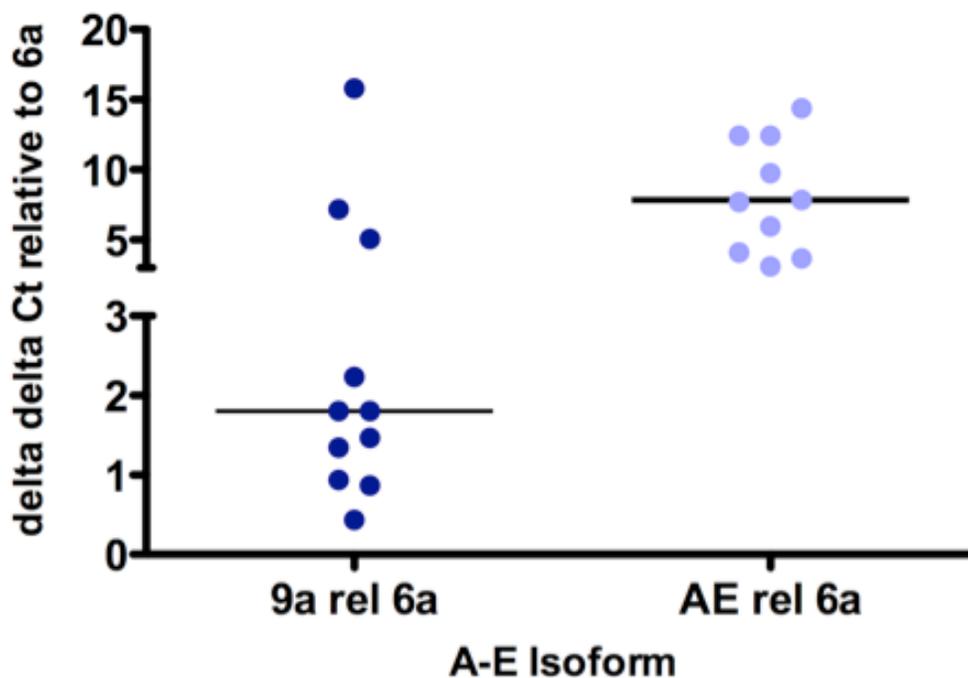
Figure 40 Expression of *AML1-ETO* isoforms in ten primary samples and Kasumi cell line determined by real-time PCR. The ΔΔCt value for each isoform relative to *18S* and normalised to that of the *AML1-ETO* 6a isoform for each individual patient is represented as a dot-plot. The median value is shown for each isoform. One value for the full-length isoform is very high and outside the axis limit and not represented. (An assumption is made that the reaction efficiencies for the amplification of the three isoforms are comparable.)

To investigate the role of the NHR1 domain in AML1-ETO induced leukaemogenesis the AML1-ETO 6a transcript was used for further functional assessment. Initially, a pMSCV AML1-ETO 6a-iresGFP (pMiG-A-E6a) and pMSCV AML1-ETO-ires hCD2 (pMihC-A-E) vectors were constructed and the cDNA insertions were confirmed by sequencing. (Appendix B, performed by author) The *AML1-ETO 9A* vector was provided by D. Gascoyne for use in some of the following experiments.

The plasmids were then incorporated into retrovirus using a packaging cell line. (performed by D. Gascoyne) Protein expression of cDNA, in packaging cell lines, was confirmed by Western blotting using lysates obtained from packaging cell lines (Figures 41 & 42). (see Chapter 2 p82)
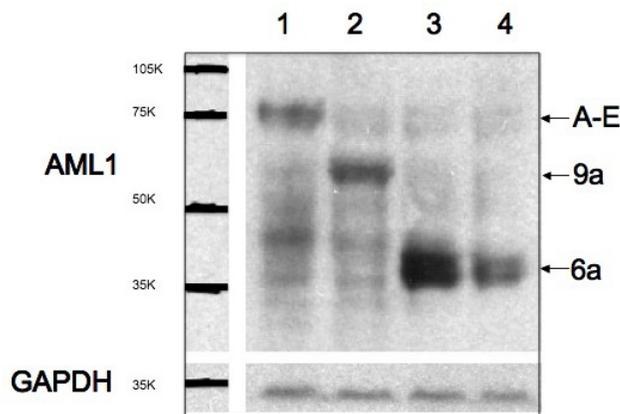
Figure 41 Western blot to confirm expression from the pMiG A-E6a vector. Lysates from cells transfected with full length, 9a or 6a *AML1-ETO* vectors (2 separate clones of the 6a vector were used) are run in lanes 1, 2, 3 and 4 respectively. The AML1 antibody detects products of the expected molecular weights: 85kDa, 65kDa and 45kDa respectively. Re-probing with anti-GAPDH showed equal loading.
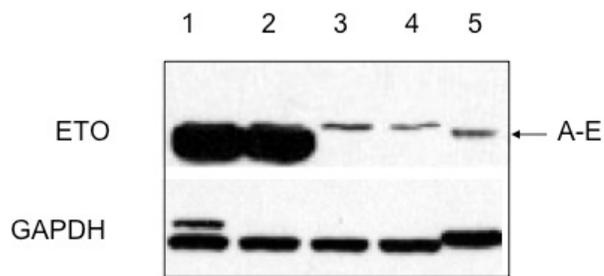
Figure 42 Western blot to confirm expression from the pMihC-A-E vector. Lysates are from cell transfected with pMihC-A-E (1 and 2), pMSCV-ires-hCD2 without transcript (pMihC) (3 and 4) and Kasumi (5). The ETO antibody detects products at the expected weight of 85K in lanes 1, 2 and 5 but not in lanes 3 and 4. The faint non-specific bands in lanes 3 and 4 correspond to similar bands in lanes 1 and 2 just above the main band corresponding to AML1-ETO.

To examine the effect of the truncated protein, murine haemopoietic stem cells were infected with AML1-ETO vectors and subjected to methylcellulose re-plating assays (See chapter 2 p89 assay performed by D. Gascoyne with transfection efficiencies of greater than 90%). The AML1-ETO 6a cDNA appears to have minimal affect on colony numbers (Figures 43 & 47) or colony morphology in first round methylcellulose plates. Similarly, cells transduced with this retrovirus did not re-plate beyond the third round, unlike the full-length and 9a vectors, indicating a lack of immortalization capacity (Figure 44).
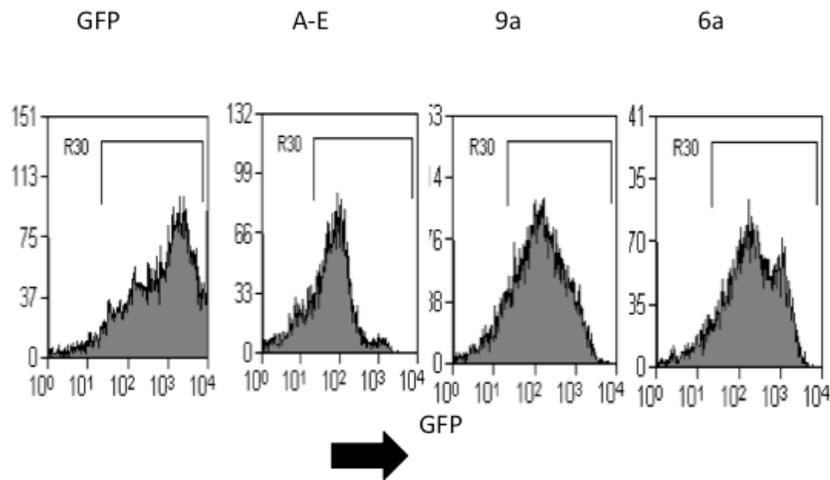
Figure 43 FACS analysis of GFP+ cells. Murine progenitor c-Kit+/Ter119- cells were transduced with retroviruses containing cDNAs for GFP alone, full length *AML1-ETO, AML1-ETO* exon 9a or *AML1-ETO* exon 6a. Experiments were performed twice and representative sample is shown. Suppression of cell numbers by A-E and 9a retrovirus are suggested although actual cell counts are shown in figure 48.
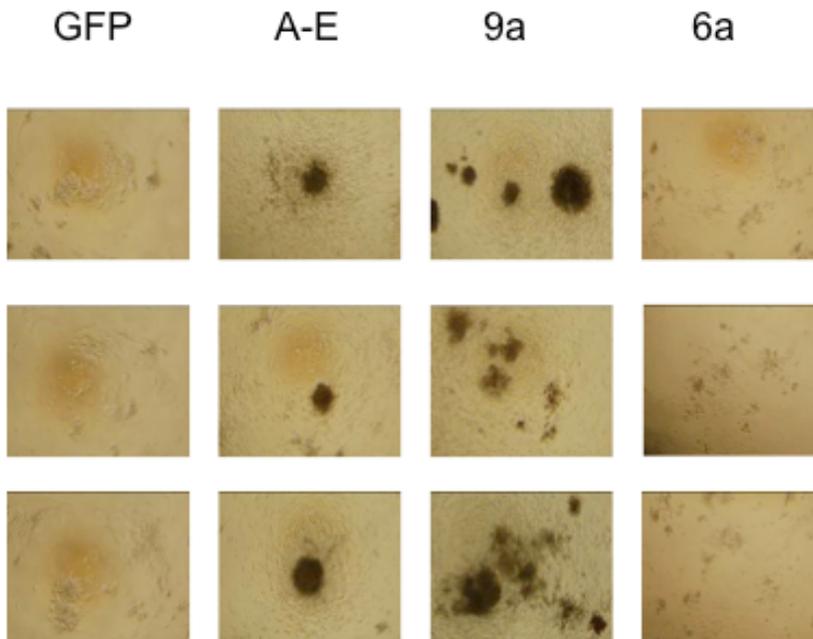
Figure 44 Effect of *AML1-ETO* 6a in murine methylcellulose re-plating assays - photos of three representative plates at second round/week of re-plating. (Cells transduced as Figure 43) Photos provided by D. Gascoyne.

Next dual transfection experiments, with full-length and 6a AML1-ETO vectors, were performed to more closely mimic in-vivo activity. Murine haemopoietic stem cells were co-infected with pMihC-A-E plus pMiG-A-E6a; pMihC-A-E plus pMiG; pMIG-A-E6a plus pMihC; or pMiG plus pMihC. Cells were analysed by FACS analysis of CD2 and GFP expression (Figure 45).
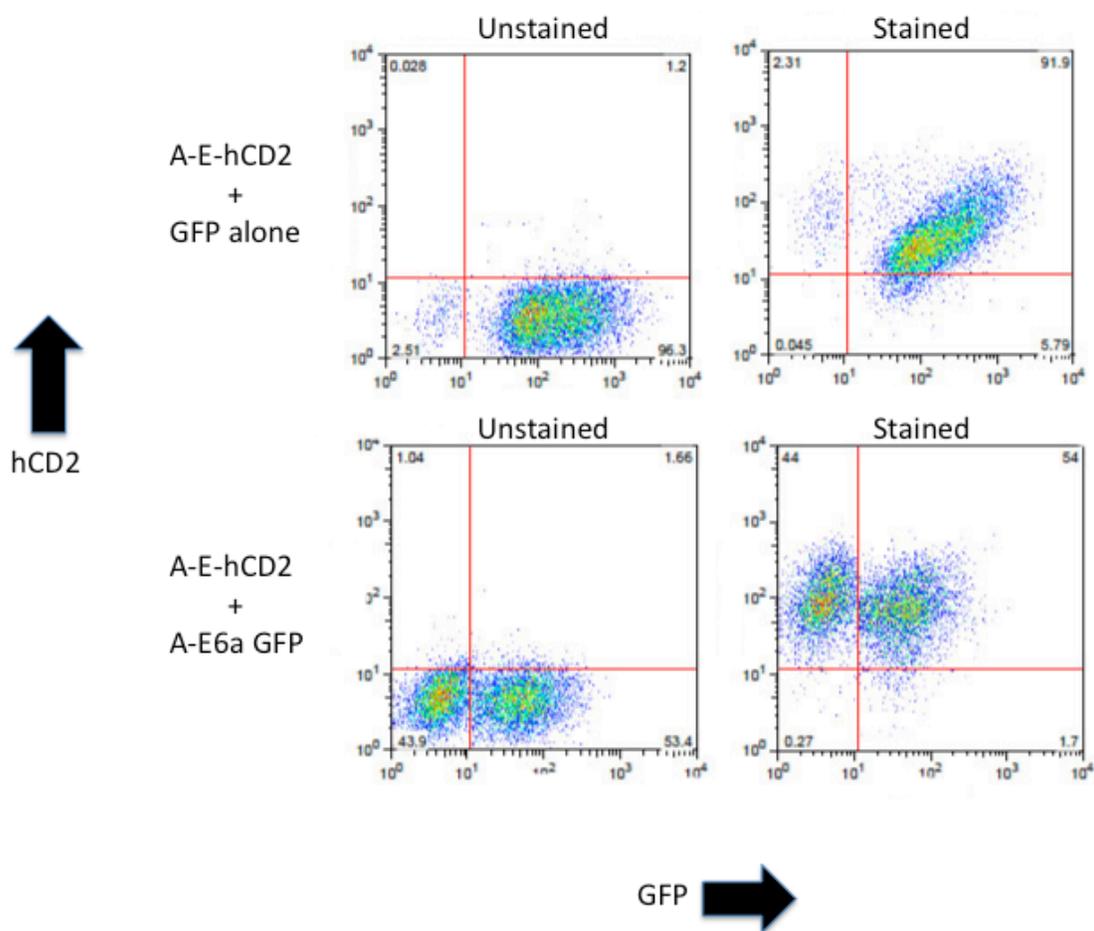
Figure 45 Representative figure of flow cytometric analysis of infection marker expression at fourth round replating. Murine progenitor cells have been co-infected with pMihC-A-E plus pMiG A-E6a; pMihC A-E plus pMiG; pMIG-A-E6a plus pMihC; or pMiG plus pMihC. Cells from the two former dual infections were analysed by FACS analysis of GFP expression and of CD2 expression (CD2 expression was performed with and without 2nd antibody resulting in both stained and unstained analysis – see Chapter 2 p84).

In contrast to the above observations, the expression of the 6a vector appears to have a suppressive effect on the ability of the full-length AML1-ETO transcript to propagate colony formation and cell proliferation (Figure 46 & 47).
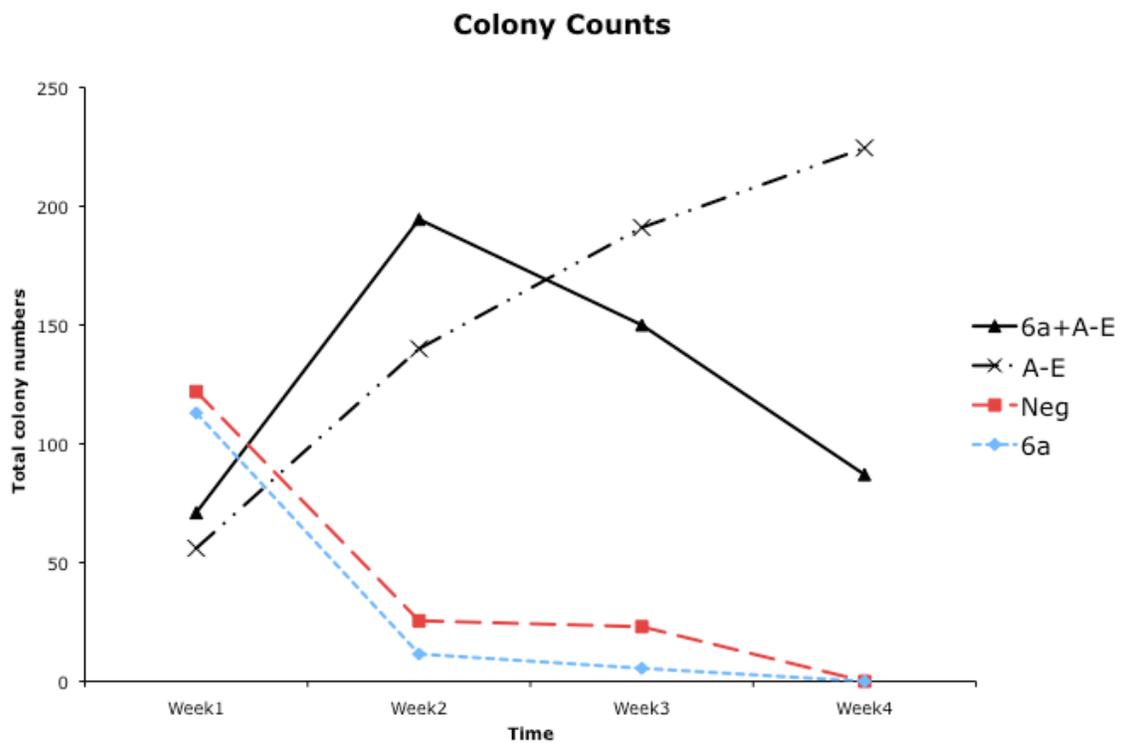
Figure 46 Graphs from methylcellulose re-plating assays. Experiments were performed twice and results of a representative sample shown. Murine progenitor cells were transduced with pMiG A-E6a and pMihC (6a) or pMihC A-E and pMiG (A-E) or pMiG A-E6a and pMihC A-E (6a+A-E) or both pMiG and pMihC retroviruses (neg). Total numbers of colonies on a plate were counted at weekly intervals and plotted against time. For details of methodology see chapter 2 p83.

Cell Numbers
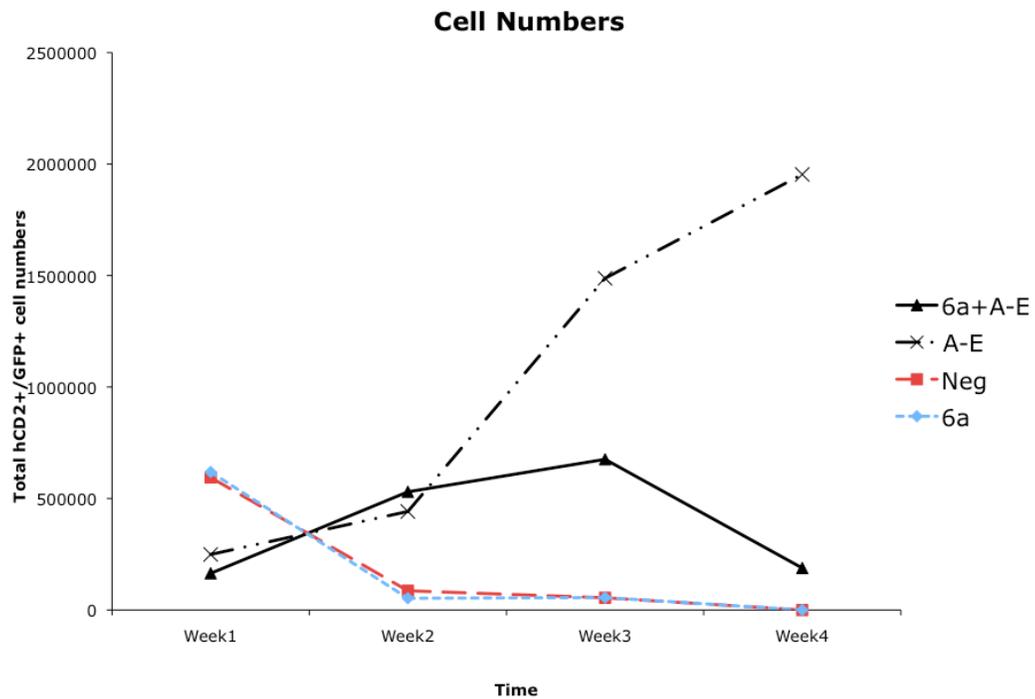
Figures 47 Graphs from methylcellulose re-plating assays. Experiments were performed twice and results of a representative sample shown. Murine progenitor cells were transduced with pMiG A-E6a and pMihC (6a) or pMihC A-E and pMiG (A-E) or pMiG A-E6a and pMihC A-E (6a+A-E) or both pMiG and pMihC retroviruses (neg). Cells were counted at weekly intervals prior to re-plating. FACS analysis of cell markers was performed to identify the percentage of dual positive cells. Dual positive cell counts are plotted against time.

Discussion

I have described a novel in-vivo AML1-ETO transcript, containing the NHR1 domain and lacking the other NHR domains, and have demonstrated that it lacks clonogenic potential in AML1-ETO induced leukaemogenesis. As the 9a transcript containing the NHRs1 and 2 domains can induce leukaemia unlike the 6a transcript, which lacks the NHR2 domain, it can be deduced that the NHR2 domain of ETO is essential in the leukaemogenic capacity of the AML1-ETO fusion protein. The homo-oligomerisation function imparted by the NHR2 domain appears to be the key mechanism for leukaemogenesis in the pathogenesis of AML with t(8;21) (Kwok et al., 2009).

By introducing specific mutations into the regions coding for the NHR1 and NHR2 domains of AML1-ETO transcripts, three recent reports have confirmed a limited role of the NHR1 domain. Although our results are in keeping with these findings, we further demonstrate that the novel 6a-containing transcript may influence the clonogenic potential of the full-length transcript. This observation highlights another potential failsafe mechanism by which cells may be able to regulate uncontrolled leukaemic growth.

In addition, the AML1-ETO 6a-sh transcript, although giving rise to a protein lacking all 4 NHR domains, may still have a role in leukaemogenesis through a dominant negative effect, competing with native AML1 or full-length AML1-ETO for DNA

binding sites. These observations provide further insights into both the genomic structure of *ETO* (Figure 48) and the role of alternative transcripts in leukaemogenesis. As a result of alternative exon usage a whole range of minor alternative transcripts are produced concurrently in t(8;21) leukaemic samples. These transcripts encode for proteins that lack the various NHR domains (Figure 49). The interactions of these proteins with each other appear to be complex affecting the leukemic potential of the main full-length AML1-ETO product. The key to these interactions may involve the role played by the NHR2 domain and the ability of the AML1-ETO molecule to form tetramers through this domain. AML1-ETO tetramers bind preferentially to DNA containing 2 AML1 binding sites (Okumura et al., 2008). In contrast AML1-ETO dimers have reduced DNA binding. It can be hypothesised that the ratio of the different proteins produced by the *AML1-ETO* oncogene can result in varying amounts of dimerisation and tetramer formation. This may then affect the binding ability of the AML1-ETO complex. Thus, disrupting the ability of the NHR2 domain gives a basis for a drug target.

In the clinical setting, the detection and quantification of 6a transcripts may result in prognostic implications where the amount of transcript may correlate with prognosis. This cohort has ten patients of which four survived less than 1 year while the remaining six survived over 5 years. Although there was no correlation between survival and transcript levels, larger numbers are required to draw any firm conclusions.

Figure 48 Schematic diagram summarising the exons of *ETO* involved in the *AML1-ETO* gene
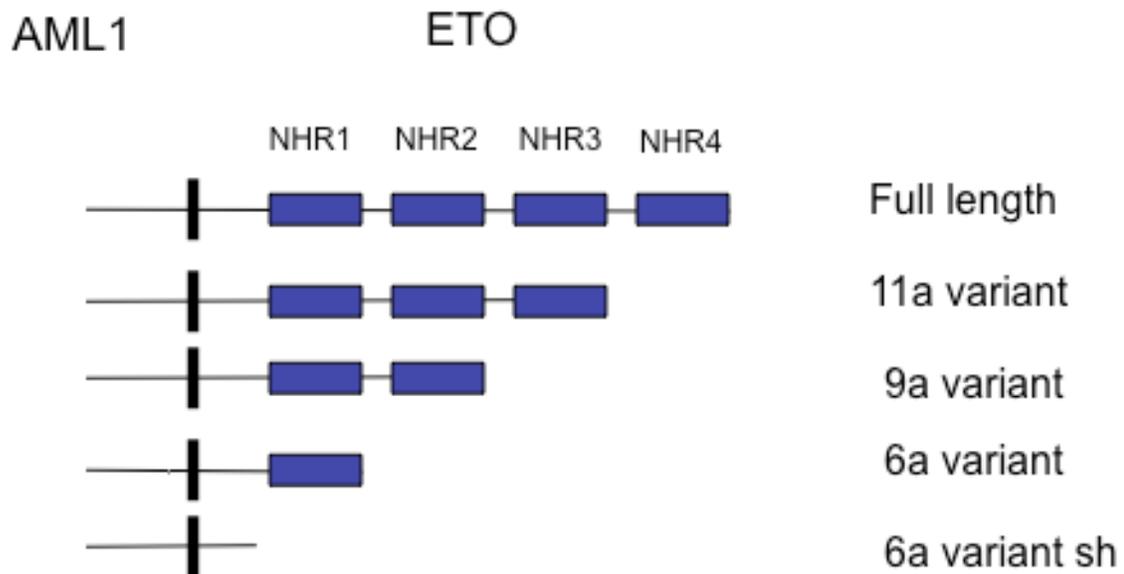
Figure 49 Schematic summary of predicted AML1-ETO proteins (highlighting the NHRs) from reported *AML1-ETO* transcripts resulting from alternative exon usage. The left of the line represents the region encoded by the DNA binding domain of *AML1,* while the right represents the region encoded by *ETO* and its NHR domains and the thick vertical bar represents the fusion junction between *AML1* and *ETO*. In addition to the major full-length AML1-ETO product containing for all four NHRs, truncated proteins lacking NHRs as result of alternative exon usage of 11a, 9a and 6a respectively are summarized. Utilization of the 6a exon can result in at least two transcripts with a short form coding for a protein lacking all four NHRs (6a sh)

In summary, a novel exon of ETO has been detected, which is utilized to form further alternative transcripts of AML1-ETO. This transcript, giving rise to a protein of 395aa and lacking NHRs 2, 3 and 4 domains is not clonogenic on its own but appears to have a negative effect on the clonogenic potential of the full-length AML1-ETO transcript.

Chapter 5 Results

ChIP-Seq Results in Acute Myeloid Leukaemia with

t(8;21)

Introduction

The advent of high throughput sequencing has opened up new ways to investigate the genome. One of the most popular techniques is the use of chromatin-immuno precipitation followed by high throughput sequencing (ChIP-Seq) to investigate global binding of transcription factors. As the onco-protein AML1-ETO acts as a transcription factor, ChIP-Seq becomes a very attractive strategy to investigate DNA binding by AML1-ETO.

ChIP

DNA and histone interactions are usually transient in nature. However, in the 1960's it was observed that applying formaldehyde to living cells covalently cross-linked DNA to histones in-situ and furthermore, these cross-links could be reversed. (Kuo and Allis, 1999) Thus, the basis for ChIP methodology began to take shape. (Figure 50) However, ChIP only became established as major technique decades later with the development of highly specific antibodies allowing identification of where on the genome specific proteins could bind. Particular interest has focused on the direct physical interaction between transcription factors and DNA, which has further increased with the development of high throughput technologies allowing global identification of immunoprecipitated DNA.

ChIP-Chip, where DNA obtained through ChIP is used to interrogate a series of probes on a tiling array, has been quickly superseded by Chip-Seq with its numerous advantages: arrays require prior knowledge of the genome; due to

constraints of hybridisation chemistry much of the DNA is not accessible for interrogation by arrays and the amount of DNA required for an array-based investigation is high especially considering the low yield of DNA produced by ChIP.
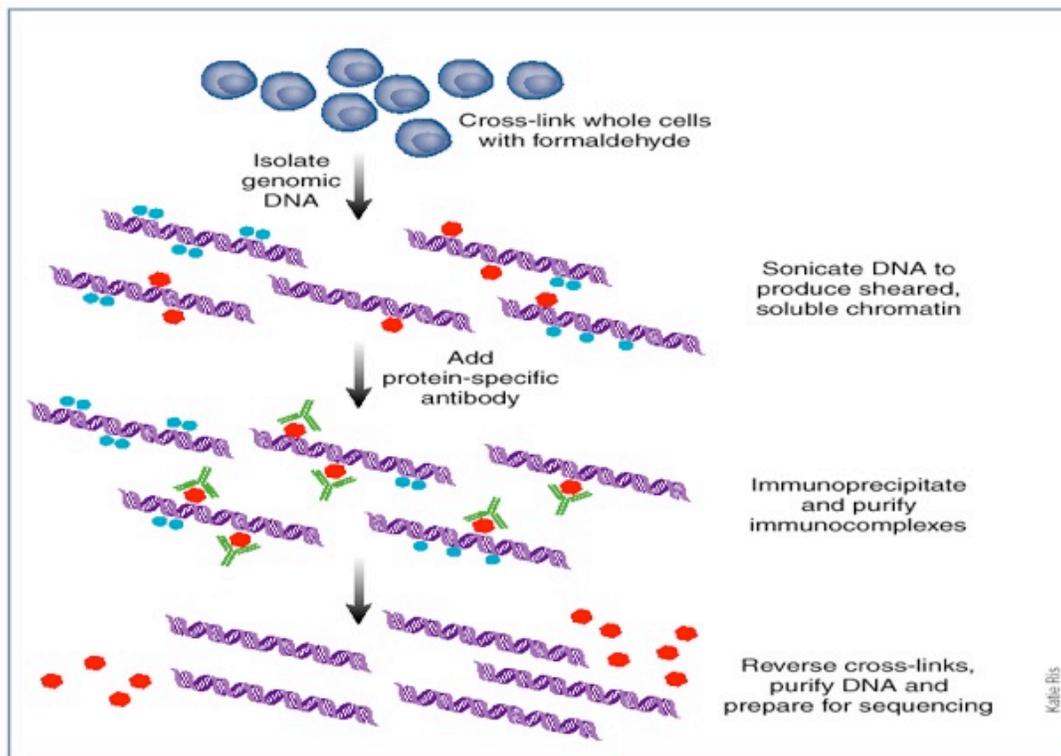
Figure 50 The three main steps required for ChIP methodology: covalently cross-linking DNA-protein interactions; using antibodies to precipitate these interactions and uncross-linking the interactions to investigate the isolated DNA. In this technique DNA must be compared to a control consisting of DNA collected without antibody addition or non-specific antibody. This is due to random protein–DNA cross-linking resulting in non-specific DNA being pulled down in the immunoprecipitation step. (Mardis, 2007).

High Throughput Sequencing

For the last 30 years the "Sanger sequencing method" has been the main tool for sequencing, proving to be powerful enough to decipher complete genomes. However, the introduction of next-generation sequencing technologies appears to be even more potent. By omitting the need to clone DNA fragments this technique is able to generate a vast of amount of sequencing data in a relatively short time frame.

Three companies have commercially introduced their systems: the 454 from Roche, Illumina genome analyser and SOLid from Applied Biosystems. All three use different technologies and have different advantages but all require a high degree of IT support and are currently expensive. The Illumina system used in this project uses a bridge amplification technique followed by sequencing-by-synthesis. There are three phases; DNA is prepared into a library suitable for applying to the sequencer, followed by cluster generation and finally sequence reading. (Figure 51) The end result is the generation of a series of reads whose DNA sequence at one end has been deciphered and tagged. These tags allow identification on the genome of the location of the DNA read (See Chapter 2 p86 for further details).
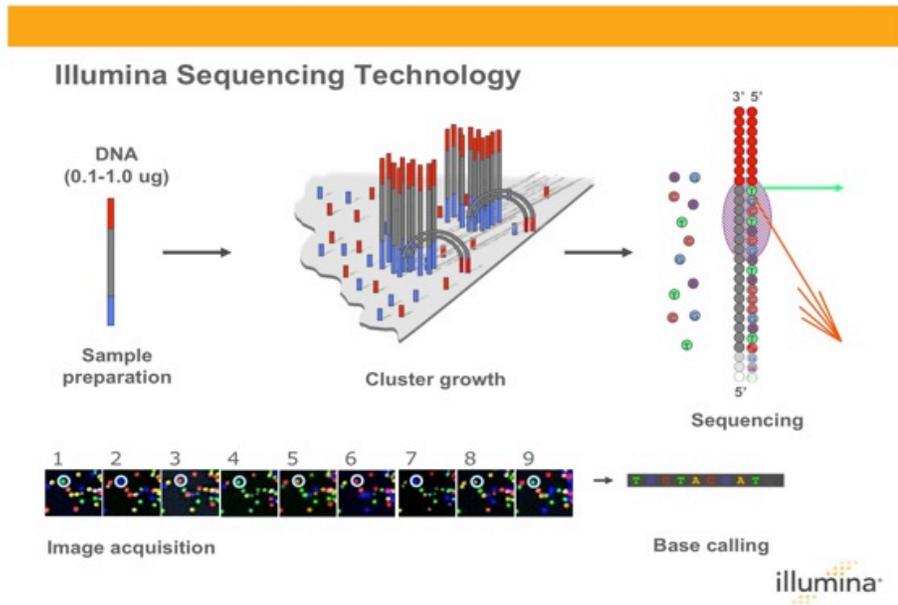
Figure 51 Illustrating the 3 phases of Sequencing. Initially DNA is prepared by adding adaptors at each end of small piece of DNA. Cluster formation allows bridge amplification and finally the ends of these reads are sequenced. (figure from www.illumina.com)

Summary

AML1-ETO acts as transcription factor binding to DNA, through its Runt domain. AML1-ETO is then able to regulate the transcription of its target genes. Identification of these target genes have traditionally been discovered through ChIP experiments followed by pull down of candidate genes using specific antibodies. Global analysis of the ChIP DNA has been more complex and most of the work regarding transcriptional control of AML1-ETO has focused on differential gene expression related to t(8;21) (Nimer and Moore, 2004). This however, does not reveal details on whether direct binding of the gene by AML1-ETO controls expression of genes. ChIP followed by high throughput sequencing overcomes these limitations enabling identification of all the sites on the genome to which AML1-ETO binds.

Using this technique on primary samples, genes that bind the oncoprotein AML1-ETO have been identified and discussed. The visualisation of binding sites on these genes provides information on the location of these binding sites. Using functional annotation software, a global assessment of the genes has identified enrichment of specific pathways that are implicated in the pathogenesis of AML-ETO. Furthermore, using the expression data provided by the exon array experiments in chapter 3 details are provided of not only which genes AML1-ETO bind but also discover if this binding controls its expression thus providing new insights into control of genes of AML1-ETO.

Patients and Methods used for ChIP-Seq Analysis

ChIP was performed using an antibody to ETO. As the *ETO* gene is transcriptionally silent in haemopoietic cells, the choice of this antibody, as opposed to using an antibody to AML1, results in pull down of DNA bound to AML1-ETO specifically. The initial control was an antibody to IgG but subsequently input DNA was used as DNA yields using the IgG antibody were unreliable. In order to establish the ChIP methodology initial experiments were performed using the Kasumi cell line. Subsequently, three AML patient samples harbouring the t(8;21) were selected for analysis by ChIP-Seq methodology. (Table 22)

Table 22 Patient samples used for ChIP-Seq

Patient	Sex	Age	Cytogenetics
1	M	18	45,X,Y,t(8;21)(q22;q22)
2	M	68	45,X,-Y,t(8;21)(q22;q22)
3	M	34	45,X,-Y,t(8;21)(q22;q22)

Sonication Experiments

Initial experiments were performed with Kasumi cell lines to optimise sonication cycles and to verify enrichment of ChIP DNA (Chapter 2 p88). 24 cycles of sonication were required to obtain the correct size of DNA fragments of 200-500bp. (Figure 52).

A) 12x
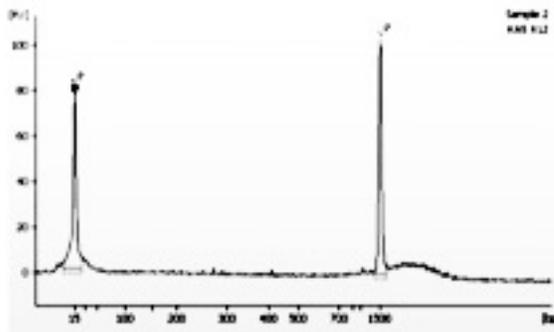
B) 18x
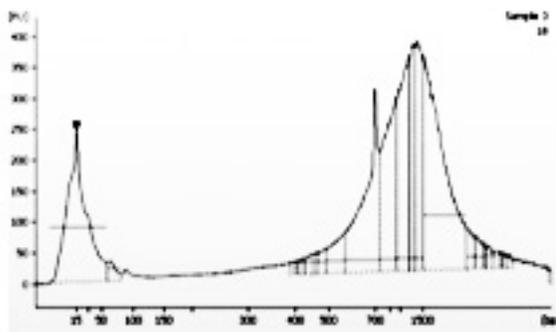
C) 24x
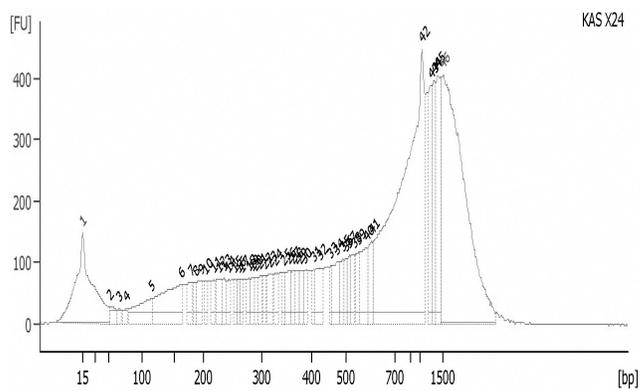
Figure 52 A,B & C -DNA was extracted from Kasumi cell lines and sonicated with protocols of 12, 18 and 24 cycles respectively. DNA was analysed on bioanalyser to analyse DNA fragment size and displayed above.

RQ experiments

To verify that the DNA enriched was specific to AML1-ETO RQ experiments were performed using the *c-fms* gene and *beta 2 microglobulin* as the negative control (Follows et al., 2003) (Chapter 2 p65). Threefold enrichment was used as the threshold to confirm the success of the ChIP technique. Unfortunately, there was loss of enrichment with sonication after 18 cycles (Figure 53 & 54). As this yields suboptimal DNA sized fragments for the sequencing technology the ChIP protocol was amended so that initial sonication was performed with limited number of cycles (x18) and subsequently an extra sonication step (x3 further cycles) was incorporated after the enrichment step to ensure 200-500bp DNA fragments.

Patient samples were then used to perform ChIP experiments using this amended protocol and enrichment for *c-fms* was verified. (Figure 55) There was no enrichment with the negative control *beta 2 microglobulin* gene (Figure 56).

Library Preparation

Enriched ChIP DNA was then prepared into libraries ready for sequencing on the Illumina system (Chapter 2 p86). Furthermore, libraries for input DNA for each patient sample were also prepared. All but one of the libraries was run on two lanes of the sequencer.
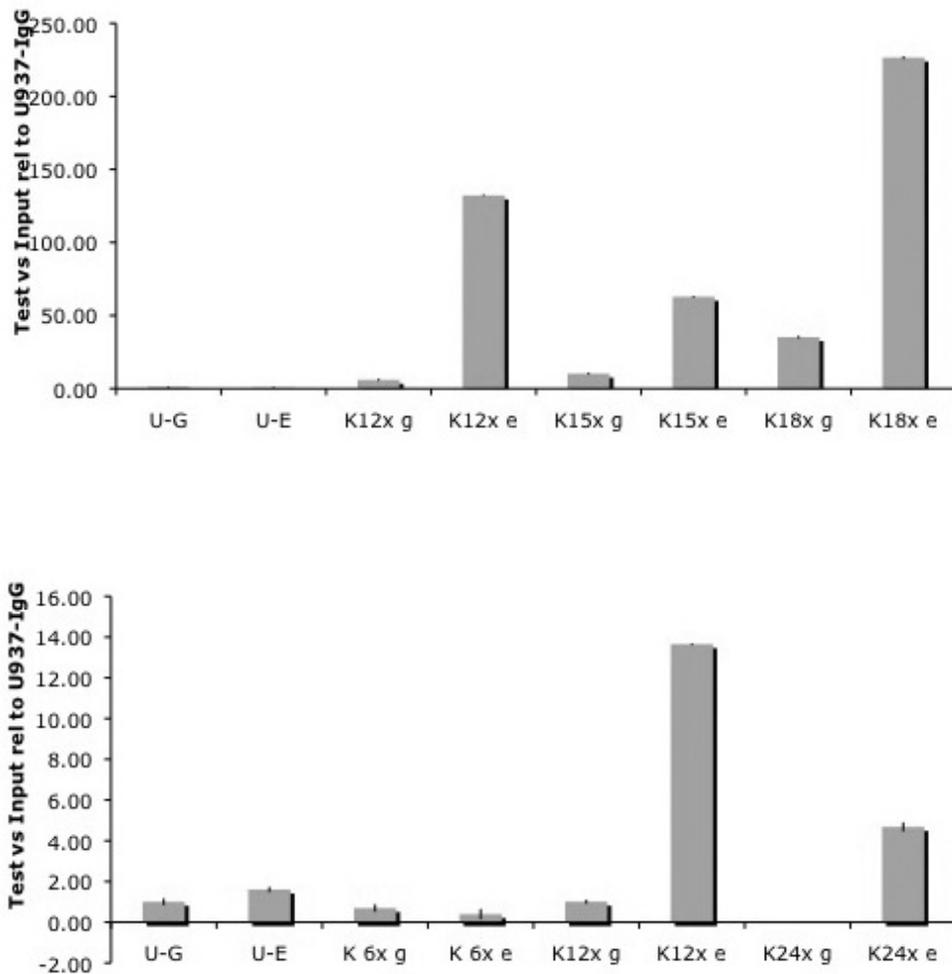
Figure 53 Initial RQ experiments on Kasumi cell lines. Kasumi (K) and the negative control U937 (U) cell lines have been used. ChIP was performed with antibodies to ETO (e) and IgG (g). Sonication cycles of 6, 12, 15, 18 and 24 have been performed. RQ to amplify c-fms has been performed and presented relative to Input relative to U937-IgG (negative control). Enrichment appeared to be less robust with cycles greater than 18. Identical negative control experiments were performed in parallel amplifying the beta 2 microglobulin gene and are represented in figure 54 below.
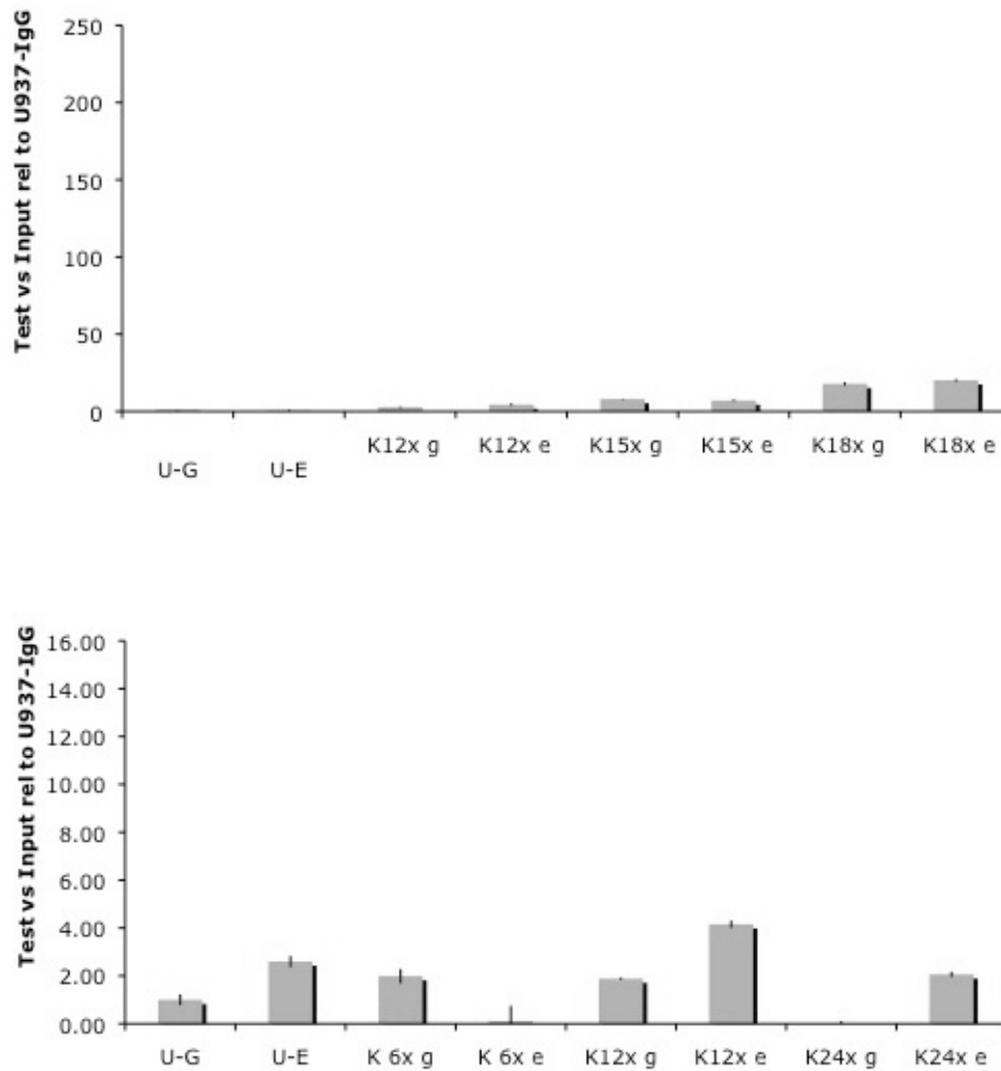
Figure 54 RQ experiments on Kasumi cell lines amplifying the beta microglobulin gene. Conditions and samples as in figure 53.
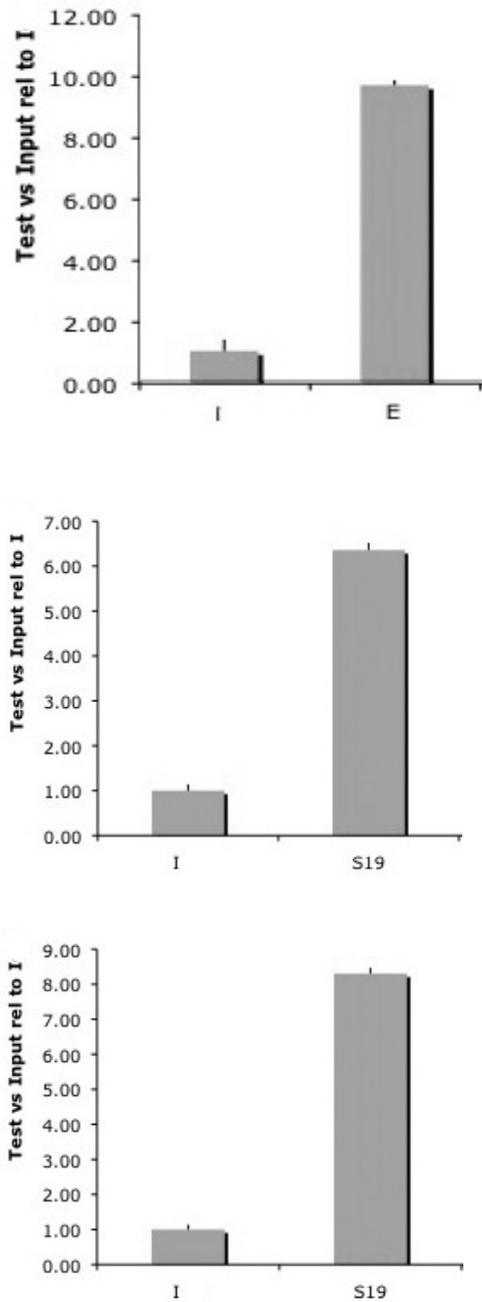
Figure 55 RQ analysis for three patients respectively. ChIP DNA was obtained using antibody to ETO (E/S19) and 18 cycles on the sonicator and then amplified for c-fms gene. Graphs presented relative to input (I) relative to IgG, showing samples had a greater than threefold enrichment relative to input.
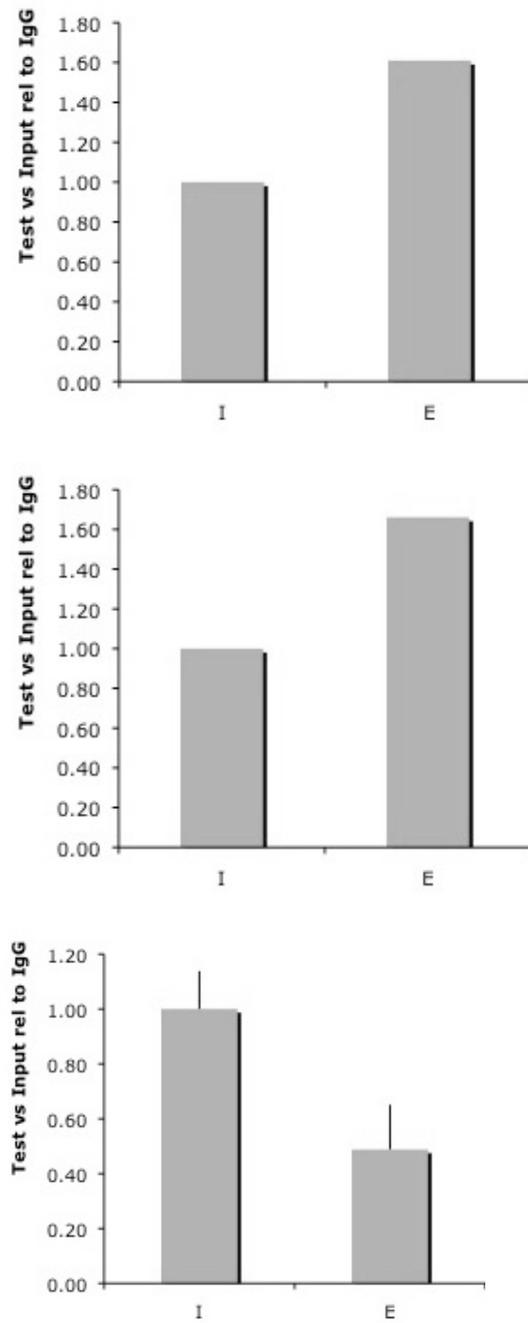
Figure 56 RQ analysis amplifying the beta2 microglobulin gene for the same three patients respectively using identical parameters as in figure 55.

Sequencing Data

Sequence reads for each patient was obtained. Alignment, quality control assessment and filtering were performed as described in materials and methods and the summaries for each primary sample tested are represented in the tables 23-25 below.

Table 23

Patient 1	RUN	Raw Reads	Filtered reads
ChIP	1	18 million	14.2 million
	2	18 million	14.7 million
Input	1	9 million	7.5 million
	2	23 million	8.2 million

Table 24

Patient 2	RUN	Raw Reads	Filtered reads
ChIP	1	34 million	26.2 million
	2	34million	25.8 million
Input	1	26 million	20.6 million
	2	25 million	19.9

Table 25

Patient 3	RUN	Raw Reads	Filtered reads
ChIP	1	24 million	18.6 million
	2	13 million	10 million
Input	1	15 million	11.3 million

Gene List Generation

For each separate sample, the total filtered reads obtained from both the ChIP and input libraries were imported into Partek. Peaks were detected by Partek, which also produced a p-value by comparing the peaks form the input and ChIP sample. Peaks detected were filtered by using an FDR of 0.05. These regions containing the peaks were used to detect genes using a Ref seq annotation file and by extending the regions by 1000bp either side of the peak. Some genes had more than one peak identified. The gene was only included once using the peak with lowest p-value and the three gene lists were obtained (Table 26)

Table 26 Numbers of genes detected from each patient obtained from Partek analysis.

Patient	Numbers of Genes
1	377
2	15,229
3	8,862

Table 26 shows a marked variation in the number of genes obtained for each patient. Initially, the three individual gene lists were compared and 338 genes common to all lists obtained. (Table 31 Appendix C).

Although using a filter of FDR p<0.05 allowed a more inclusive initial gene list to be generated, this obviously results in a large number of false positives as, for example for patient 2, it is unlikely that 15,229 genes bind AML1-ETO. Therefore a further

195

more stringent criterion was set using visual inspection. By individually inspecting the shape and the location of the peaks for all these 338 common genes on the visualisation browser I was able to refine and rationalise the list to include only the top 86 genes with the largest p-value of 1×10^{-20}. (Table 27)

Subsequently, the top individual genes were then inspected and their function analysed through PUBMED searches. (http://www.ncbi.nlm.nih.gov/pubmed) This enabled us to confirm that some of these genes had been previously associated with AML1-ETO, giving confidence to our methodology, whilst novel gene associations were made. For a global assessment of the gene lists the gene ontology tool, GeneGo was used to identify enriched pathways. Finally the gene list was compared to the expression data obtained previously in chapter 3 (Table 30 Appendix A). This further validates the experiments and gives an insight into the relationship between DNA binding and effects on gene expression.

Results

Primary assessment of the sequencing data was performed using the visualisation browser on Partek. Reads obtained from the sequencer are mapped to the genome and depicted as peaks in the following figures. These figures visualise the location of the peaks for the top genes and validate the methodology.
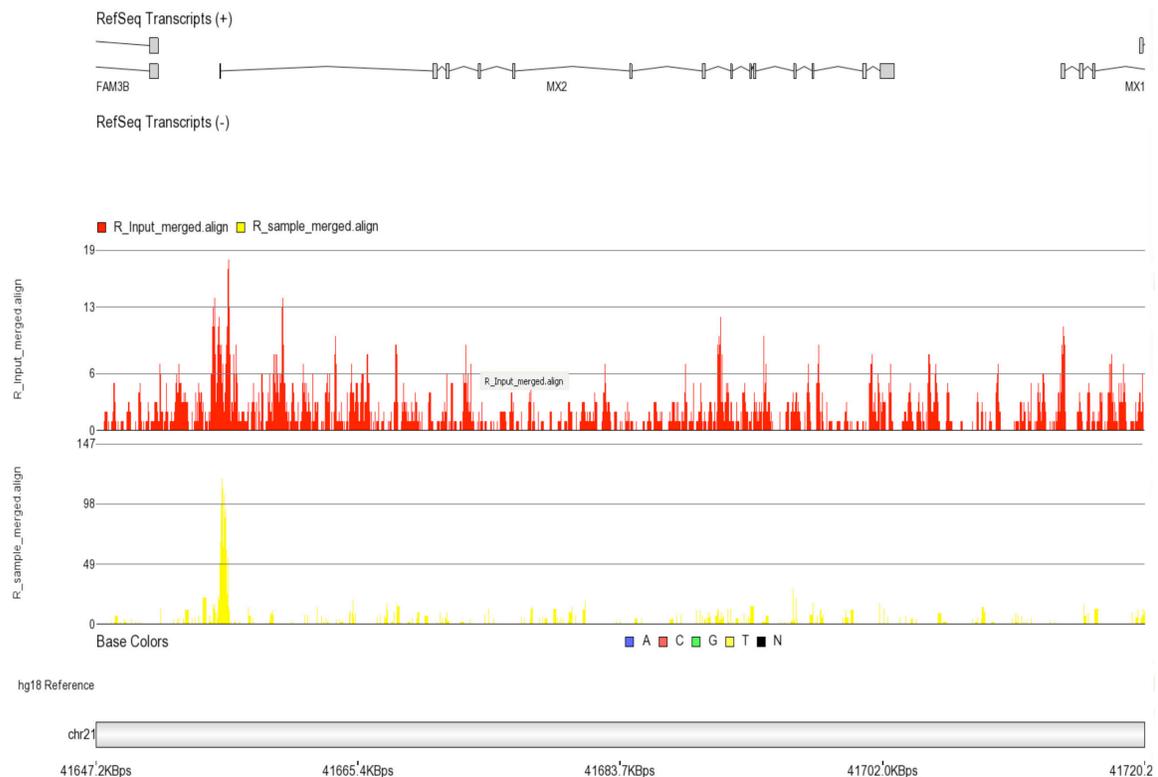
Figure 57 *MX2* gene showing peaks at the 5'end of gene. At the top is a schematic picture of the gene. The next two tracks depict the reads mapping to that genomic location – the upper red track for input DNA and the lower yellow track for enriched DNA. (note different scales for each track) The genomic location marker is at the bottom of the figure. Subsequent figures all have the same configuration.
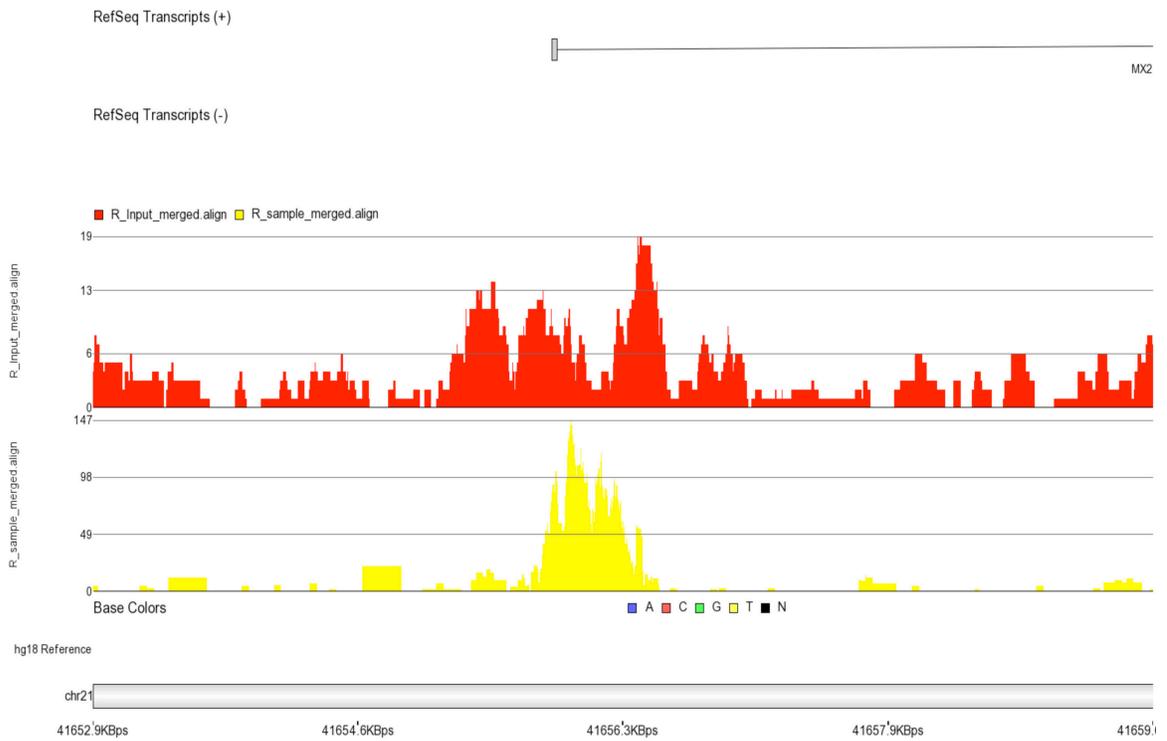
Figure 58 *MX2* gene. This focuses in on the peak detected in figure 57.
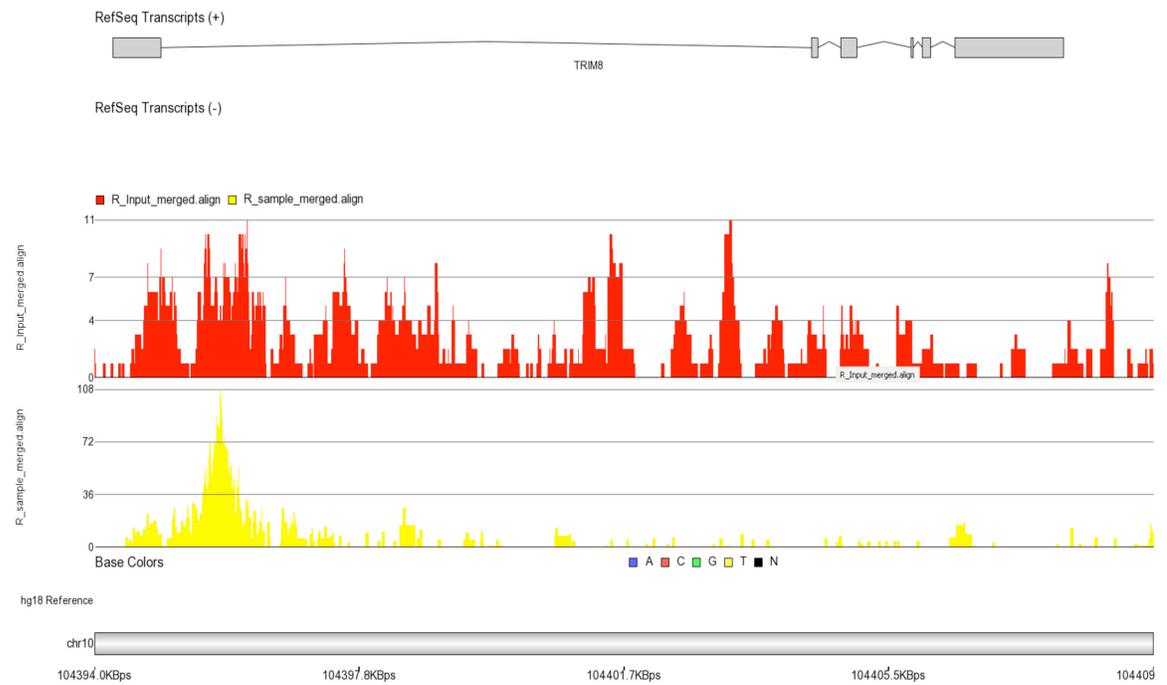
Figure 59 *TRIM8* gene showing peaks at the 5'end of gene.
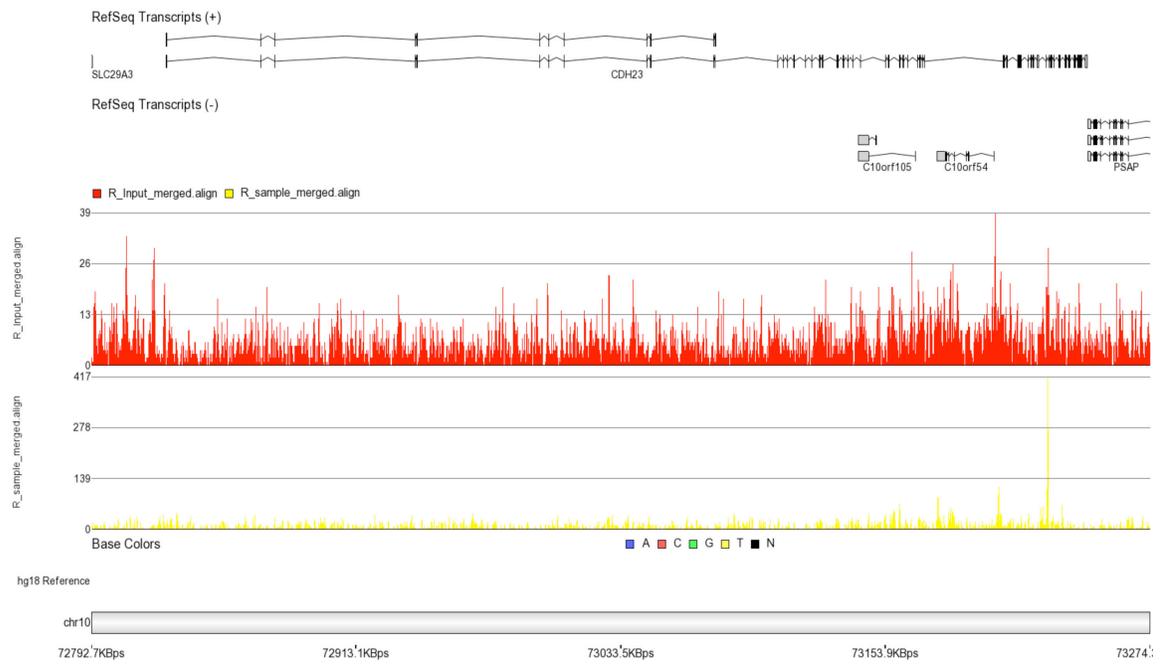
Figure 60 *CDH23* gene showing peaks at the 3'end of gene
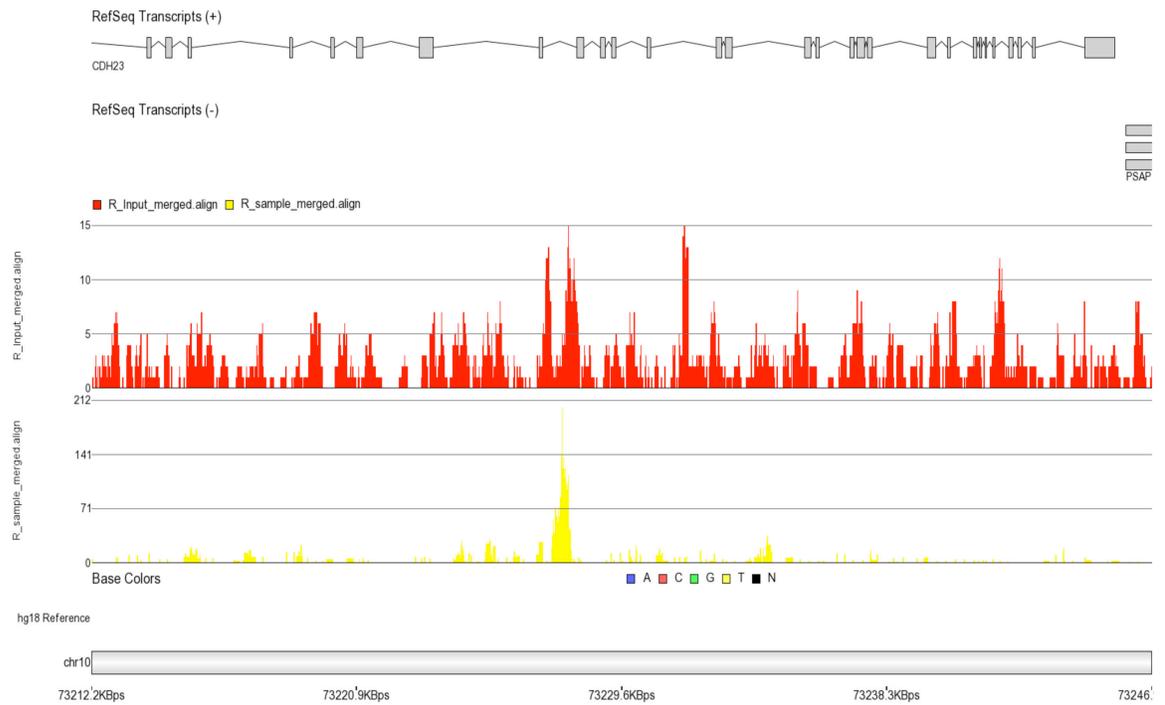
Figure 61 *CDH23* gene This focuses in on the peak detected on figure 60

199
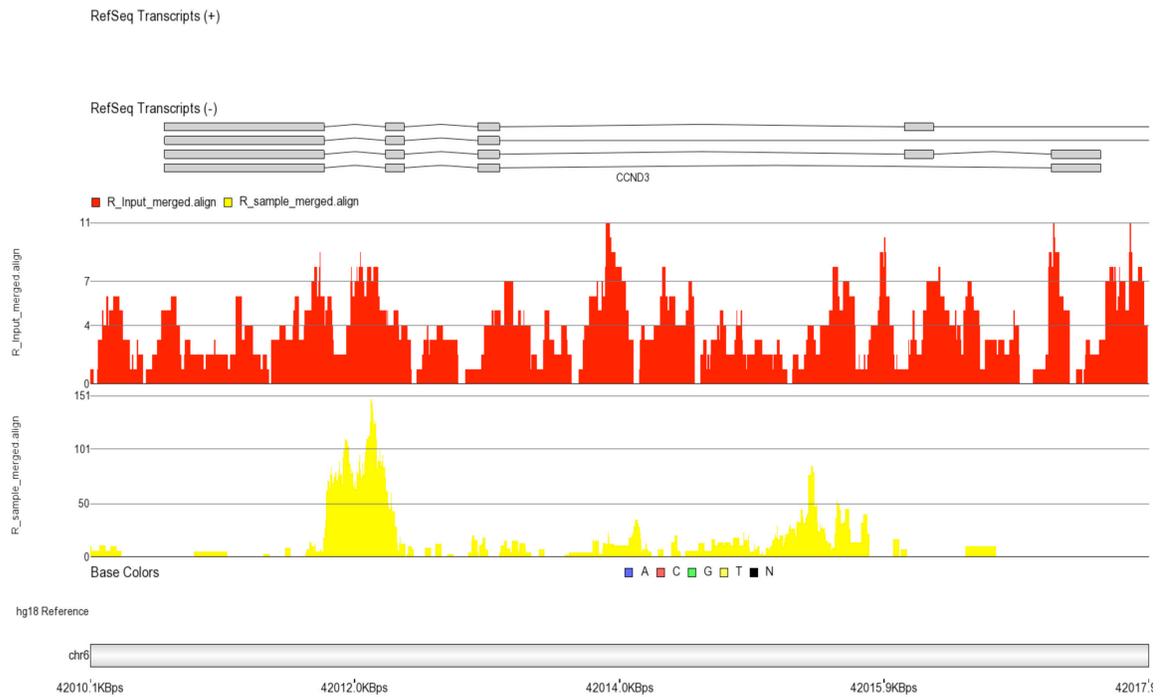
Figure 62 *CCND3* gene showing peaks at the 5'end of gene
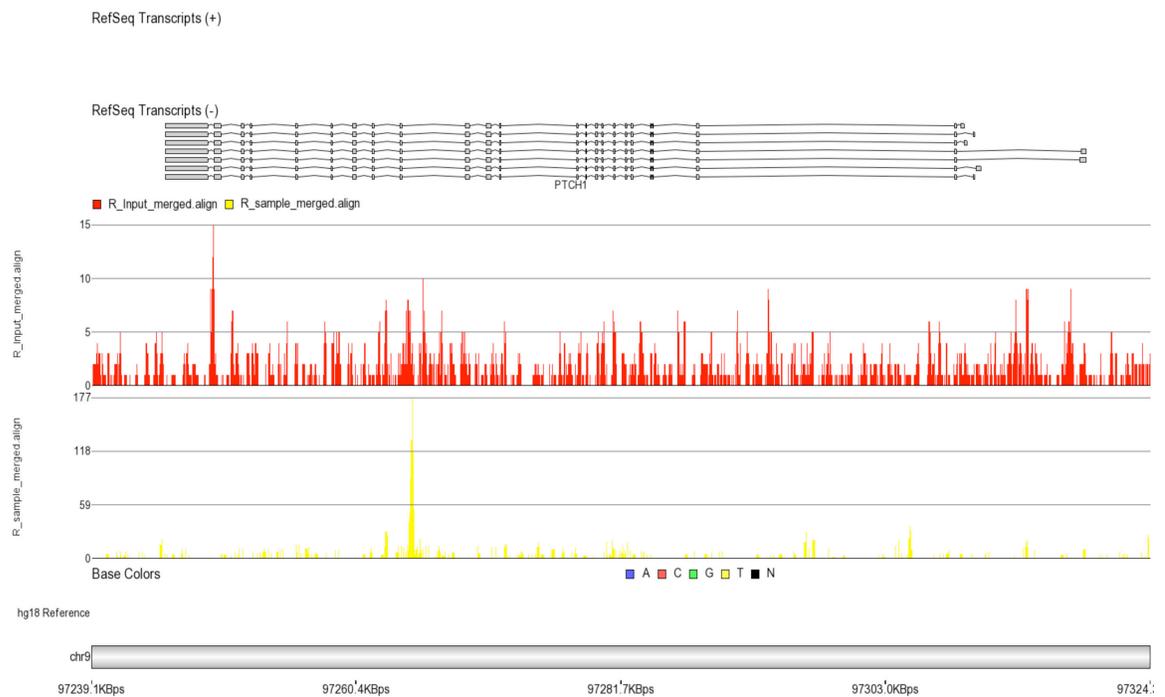
Figure 63 *PTCH1* gene showing peaks in the middle of gene

Visual inspection of the binding sites for these top genes shows conclusively high numbers of reads at specific sites on the genome compared to the control sample. This gives confidence that enrichment by ChIP and peak detection is a feasible technique and to an extent validates our data, especially for genes that had low p-values. Conversely, for some binding sites with higher p-values there was less confidence that these results represent true binding sites. As non-stringent criteria of FDR of p<0.05 was deliberately used in the initial analysis to generate these inclusive lists, the use of this visual analysis was used to produce a more stringent criterion, omitting genes whose p-value was higher than $1x10^{-20}$. Using this p-value to limit the numbers of genes generated resulted in a definitive list of 86 genes (Table 27).

Most AML1-ETO Binding sites are at 5' end of Genes

Inspection of the above figures demonstrates that the main AML1 binding sites for the genes depicted in the above figures are located at the 5' end. This would be consistent with the concept that AML1-ETO binds to promoter sites, which are normally located at this end of the gene. However, this does not appear to be exclusive with some binding regions located at the 3' end and some located in introns in the middle of the gene suggesting that control elements of genes can be found anywhere along the gene.

Gene List

Table 27 Top 86 Genes ordered by lowest p-values from ChIP-Seq experiments

GENE	CHROMOSOME	PEAKS INPUT	PEAKS CHIP	p-VALUE
MX2	21	121	776	0
CDH23	10	74	651	0
TRIM8	10	125	714	0
CCND3	6	141	1123	0
PTCH1	9	77	576	0
STAB1	3	99	974	0
PRKCD	3	129	709	0
HCCA2	11	71	636	0
RPS6KA1	1	126	1011	0
INPP5A	10	74	617	0
GFOD1	6	110	693	0
ZBTB16	11	56	459	0
NEK8	17	88	602	0
MAP2K2	19	29	420	0
BAHCC1	17	117	936	0
USP20	9	30	356	9.81E-45
SH3BP5	3	70	482	7.26E-43
KIAA0427	18	83	519	1.89E-42
LOC374443	12	29	338	3.63E-42
FAM129B	9	22	310	3.92E-42
PTTG1IP	21	80	507	5.35E-42
GALNT6	12	104	567	1.50E-40
BCL3	19	86	512	4.44E-40
ADCY4	14	72	470	4.70E-40
HIP1	7	69	459	8.93E-40
LMNA	1	108	564	1.48E-38
SULF2	20	71	452	8.99E-38
LAPTM5	1	108	558	1.01E-37
C6orf103	6	64	417	1.13E-35
HNRNPM	19	99	517	1.74E-35
CXXC5	5	177	711	2.07E-34
PPP4R1L	20	72	431	3.54E-34
CYP51A1	7	88	474	9.30E-34
GLIPR2	9	70	419	2.86E-33
KIAA0652	11	56	376	3.93E-33
VAV1	19	95	485	1.38E-32
JMJD6	17	84	452	3.09E-32
GRK7	3	165	661	5.52E-32
GNB5	15	110	520	7.40E-32
TADA3	3	98	485	1.80E-31
ACSL1	4	71	407	4.40E-31
RASSF2	20	73	411	7.87E-31
TPM4	19	63	374	1.19E-29
RAI1	17	77	414	1.21E-29

Gene	Chr	Start	End	p-value
EEPD1	7	54	347	1.29E-29
RARA	17	173	659	1.72E-29
LAT2	7	43	306	1.60E-28
HSH2D	19	32	269	2.47E-28
MSI2	17	105	473	1.90E-27
SLC2A1	1	120	507	6.61E-27
SH3BP2	4	67	364	1.53E-26
TSPAN32	11	103	460	1.90E-26
ZFYVE28	4	98	446	2.59E-26
KIAA0182	16	75	384	3.13E-26
NHSL2	X	14	190	5.31E-26
ALDH16A1	19	50	310	7.87E-26
RASSF5	1	119	494	1.27E-25
C5orf56	5	121	497	2.21E-25
ABCG1	21	33	254	2.53E-25
GPR114	16	99	437	7.84E-25
TOX2	20	77	377	1.55E-24
CD276	15	75	371	1.87E-24
TBC1D14	4	120	486	2.33E-24
SPNS3	17	38	263	3.20E-24
SMG6	17	154	561	1.43E-23
MBP	18	171	599	2.15E-23
STARD9	15	20	198	2.35E-23
PKIB	6	80	376	2.44E-23
MAX	14	66	337	3.74E-23
FNTB	14	66	337	3.74E-23
IL21R	16	55	305	5.82E-23
CCDC135	16	48	283	1.09E-22
TK2	16	134	504	1.68E-22
LOC100130987	11	84	380	1.69E-22
IL17RA	22	101	423	1.86E-22
SORBS1	10	68	335	3.87E-22
ADORA2A	22	92	396	6.34E-22
C22orf45	22	92	396	6.34E-22
MGAT1	5	198	646	6.38E-22
COL23A1	5	52	287	1.33E-21
SIK3	11	152	538	1.41E-21
CNIH3	1	82	365	3.07E-21
SIPA1L3	19	82	362	7.39E-21
LST1	6	100	407	8.11E-21
CD82	11	109	429	8.32E-21

AML1-ETO binds Genes involved in Cell Growth

Many of these genes appear to play a key role in cell cycle or tumour suppressor functions. *MX2* encodes for a protein that is part of the dynamin and GTPase family (Melen et al., 1996). *CDH23* encodes a cadherin and involved in cell-cell adhesion (Di Palma et al., 2001). *TRIM8* is thought to be a tumour-suppressor gene working through enhancement of the Stat3 pathway (Okumura et al.). *CCND3* encodes for a cyclin involved in cell cycle and mutations in this gene have been implicated in the pathogenesis of AML (Smith et al., 2005). *PTCH1* encodes for the receptor for sonic hedgehog and functions as tumour suppressor gene (Katoh and Katoh, 2009). *PRKCD* is involved in the regulation of cell growth and apoptosis (Brodie and Blumberg, 2003). *RPS6KA1* encodes for a kinase involved in cell growth, differentiation and apoptosis (Frodin and Gammeltoft, 1999). *PLZF* (*ZBTB16*) has been shown to interact with AML1-ETO, which causes its repression (Yeyati et al., 1999). *CXXC5* is involved in Wnt pathway and p53 regulation (Kim et al.). These molecules may have a key role to play in the pathogenesis of t(8;21) related acute myeloid leukaemia.

Gene Ontology Functional Analysis

Further analysis to inspect the gene list on a global rather than individual basis uses commercial software designed to look for canonical pathways. GeneGo was used to inspect the initial list of 338 genes as previously generated, with the data summarised in table 28.

Table 28 Significant processes as detected by GeneGo generated from the definitive gene list of 86. Genes involved in the processes are referred to as network objects.

	NAME	p-VALUE	NETWORK OBJECTS
1	Signal transduction_CREM pathway	2.60E-04	5/98
2	Reproduction_FSH-beta signaling pathway	3.14E-04	6/159
3	Development_Hemopoiesis, Erythropoietin pathway	1.13E-03	5/135
4	Cell cycle_G2-M	1.20E-03	6/205
5	Inflammation_MIF signaling	1.33E-03	5/140
6	Signal transduction_ESR1-nuclear pathway	1.57E-03	6/216
7	Inflammation_Neutrophil activation	1.85E-03	6/223
8	Immune response_IL-5 signaling	2.17E-03	3/44
9	Reproduction_Male sex differentiation	3.04E-03	6/246
10	Apoptosis_Anti-Apoptosis mediated by external signals via MAPK and JAK/STAT	3.12E-03	5/170
11	Signal Transduction_Cholecystokinin signaling	3.43E-03	4/106
12	Inflammation_Interferon signaling	3.92E-03	4/110
13	Cell cycle_G1-S Growth factor regulation	5.60E-03	5/195
14	Reproduction_Gonadotropin regulation	6.10E-03	5/199

AML1-ETO binds genes related to cell cycle and apoptosis

Notable pathways from table 24 include CREM (cAMP response-element modulator) pathway, genes involved in cell cycle and apoptosis. Although the CREB and CREM pathways are associated with spermatogenesis it has been shown that there is increased expression of genes encoding CREB in AML cells (Shankar et al., 2005). These observations once again offer glimpses of the pathways that appear to be dysregulated in the pathogenesis of AML with t(8;21).

ChIP-Seq correlation with Gene Expression

The initial list of 338 genes obtained from the ChIP-seq analysis was compared to the 448 genes obtained from the expression data from the exon arrays. This results in a list of 30 genes that are common to both lists. Similarly, comparisons of the expression data to the ChIP-Seq list of 86 genes generated using the stricter criteria, resulted in 15 common genes. (Table 29) The probability of randomly obtaining these common genes from two lists is small. For example, using appropriate consideration of permutation and combinatorial theory and assuming the presence of 25,000 genes in the human genome, the likelihood of randomly generating more than 30 matches from the initial comparison is 7×10^{-13}. This statistical analysis provides further evidence of the robustness of the ChIP-Seq technique.

Many of the genes that have been detected, by the ChIP-Seq methodology, to bind AML1-ETO do not have an effect on gene expression as detected by the exon array data. However, there are some genes that are enriched in both expression and ChIP-Seq methodologies with the majority of these genes appearing to be repressed. (Table 29) This suggests that the traditional hypothesis that AML1-ETO acts by binding to genes and causing their down-regulation due to its repressor domains of ETO are valid. In contrast, this data suggests that the up regulation of genes associated with AML1-ETO occurs through indirect associations.

Table 29 Table of Genes common to Expression and ChIP-Seq analysis (List 1). The top15 genes are derived when the more stringent Chip-Seq list of 86 is used for comparison.

GENE	CHROMOSOME	INPUT	CHIP	p-VALUE	EXPRESSSION
TRIM8	10	125	714	0	-2.62267
CCND3	6	141	1123	0	-2.17201
PRKCD	3	129	709	0	-2.75581
LAPTM5	1	108	558	1.01E-37	-2.34841
VAV1	19	95	485	1.38E-32	-1.66865
GNB5	15	110	520	7.40E-32	-1.98786
LAT2	7	43	306	1.60E-28	-6.46607
SH3BP2	4	67	364	1.53E-26	-1.72992
TSPAN32	11	103	460	1.90E-26	-4.58844
GPR114	16	99	437	7.84E-25	-2.49612
TBC1D14	4	120	486	2.33E-24	1.45898
SPNS3	17	38	263	3.20E-24	-4.40062
MBP	18	171	599	2.15E-23	-1.77519
LST1	6	100	407	8.11E-21	-2.69105
CD82	11	109	429	8.32E-21	-2.15598

GENE	CHROMOSOME	INPUT	CHIP	p-VALUE	EXPRESSSION
HSPG2	1	70	319	2.54E-19	2.3495
FUT7	9	93	376	3.56E-19	-2.65019
KIAA0513	16	81	324	1.70E-16	-2.12616
ZC3H3	8	82	323	4.35E-16	1.46274
MTSS1	8	4	83	8.35E-14	3.70334
NTNG2	9	43	198	9.80E-13	2.08723
C9orf89	9	99	326	4.41E-12	-2.1954
PPCDC	15	15	112	9.61E-12	-1.58971
UHRF1	19	18	121	1.09E-11	3.08633
SMAD3	15	60	217	5.52E-10	-1.65456
PI4KA	22	20	115	6.91E-10	-1.4591
FGD3	9	4	53	4.28E-08	-1.50351
SIPA1	11	118	320	7.47E-08	-1.6328
MAN1A1	6	297	650	5.97E-07	4.82998
WIPF1	2	11	59	2.04E-05	1.40696

AML1-ETO controls genes that regulate haemopoiesis

Whilst many of these genes in this list have not been extensively investigated others have been shown to play a role in haemopoiesis. The *VAV1* proto-oncogene encodes a protein that is a guanine nucleotide exchange factor for the Rho family of GTP binding proteins and has an important role in T-cell and B-cell development and activation. It is also able to affect JNK/SAPK signalling cascade (Katzav, 2007). The *LAPTM5* gene encodes a transmembrane receptor, also known as E3 protein, that plays a role in haematopoiesis (Scott et al., 1996). The protein encoded by *SH3BP2* binds to the SH3 domains of several proteins including the ABL1 and SYK protein tyrosine kinases, and functions as a cytoplasmic adaptor protein to positively regulate transcriptional activity in T, natural killer (NK), and basophilic cells (Chen et al., 2007). *TSPAN32* is a member of the tetraspanin superfamily and behaves as tumour suppressor gene. It is also involved in hematopoietic cell function (Robb et al., 2001). These findings give a rational hypothesis that AML1-ETO binds to and represses the transcription of genes involved in haemopoietic control leading to dysregulation of normal haemopoiesis.

Discussion

ChIP-Seq although relatively new appears to be a very powerful method to locate transcription factor binding to the genome. Using an antibody to ETO, DNA that is enriched for genes that bind AML1-ETO is pulled down. A number of novel genes that bind AML1-ETO are identified. It appears that AML1-ETO binding does not lead to differential gene transcription in all cases. However, where it does have a direct effect on gene expression it appears, as has traditionally been thought, to cause transcriptional repression. Conversely, the up-regulation of genes through AML1-ETO appears to be an indirect effect in most cases.

AML1-ETO binds to genes involved in cell cycle and apoptosis as well as down regulating genes related to control of haemopoiesis. These observations validate our approach to investigate AML1-ETO. Some specific pathways implicated include the WNT, Notch and Hedgehog pathways suggesting that these pathways may be appropriate targets for further treatment strategies.

The ChIP-methodology whilst being very powerful does result in a vast amount of data. There are further possibilities for both analysing and interrogating the data. Due to time constraints only limited analysis has been performed and there are many areas that have been left unexplored.

More primary AML samples may be used to give more confidence to the validity of the data. The sample preparation methodology needs to be optimised to enrich for DNA at sonication levels to produce 500bp lengths. To obtain more data, further runs on the sequencer for the 3 samples already obtained would deliver more sequence reads. This would be particularly useful for sample 1 which resulted in low numbers of genes discovered and out of keeping with the other samples. Inspection of this sequence run revealed that its quality was poor and consequently affected the data obtained from the sequencer. In fact, a further run of good quality has been subsequently performed with excellent quantities and quality of reads but not yet analysed. This may be particularly pertinent as this sample drastically limited the numbers of common genes and it was felt that this might have introduced bias excluding some important genes. In fact comparing only genes obtained from the analysis of samples 2 and 3 and then using the strict p-value criteria of $p<1\times10^{-20}$ resulted in 186 common genes. (Appendix C Table 28) GeneGo analysis of this list highlighted the enrichment of molecules involved in the WNT, Notch and Hedgehog pathways. (Appendix C Table 29) Re-analysing sample 1 with data from the new sequencing run may confirm these findings.

Analysis of the data can also be altered by changing some of the statistical parameters used in setting peak detection to deliver more stringent criteria. The use of different software for analysis would also provide greater confidence if similar gene lists were produced. In this analysis each patient sample was analysed separately and only then compared to detect common genes to all three lists. A

more robust analysis may be obtained if the peaks detection were performed with all samples concurrently.

Due to the extra sonication step in the ChIP methodology used here it is difficult to perform a motif analysis. This analysis would be invaluable to confirm some of the findings by detecting the recognised AML1 binding motifs at sites of peak detection. Motif analysis can also be used to identify novel AML1-ETO binding motifs. In this analysis genes that may have more than one peak detected were only recognised once and other peaks ignored so that only one binding site for each gene was detected. It is known that AML1-ETO tetramers preferentially bind genes with 2 AML1 binding sites . Motif analysis analysis would aid in confirming this observation. A further analysis that is outstanding is comparing this ChIP-Seq data to data obtained in our laboratory regarding methylation status in t(8;21) samples, particularly as epigenetics are a key feature of the pathogenesis in t(8;21). Furthermore, current work to confirm some of these enriched genes by ChIP followed by RQ-PCR is ongoing.

In summary, ChIP-Seq appears a very powerful technique and produces a vast amount of data. Our analysis is preliminary and further work is required to confirm our findings. However, the data appears valid and interesting observations have been made giving further clues on the mechanisms of the pathogenesis of AML1-ETO.

Chapter 6   Conclusions & Future Work

AML is characterised by the occurrence of non-random translocations. The commonest is the t(8;21)(q22;p22). The pathogenesis of this form of AML has been extensively investigated. The t(8;21) results in the *AML1* gene being placed alongside most of the *ETO* gene. AML1 acts as a master regulator of haemopoiesis while ETO through its four nervy domains is involved in recruiting co-repressors. The first five exons of *AML1* containing the DNA binding domain are involved in the translocation while nearly the whole of *ETO* is involved. The t(8;21) occurs early in stem cells and gives rise to a product of 752aa with the RHD at the NH2 end and the ETO domains at the C-terminal end. AML1-ETO loses the activation domains of AML1 that recruit HATs, and gains repressor domains associated with ETO, which are involved in recruiting HDACs. Deacetylation of histones causes a closed chromatin structure and repression of transcription of a number of genes. In cell lines AML1-ETO blocks differentiation and proliferation and increases apoptosis. In stem cells it causes cell proliferation and renewal. Self-renewal may occur through down regulation of lineage-programming transcription factors such as GATA1 or through up-regulation of pathways such as Wnt or Notch. In mouse models introduction of AML1-ETO does not lead to the development of leukaemia on its own but can be seen when mice are treated with alkylating agents. This suggests that leukaemogenesis follows a multi-hit model with t(8;21) being an early and necessary mutation but requiring a second hit such as *Kit* mutations. AML1-ETO favours genetic instability through repression of genes involved in DNA excision repair. Alternative splice variants, which may themselves result from dysregulated splicing of AML1-ETO, also aid in loss or gain of genetic material by deregulating the mitotic

checkpoint. In addition to the acquisition of further mutations, the alternative splice variants on their own or with co-operation of full length AML1-ETO may themselves act as the "second hit" which results in leukaemia.

The observations in this study adds further knowledge to the current understanding of AML1-ETO induced leukaemia as well as providing a description of how the newer high throughput technologies may be used to investigate the genome. The results also highlight the importance of the role that alternative transcripts play in the pathogenesis of leukaemia, which may also be extrapolated to genomics of cancer in general.

The number of genes in the human genome is less than anticipated. Alternative transcripts have a function in increasing the diversity of the products the genome produces. Here, further evidence is provided for the key role alternative transcripts have in leukaemogenesis. The *AML1-ETO* gene has alternative exons and together with alternative splicing results in a number of transcripts all produced concurrently. These transcripts are translated into different sized protein products, which appear to have different roles and may affect the overall balance of the function of the oncogene. This mechanism may play a role in other malignancies and may even explain the different prognosis in patients with essentially the same disease. Specifically in the t(8;21) the key region appears to be the NHR2 domain of ETO. Its ability to interact with other ETO proteins through its dimerisation function is crucial to the pathogenesis of the t(8;21) in a similar mechanism seen in AML with t(15;17).

The shorter proteins produced by alternative transcripts of *AML1-ETO* may be more susceptible to produce tetramers of AML1-ETO which result in more effective function of AML1-ETO protein in its role as a transcription factor.

AML1-ETO regulates the transcription of a number of genes as evidenced by expression arrays. Many of these genes are related to key molecules in G-protein pathways as well as molecules in involved in cell adhesion pathways suggesting that AML1-ETO disrupts cell control through these pathways and may offer targets for intervention. The effect of AML1-ETO on transcription results in both up and down regulation of its target gene. ChIP-Seq experiments highlight the different mechanisms for this differential gene regulation. Down regulation of genes result from a direct effect of AML1-ETO binding to its target gene and causing gene repression through its ETO domains as traditionally hypothesised. Up regulation of genes is likely in many cases to result from an indirect effect of AML1-ETO rather directly binding of AML1-ETO to these genes. In addition, AML1-ETO binds to a number of genes without having an effect on transcriptional control. This observation may simply stem from the limitations of expression arrays. Alternatively the effects in transcription are not detected because of negative feedback or through pleiotropy of transcriptional control.

Expression arrays appear to be a very robust method for obtaining expression data. Its use for detecting alternative transcripts however, is more limited currently but

may become more useful as we build up our understanding of alternative transcripts particularly through sequencing technologies.

The ChIP-Seq technology is extremely powerful and generates vast amounts of data. The output we have generated remains incredibly exciting and regarding transcriptional control will provide a potent way to unravel key pathways. Our analyses due to time constraints have been limited and require further analysis. In this regard a great deal of future work is possible.

Future Work

The data from the ChIP-Seq needs further mining and validating to ensure robustness of the data. As discussed in chapter 5 further samples, further analysis of the sequence runs already performed and using different software other than Partek would give more confidence to the gene list produced. Furthermore ChIP experiments to optimise enrichment at appropriate sonication parameters without the need to add additional sonication steps allows for analysis of binding motifs, both to confirm that discovered peaks have recognised AML1 binding motifs as well as discovering new motifs. RQ PCR for enrichment of specific genes that are highlighted from the ChIP-Seq experiments should be performed. While the ChIp-Seq data has been compared to my own expression data I would also compare it with published AML expression data available from the GEO omnibus. This analysis would also highlight further any specific genes, which may be a suitable candidate

for further functional assessment. In this regard the expression data did highlight two genes *UHRF1* and *CD82* that are potential candidates for further assessment. Intriguingly the work investigating differential exon usage revealed two genes *WNK2* and *DHRS2* that provide further avenues for additional analysis. In this regard work with mRNA-Seq techniques investigating transcriptomes may be appropriate to clarify the significance of this observation.

The underlying drive to unravel the genomics of this heterogeneous disease is initially to understand the genomic mechanisms of disease states but also eventually to produce treatments, which are more effective for our patients. Modern technology has made designer drugs aimed at a molecular target a possibility and in some diseases a reality. For most malignancies a single targeted therapy is unlikely to prove a cure due to the pleiotropy and redundancy in biological systems that makes it possible for many cancer cells to circumvent one specific form of attack. However, the current global analysis techniques provided, gives us insights on multiple and global targets that appear to work simultaneously.

Since Velpeau published the first case of leukaemia almost 200 years ago there have been huge advances in the understanding of the pathogenesis of AML. The newer techniques, exploring the genome, offer even greater potential to unravel the mysteries of human biology and may eventually provide safe and effective treatment for our patients.

References

"GROUPEFRANCAIS" 1990. Acute myelogenous leukemia with an 8;21 translocation. A report on 148 cases from the Groupe Francais de Cytogenetique Hematologique. *Cancer Genet Cytogenet,* 44**,** 169-79.

ABU-DUHIER, F. M., GOODEVE, A. C., WILSON, G. A., PEAKE, I. R. & REILLY, J. T. 2003. c-FMS mutational analysis in acute myeloid leukaemia. *Br J Haematol,* 123**,** 749-50.

AHN, E. Y., YAN, M., MALAKHOVA, O. A., LO, M. C., BOYAPATI, A., OMMEN, H. B., HINES, R., HOKLAND, P. & ZHANG, D. E. 2008. Disruption of the NHR4 domain structure in AML1-ETO abrogates SON binding and promotes leukemogenesis. *Proc Natl Acad Sci U S A,* 105**,** 17103-8.

AHN, M. Y., HUANG, G., BAE, S. C., WEE, H. J., KIM, W. Y. & ITO, Y. 1998. Negative regulation of granulocytic differentiation in the myeloid precursor cell line 32Dcl3 by ear-2, a mammalian homolog of Drosophila seven-up, and a chimeric leukemogenic gene, AML1/ETO. *Proc Natl Acad Sci U S A,* 95**,** 1812-7.

ALCALAY, M., MEANI, N., GELMETTI, V., FANTOZZI, A., FAGIOLI, M., ORLETH, A., RIGANELLI, D., SEBASTIANI, C., CAPPELLI, E., CASCIARI, C., SCIURPI, M. T., MARIANO, A. R., MINARDI, S. P., LUZI, L., MULLER, H., DI FIORE, P. P., FROSINA, G. & PELICCI, P. G. 2003. Acute myeloid leukemia fusion proteins deregulate genes involved in stem cell maintenance and DNA repair. *J Clin Invest,* 112**,** 1751-61.

AMANN, J. M., NIP, J., STROM, D. K., LUTTERBACH, B., HARADA, H., LENNY, N., DOWNING, J. R., MEYERS, S. & HIEBERT, S. W. 2001. ETO, a target of t(8;21) in acute leukemia, makes distinct contacts with multiple histone deacetylases and binds mSin3A through its oligomerization domain. *Mol Cell Biol,* 21**,** 6470-83.

APPELBAUM, F. R., KOPECKY, K. J., TALLMAN, M. S., SLOVAK, M. L., GUNDACKER, H. M., KIM, H. T., DEWALD, G. W., KANTARJIAN, H. M., PIERCE, S. R. & ESTEY, E. H. 2006. The clinical spectrum of adult acute myeloid leukaemia associated with core binding factor translocations. *Br J Haematol,* 135**,** 165-73.

ARONSON, B. D., FISHER, A. L., BLECHMAN, K., CAUDY, M. & GERGEN, J. P. 1997. Groucho-dependent and -independent repression activities of Runt domain proteins. *Mol Cell Biol,* 17**,** 5581-7.

ASOU, H., TASHIRO, S., HAMAMOTO, K., OTSUJI, A., KITA, K. & KAMADA, N. 1991. Establishment of a human acute myeloid leukemia cell line (Kasumi-1) with 8;21 chromosome translocation. *Blood,* 77**,** 2031-6.

ASOU, N. 2003. The role of a Runt domain transcription factor AML1/RUNX1 in leukemogenesis and its clinical implications. *Crit Rev Oncol Hematol,* 45**,** 129-50.

ASOU, N., YANAGIDA, M., HUANG, L., YAMAMOTO, M., SHIGESADA, K., MITSUYA, H., ITO, Y. & OSATO, M. 2007. Concurrent transcriptional

deregulation of AML1/RUNX1 and GATA factors by the AML1-TRPS1 chimeric gene in t(8;21)(q24;q22) acute myeloid leukemia. *Blood,* 109**,** 4023-7.

BACHER, U., HAFERLACH, C., SCHNITTGER, S., KOHLMANN, A., KERN, W. & HAFERLACH, T. 2010. Mutations of the TET2 and CBL genes: novel molecular markers in myeloid malignancies. *Ann Hematol,* 89**,** 643-52.

BACHER, U., HAFERLACH, T., SCHOCH, C., KERN, W. & SCHNITTGER, S. 2006. Implications of NRAS mutations in AML: a study of 2502 patients. *Blood,* 107**,** 3847-53.

BAE, S. C., YAMAGUCHI-IWAI, Y., OGAWA, E., MARUYAMA, M., INUZUKA, M., KAGOSHIMA, H., SHIGESADA, K., SATAKE, M. & ITO, Y. 1993. Isolation of PEBP2 alpha B cDNA representing the mouse homolog of human acute myeloid leukemia gene, AML1. *Oncogene,* 8**,** 809-14.

BAKSHI, R., ZAIDI, S. K., PANDE, S., HASSAN, M. Q., YOUNG, D. W., MONTECINO, M., LIAN, J. B., VAN WIJNEN, A. J., STEIN, J. L. & STEIN, G. S. 2008. The leukemogenic t(8;21) fusion protein AML1-ETO controls rRNA genes and associates with nucleolar-organizing regions at mitotic chromosomes. *J Cell Sci,* 121**,** 3981-90.

BEMMO, A., BENOVOY, D., KWAN, T., GAFFNEY, D. J., JENSEN, R. V. & MAJEWSKI, J. 2008. Gene expression and isoform variation analysis using Affymetrix Exon Arrays. *BMC Genomics,* 9**,** 529.

BENNETT, J. M., CATOVSKY, D., DANIEL, M. T., FLANDRIN, G., GALTON, D. A., GRALNICK, H. R. & SULTAN, C. 1976. Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. *Br J Haematol,* 33**,** 451-8.

BLOOMFIELD, C. D., LAWRENCE, D., BYRD, J. C., CARROLL, A., PETTENATI, M. J., TANTRAVAHI, R., PATIL, S. R., DAVEY, F. R., BERG, D. T., SCHIFFER, C. A., ARTHUR, D. C. & MAYER, R. J. 1998. Frequency of prolonged remission duration after high-dose cytarabine intensification in acute myeloid leukemia varies by cytogenetic subtype. *Cancer Res,* 58**,** 4173-9.

BOYAPATI, A., YAN, M., PETERSON, L. F., BIGGS, J. R., LE BEAU, M. M. & ZHANG, D. E. 2007. A leukemia fusion protein attenuates the spindle checkpoint and promotes aneuploidy. *Blood,* 109**,** 3963-71.

BRITOS-BRAY, M. & FRIEDMAN, A. D. 1997. Core binding factor cannot synergistically activate the myeloperoxidase proximal enhancer in immature myeloid cells without c-Myb. *Mol Cell Biol,* 17**,** 5127-35.

BRODIE, C. & BLUMBERG, P. M. 2003. Regulation of cell apoptosis by protein kinase c delta. *Apoptosis,* 8**,** 19-27.

BRUHN, L., MUNNERLYN, A. & GROSSCHEDL, R. 1997. ALY, a context-dependent coactivator of LEF-1 and AML-1, is required for TCRalpha enhancer function. *Genes Dev,* 11**,** 640-53.

BUCHHOLZ, F., REFAELI, Y., TRUMPP, A. & BISHOP, J. M. 2000. Inducible chromosomal translocation of AML1 and ETO genes through Cre/loxP-mediated recombination in the mouse. *EMBO Rep,* 1**,** 133-9.

BULLINGER, L., RUCKER, F. G., KURZ, S., DU, J., SCHOLL, C., SANDER, S., CORBACIOGLU, A., LOTTAZ, C., KRAUTER, J., FROHLING, S., GANSER, A., SCHLENK, R. F., DOHNER, K., POLLACK, J. R. & DOHNER, H. 2007. Gene-expression profiling identifies distinct subclasses of core binding factor acute myeloid leukemia. *Blood,* 110**,** 1291-300.

BYRD, J. C., DODGE, R. K., CARROLL, A., BAER, M. R., EDWARDS, C., STAMBERG, J., QUMSIYEH, M., MOORE, J. O., MAYER, R. J., DAVEY, F., SCHIFFER, C. A. & BLOOMFIELD, C. D. 1999. Patients with t(8;21)(q22;q22) and acute myeloid leukemia have superior failure-free and overall survival when repetitive cycles of high-dose cytarabine are administered. *J Clin Oncol,* 17**,** 3767-75.

CHEN, G., DIMITRIOU, I. D., LA ROSE, J., ILANGUMARAN, S., YEH, W. C., DOODY, G., TURNER, M., GOMMERMAN, J. & ROTTAPEL, R. 2007. The 3BP2 adapter protein is required for optimal B-cell activation and thymus-independent type 2 humoral response. *Mol Cell Biol,* 27**,** 3109-22.

CHOU, W. C., CHOU, S. C., LIU, C. Y., CHEN, C. Y., HOU, H. A., KUO, Y. Y., LEE, M. C., KO, B. S., TANG, J. L., YAO, M., TSAY, W., WU, S. J., HUANG, S. Y., HSU, S. C., CHEN, Y. C., CHANG, Y. C., KUO, Y. Y., KUO, K. T., LEE, F. Y., LIU, M. C., LIU, C. W., TSENG, M. H., HUANG, C. F. & TIEN, H. F. 2011. TET2 mutation is an unfavorable prognostic factor in acute myeloid leukemia patients with intermediate-risk cytogenetics. *Blood,* 118**,** 3803-10.

DAVIS, J. N., MCGHEE, L. & MEYERS, S. 2003. The ETO (MTG8) gene family. *Gene,* 303**,** 1-10.

DAVIS, J. N., WILLIAMS, B. J., HERRON, J. T., GALIANO, F. J. & MEYERS, S. 1999. ETO-2, a new member of the ETO-family of nuclear proteins. *Oncogene,* 18**,** 1375-83.

DE GUZMAN, C. G., WARREN, A. J., ZHANG, Z., GARTLAND, L., ERICKSON, P., DRABKIN, H., HIEBERT, S. W. & KLUG, C. A. 2002. Hematopoietic stem cell expansion and distinct myeloid developmental abnormalities in a murine model of the AML1-ETO translocation. *Mol Cell Biol,* 22**,** 5506-17.

DEBERNARDI, S., LILLINGTON, D. M., CHAPLIN, T., TOMLINSON, S., AMESS, J., ROHATINER, A., LISTER, T. A. & YOUNG, B. D. 2003. Genome-wide analysis of acute myeloid leukemia with normal karyotype reveals a unique pattern of homeobox gene expression distinct from those with translocation-mediated fusion events. *Genes Chromosomes Cancer,* 37**,** 149-58.

DEFEA, K. 2008. Beta-arrestins and heterotrimeric G-proteins: collaborators and competitors in signal transduction. *Br J Pharmacol,* 153 Suppl 1**,** S298-309.

DEGOS, L. 2001. John Hughes Bennett, Rudolph Virchow... and Alfred Donne: the first description of leukemia. *Hematol J,* 2**,** 1.

DI PALMA, F., HOLME, R. H., BRYDA, E. C., BELYANTSEVA, I. A., PELLEGRINO, R., KACHAR, B., STEEL, K. P. & NOBEN-TRAUTH, K. 2001. Mutations in Cdh23, encoding a new type of cadherin, cause stereocilia disorganization in waltzer, the mouse model for Usher syndrome type 1D. *Nat Genet,* 27**,** 103-7.

DOWTON, S. B., BEARDSLEY, D., JAMISON, D., BLATTNER, S. & LI, F. P. 1985. Studies of a familial platelet disorder. *Blood,* 65**,** 557-63.

DUNNE, J., CULLMANN, C., RITTER, M., SORIA, N. M., DRESCHER, B., DEBERNARDI, S., SKOULAKIS, S., HARTMANN, O., KRAUSE, M., KRAUTER, J., NEUBAUER, A., YOUNG, B. D. & HEIDENREICH, O. 2006. siRNA-mediated AML1/MTG8 depletion affects differentiation and proliferation-associated gene expression in t(8;21)-positive cell lines and primary AML blasts. *Oncogene,* 25**,** 6067-78.

ELAGIB, K. E. & GOLDFARB, A. N. 2007a. Oncogenic pathways of AML1-ETO in acute myeloid leukemia: multifaceted manipulation of marrow maturation. *Cancer Lett,* 251**,** 179-86.

ELAGIB, K. E. & GOLDFARB, A. N. 2007b. Regulation of RUNX1 transcriptional function by GATA-1. *Crit Rev Eukaryot Gene Expr,* 17**,** 271-80.

ELSASSER, A., FRANZEN, M., KOHLMANN, A., WEISSER, M., SCHNITTGER, S., SCHOCH, C., REDDY, V. A., BUREL, S., ZHANG, D. E., UEFFING, M., TENEN, D. G., HIDDEMANN, W. & BEHRE, G. 2003. The fusion protein AML1-ETO in acute myeloid leukemia with translocation t(8;21) induces c-jun protein expression via the proximal AP-1 site of the c-jun promoter in an indirect, JNK-dependent manner. *Oncogene,* 22**,** 5646-57.

ERA, T., ASOU, N., KUNISADA, T., YAMASAKI, H., ASOU, H., KAMADA, N., NISHIKAWA, S., YAMAGUCHI, K. & TAKATSUKI, K. 1995. Identification of two transcripts of AML1/ETO-fused gene in t(8;21) leukemic cells and expression of wild-type ETO gene in hematopoietic cells. *Genes Chromosomes Cancer,* 13**,** 25-33.

ERICKSON, P., GAO, J., CHANG, K. S., LOOK, T., WHISENANT, E., RAIMONDI, S., LASHER, R., TRUJILLO, J., ROWLEY, J. & DRABKIN, H. 1992. Identification of breakpoints in t(8;21) acute myelogenous leukemia and isolation of a fusion transcript, AML1/ETO, with similarity to Drosophila segmentation gene, runt. *Blood,* 80**,** 1825-31.

ERICKSON, P. F., DESSEV, G., LASHER, R. S., PHILIPS, G., ROBINSON, M. & DRABKIN, H. A. 1996. ETO and AML1 phosphoproteins are expressed in CD34+ hematopoietic progenitors: implications for t(8;21) leukemogenesis and monitoring residual disease. *Blood,* 88**,** 1813-23.

FALINI, B., MECUCCI, C., TIACCI, E., ALCALAY, M., ROSATI, R., PASQUALUCCI, L., LA STARZA, R., DIVERIO, D., COLOMBO, E., SANTUCCI, A., BIGERNA, B., PACINI, R., PUCCIARINI, A., LISO, A., VIGNETTI, M., FAZI, P., MEANI, N., PETTIROSSI, V., SAGLIO, G., MANDELLI, F., LO-COCO, F., PELICCI, P. G., MARTELLI, M. F. & PARTY, G. A. L. W. 2005. Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *N Engl J Med,* 352**,** 254-66.

FAZI, F., ZARDO, G., GELMETTI, V., TRAVAGLINI, L., CIOLFI, A., DI CROCE, L., ROSA, A., BOZZONI, I., GRIGNANI, F., LO-COCO, F., PELICCI, P. G. & NERVI, C. 2007. Heterochromatic gene repression of the retinoic acid pathway in acute myeloid leukemia. *Blood,* 109**,** 4432-40.

FENSKE, T. S., PENGUE, G., MATHEWS, V., HANSON, P. T., HAMM, S. E., RIAZ, N. & GRAUBERT, T. A. 2004. Stem cell expression of the AML1/ETO

fusion protein induces a myeloproliferative disorder in mice. *Proc Natl Acad Sci U S A,* 101**,** 15184-9.

FOLLOWS, G. A., TAGOH, H., LEFEVRE, P., HODGE, D., MORGAN, G. J. & BONIFER, C. 2003. Epigenetic consequences of AML1-ETO action at the human c-FMS locus. *EMBO J,* 22**,** 2798-809.

FRANK, R., ZHANG, J., UCHIDA, H., MEYERS, S., HIEBERT, S. W. & NIMER, S. D. 1995. The AML1/ETO fusion protein blocks transactivation of the GM-CSF promoter by AML1B. *Oncogene,* 11**,** 2667-74.

FRODIN, M. & GAMMELTOFT, S. 1999. Role and regulation of 90 kDa ribosomal S6 kinase (RSK) in signal transduction. *Mol Cell Endocrinol,* 151**,** 65-77.

GAIDATZIS, D., JACOBEIT, K., OAKELEY, E. J. & STADLER, M. B. 2009. Overestimation of alternative splicing caused by variable probe characteristics in exon arrays. *Nucleic Acids Res,* 37**,** e107.

GAIDZIK, V. I., BULLINGER, L., SCHLENK, R. F., ZIMMERMANN, A. S., ROCK, J., PASCHKA, P., CORBACIOGLU, A., KRAUTER, J., SCHLEGELBERGER, B., GANSER, A., SPATH, D., KUNDGEN, A., SCHMIDT-WOLF, I. G., GOTZE, K., NACHBAUR, D., PFREUNDSCHUH, M., HORST, H. A., DOHNER, H. & DOHNER, K. 2011. RUNX1 mutations in acute myeloid leukemia: results from a comprehensive genetic and clinical analysis from the AML study group. *J Clin Oncol,* 29**,** 1364-72.

GAMOU, T., KITAMURA, E., HOSODA, F., SHIMIZU, K., SHINOHARA, K., HAYASHI, Y., NAGASE, T., YOKOYAMA, Y. & OHKI, M. 1998. The partner gene of AML1 in t(16;21) myeloid malignancies is a novel member of the MTG8(ETO) family. *Blood,* 91**,** 4028-37.

GELMETTI, V., ZHANG, J., FANELLI, M., MINUCCI, S., PELICCI, P. G. & LAZAR, M. A. 1998. Aberrant recruitment of the nuclear receptor corepressor-histone deacetylase complex by the acute myeloid leukemia fusion partner ETO. *Mol Cell Biol,* 18**,** 7185-91.

GIESE, K., KINGSLEY, C., KIRSHNER, J. R. & GROSSCHEDL, R. 1995. Assembly and function of a TCR alpha enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein-protein interactions. *Genes Dev,* 9**,** 995-1008.

GILLILAND, D. G. 2002. Molecular genetics of human leukemias: new insights into therapy. *Semin Hematol,* 39**,** 6-11.

GOEMANS, B. F., ZWAAN, C. M., MILLER, M., ZIMMERMANN, M., HARLOW, A., MESHINCHI, S., LOONEN, A. H., HAHLEN, K., REINHARDT, D., CREUTZIG, U., KASPERS, G. J. & HEINRICH, M. C. 2005. Mutations in KIT and RAS are frequent events in pediatric core-binding factor acute myeloid leukemia. *Leukemia,* 19**,** 1536-42.

GOETZ, T. L., GU, T. L., SPECK, N. A. & GRAVES, B. J. 2000. Auto-inhibition of Ets-1 is counteracted by DNA binding cooperativity with core-binding factor alpha2. *Mol Cell Biol,* 20**,** 81-90.

GOLLING, G., LI, L., PEPLING, M., STEBBINS, M. & GERGEN, J. P. 1996. Drosophila homologs of the proto-oncogene product PEBP2/CBF beta regulate the DNA-binding properties of Runt. *Mol Cell Biol,* 16**,** 932-42.

GOLUB, T. R., BARKER, G. F., BOHLANDER, S. K., HIEBERT, S. W., WARD, D. C., BRAY-WARD, P., MORGAN, E., RAIMONDI, S. C., ROWLEY, J. D. & GILLILAND, D. G. 1995. Fusion of the TEL gene on 12p13 to the AML1 gene on 21q22 in acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A,* 92**,** 4917-21.

GONCALVES, J. E., J.D. 2007. Use of an Exon Software Package to Identify Global Gene Splicing and Gene Expression Regulation. *American Biotechnolog Laboratory***,** 24-26.

GREEN, C. L., EVANS, C. M., ZHAO, L., HILLS, R. K., BURNETT, A. K., LINCH, D. C. & GALE, R. E. 2011. The prognostic significance of IDH2 mutations in AML depends on the location of the mutation. *Blood,* 118**,** 409-12.

GRIMWADE, D., HILLS, R. K., MOORMAN, A. V., WALKER, H., CHATTERS, S., GOLDSTONE, A. H., WHEATLEY, K., HARRISON, C. J. & BURNETT, A. K. 2010. Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. *Blood,* 116**,** 354-65.

GRIMWADE, D., WALKER, H., OLIVER, F., WHEATLEY, K., HARRISON, C., HARRISON, G., REES, J., HANN, I., STEVENS, R., BURNETT, A. & GOLDSTONE, A. 1998. The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties. *Blood,* 92**,** 2322-33.

GRISOLANO, J. L., O'NEAL, J., CAIN, J. & TOMASSON, M. H. 2003. An activated receptor tyrosine kinase, TEL/PDGFbetaR, cooperates with AML1/ETO to induce acute myeloid leukemia in mice. *Proc Natl Acad Sci U S A,* 100**,** 9506-11.

GU, T. L., GOETZ, T. L., GRAVES, B. J. & SPECK, N. A. 2000. Auto-inhibition and partner proteins, core-binding factor beta (CBFbeta) and Ets-1, modulate DNA binding by CBFalpha2 (AML1). *Mol Cell Biol,* 20**,** 91-103.

GUERRASIO, A., ROSSO, C., MARTINELLI, G., LO COCO, F., PAMPINELLA, M., SANTORO, A., LANZA, C., ALLIONE, B., RESEGOTTI, L. & SAGLIO, G. 1995. Polyclonal haemopoieses associated with long-term persistence of the AML1-ETO transcript in patients with FAB M2 acute myeloid leukaemia in continous clinical remission. *Br J Haematol,* 90**,** 364-8.

HARRISON, C. J., HILLS, R. K., MOORMAN, A. V., GRIMWADE, D. J., HANN, I., WEBB, D. K., WHEATLEY, K., DE GRAAF, S. S., VAN DEN BERG, E., BURNETT, A. K. & GIBSON, B. E. 2010. Cytogenetics of childhood acute myeloid leukemia: United Kingdom Medical Research Council Treatment trials AML 10 and 12. *J Clin Oncol,* 28**,** 2674-81.

HASHIMOTO, H., HORTON, J. R., ZHANG, X. & CHENG, X. 2009. UHRF1, a modular multi-domain protein, regulates replication-coupled crosstalk between DNA methylation and histone modifications. *Epigenetics,* 4**,** 8-14.

HEINZ, S., KRAUSE, S. W., GABRIELLI, F., WAGNER, H. M., ANDREESEN, R. & REHLI, M. 2002. Genomic organization of the human gene HEP27:

alternative promoter usage in HepG2 cells and monocyte-derived dendritic cells. *Genomics,* 79**,** 608-15.

HIEBERT, S. W., LUTTERBACH, B. & AMANN, J. 2001. Role of co-repressors in transcriptional repression mediated by the t(8;21), t(16;21), t(12;21), and inv(16) fusion proteins. *Curr Opin Hematol,* 8**,** 197-200.

HIEBERT, S. W., REED-INDERBITZIN, E. F., AMANN, J., IRVIN, B., DURST, K. & LINGGI, B. 2003. The t(8;21) fusion protein contacts co-repressors and histone deacetylases to repress the transcription of the p14ARF tumor suppressor. *Blood Cells Mol Dis,* 30**,** 177-83.

HONG, C., MOOREFIELD, K. S., JUN, P., ALDAPE, K. D., KHARBANDA, S., PHILLIPS, H. S. & COSTELLO, J. F. 2007. Epigenome scans and cancer genome sequencing converge on WNK2, a kinase-independent suppressor of cell growth. *Proc Natl Acad Sci U S A,* 104**,** 10974-9.

HROMAS, R., BUSSE, T., CARROLL, A., MACK, D., SHOPNICK, R., ZHANG, D. E., NAKSHATRI, H. & RICHKIND, K. 2001. Fusion AML1 transcript in a radiation-associated leukemia results in a truncated inhibitory AML1 protein. *Blood,* 97**,** 2168-70.

HUANG, G., SHIGESADA, K., ITO, K., WEE, H. J., YOKOMIZO, T. & ITO, Y. 2001. Dimerization with PEBP2beta protects RUNX1/AML1 from ubiquitin-proteasome-mediated degradation. *Embo J,* 20**,** 723-33.

HUG, B. A. & LAZAR, M. A. 2004. ETO interacting proteins. *Oncogene,* 23**,** 4270-4.

HUG, B. A., LEE, S. Y., KINSLER, E. L., ZHANG, J. & LAZAR, M. A. 2002. Cooperative function of Aml1-ETO corepressor recruitment domains in the expansion of primary bone marrow cells. *Cancer Res,* 62**,** 2906-12.

IIZUKA, M. & SMITH, M. M. 2003. Functional consequences of histone modifications. *Curr Opin Genet Dev,* 13**,** 154-60.

IMAI, Y., KUROKAWA, M., TANAKA, K., FRIEDMAN, A. D., OGAWA, S., MITANI, K., YAZAKI, Y. & HIRAI, H. 1998. TLE, the human homolog of groucho, interacts with AML1 and acts as a repressor of AML1-induced transactivation. *Biochem Biophys Res Commun,* 252**,** 582-9.

JAKUBOWIAK, A., POUPONNOT, C., BERGUIDO, F., FRANK, R., MAO, S., MASSAGUE, J. & NIMER, S. D. 2000. Inhibition of the transforming growth factor beta 1 signaling pathway by the AML1/ETO leukemia-associated fusion protein. *J Biol Chem,* 275**,** 40282-7.

KAMACHI, Y., OGAWA, E., ASANO, M., ISHIDA, S., MURAKAMI, Y., SATAKE, M., ITO, Y. & SHIGESADA, K. 1990. Purification of a mouse nuclear factor that binds to both the A and B cores of the polyomavirus enhancer. *J Virol,* 64**,** 4808-19.

KANIA, M. A., BONNER, A. S., DUFFY, J. B. & GERGEN, J. P. 1990. The Drosophila segmentation gene runt encodes a novel nuclear regulatory protein that is also expressed in the developing nervous system. *Genes Dev,* 4**,** 1701-13.

KATOH, Y. & KATOH, M. 2009. Hedgehog target genes: mechanisms of carcinogenesis induced by aberrant hedgehog signaling activation. *Curr Mol Med,* 9**,** 873-86.

KATZAV, S. 2007. Flesh and blood: the story of Vav1, a gene that signals in hematopoietic cells but can be transforming in human malignancies. *Cancer Lett,* 255**,** 241-54.

KELLER, G., LACAUD, G. & ROBERTSON, S. 1999. Development of the hematopoietic system in the mouse. *Exp Hematol,* 27**,** 777-87.

KIM, M. S., YOON, S. K., BOLLIG, F., KITAGAKI, J., HUR, W., WHYE, N. J., WU, Y. P., RIVERA, M. N., PARK, J. Y., KIM, H. S., MALIK, K., BELL, D. W., ENGLERT, C., PERANTONI, A. O. & LEE, S. B. A novel Wilms tumor 1 (WT1) target gene negatively regulates the WNT signaling pathway. *J Biol Chem,* 285**,** 14585-93.

KIM, W. Y., SIEWEKE, M., OGAWA, E., WEE, H. J., ENGLMEIER, U., GRAF, T. & ITO, Y. 1999. Mutual activation of Ets-1 and AML1 DNA binding by direct interaction of their autoinhibitory domains. *Embo J,* 18**,** 1609-20.

KITABAYASHI, I., IDA, K., MOROHOSHI, F., YOKOYAMA, A., MITSUHASHI, N., SHIMIZU, K., NOMURA, N., HAYASHI, Y. & OHKI, M. 1998a. The AML1-MTG8 leukemic fusion protein forms a complex with a novel member of the MTG8(ETO/CDR) family, MTGR1. *Mol Cell Biol,* 18**,** 846-58.

KITABAYASHI, I., YOKOYAMA, A., SHIMIZU, K. & OHKI, M. 1998b. Interaction and functional cooperation of the leukemia-associated factors AML1 and p300 in myeloid cell differentiation. *Embo J,* 17**,** 2994-3004.

KLAMPFER, L., ZHANG, J., ZELENETZ, A. O., UCHIDA, H. & NIMER, S. D. 1996. The AML1/ETO fusion protein activates transcription of BCL-2. *Proc Natl Acad Sci U S A,* 93**,** 14059-64.

KLISOVIC, M. I., MAGHRABY, E. A., PARTHUN, M. R., GUIMOND, M., SKLENAR, A. R., WHITMAN, S. P., CHAN, K. K., MURPHY, T., ANON, J., ARCHER, K. J., RUSH, L. J., PLASS, C., GREVER, M. R., BYRD, J. C. & MARCUCCI, G. 2003. Depsipeptide (FR 901228) promotes histone acetylation, gene transcription, apoptosis and its activity is enhanced by DNA methyltransferase inhibitors in AML1/ETO-positive leukemic cells. *Leukemia,* 17**,** 350-8.

KOHZAKI, H., ITO, K., HUANG, G., WEE, H. J., MURAKAMI, Y. & ITO, Y. 1999. Block of granulocytic differentiation of 32Dcl3 cells by AML1/ETO(MTG8) but not by highly expressed Bcl-2. *Oncogene,* 18**,** 4055-62.

KOJIMA, K., OMOTO, E., HARA, M., SASAKI, K., KATAYAMA, Y., NAWA, Y., KIMURA, Y., AZUMA, T., TAKIMOTO, H. & HARADA, M. 1998. Myelodysplastic syndrome with translocation (8;21): a distinct myelodysplastic syndrome entity or M2-acute myeloid leukemia with extensive myeloid maturation? *Ann Hematol,* 76**,** 279-82.

KOTTARIDIS, P. D., GALE, R. E., FREW, M. E., HARRISON, G., LANGABEER, S. E., BELTON, A. A., WALKER, H., WHEATLEY, K., BOWEN, D. T., BURNETT, A. K., GOLDSTONE, A. H. & LINCH, D. C. 2001. The presence of a FLT3 internal tandem duplication in patients with acute myeloid leukemia (AML) adds important prognostic information to cytogenetic risk group and response to the first cycle of chemotherapy: analysis of 854 patients from the United Kingdom Medical Research Council AML 10 and 12 trials. *Blood,* 98**,** 1752-9.

KOZU, T., FUKUYAMA, T., YAMAMI, T., AKAGI, K. & KANEKO, Y. 2005. MYND-less splice variants of AML1-MTG8 (RUNX1-CBFA2T1) are expressed in leukemia with t(8;21). *Genes Chromosomes Cancer,* 43**,** 45-53.

KOZU, T., MIYOSHI, H., SHIMIZU, K., MASEKI, N., KANEKO, Y., ASOU, H., KAMADA, N. & OHKI, M. 1993. Junctions of the AML1/MTG8(ETO) fusion are constant in t(8;21) acute myeloid leukemia detected by reverse transcription polymerase chain reaction. *Blood,* 82**,** 1270-6.

KREJCI, O., WUNDERLICH, M., GEIGER, H., CHOU, F. S., SCHLEIMER, D., JANSEN, M., ANDREASSEN, P. R. & MULLOY, J. C. 2008. p53 signaling in response to increased DNA damage sensitizes AML1-ETO cells to stress-induced death. *Blood,* 111**,** 2190-9.

KUO, M. H. & ALLIS, C. D. 1999. In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. *Methods,* 19**,** 425-33.

KUWATSUKA, Y., MIYAMURA, K., SUZUKI, R., KASAI, M., MARUTA, A., OGAWA, H., TANOSAKI, R., TAKAHASHI, S., KODA, K., YAGO, K., ATSUTA, Y., YOSHIDA, T., SAKAMAKI, H. & KODERA, Y. 2009. Hematopoietic stem cell transplantation for core binding factor acute myeloid leukemia: t(8;21) and inv(16) represent different clinical outcomes. *Blood,* 113**,** 2096-103.

KWOK, C., ZEISIG, B. B., QIU, J., DONG, S. & SO, C. W. 2009. Transforming activity of AML1-ETO is independent of CBFbeta and ETO interaction but requires formation of homo-oligomeric complexes. *Proc Natl Acad Sci U S A,* 106**,** 2853-8.

LAFIURA, K. M., EDWARDS, H., TAUB, J. W., MATHERLY, L. H., FONTANA, J. A., MOHAMED, A. N., RAVINDRANATH, Y. & GE, Y. 2008. Identification and characterization of novel AML1-ETO fusion transcripts in pediatric t(8;21) acute myeloid leukemia: a report from the Children's Oncology Group. *Oncogene,* 27**,** 4933-42.

LASA, A., NOMDEDEU, J. F., CARNICER, M. J., LLORENTE, A. & SIERRA, J. 2002. ETO sequence may be dispensable in some AML1-ETO leukemias. *Blood,* 100**,** 4243-4.

LATAGLIATA, R., BONGARZONI, V., CARMOSINO, I., MENGARELLI, A., BRECCIA, M., BORZA, P. A., D'ANDREA, M., D'ELIA, G. M., MECAROCCI, S., MORANO, S. G., PETTI, M. C., MANDELLI, F. & ALIMENA, G. 2006. Acute myelogenous leukemia in elderly patients not eligible for intensive chemotherapy: the dark side of the moon. *Ann Oncol,* 17**,** 281-5.

LEE, C. & ROY, M. 2004. Analysis of alternative splicing with microarrays: successes and challenges. *Genome Biol,* 5**,** 231.

LEE, T. L., SHYU, Y. C., HSU, T. Y. & SHEN, C. K. 2008. Itch regulates p45/NF-E2 in vivo by Lys63-linked ubiquitination. *Biochem Biophys Res Commun,* 375**,** 326-30.

LEVANON, D., GLUSMAN, G., BANGSOW, T., BEN-ASHER, E., MALE, D. A., AVIDAN, N., BANGSOW, C., HATTORI, M., TAYLOR, T. D., TAUDIEN, S., BLECHSCHMIDT, K., SHIMIZU, N., ROSENTHAL, A., SAKAKI, Y., LANCET, D. & GRONER, Y. 2001. Architecture and anatomy of the

genomic locus encoding the human leukemia-associated transcription factor RUNX1/AML1. *Gene,* 262**,** 23-33.

LEVANON, D., GOLDSTEIN, R. E., BERNSTEIN, Y., TANG, H., GOLDENBERG, D., STIFANI, S., PAROUSH, Z. & GRONER, Y. 1998. Transcriptional repression by AML1 and LEF-1 is mediated by the TLE/Groucho corepressors. *Proc Natl Acad Sci U S A,* 95**,** 11590-5.

LEY, T. J., DING, L., WALTER, M. J., MCLELLAN, M. D., LAMPRECHT, T., LARSON, D. E., KANDOTH, C., PAYTON, J. E., BATY, J., WELCH, J., HARRIS, C. C., LICHTI, C. F., TOWNSEND, R. R., FULTON, R. S., DOOLING, D. J., KOBOLDT, D. C., SCHMIDT, H., ZHANG, Q., OSBORNE, J. R., LIN, L., O'LAUGHLIN, M., MCMICHAEL, J. F., DELEHAUNTY, K. D., MCGRATH, S. D., FULTON, L. A., MAGRINI, V. J., VICKERY, T. L., HUNDAL, J., COOK, L. L., CONYERS, J. J., SWIFT, G. W., REED, J. P., ALLDREDGE, P. A., WYLIE, T., WALKER, J., KALICKI, J., WATSON, M. A., HEATH, S., SHANNON, W. D., VARGHESE, N., NAGARAJAN, R., WESTERVELT, P., TOMASSON, M. H., LINK, D. C., GRAUBERT, T. A., DIPERSIO, J. F., MARDIS, E. R. & WILSON, R. K. 2010. DNMT3A mutations in acute myeloid leukemia. *N Engl J Med,* 363**,** 2424-33.

LI, J. & WANG, C. Y. 2008. TBL1-TBLR1 and beta-catenin recruit each other to Wnt target-gene promoter for transcription activation and oncogenesis. *Nat Cell Biol,* 10**,** 160-9.

LICHT, J. D. 2001. AML1 and the AML1-ETO fusion protein in the pathogenesis of t(8;21) AML. *Oncogene,* 20**,** 5660-79.

LINGGI, B., MULLER-TIDOW, C., VAN DE LOCHT, L., HU, M., NIP, J., SERVE, H., BERDEL, W. E., VAN DER REIJDEN, B., QUELLE, D. E., ROWLEY, J. D., CLEVELAND, J., JANSEN, J. H., PANDOLFI, P. P. & HIEBERT, S. W. 2002. The t(8;21) fusion protein, AML1 ETO, specifically represses the transcription of the p14(ARF) tumor suppressor in acute myeloid leukemia. *Nat Med,* 8**,** 743-50.

LIU, P., TARLE, S. A., HAJRA, A., CLAXTON, D. F., MARLTON, P., FREEDMAN, M., SICILIANO, M. J. & COLLINS, F. S. 1993. Fusion between transcription factor CBF beta/PEBP2 beta and a myosin heavy chain in acute myeloid leukemia. *Science,* 261**,** 1041-4.

LIU, S., SHEN, T., HUYNH, L., KLISOVIC, M. I., RUSH, L. J., FORD, J. L., YU, J., BECKNELL, B., LI, Y., LIU, C., VUKOSAVLJEVIC, T., WHITMAN, S. P., CHANG, K. S., BYRD, J. C., PERROTTI, D., PLASS, C. & MARCUCCI, G. 2005. Interplay of RUNX1/MTG8 and DNA methyltransferase 1 in acute myeloid leukemia. *Cancer Res,* 65**,** 1277-84.

LIU, W. M. & ZHANG, X. A. 2006. KAI1/CD82, a tumor metastasis suppressor. *Cancer Lett,* 240**,** 183-94.

LIU, Y., CHENEY, M. D., GAUDET, J. J., CHRUSZCZ, M., LUKASIK, S. M., SUGIYAMA, D., LARY, J., COLE, J., DAUTER, Z., MINOR, W., SPECK, N. A. & BUSHWELLER, J. H. 2006. The tetramer structure of the Nervy homology two domain, NHR2, is critical for AML1/ETO's activity. *Cancer Cell,* 9**,** 249-60.

LOOK, A. T. 1997. Oncogenic transcription factors in the human acute leukemias. *Science,* 278**,** 1059-64.

LU, J., MARUYAMA, M., SATAKE, M., BAE, S. C., OGAWA, E., KAGOSHIMA, H., SHIGESADA, K. & ITO, Y. 1995. Subcellular localization of the alpha and beta subunits of the acute myeloid leukemia-linked transcription factor PEBP2/CBF. *Mol Cell Biol,* 15**,** 1651-61.

LUCK, S. C., RUSS, A. C., DU, J., GAIDZIK, V., SCHLENK, R. F., POLLACK, J. R., DOHNER, K., DOHNER, H. & BULLINGER, L. 2010. KIT mutations confer a distinct gene expression signature in core binding factor leukaemia. *Br J Haematol,* 148**,** 925-37.

LUTTERBACH, B. & HIEBERT, S. W. 2000. Role of the transcription factor AML-1 in acute leukemia and hematopoietic differentiation. *Gene,* 245**,** 223-35.

LUTTERBACH, B., WESTENDORF, J. J., LINGGI, B., ISAAC, S., SETO, E. & HIEBERT, S. W. 2000. A mechanism of repression by acute myeloid leukemia-1, the target of multiple chromosomal translocations in acute leukemia. *J Biol Chem,* 275**,** 651-6.

LUTTERBACH, B., WESTENDORF, J. J., LINGGI, B., PATTEN, A., MONIWA, M., DAVIE, J. R., HUYNH, K. D., BARDWELL, V. J., LAVINSKY, R. M., ROSENFELD, M. G., GLASS, C., SETO, E. & HIEBERT, S. W. 1998. ETO, a target of t(8;21) in acute leukemia, interacts with the N-CoR and mSin3 corepressors. *Mol Cell Biol,* 18**,** 7176-84.

MAO, S., FRANK, R. C., ZHANG, J., MIYAZAKI, Y. & NIMER, S. D. 1999. Functional and physical interactions between AML1 proteins and an ETS protein, MEF: implications for the pathogenesis of t(8;21)-positive leukemias. *Mol Cell Biol,* 19**,** 3635-44.

MARCUCCI, G., MROZEK, K., RUPPERT, A. S., MAHARRY, K., KOLITZ, J. E., MOORE, J. O., MAYER, R. J., PETTENATI, M. J., POWELL, B. L., EDWARDS, C. G., STERLING, L. J., VARDIMAN, J. W., SCHIFFER, C. A., CARROLL, A. J., LARSON, R. A. & BLOOMFIELD, C. D. 2005. Prognostic factors and outcome of core binding factor acute myeloid leukemia patients with t(8;21) differ from those of patients with inv(16): a Cancer and Leukemia Group B study. *J Clin Oncol,* 23**,** 5705-17.

MARDIS, E. R. 2007. ChIP-seq: welcome to the new frontier. *Nat Methods,* 4**,** 613-4.

MARKOVA, J., MICHKOVA, P., BURCKOVA, K., BREZINOVA, J., MICHALOVA, K., DOHNALOVA, A., MAALOUFOVA, J. S., SOUKUP, P., VITEK, A., CETKOVSKY, P. & SCHWARZ, J. 2011. Prognostic impact of DNMT3A mutations in patients with intermediate cytogenetic risk profile acute myeloid leukemia. *Eur J Haematol*.

MATOZAKI, S., NAKAGAWA, T., KAWAGUCHI, R., AOZAKI, R., TSUTSUMI, M., MURAYAMA, T., KOIZUMI, T., NISHIMURA, R., ISOBE, T. & CHIHARA, K. 1995. Establishment of a myeloid leukaemic cell line (SKNO-1) from a patient with t(8;21) who acquired monosomy 17 during disease progression. *Br J Haematol,* 89**,** 805-11.

MEAD, A. J., LINCH, D. C., HILLS, R. K., WHEATLEY, K., BURNETT, A. K. & GALE, R. E. 2007. FLT3 tyrosine kinase domain mutations are biologically

distinct from and have a significantly more favorable prognosis than FLT3 internal tandem duplications in patients with acute myeloid leukemia. *Blood,* 110**,** 1262-70.

MELEN, K., KESKINEN, P., RONNI, T., SARENEVA, T., LOUNATMAA, K. & JULKUNEN, I. 1996. Human MxB protein, an interferon-alpha-inducible GTPase, contains a nuclear targeting signal and is localized in the heterochromatin region beneath the nuclear envelope. *J Biol Chem,* 271**,** 23478-86.

METZELER, K. H., BECKER, H., MAHARRY, K., RADMACHER, M. D., KOHLSCHMIDT, J., MROZEK, K., NICOLET, D., WHITMAN, S. P., WU, Y. Z., SCHWIND, S., POWELL, B. L., CARTER, T. H., WETZLER, M., MOORE, J. O., KOLITZ, J. E., BAER, M. R., CARROLL, A. J., LARSON, R. A., CALIGIURI, M. A., MARCUCCI, G. & BLOOMFIELD, C. D. 2011. ASXL1 mutations identify a high-risk subgroup of older patients with primary cytogenetically normal AML within the ELN "favorable" genetic category. *Blood*.

MEYERS, S., DOWNING, J. R. & HIEBERT, S. W. 1993. Identification of AML-1 and the (8;21) translocation protein (AML-1/ETO) as sequence-specific DNA-binding proteins: the runt homology domain is required for DNA binding and protein-protein interactions. *Mol Cell Biol,* 13**,** 6336-45.

MEYERS, S., LENNY, N. & HIEBERT, S. W. 1995. The t(8;21) fusion protein interferes with AML-1B-dependent transcriptional activation. *Mol Cell Biol,* 15**,** 1974-82.

MILLIGAN, D. W., GRIMWADE, D., CULLIS, J. O., BOND, L., SWIRSKY, D., CRADDOCK, C., KELL, J., HOMEWOOD, J., CAMPBELL, K., MCGINLEY, S., WHEATLEY, K. & JACKSON, G. 2006. Guidelines on the management of acute myeloid leukaemia in adults. *Br J Haematol,* 135**,** 450-74.

MIYAMOTO, T., NAGAFUJI, K., AKASHI, K., HARADA, M., KYO, T., AKASHI, T., TAKENAKA, K., MIZUNO, S., GONDO, H., OKAMURA, T., DOHY, H. & NIHO, Y. 1996. Persistence of multipotent progenitors expressing AML1/ETO transcripts in long-term remission patients with t(8;21) acute myelogenous leukemia. *Blood,* 87**,** 4789-96.

MIYAMOTO, T., WEISSMAN, I. L. & AKASHI, K. 2000. AML1/ETO-expressing nonleukemic stem cells in acute myelogenous leukemia with 8;21 chromosomal translocation. *Proc Natl Acad Sci U S A,* 97**,** 7521-6.

MIYOSHI, H., KOZU, T., SHIMIZU, K., ENOMOTO, K., MASEKI, N., KANEKO, Y., KAMADA, N. & OHKI, M. 1993. The t(8;21) translocation in acute myeloid leukemia results in production of an AML1-MTG8 fusion transcript. *Embo J,* 12**,** 2715-21.

MIYOSHI, H., OHIRA, M., SHIMIZU, K., MITANI, K., HIRAI, H., IMAI, T., YOKOYAMA, K., SOEDA, E. & OHKI, M. 1995. Alternative splicing and genomic structure of the AML1 gene involved in acute myeloid leukemia. *Nucleic Acids Res,* 23**,** 2762-9.

MIYOSHI, H., SHIMIZU, K., KOZU, T., MASEKI, N., KANEKO, Y. & OHKI, M. 1991. t(8;21) breakpoints on chromosome 21 in acute myeloid leukemia are

clustered within a limited region of a single gene, AML1. *Proc Natl Acad Sci U S A,* 88**,** 10431-4.

MONIZ, S., MATOS, P. & JORDAN, P. 2008. WNK2 modulates MEK1 activity through the Rho GTPase pathway. *Cell Signal,* 20**,** 1762-8.

MONIZ, S., VERISSIMO, F., MATOS, P., BRAZAO, R., SILVA, E., KOTELEVETS, L., CHASTRE, E., GESPACH, C. & JORDAN, P. 2007. Protein kinase WNK2 inhibits cell proliferation by negatively modulating the activation of MEK1/ERK1/2. *Oncogene,* 26**,** 6071-81.

MULLER-TIDOW, C., STEFFEN, B., CAUVET, T., TICKENBROCK, L., JI, P., DIEDERICHS, S., SARGIN, B., KOHLER, G., STELLJES, M., PUCCETTI, E., RUTHARDT, M., DEVOS, S., HIEBERT, S. W., KOEFFLER, H. P., BERDEL, W. E. & SERVE, H. 2004. Translocation products in acute myeloid leukemia activate the Wnt signaling pathway in hematopoietic cells. *Mol Cell Biol,* 24**,** 2890-904.

MULLOY, J. C., CAMMENGA, J., BERGUIDO, F. J., WU, K., ZHOU, P., COMENZO, R. L., JHANWAR, S., MOORE, M. A. & NIMER, S. D. 2003. Maintaining the self-renewal and differentiation potential of human CD34+ hematopoietic cells using a single genetic element. *Blood,* 102**,** 4369-76.

MULLOY, J. C., CAMMENGA, J., MACKENZIE, K. L., BERGUIDO, F. J., MOORE, M. A. & NIMER, S. D. 2002. The AML1-ETO fusion protein promotes the expansion of human hematopoietic stem cells. *Blood,* 99**,** 15-23.

MULLOY, J. C., JANKOVIC, V., WUNDERLICH, M., DELWEL, R., CAMMENGA, J., KREJCI, O., ZHAO, H., VALK, P. J., LOWENBERG, B. & NIMER, S. D. 2005. AML1-ETO fusion protein up-regulates TRKA mRNA expression in human CD34+ cells, allowing nerve growth factor-induced expansion. *Proc Natl Acad Sci U S A,* 102**,** 4016-21.

MUNUGALAVADLA, V., DORE, L. C., TAN, B. L., HONG, L., VISHNU, M., WEISS, M. J. & KAPUR, R. 2005. Repression of c-kit and its downstream substrates by GATA-1 inhibits cell proliferation during erythroid maturation. *Mol Cell Biol,* 25**,** 6747-59.

NIMER, S. D. & MOORE, M. A. 2004. Effects of the leukemia-associated AML1-ETO protein on hematopoietic stem and progenitor cells. *Oncogene,* 23**,** 4249-54.

NISHIDA, S., HOSEN, N., SHIRAKATA, T., KANATO, K., YANAGIHARA, M., NAKATSUKA, S., HOSHIDA, Y., NAKAZAWA, T., HARADA, Y., TATSUMI, N., TSUBOI, A., KAWAKAMI, M., OKA, Y., OJI, Y., AOZASA, K., KAWASE, I. & SUGIYAMA, H. 2006. AML1-ETO rapidly induces acute myeloblastic leukemia in cooperation with the Wilms tumor gene, WT1. *Blood,* 107**,** 3303-12.

NISSON, P. E., WATKINS, P. C. & SACCHI, N. 1992. Transcriptionally active chimeric gene derived from the fusion of the AML1 gene and a novel gene on chromosome 8 in t(8;21) leukemic cells. *Cancer Genet Cytogenet,* 63**,** 81-8.

NORTH, T., GU, T. L., STACY, T., WANG, Q., HOWARD, L., BINDER, M., MARIN-PADILLA, M. & SPECK, N. A. 1999. Cbfa2 is required for the formation of intra-aortic hematopoietic clusters. *Development,* 126**,** 2563-75.

NUCIFORA, G., BEGY, C. R., ERICKSON, P., DRABKIN, H. A. & ROWLEY, J. D. 1993. The 3;21 translocation in myelodysplasia results in a fusion transcript between the AML1 gene and the gene for EAP, a highly conserved protein associated with the Epstein-Barr virus small RNA EBER 1. *Proc Natl Acad Sci U S A,* 90**,** 7784-8.

O'CONNELL, R. M., RAO, D. S., CHAUDHURI, A. A., BOLDIN, M. P., TAGANOV, K. D., NICOLL, J., PAQUETTE, R. L. & BALTIMORE, D. 2008. Sustained expression of microRNA-155 in hematopoietic stem cells causes a myeloproliferative disorder. *J Exp Med,* 205**,** 585-94.

ODAKA, Y., MALLY, A., ELLIOTT, L. T. & MEYERS, S. 2000. Nuclear import and subnuclear localization of the proto-oncoprotein ETO (MTG8). *Oncogene,* 19**,** 3584-97.

OGAWA, E., INUZUKA, M., MARUYAMA, M., SATAKE, M., NAITO-FUJIMOTO, M., ITO, Y. & SHIGESADA, K. 1993. Molecular cloning and characterization of PEBP2 beta, the heterodimeric partner of a novel Drosophila runt-related DNA binding protein PEBP2 alpha. *Virology,* 194**,** 314-31.

OKUDA, T., CAI, Z., YANG, S., LENNY, N., LYU, C. J., VAN DEURSEN, J. M., HARADA, H. & DOWNING, J. R. 1998. Expression of a knocked-in AML1-ETO leukemia gene inhibits the establishment of normal definitive hematopoiesis and directly generates dysplastic hematopoietic progenitors. *Blood,* 91**,** 3134-43.

OKUDA, T., TAKEDA, K., FUJITA, Y., NISHIMURA, M., YAGYU, S., YOSHIDA, M., AKIRA, S., DOWNING, J. R. & ABE, T. 2000. Biological characteristics of the leukemia-associated transcriptional factor AML1 disclosed by hematopoietic rescue of AML1-deficient embryonic stem cells by using a knock-in strategy. *Mol Cell Biol,* 20**,** 319-28.

OKUDA, T., VAN DEURSEN, J., HIEBERT, S. W., GROSVELD, G. & DOWNING, J. R. 1996. AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell,* 84**,** 321-30.

OKUMURA, A. J., PETERSON, L. F., OKUMURA, F., BOYAPATI, A. & ZHANG, D. E. 2008. t(8;21)(q22;q22) Fusion proteins preferentially bind to duplicated AML1/RUNX1 DNA-binding sequences to differentially regulate gene expression. *Blood,* 112**,** 1392-401.

OKUMURA, F., MATSUNAGA, Y., KATAYAMA, Y., NAKAYAMA, K. I. & HATAKEYAMA, S. TRIM8 modulates STAT3 activity through negative regulation of PIAS3. *J Cell Sci,* 123**,** 2238-45.

OWEN, C., BARNETT, M. & FITZGIBBON, J. 2008. Familial myelodysplasia and acute myeloid leukaemia--a review. *Br J Haematol,* 140**,** 123-32.

PARDALI, E., XIE, X. Q., TSAPOGAS, P., ITOH, S., ARVANITIDIS, K., HELDIN, C. H., TEN DIJKE, P., GRUNDSTROM, T. & SIDERAS, P. 2000. Smad and AML proteins synergistically confer transforming growth factor beta1 responsiveness to human germ-line IgA genes. *J Biol Chem,* 275**,** 3552-60.

PASCHKA, P. 2008. Core binding factor acute myeloid leukemia. *Semin Oncol,* 35**,** 410-7.

PASCHKA, P., DU, J., SCHLENK, R. F., BULLINGER , L., GAIDZIK, V. I., CORBACIOGLU, A., BENNER, A., ZUCKNICK, M., STEGELMANN, F., SPÄTH, D., KAYSER, S., SCHLEGELBERGER, B., KRAUTER, J., GANSER, A., DÖHNER, H. & DÖHNER, K. 2009. Mutations in the Fms-Related Tyrosine Kinase 3 (FLT3) Gene Independently Predict Poor Outcome in Acute Myeloid Leukemia (AML) with t(8;21): A Study of the German-Austrian AML Study Group (AMLSG). *Blood (ASH Annual Meeting Abstracts),* 114.

PASCHKA, P., MARCUCCI, G., RUPPERT, A. S., MROZEK, K., CHEN, H., KITTLES, R. A., VUKOSAVLJEVIC, T., PERROTTI, D., VARDIMAN, J. W., CARROLL, A. J., KOLITZ, J. E., LARSON, R. A., BLOOMFIELD, C. D., CANCER & LEUKEMIA GROUP, B. 2006. Adverse prognostic significance of KIT mutations in adult acute myeloid leukemia with inv(16) and t(8;21): a Cancer and Leukemia Group B Study. *J Clin Oncol,* 24**,** 3904-11.

PASCHKA, P., RADMACHER, M. D., MARCUCCI, G., RUPPERT, A. S., VUKOSAVLJEVIC, T., WHITMAN, S. P., MROZEK, K., LIU, C., R. A. LARSON, R. A. & BLOOMFIELD, C. D. 2007. Outcome prediction in adult core binding factor (CBF) acute myeloid leukemia (AML) with gene expression profiling: A Cancer and Leukemia Group B (CALGB) study. *ASCO MEETING ABSTRACTS.*

PELLEGRINI, S., CENSINI, S., GUIDOTTI, S., IACOPETTI, P., ROCCHI, M., BIANCHI, M., COVACCI, A. & GABRIELLI, F. 2002. A human short-chain dehydrogenase/reductase gene: structure, chromosomal localization, tissue expression and subcellular localization of its product. *Biochim Biophys Acta,* 1574**,** 215-22.

PETERSON, L. F., BOYAPATI, A., AHN, E. Y., BIGGS, J. R., OKUMURA, A. J., LO, M. C., YAN, M. & ZHANG, D. E. 2007a. Acute myeloid leukemia with the 8q22;21q22 translocation: secondary mutational events and alternative t(8;21) transcripts. *Blood,* 110**,** 799-805.

PETERSON, L. F., YAN, M. & ZHANG, D. E. 2007b. The p21Waf1 pathway is involved in blocking leukemogenesis by the t(8;21) fusion protein AML1-ETO. *Blood,* 109**,** 4392-8.

PETERSON, L. F. & ZHANG, D. E. 2004. The 8;21 translocation in leukemogenesis. *Oncogene,* 23**,** 4255-62.

PETROVICK, M. S., HIEBERT, S. W., FRIEDMAN, A. D., HETHERINGTON, C. J., TENEN, D. G. & ZHANG, D. E. 1998. Multiple functional domains of AML1: PU.1 and C/EBPalpha synergize with different regions of AML1. *Mol Cell Biol,* 18**,** 3915-25.

PHAN, V. T., SHULTZ, D. B., TRUONG, B. T., BLAKE, T. J., BROWN, A. L., GONDA, T. J., LE BEAU, M. M. & KOGAN, S. C. 2003. Cooperation of cytokine signaling with chimeric transcription factors in leukemogenesis: PML-retinoic acid receptor alpha blocks growth factor-mediated differentiation. *Mol Cell Biol,* 23**,** 4573-85.

REUTHER, G. W., LAMBERT, Q. T., CALIGIURI, M. A. & DER, C. J. 2000. Identification and characterization of an activating TrkA deletion mutation in acute myeloid leukemia. *Mol Cell Biol,* 20**,** 8655-66.

RHOADES, K. L., HETHERINGTON, C. J., HARAKAWA, N., YERGEAU, D. A., ZHOU, L., LIU, L. Q., LITTLE, M. T., TENEN, D. G. & ZHANG, D. E. 2000. Analysis of the role of AML1-ETO in leukemogenesis, using an inducible transgenic mouse model. *Blood,* 96**,** 2108-15.

RHOADES, K. L., HETHERINGTON, C. J., ROWLEY, J. D., HIEBERT, S. W., NUCIFORA, G., TENEN, D. G. & ZHANG, D. E. 1996. Synergistic up-regulation of the myeloid-specific promoter for the macrophage colony-stimulating factor receptor by AML1 and the t(8;21) fusion protein may contribute to leukemogenesis. *Proc Natl Acad Sci U S A,* 93**,** 11895-900.

ROBB, L., TARRANT, J., GROOM, J., IBRAHIM, M., LI, R., BOROBAKAS, B. & WRIGHT, M. D. 2001. Molecular characterisation of mouse and human TSSC6: evidence that TSSC6 is a genuine member of the tetraspanin superfamily and is expressed specifically in haematopoietic organs. *Biochim Biophys Acta,* 1522**,** 31-41.

ROSENBAUER, F., KOSCHMIEDER, S., STEIDL, U. & TENEN, D. G. 2005. Effect of transcription-factor concentrations on leukemic stem cells. *Blood,* 106**,** 1519-24.

ROSENBAUER, F., WAGNER, K., KUTOK, J. L., IWASAKI, H., LE BEAU, M. M., OKUNO, Y., AKASHI, K., FIERING, S. & TENEN, D. G. 2004. Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1. *Nat Genet,* 36**,** 624-30.

ROUDAIA, L., CHENEY, M. D., MANUYLOVA, E., CHEN, W., MORROW, M., PARK, S., LEE, C. T., KAUR, P., WILLIAMS, O., BUSHWELLER, J. H. & SPECK, N. A. 2009. CBFbeta is critical for AML1-ETO and TEL-AML1 activity. *Blood,* 113**,** 3070-9.

ROWLEY, J. D. 1973. Identificaton of a translocation with quinacrine fluorescence in a patient with acute leukemia. *Ann Genet,* 16**,** 109-12.

RUCKER, F. G., SANDER, S., DOHNER, K., DOHNER, H., POLLACK, J. R. & BULLINGER, L. 2006. Molecular profiling reveals myeloid leukemia cell lines to be faithful model systems characterized by distinct genomic aberrations. *Leukemia,* 20**,** 994-1001.

SALOMON-NGUYEN, F., BUSSON-LE CONIAT, M., LAFAGE POCHITALOFF, M., MOZZICONACCI, J., BERGER, R. & BERNARD, O. A. 2000. AML1-MTG16 fusion gene in therapy-related acute leukemia with t(16;21)(q24;q22): two new cases. *Leukemia,* 14**,** 1704-5.

SCHESSL, C., RAWAT, V. P., CUSAN, M., DESHPANDE, A., KOHL, T. M., ROSTEN, P. M., SPIEKERMANN, K., HUMPHRIES, R. K., SCHNITTGER, S., KERN, W., HIDDEMANN, W., QUINTANILLA-MARTINEZ, L., BOHLANDER, S. K., FEURING-BUSKE, M. & BUSKE, C. 2005. The AML1-ETO fusion gene and the FLT3 length mutation collaborate in inducing acute leukemia in mice. *J Clin Invest,* 115**,** 2159-68.

SCHLENK, R. F., BENNER, A., KRAUTER, J., BUCHNER, T., SAUERLAND, C., EHNINGER, G., SCHAICH, M., MOHR, B., NIEDERWIESER, D., KRAHL,

R., PASOLD, R., DOHNER, K., GANSER, A., DOHNER, H. & HEIL, G. 2004. Individual patient data-based meta-analysis of patients aged 16 to 60 years with core binding factor acute myeloid leukemia: a survey of the German Acute Myeloid Leukemia Intergroup. *J Clin Oncol,* 22**,** 3741-50.

SCHLENK, R. F., PASQUINI, M. C., PEREZ, W. S., ZHANG, M. J., KRAUTER, J., ANTIN, J. H., BASHEY, A., BOLWELL, B. J., BUCHNER, T., CAHN, J. Y., CAIRO, M. S., COPELAN, E. A., CUTLER, C. S., DOHNER, H., GALE, R. P., ILHAN, O., LAZARUS, H. M., LIESVELD, J. L., LITZOW, M. R., MARKS, D. I., MAZIARZ, R. T., MCCARTHY, P. L., NIMER, S. D., SIERRA, J., TALLMAN, M. S., WEISDORF, D. J., HOROWITZ, M. M. & GANSER, A. 2008. HLA-identical sibling allogeneic transplants versus chemotherapy in acute myelogenous leukemia with t(8;21) in first complete remission: collaborative study between the German AML Intergroup and CIBMTR. *Biol Blood Marrow Transplant,* 14**,** 187-96.

SCHLESSINGER, K., HALL, A. & TOLWINSKI, N. 2009. Wnt signaling pathways meet Rho GTPases. *Genes Dev,* 23**,** 265-77.

SCHNITTGER, S., HAFERLACH, C., ULKE, M., ALPERMANN, T., KERN, W. & HAFERLACH, T. 2010. IDH1 mutations are detected in 6.6% of 1414 AML patients and are associated with intermediate risk karyotype and unfavorable prognosis in adults younger than 60 years and unmutated NPM1 status. *Blood,* 116**,** 5486-96.

SCHWIEGER, M., LOHLER, J., FRIEL, J., SCHELLER, M., HORAK, I. & STOCKING, C. 2002. AML1-ETO inhibits maturation of multiple lymphohematopoietic lineages and induces myeloblast transformation in synergy with ICSBP deficiency. *J Exp Med,* 196**,** 1227-40.

SCOTT, L. M., MUELLER, L. & COLLINS, S. J. 1996. E3, a hematopoietic-specific transcript directly regulated by the retinoic acid receptor alpha. *Blood,* 88**,** 2517-30.

SHANKAR, D. B., CHENG, J. C. & SAKAMOTO, K. M. 2005. Role of cyclic AMP response element binding protein in human leukemias. *Cancer,* 104**,** 1819-24.

SHAPIRO, M. B. & SENAPATHY, P. 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res,* 15**,** 7155-74.

SHIMADA, H., ICHIKAWA, H., NAKAMURA, S., KATSU, R., IWASA, M., KITABAYASHI, I. & OHKI, M. 2000. Analysis of genes under the downstream control of the t(8;21) fusion protein AML1-MTG8: overexpression of the TIS11b (ERF-1, cMG1) gene induces myeloid cell proliferation in response to G-CSF. *Blood,* 96**,** 655-63.

SHIMIZU, K., KITABAYASHI, I., KAMADA, N., ABE, T., MASEKI, N., SUZUKAWA, K. & OHKI, M. 2000. AML1-MTG8 leukemic protein induces the expression of granulocyte colony-stimulating factor (G-CSF) receptor through the up-regulation of CCAAT/enhancer binding protein epsilon. *Blood,* 96**,** 288-96.

SHIMIZU, R., KUROHA, T., OHNEDA, O., PAN, X., OHNEDA, K., TAKAHASHI, S., PHILIPSEN, S. & YAMAMOTO, M. 2004. Leukemogenesis caused by incapacitated GATA-1 function. *Mol Cell Biol,* 24**,** 10814-25.

SLOVAK, M. L., KOPECKY, K. J., CASSILETH, P. A., HARRINGTON, D. H., THEIL, K. S., MOHAMED, A., PAIETTA, E., WILLMAN, C. L., HEAD, D. R., ROWE, J. M., FORMAN, S. J. & APPELBAUM, F. R. 2000. Karyotypic analysis predicts outcome of preremission and postremission therapy in adult acute myeloid leukemia: a Southwest Oncology Group/Eastern Cooperative Oncology Group Study. *Blood,* 96**,** 4075-83.

SMITH, M. L., ARCH, R., SMITH, L. L., BAINTON, N., NEAT, M., TAYLOR, C., BONNET, D., CAVENAGH, J. D., ANDREW LISTER, T. & FITZGIBBON, J. 2005. Development of a human acute myeloid leukaemia screening panel and consequent identification of novel gene mutation in FLT3 and CCND3. *Br J Haematol,* 128**,** 318-23.

SMITH, S. C. & THEODORESCU, D. 2009. Learning therapeutic lessons from metastasis suppressor proteins. *Nat Rev Cancer,* 9**,** 253-64.

SROCZYNSKA, P., LANCRIN, C., KOUSKOFF, V. & LACAUD, G. 2009. The differential activities of Runx1 promoters define milestones during embryonic hematopoiesis. *Blood,* 114**,** 5279-89.

STAMATOPOULOS, B., MEULEMAN, N., HAIBE-KAINS, B., SAUSSOY, P., VAN DEN NESTE, E., MICHAUX, L., HEIMANN, P., MARTIAT, P., BRON, D. & LAGNEAUX, L. 2009. microRNA-29c and microRNA-223 down-regulation has in vivo significance in chronic lymphocytic leukemia and improves disease risk stratification. *Blood,* 113**,** 5237-45.

TAKAHASHI, S. 2011. Current findings for recurring mutations in acute myeloid leukemia. *J Hematol Oncol,* 4**,** 36.

TAKAKURA, N., WATANABE, T., SUENOBU, S., YAMADA, Y., NODA, T., ITO, Y., SATAKE, M. & SUDA, T. 2000. A role for hematopoietic stem cells in promoting angiogenesis. *Cell,* 102**,** 199-209.

TAKUWA, Y. 2002. Subtype-specific differential regulation of Rho family G proteins and cell migration by the Edg family sphingosine-1-phosphate receptors. *Biochim Biophys Acta,* 1582**,** 112-20.

TANAKA, K., TANAKA, T., KUROKAWA, M., IMAI, Y., OGAWA, S., MITANI, K., YAZAKI, Y. & HIRAI, H. 1998. The AML1/ETO(MTG8) and AML1/Evi-1 leukemia-associated chimeric oncoproteins accumulate PEBP2beta(CBFbeta) in the nucleus more efficiently than wild-type AML1. *Blood,* 91**,** 1688-99.

TANAKA, Y., WATANABE, T., CHIBA, N., NIKI, M., KUROIWA, Y., NISHIHIRA, T., SATOMI, S., ITO, Y. & SATAKE, M. 1997. The protooncogene product, PEBP2beta/CBFbeta, is mainly located in the cytoplasm and has an affinity with cytoskeletal structures. *Oncogene,* 15**,** 677-83.

TASKESEN, E., BULLINGER, L., CORBACIOGLU, A., SANDERS, M. A., ERPELINCK, C. A., WOUTERS, B. J., VAN DER POEL-VAN DE LUYTGAARDE, S. C., DAMM, F., KRAUTER, J., GANSER, A., SCHLENK, R. F., LOWENBERG, B., DELWEL, R., DOHNER, H., VALK, P. J. & DOHNER, K. Prognostic impact, concurrent genetic mutations, and gene expression features of AML with CEBPA mutations in a cohort of 1182 cytogenetically normal AML patients: further evidence for CEBPA double mutant AML as a distinctive disease entity. *Blood,* 117**,** 2469-75.

TIGHE, J. E. & CALABI, F. 1994. Alternative, out-of-frame runt/MTG8 transcripts are encoded by the derivative (8) chromosome in the t(8;21) of acute myeloid leukemia M2. *Blood,* 84**,** 2115-21.

TONKS, A., PEARN, L., TONKS, A. J., PEARCE, L., HOY, T., PHILLIPS, S., FISHER, J., DOWNING, J. R., BURNETT, A. K. & DARLEY, R. L. 2003. The AML1-ETO fusion gene promotes extensive self-renewal of human primary erythroid cells. *Blood,* 101**,** 624-32.

UNOKI, M., NISHIDATE, T. & NAKAMURA, Y. 2004. ICBP90, an E2F-1 target, recruits HDAC1 and binds to methyl-CpG through its SRA domain. *Oncogene,* 23**,** 7601-10.

VAN DE LOCHT, L. T., SMETSERS, T. F., WITTEBOL, S., RAYMAKERS, R. A. & MENSINK, E. J. 1994. Molecular diversity in AML1/ETO fusion transcripts in patients with t(8;21) positive acute myeloid leukaemia. *Leukemia,* 8**,** 1780-4.

VANGALA, R. K., HEISS-NEUMANN, M. S., RANGATIA, J. S., SINGH, S. M., SCHOCH, C., TENEN, D. G., HIDDEMANN, W. & BEHRE, G. 2003. The myeloid master regulator transcription factor PU.1 is inactivated by AML1-ETO in t(8;21) myeloid leukemia. *Blood,* 101**,** 270-7.

VARDIMAN, J. W., THIELE, J., ARBER, D. A., BRUNNING, R. D., BOROWITZ, M. J., PORWIT, A., HARRIS, N. L., LE BEAU, M. M., HELLSTROM-LINDBERG, E., TEFFERI, A. & BLOOMFIELD, C. D. 2009. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood,* 114**,** 937-51.

VASSEN, L., KHANDANPOUR, C., EBELING, P., VAN DER REIJDEN, B. A., JANSEN, J. H., MAHLMANN, S., DUHRSEN, U. & MOROY, T. 2009. Growth factor independent 1b (Gfi1b) and a new splice variant of Gfi1b are highly expressed in patients with acute and chronic leukemia. *Int J Hematol,* 89**,** 422-30.

VIRAPPANE, P., GALE, R., HILLS, R., KAKKAS, I., SUMMERS, K., STEVENS, J., ALLEN, C., GREEN, C., QUENTMEIER, H., DREXLER, H., BURNETT, A., LINCH, D., BONNET, D., LISTER, T. A. & FITZGIBBON, J. 2008. Mutation of the Wilms' tumor 1 gene is a poor prognostic factor associated with chemotherapy resistance in normal karyotype acute myeloid leukemia: the United Kingdom Medical Research Council Adult Leukaemia Working Party. *J Clin Oncol,* 26**,** 5429-35.

WANG, J., HOSHINO, T., REDNER, R. L., KAJIGAYA, S. & LIU, J. M. 1998. ETO, fusion partner in t(8;21) acute myeloid leukemia, represses transcription by interaction with the human N-CoR/mSin3/HDAC1 complex. *Proc Natl Acad Sci U S A,* 95**,** 10860-5.

WANG, J., SAUNTHARARAJAH, Y., REDNER, R. L. & LIU, J. M. 1999. Inhibitors of histone deacetylase relieve ETO-mediated repression and induce differentiation of AML1-ETO leukemia cells. *Cancer Res,* 59**,** 2766-9.

WANG, Q., STACY, T., BINDER, M., MARIN-PADILLA, M., SHARPE, A. H. & SPECK, N. A. 1996a. Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. *Proc Natl Acad Sci U S A,* 93**,** 3444-9.

WANG, Q., STACY, T., MILLER, J. D., LEWIS, A. F., GU, T. L., HUANG, X., BUSHWELLER, J. H., BORIES, J. C., ALT, F. W., RYAN, G., LIU, P. P., WYNSHAW-BORIS, A., BINDER, M., MARIN-PADILLA, M., SHARPE, A. H. & SPECK, N. A. 1996b. The CBFbeta subunit is essential for CBFalpha2 (AML1) function in vivo. *Cell,* 87**,** 697-708.

WANG, S., WANG, Q., CRUTE, B. E., MELNIKOVA, I. N., KELLER, S. R. & SPECK, N. A. 1993. Cloning and characterization of subunits of the T-cell receptor and murine leukemia virus enhancer core-binding factor. *Mol Cell Biol,* 13**,** 3324-39.

WANG, S. W. & SPECK, N. A. 1992. Purification of core-binding factor, a protein that binds the conserved core site in murine leukemia virus enhancers. *Mol Cell Biol,* 12**,** 89-102.

WANG, Y. Y., ZHOU, G. B., YIN, T., CHEN, B., SHI, J. Y., LIANG, W. X., JIN, X. L., YOU, J. H., YANG, G., SHEN, Z. X., CHEN, J., XIONG, S. M., CHEN, G. Q., XU, F., LIU, Y. W., CHEN, Z. & CHEN, S. J. 2005. AML1-ETO and C-KIT mutation/overexpression in t(8;21) leukemia: implication in stepwise leukemogenesis and response to Gleevec. *Proc Natl Acad Sci U S A,* 102**,** 1104-9.

WARREN, A. J., BRAVO, J., WILLIAMS, R. L. & RABBITTS, T. H. 2000. Structural basis for the heterodimeric interaction between the acute leukaemia-associated transcription factors AML1 and CBFbeta. *Embo J,* 19**,** 3004-15.

WESTENDORF, J. J., YAMAMOTO, C. M., LENNY, N., DOWNING, J. R., SELSTED, M. E. & HIEBERT, S. W. 1998. The t(8;21) fusion product, AML-1-ETO, associates with C/EBP-alpha, inhibits C/EBP-alpha-dependent transcription, and blocks granulocytic differentiation. *Mol Cell Biol,* 18**,** 322-33.

WICHMANN, C., CHEN, L., HEINRICH, M., BAUS, D., PFITZNER, E., ZORNIG, M., OTTMANN, O. G. & GREZ, M. 2007. Targeting the oligomerization domain of ETO interferes with RUNX1/ETO oncogenic activity in t(8;21)-positive leukemic cells. *Cancer Res,* 67**,** 2280-9.

WIEMELS, J. L., XIAO, Z., BUFFLER, P. A., MAIA, A. T., MA, X., DICKS, B. M., SMITH, M. T., ZHANG, L., FEUSNER, J., WIENCKE, J., PRITCHARD-JONES, K., KEMPSKI, H. & GREAVES, M. 2002. In utero origin of t(8;21) AML1-ETO translocations in childhood acute myeloid leukemia. *Blood,* 99**,** 3801-5.

WOLFORD, J. K. & PROCHAZKA, M. 1998. Structure and expression of the human MTG8/ETO gene. *Gene,* 212**,** 103-9.

WOO, H. N., HONG, G. S., JUN, J. I., CHO, D. H., CHOI, H. W., LEE, H. J., CHUNG, C. W., KIM, I. K., JO, D. G., PYO, J. O., BERTIN, J. & JUNG, Y. K. 2004. Inhibition of Bcl10-mediated activation of NF-kappa B by BinCARD, a Bcl10-interacting CARD protein. *FEBS Lett,* 578**,** 239-44.

YAGI, R., CHEN, L. F., SHIGESADA, K., MURAKAMI, Y. & ITO, Y. 1999. A WW domain-containing yes-associated protein (YAP) is a novel transcriptional co-activator. *Embo J,* 18**,** 2551-62.

YAN, M., BUREL, S. A., PETERSON, L. F., KANBE, E., IWASAKI, H., BOYAPATI, A., HINES, R., AKASHI, K. & ZHANG, D. E. 2004. Deletion of an AML1-

ETO C-terminal NcoR/SMRT-interacting region strongly induces leukemia development. *Proc Natl Acad Sci U S A,* 101**,** 17186-91.

YAN, M., KANBE, E., PETERSON, L. F., BOYAPATI, A., MIAO, Y., WANG, Y., CHEN, I. M., CHEN, Z., ROWLEY, J. D., WILLMAN, C. L. & ZHANG, D. E. 2006. A previously unidentified alternatively spliced isoform of t(8;21) transcript promotes leukemogenesis. *Nat Med,* 12**,** 945-9.

YERGEAU, D. A., HETHERINGTON, C. J., WANG, Q., ZHANG, P., SHARPE, A. H., BINDER, M., MARIN-PADILLA, M., TENEN, D. G., SPECK, N. A. & ZHANG, D. E. 1997. Embryonic lethality and impairment of haematopoiesis in mice heterozygous for an AML1-ETO fusion gene. *Nat Genet,* 15**,** 303-6.

YEYATI, P. L., SHAKNOVICH, R., BOTERASHVILI, S., LI, J., BALL, H. J., WAXMAN, S., NASON-BURCHENAL, K., DMITROVSKY, E., ZELENT, A. & LICHT, J. D. 1999. Leukemia translocation protein PLZF inhibits cell growth and expression of cyclin A. *Oncogene,* 18**,** 925-34.

YIN, C. C., CORTES, J., BARKOH, B., HAYES, K., KANTARJIAN, H. & JONES, D. 2006. t(3;21)(q26;q22) in myeloid leukemia: an aggressive syndrome of blast transformation associated with hydroxyurea or antimetabolite therapy. *Cancer,* 106**,** 1730-8.

YUAN, Y., ZHOU, L., MIYAMOTO, T., IWASAKI, H., HARAKAWA, N., HETHERINGTON, C. J., BUREL, S. A., LAGASSE, E., WEISSMAN, I. L., AKASHI, K. & ZHANG, D. E. 2001. AML1-ETO expression is directly involved in the development of acute myeloid leukemia in the presence of additional mutations. *Proc Natl Acad Sci U S A,* 98**,** 10398-403.

ZENG, C., VAN WIJNEN, A. J., STEIN, J. L., MEYERS, S., SUN, W., SHOPLAND, L., LAWRENCE, J. B., PENMAN, S., LIAN, J. B., STEIN, G. S. & HIEBERT, S. W. 1997. Identification of a nuclear matrix targeting signal in the leukemia and bone-related AML/CBF-alpha transcription factors. *Proc Natl Acad Sci U S A,* 94**,** 6746-51.

ZHANG, D. E., HETHERINGTON, C. J., MEYERS, S., RHOADES, K. L., LARSON, C. J., CHEN, H. M., HIEBERT, S. W. & TENEN, D. G. 1996. CCAAT enhancer-binding protein (C/EBP) and AML1 (CBF alpha2) synergistically activate the macrophage colony-stimulating factor receptor promoter. *Mol Cell Biol,* 16**,** 1231-40.

ZHANG, J., HUG, B. A., HUANG, E. Y., CHEN, C. W., GELMETTI, V., MACCARANA, M., MINUCCI, S., PELICCI, P. G. & LAZAR, M. A. 2001. Oligomerization of ETO is obligatory for corepressor interaction. *Mol Cell Biol,* 21**,** 156-63.

ZHANG, J., KALKUM, M., YAMAMURA, S., CHAIT, B. T. & ROEDER, R. G. 2004. E protein silencing by the leukemogenic AML1-ETO fusion protein. *Science,* 305**,** 1286-9.

ZHANG, Y., STRISSEL, P., STRICK, R., CHEN, J., NUCIFORA, G., LE BEAU, M. M., LARSON, R. A. & ROWLEY, J. D. 2002. Genomic DNA breakpoints in AML1/RUNX1 and ETO cluster with topoisomerase II DNA cleavage and DNase I hypersensitive sites in t(8;21) leukemia. *Proc Natl Acad Sci U S A,* 99**,** 3070-5.

ZHANG, Y. E. 2009. Non-Smad pathways in TGF-beta signaling. *Cell Res,* 19**,** 128-39.

Appendix A

Gene Lists derived from Exon Arrays

Table 30 Gene List of 448 differentially expressed genes with FDR p<0.05 and FC +/-1 using the Core metaprobeset file on the Human Exon 1.0 ST array.

GENE SYMBOL	GO BIOLOGICAL PROCESS	GO CELLULAR COMPONENT	GO MOLECULAR FUNCTION	p-VALUE	FOLD-CHANGE
MEIS1	regulation of transcription, DNA-dependent	nucleus	transcription factor activity	4.30E-06	-21.7739
HOXA9	regulation of transcription, DNA-dependent	nucleus	transcription factor activity	7.75E-06	-14.269
CPNE8				4.92E-06	-13.8429
SLC40A1	ion transport	cytoplasm	iron ion transmembrane transporter activity	5.93E-05	-9.27787
KIAA0125				1.68E-06	-9.23424
CAPG	protein complex assembly	nucleus	actin binding	8.71E-06	-7.37208
C1orf150				2.27E-06	-7.3061
ANGPT1	angiogenesis	extracellular region	receptor binding	0.00035529	-6.64078
HOXA5	regulation of transcription, DNA-dependent	nucleus	transcription factor activity	0.00010866 9	-6.48519
LAT2	immune response	plasma membrane	protein binding	2.80E-10	-6.46607
NRG4		extracellular region	growth factor activity	1.57E-05	-6.3083
PLXNC1	cell adhesion	intracellular	receptor activity	2.44E-06	-5.87562
KIAA0125				5.81E-08	-5.72191
PBX3	regulation of transcription, DNA-dependent	nucleus	transcription factor activity	0.00051959 2	-5.61075
COL4A5		extracellular region	extracellular matrix structural constituent	0.00015335 4	-5.18577
TSPAN2		membrane		0.00077091 2	-5.16152
WBP5			protein binding	2.77E-05	-5.08686
hCG_1645 516	regulation of transcription, DNA-dependent		DNA binding	6.39E-07	-5.04596
HOXB3	regulation of transcription, DNA-dependent	nucleus	transcription factor activity	0.00046102 4	-4.92716
LTBP1	biological_process	extracellular region	transforming growth factor beta receptor activity	0.00048890 4	-4.92319
HOXB6	regulation of transcription, DNA-dependent	nucleus	transcription factor activity	0.00011744 9	-4.90793

CTSW	proteolysis		cysteine-type endopeptidase activity	1.36E-05	-4.80151
TSPAN32	cell-cell signaling	membrane	molecular function	6.02E-12	-4.58844
NGFRAP1	apoptosis	nucleus	protein binding	9.85E-06	-4.53085
SPNS3	lipid transport	membrane		8.97E-08	-4.40062
MED12L	transcription	nucleus	transcription regulator activity	8.66E-06	-4.38581
HOXB2	regulation of transcription, DNA-dependent	nucleus	transcription factor activity	7.82E-05	-4.27036
MPP7		tight junction	protein binding	7.15E-06	-4.18498
BTBD11		membrane	DNA binding	0.000161817	-4.13874
UGCGL2	protein amino acid glycosylation	endoplasmic reticulum	UDP-glucose: glycoprotein glucosyltransferase activity	0.000196265	-3.99205
HNRPLL	mRNA processing	nucleus	nucleotide binding	1.59E-05	-3.90223
HOXB8	regulation of transcription, DNA-dependent	nucleus	transcription factor activity	0.00118619	-3.78397
EMR1	cell adhesion	plasma membrane	G-protein coupled receptor activity	6.08E-05	-3.78063
SLC44A1	transport	membrane	choline transmembrane transporter activity	4.46E-05	-3.75674
HOXB5	regulation of transcription, DNA-dependent	nucleus	transcription factor activity	0.00100033	-3.72638
				0.000549725	-3.50534
HOXA7	negative regulation of transcription from RNA polymer	nucleus	transcription factor activity	0.000106895	-3.49503
SCHIP1	Biological process	cellular_component	molecular_function	0.000470938	-3.41816
LGALS1	regulation of apoptosis	extracellular space	signal transducer activity	0.000408463	-3.41033
HOXA4	regulation of transcription, DNA-dependent	nucleus	transcription factor activity	0.000801101	-3.4083
SSX2IP	cell adhesion	nucleus	protein binding	1.35E-07	-3.36613
SELPLG	cell adhesion	membrane fraction	receptor binding	1.91E-05	-3.3538
HOXA2	negative regulation of transcription from RNA polymer	nucleus	transcription factor activity	0.000389982	-3.28137
GFI1B	regulation of transcription	intracellular	DNA binding	4.59E-06	-3.22941

	during G1 phase of mitoti				
HOXA3	blood vessel remodeling	nucleus	transcription factor activity	0.000329629	-3.17032
CAT	response to reactive oxygen species	mitochondrion	aminoacylase activity	2.30E-06	-3.15249
CXCL2	chemotaxis	extracellular region	chemokine activity	0.000444284	-3.14042
CYP7B1	lipid metabolic process	endoplasmic reticulum	monooxygenase activity	0.000607808	-3.09312
SDPR		cytoplasm	phosphatidylserine binding	3.32E-05	-3.09288
PTPN22	protein amino acid dephosphorylation	cytoplasm	protein tyrosine phosphatase activity	2.40E-05	-3.08947
LSP1	cell motion	plasma membrane	actin binding	0.000273095	-3.07237
ADCY7	cAMP biosynthetic process	plasma membrane	magnesium ion binding	5.93E-06	-3.02996
NFIL3	transcription	nucleus	transcription factor activity	3.16E-05	-2.99889
JAG1	angiogenesis	extracellular region	Notch binding	0.000149946	-2.99747
CMAH	biological_process	cytoplasm	electron carrier activity	3.64E-05	-2.96105
ALCAM	cell adhesion	membrane	receptor binding	0.00106233	-2.95225
C1orf162		membrane		0.000387038	-2.94159
PDGFD	cell proliferation	extracellular region	growth factor activity	0.000430734	-2.93266
CD300A	immune response	plasma membrane	receptor activity	8.87E-05	-2.92134
C5orf53		extracellular region		0.000154674	-2.88712
SLC46A3		membrane		0.000417741	-2.85646
C11orf21		cytoplasm		1.05E-07	-2.83386
PRKCD	protein amino acid phosphorylation	nucleus	nucleotide binding	1.13E-05	-2.75581
KCNK17	ion transport	membrane	voltage-gated ion channel activity	0.000555368	-2.75185
ABCA1	transport	membrane fraction	nucleotide binding	6.77E-05	-2.7176
C15orf39				3.18E-09	-2.70099
LST1	cell morphogenesis	Golgi membrane	molecular_function	8.74E-05	-2.69105
CPEB2	regulation of translation	cytoplasm	nucleotide binding	0.000155393	-2.6822
BST2	humoral immune response	Golgi apparatus	signal transducer activity	0.00109661	-2.68131
RECK	negative	peripheral to	serine-type	5.03E-05	-2.66901

	regulation of cell cycle	membrane of membrane fraction	endopeptidase inhibitor activity		
MAP7	microtubule cytoskeleton organization	cytoplasm	structural molecule activity	0.000278934	-2.66847
FUT7	protein amino acid glycosylation	Golgi apparatus	transferase activity, transferring glycosyl groups	0.00035261	-2.65019
TRIM8	biological_proces	cellular_compo nent	protein binding	2.65E-06	-2.62267
AGPAT9	metabolic process	endoplasmic reticulum	1-acylglycerol-3-phosphate O-acyltransferase activity	0.000449553	-2.60573
CAPRIN2	negative regulation of cell growth	cytoplasm		3.83E-06	-2.59902
NIPA1		membrane		9.20E-05	-2.54478
ITGB7	cell adhesion	integrin complex	magnesium ion binding	0.00110695	-2.54408
NFE2	nucleosome disassembly	nucleus	transcription factor activity	1.23E-05	-2.53148
GABARAP L1		intracellular	protein binding	6.77E-06	-2.52939
SASH3	positive regulation of immunoglobulin production	nucleus		1.71E-06	-2.52829
HOXB4	negative regulation of transcription from RNA polymer	nucleus	transcription factor activity	0.000791581	-2.52815
GPR114	neuropeptide signaling pathway	plasma membrane	G-protein coupled receptor activity	3.57E-06	-2.49612
NCRNA001 73				0.000533686	-2.4669
SLA		cytoplasm	SH3	0.00100925	-2.44588
C10orf54		membrane	receptor activity	0.00116098	-2.44222
ECOP				0.000111701	-2.43758
GBGT1	carbohydrate metabolic process	Golgi membrane	transferase activity, transferring hexosyl groups	8.19E-05	-2.43647
GCLM	cysteine metabolic process	soluble fraction	glutamate-cysteine ligase activity	2.10E-05	-2.42582
BEX2	multicellular organismal development	nucleus		0.000699113	-2.38314
ATXN1	RNA processing	nucleus	RNA binding	0.000279414	-2.38013

CTSD	autophagic vacuole formation	extracellular region	aspartic-type endopeptidase activity	0.000101984	-2.36187
ITGAX	cell adhesion	integrin complex	magnesium ion binding	2.95E-05	-2.36
FLJ11710				0.00107503	-2.35173
LAPTM5	transport	lysosomal membrane		1.29E-05	-2.34841
				0.000496448	-2.31836
TBC1D8B	regulation of Rab GTPase activity	intracellular	GTPase activator activity	3.22E-05	-2.29215
ADARB1	mRNA processing	intracellular	double-stranded RNA binding	5.73E-05	-2.24964
ELL2	transcription	nucleus	RNA polymerase II transcription factor activity	0.000968956	-2.23995
TPM1	positive regulation of heart rate by epinephrine	stress fiber	actin binding	3.70E-05	-2.22996
C9orf89	negative regulation of I-kappaB kinase	nucleus	CARD domain binding	1.03E-06	-2.1954
SH2D3C	JNK cascade	intracellular	SH3	8.18E-06	-2.18833
CCND3	positive regulation of protein amino acid phosphoryla	cyclin-dependent protein kinase holoenzyme complex	protein binding	0.000284016	-2.17201
GPR132	G1	plasma membrane	receptor activity	0.000408561	-2.16618
LPXN	protein complex assembly	cytoplasm	zinc ion binding	0.000118028	-2.16546
CD82		plasma membrane	protein binding	0.000180677	-2.15598
PLCL1	lipid metabolic process	cytoplasm	phosphoinositide phospholipase C activity	0.000807905	-2.13828
KIAA0513		cytoplasm		3.33E-06	-2.12616
SYTL4	transport	plasma membrane	transporter activity	0.000904193	-2.11296
ELF4	natural killer cell proliferation	nucleus	transcription factor activity	1.03E-05	-2.08202
ABCA7	transport	Golgi membrane	nucleotide binding	2.56E-06	-2.06555
C9orf95	pyridine nucleotide biosynthetic process		nucleotide binding	0.00103088	-2.03839
NINJ1	cell adhesion	membrane	protein binding	6.61E-05	-2.03284
GAS2L3	cell cycle arrest	glycosylphosphatidylinositol-N-acetylglucosaminyltran	protein binding	0.000932724	-2.03145

				0.0004546	-2.02643
C1orf217				0.0004546	-2.02643
MAP2K6	activation of MAPK activity		nucleotide binding	0.000262006	-2.0132
C17orf91				0.000604455	-2.00719
BBS2	sperm axoneme assembly	cytoplasm	protein binding	6.43E-05	-1.99647
GNB5	signal transduction	nucleus	GTPase activity	0.000141555	-1.98786
WDFY4		membrane	binding	0.000263963	-1.96759
CYP4V2	visual perception	endoplasmic reticulum	monooxygenase activity	0.000273131	-1.96286
PPM1M	protein amino acid dephosphorylation	nucleus	magnesium ion binding	0.000725513	-1.94834
AIF1	response to stress	nucleus	calcium ion binding	0.000684177	-1.93043
BTG2	DNA repair	nucleus	protein binding	0.000326775	-1.92795
C11orf63				0.000490469	-1.91537
TMEM65		membrane		0.000475549	-1.91292
RAB24	autophagy	late endosome	nucleotide binding	0.000171693	-1.90865
SRXN1	response to oxidative stress	cytoplasm	nucleotide binding	0.000982309	-1.89357
HEBP1	circadian rhythm	cytoplasm	binding	0.000496086	-1.89158
ICAM2	cell-cell adhesion	plasma membrane	integrin binding	0.000282062	-1.88933
BLOC1S1	melanosome organization	cytosol	protein binding	0.000430625	-1.87382
SLC22A20	ion transport	membrane		0.000328347	-1.87361
CD97	cell motion	extracellular region	G-protein coupled receptor activity	0.000203927	-1.86541
MVP	protein transport	nucleus	protein binding	0.000201115	-1.86042
SLC24A6	ion transport	plasma membrane	calcium ion binding	3.35E-05	-1.84444
RMND5B				0.000175816	-1.84332
WDFY4		membrane	binding	0.00106092	-1.83932
ARHGAP22	angiogenesis	intracellular	GTPase activator activity	0.000128199	-1.83507
MTMR1	dephosphorylation	cellular_component	protein tyrosine phosphatase activity	0.000248383	-1.82994
FNBP1	endocytosis	cytoplasm	lipid binding	2.02E-05	-1.82357
SYTL1	intracellular protein transport	soluble fraction	Rab GTPase binding	7.97E-05	-1.82339

SKIL	transcription from RNA polymerase II promoter	nucleus	nucleotide binding	0.000913168	-1.82193
PARVG	cell adhesion	cytoplasm	actin binding	3.96E-06	-1.81451
DNMT3B	DNA methylation	nucleus	DNA binding	0.00100801	-1.8027
CNNM1	ion transport	plasma membrane		0.000617148	-1.7961
C20orf197				0.00124074	-1.79484
TGFB1	skeletal system development	extracellular region	type II transforming growth factor beta receptor bind	2.76E-05	-1.79352
PLBD2	lipid catabolic process	lysosomal lumen	hydrolase activity	0.000537488	-1.79336
MBP	immune response	plasma membrane	structural constituent of myelin sheath	6.41E-05	-1.77519
IDE	proteolysis	extracellular space	metalloendopeptidase activity	5.87E-05	-1.76771
EIF4A1	translation	cytosol	nucleotide binding	0.000630277	-1.75197
CDRT4				0.000751491	-1.75144
MYO7B		cytoskeleton	nucleotide binding	0.000979686	-1.74874
SH3BP2	signal transduction		SH3	0.000548766	-1.72992
ACP2	skeletal system development	lysosomal membrane	phosphotyrosine binding	0.000178874	-1.70914
DPEP2	proteolysis	membrane	peptidase activity	2.22E-05	-1.69009
OPHN1	substrate-bound cell migration, cell extension	intracellular	actin binding	0.000330652	-1.68806
CCDC122				1.80E-05	-1.67325
VAV1	phagocytosis	intracellular	transcription factor activity	5.22E-05	-1.66865
SMAD3	negative regulation of transcription from RNA polymer	intracellular	double-stranded DNA binding	0.000262273	-1.65456
OR2G3	signal transduction	plasma membrane	receptor activity	0.000931241	-1.64773
AMPD3	purine base metabolic process		AMP deaminase activity	0.000349534	-1.6454
SIPA1	cytoskeleton organization	intracellular	GTPase activator activity	2.70E-06	-1.6328
CEBPA	urea cycle	nucleus	RNA polymerase II transcription factor activity, enha	0.000710649	-1.62943
TALDO1	carbohydrate metabolic process	cytoplasm	transaldolase activity	8.10E-05	-1.62352

GARNL4	regulation of small GTPase mediated signal transducti	intracellular	GTPase activator activity	0.000681654	-1.61798
PLCB2	phospholipid metabolic process		phosphoinositide phospholipase C activity	0.000534088	-1.61757
SPON2	immune response	extracellular region	protein binding	0.000118114	-1.61712
NUP210	protein transport	nucleus	protein binding	0.000223543	-1.61167
PPCDC	coenzyme A biosynthetic process		phosphopantothe noylcysteine decarboxylase activity	4.65E-05	-1.58971
NINJ2	cell adhesion	integral to plasma membrane	protein binding	0.000378429	-1.58717
VCL	cell motion	cytoplasm	actin binding	0.000445131	-1.58274
LRRC49		cytoplasm	protein binding	0.000430923	-1.56565
SLC25A28	ion transport	mitochondrion	binding	0.000278867	-1.5616
ZNF133	transcription	intracellular	transcription factor activity	0.00053971	-1.55992
SEMA4D	anti-apoptosis	membrane	receptor activity	0.000491748	-1.52265
C14orf83		membrane		2.89E-05	-1.51411
STK10	protein amino acid phosphorylation		nucleotide binding	3.62E-06	-1.51347
FGD3	regulation of cell shape	ruffle	guanyl-nucleotide exchange factor activity	0.00037683	-1.50351
ENKUR		cilium	calmodulin binding	0.000227185	-1.4962
RIPK3	protein modification process	intracellular	nucleotide binding	0.000138492	-1.49389
GPT2	biosynthetic process		L-alanine:2-oxoglutarate aminotransferase activity	0.000718714	-1.47947
SERGEF	signal transduction	nucleus	guanyl-nucleotide exchange factor activity	0.00112327	-1.46573
LTB4R2	chemotaxis	membrane fraction	leukotriene B4 receptor activity	0.00120594	-1.4653
CLNK	immune response	intracellular	SH3	0.000263347	-1.46473
PTPN6	protein amino acid dephosphorylation	nucleus	protein tyrosine phosphatase activity	0.000535094	-1.46037
AACS	lipid metabolic	cytoplasm	ligase activity	0.00074441	-1.45965

	process			7	
PI4KA	phosphatidylinosit ol biosynthetic process	Golgi-associated vesicle	1-phosphatidylinosit ol 4-kinase activity	0.00114092	-1.4591
TRABD				0.00014540 1	-1.45805
SH3GLB2		cytoplasm	protein binding	0.00065121 2	-1.45739
GPR107		membrane		0.00093797	-1.45449
RGS14	regulation of G-protein coupled receptor protein sign	membrane	receptor signaling protein activity	0.00016773 2	-1.45232
ANKMY1			zinc ion binding	0.00044075	-1.44591
NDNL2	regulation of transcription, DNA-dependent	nucleus	protein binding	0.00042686 6	-1.44225
TMBIM4		membrane	protein binding	0.00011066 7	-1.43443
FLJ37396		membrane		0.00123062	-1.43338
GRK6	protein amino acid phosphorylation	membrane	nucleotide binding	2.27E-05	-1.43099
CDK9	transcription	nucleus	nucleotide binding	0.00015557 4	-1.43001
TM2D3		membrane		0.00012749 1	-1.42315
REEP6		membrane		0.00121025	-1.42281
KLC1	axon cargo transport	cytoplasm	microtubule motor activity	0.00038707 1	-1.41655
SH3BP1	signal transduction	intracellular	GTPase activator activity	0.00118531	-1.40241
TSLP		extracellular region	cytokine activity	0.00085544 1	-1.40176
NUDT13			hydrolase activity	0.00033758 2	-1.38994
PSD4	regulation of ARF protein signal transduction	intracellular	ARF guanyl-nucleotide exchange factor activity	0.00031228 8	-1.37548
C12orf59		membrane		0.00069524	-1.35451
GFER	spermatogenesis	cellular_compo nent	protein binding	0.00080958 2	-1.34615
TICAM1	apoptosis	cytosol	signal transducer activity	0.00056860 5	-1.34504
CCDC84				0.00041333	-1.34298
HTT	urea cycle	soluble fraction	transcription corepressor activity	0.00082696 9	-1.33302
ELMOD3	phagocytosis	cytoskeleton		0.00087833 1	-1.33282
UBAC2		membrane		0.00115142	-1.32839

FAM78A				0.00040379	-1.32732
DEF6		nucleus		3.01E-06	-1.30322
HPS1	lysosome organization	membrane fraction	protein dimerization activity	5.66E-06	-1.25301
TOP3A	DNA topological change	nucleus	DNA topoisomerase type I activity	0.00017297 2	1.22896
DDX56	rRNA processing	nucleus	nucleotide binding	0.00120209	1.30068
USP10	ubiquitin-dependent protein catabolic process		ubiquitin thiolesterase activity	0.00085000 3	1.33409
NECAB3	protein secretion	Golgi cis cisterna	calcium ion binding	0.00024576 1	1.33518
YTHDF2	humoral immune response			0.00057162 3	1.33732
RAB40C	small GTPase mediated signal transduction	plasma membrane	nucleotide binding	0.00023237 4	1.33889
AHCTF1	transcription	nucleus	DNA binding	0.00090155 6	1.35375
TPO	response to oxidative stress	cytoplasm	iodide peroxidase activity	0.00087567 1	1.36082
CCNG2	cell cycle checkpoint	cytoplasm		0.00121933	1.37762
LPCAT4	metabolic process	endoplasmic reticulum	1-acylglycerol-3-phosphate O-acyltransferase activity	0.00095470 2	1.38109
C20orf4			protein binding	0.00023398 8	1.38656
C7orf68	response to stress	cellular_compo nent	molecular_functio n	0.00047319 3	1.39432
CTNNB1	negative regulation of transcription from RNA polymer	membrane fraction	chromatin binding	0.00114135	1.39591
DKFZp451 A211				0.0011624	1.39641
TMEM11	signal transduction	integral to plasma membrane	G-protein coupled receptor activity	0.00048581 5	1.39782
LMBR1		membrane	receptor activity	0.00083940 8	1.40494
WIPF1	protein complex assembly	cytoplasm	actin binding	0.00096310 8	1.40696
CDH26	cell adhesion	plasma membrane	calcium ion binding	0.00037851 6	1.40968
C8orf41		nucleus		0.00010168 1	1.42226
MGAT5	protein amino acid N-linked glycosylation	Golgi membrane	transferase activity, transferring	7.83E-05	1.42842

			glycosyl groups		
NR2F6	transcription	nucleus	transcription factor activity	5.89E-05	1.43035
TMEM41A		membrane		3.45E-05	1.4413
SCGB3A1	negative regulation of cell growth	extracellular region	cytokine activity	0.00114653	1.44361
BLMH	proteolysis	soluble fraction	aminopeptidase activity	0.00024046	1.44732
RPS6KA4	regulation of transcription, DNA-dependent	nucleus	nucleotide binding	6.78E-06	1.44786
TBC1D14	regulation of Rab GTPase activity	intracellular	GTPase activator activity	0.000368966	1.45898
ZC3H3			nucleic acid binding	0.000488867	1.46274
C1orf96				0.000674685	1.4632
ZNF662	transcription	intracellular	DNA binding	0.000368172	1.46757
CLN6	ganglioside metabolic process	endoplasmic reticulum	protein homodimerization activity	0.00106363	1.47062
PARD6G	cell cycle	cytoplasm	protein binding	0.000207856	1.47234
GNG7	behavioral fear response	heterotrimeric G-protein complex	signal transducer activity	0.000287043	1.4755
ZNF280B	transcription	intracellular	DNA binding	0.000978273	1.48491
LSM14B				0.000722411	1.48495
KIAA1383				0.000719628	1.48661
CRBN	RNA processing	cytoplasm	ATP-dependent peptidase activity	0.000337158	1.48972
LOC439914				0.000346924	1.4901
C11orf84				0.00036749	1.49155
CCND1	G1	cyclin-dependent protein kinase holoenzyme complex	protein kinase activity	0.000981673	1.49296
ATF4	gluconeogenesis	nucleus	transcription factor activity	0.000486023	1.49855
FMNL2	cellular component organization	cytoplasm	actin binding	0.0011962	1.50194
MEST	mesoderm development	Golgi apparatus	catalytic activity	0.000586476	1.50222
DOK1	cell surface receptor linked		insulin receptor binding	0.000954308	1.503

	signal transduction				
IPO7	protein import into nucleus, docking	soluble fraction	small GTPase regulator activity	0.000137229	1.50949
PLXND1	signal transduction	intracellular	receptor activity	0.000807744	1.52517
LRP5	gastrulation with mouth forming second	mitochondrion	receptor activity	0.000892866	1.52563
KDM4B	transcription	nucleus	nucleic acid binding	0.00031177	1.52597
NT5C2	nucleotide metabolic process	cytoplasm	nucleotide binding	0.000540412	1.52624
SRrp35	assembly of spliceosomal tri-snRNP	nucleus	nucleotide binding	1.16E-05	1.52657
L2HGDH	cellular protein metabolic process	mitochondrion	oxidoreductase activity	0.000206904	1.53355
PLAU	proteolysis	extracellular region	serine-type endopeptidase activity	0.000254102	1.53373
F11R	inflammatory response	plasma membrane	protein binding	0.0011187	1.53788
MIXL1	endoderm formation	nucleus	transcription factor activity	0.000552597	1.54255
SYNGR1		integral to plasma membrane		0.000390489	1.54497
CMTM7	chemotaxis	extracellular space	cytokine activity	0.000156486	1.5485
RNF219			protein binding	0.000838572	1.56961
NBL1	negative regulation of cell cycle	extracellular region		0.000304328	1.57059
SHB	angiogenesis	cytoplasm	SH3	0.000945564	1.57328
LPO	response to oxidative stress	extracellular region	peroxidase activity	8.88E-05	1.57529
RAB43	small GTPase mediated signal transduction	plasma membrane	nucleotide binding	0.000255381	1.57679
TTC28			binding	0.00103917	1.58417
CORO6				0.0010704	1.59034
CRBN	RNA processing	cytoplasm	ATP-dependent peptidase activity	0.000986728	1.60102
GALNT1	protein amino acid O-linked glycosylation via serine	extracellular region	polypeptide N-acetylgalactosaminyltransferase activit	0.000794258	1.60304
SCMH1	regulation of transcription, DNA-dependent	nucleus	transcription factor activity	0.000371078	1.60737
PITPNM2	transport	intracellular	calcium ion binding	0.000641899	1.60756

FAM129A	negative regulation of protein amino acid phosphoryla	cytoplasm	molecular_function	0.000352307	1.6113
TRIM47		intracellular	protein binding	0.00043128	1.61221
SYNGAP1	signal transduction	intracellular	GTPase activator activity	0.0009485	1.6126
PDE4B	signal transduction	soluble fraction	3',5'-cyclic-AMP phosphodiesterase activity	0.000279405	1.61386
TGFBRAP1	signal transduction	cytoplasm	transforming growth factor beta receptor binding	0.000177883	1.62448
FLT4	protein amino acid phosphorylation	integral to plasma membrane	nucleotide binding	0.000179659	1.63524
AGPAT2	phosphatidic acid biosynthetic process	endoplasmic reticulum	1-acylglycerol-3-phosphate O-acyltransferase activity	0.000149428	1.63595
GALNT2	immunoglobulin biosynthetic process	extracellular region	polypeptide N-acetylgalactosaminyltransferase activit	1.42E-06	1.638
NR3C1	regulation of gluconeogenesis	intracellular	transcription factor activity	0.000515903	1.6431
C12orf42				0.00107498	1.65198
TNFRSF10D	apoptosis	membrane	transmembrane receptor activity	4.09E-05	1.66791
HK2	glucose metabolic process	mitochondrion	nucleotide binding	5.47E-05	1.67684
RAB44	small GTPase mediated signal transduction	plasma membrane	nucleotide binding	0.000995276	1.67714
TEX2	sphingolipid metabolic process	membrane	molecular_function	0.00115724	1.6788
C2orf65				0.000338471	1.68211
LOC93622				0.000929145	1.68785
HYAL2	carbohydrate metabolic process	lysosome	hyaluronoglucosaminidase activity	0.000299496	1.68874
PLCB3	intracellular signaling cascade	insoluble fraction	phosphoinositide phospholipase C activity	8.09E-05	1.6904
ALAS1	heme biosynthetic process	mitochondrion	5-aminolevulinate synthase activity	0.000870589	1.69236
MARCKSL1	positive regulation of cell proliferation	plasma membrane	calmodulin binding	0.0012456	1.69944
PAFAH2	lipid catabolic process	cytoplasm	1-alkyl-2-acetylglycerophosphocholine esterase activi	2.31E-06	1.71182

GIPC3			protein binding	0.000197059	1.7134
PGAM5		nucleus		7.39E-05	1.71411
SLC5A3	inositol metabolic process	integral to plasma membrane	transporter activity	0.000853386	1.72229
PKN3	protein amino acid phosphorylation	intracellular	nucleotide binding	0.00115968	1.7251
GNPTAB	cell differentiation	nucleus	UDP-N-acetylglucosamine-lysosomal-enzyme N-acetylgluc	9.45E-05	1.73182
MMP28	proteolysis	extracellular region	metalloendopeptidase activity	0.000128528	1.73511
ZHX2	negative regulation of transcription from RNA polymer	intracellular	transcription factor activity	0.000226656	1.75381
PLEKHG2	regulation of Rho protein signal transduction	intracellular	guanyl-nucleotide exchange factor activity	8.15E-05	1.76193
WEE1	protein amino acid phosphorylation	nucleus	nucleotide binding	1.13E-05	1.76233
TRIO	protein amino acid phosphorylation	intracellular	nucleotide binding	0.000206609	1.79003
SYTL2	intracellular protein transport	membrane fraction	Rab GTPase binding	0.00064487	1.79181
NEDD4	protein modification process	ubiquitin ligase complex	ubiquitin-protein ligase activity	0.000920508	1.81864
ACADSB	lipid metabolic process	mitochondrion	acyl-CoA dehydrogenase activity	0.000861945	1.82738
DAGLB	lipid catabolic process	plasma membrane	triacylglycerol lipase activity	4.87E-05	1.83607
PXMP2		peroxisome	protein binding	0.00093277	1.85357
RASSF8	signal transduction			4.49E-05	1.8545
FNDC3B		endoplasmic reticulum		0.000710654	1.85746
DISC1	biological_process	cellular_component	protein binding	8.70E-06	1.85893
RGS9	intracellular signaling cascade	heterotrimeric G-protein complex	signal transducer activity	0.000847298	1.88556
SHANK1	cytoskeletal anchoring at plasma membrane	intracellular	protein binding	1.27E-05	1.89216
CKB	creatine metabolic process	cytoplasm	nucleotide binding	0.000306765	1.89245
ZMYM1		nucleus	zinc ion binding	0.00111921	1.89854
AP1B1	intracellular protein transport	Golgi apparatus	transporter activity	0.000324249	1.90275

EPN2	endocytosis	cytoplasm	lipid binding	3.93E-06	1.90633
BAIAP3	neurotransmitter secretion		protein C-terminus binding	0.00065751 1	1.91359
HOMER3	protein targeting	cytoplasm	protein binding	0.00011675 6	1.92102
RRAGD		nucleus	nucleotide binding	0.00031415 5	1.92733
P2RY2	cellular ion homeostasis	plasma membrane	receptor activity	0.00104812	1.93398
AIF1L		focal adhesion	calcium ion binding	0.00078229 5	1.93842
C8orf56				0.00116542	1.93973
ZNF440	transcription	intracellular	DNA binding	0.00078767	1.95459
DLL3	skeletal system development	membrane	Notch binding	0.00036366 7	1.97945
TRAF3IP2	intracellular signaling cascade	cellular_compo nent		9.85E-05	1.98066
EVC2		membrane		0.00024081 6	1.98619
EGFL7	angiogenesis	extracellular region	calcium ion binding	0.00010015 6	2.02239
BCAT1	G1	cytoplasm	branched-chain-amino-acid transaminase activity	6.87E-05	2.04554
FLJ46020				0.00057356 6	2.05882
PDGFC	activation of transmembrane receptor protein tyrosine	extracellular region	platelet-derived growth factor receptor binding	7.67E-05	2.07071
KIFAP3	protein complex assembly	endoplasmic reticulum	protein binding	0.00103572	2.07599
GLT25D1	lipopolysaccharide biosynthetic process	endoplasmic reticulum	transferase activity, transferring glycosyl groups	0.00029838 9	2.0786
NTNG2	multicellular organismal development	proteinaceous extracellular matrix	molecular_functio n	0.00046959 3	2.08723
IFRG15				0.0007004	2.09089
ITGB4	cell communication	integrin complex	receptor activity	1.64E-06	2.11938
SPRY1	ureteric bud development	cytoplasm	protein binding	0.00053793 1	2.12373
RCBTB1	transcription	nucleus	protein binding	1.98E-05	2.13903
KIAA1841				0.00115897	2.14592
MPL	cell surface receptor linked signal transduction	plasma membrane	cytokine receptor activity	0.00112164	2.15246
CPNE7	lipid metabolic process		transporter activity	0.00019641	2.17656

DNM1	receptor-mediated endocytosis	cytoplasm	nucleotide binding	0.00062269	2.17671
STX2	cytokinesis	membrane fraction	SNARE binding	0.00028982	2.19314
LRP3	receptor-mediated endocytosis	coated pit	receptor activity	0.000342964	2.20438
SLC22A16	ion transport	plasma membrane	carnitine transporter activity	3.94E-05	2.20807
PRELID2				0.000436084	2.22485
JUP	cell-cell adhesion	membrane fraction	cytoskeletal protein binding	0.000210495	2.22614
CIITA	regulation of transcription, DNA-dependent	nucleus	nucleotide binding	0.000167645	2.22787
SCML2	regulation of transcription, DNA-dependent	nucleus	transcription factor activity	0.0011144	2.25451
RELL1		plasma membrane		2.44E-07	2.25629
PXK	protein amino acid phosphorylation	cytoplasm	nucleotide binding	0.000204814	2.29579
LASS4	regulation of transcription, DNA-dependent	nucleus	transcription factor activity	5.65E-05	2.31841
GFI1	regulation of transcription of G1	intracellular	DNA binding	0.000756961	2.32281
Sep-11	cell cycle	stress fiber	nucleotide binding	2.48E-06	2.33785
KCNH8	two-component signal transduction system (phosphorela	membrane	two-component sensor activity	0.000229385	2.34888
HSPG2	cell adhesion	extracellular region	protein binding	3.40E-05	2.3495
C6orf150				0.000204209	2.36238
SLCO4A1	ion transport	membrane	transporter activity	6.59E-05	2.36246
CCL28	chemotaxis	extracellular region	chemokine activity	0.000835078	2.37549
SLC25A1	transport	mitochondrion	transporter activity	5.56E-06	2.37864
ADRA2C	activation of MAPK activity	endosome	receptor activity	0.000144705	2.40161
TRERF1	transcription	intracellular	transcription factor activity	0.00100373	2.40251
LAMB2	cell adhesion	extracellular region	structural molecule activity	0.000699863	2.41525
ADCY9	cAMP biosynthetic process	integral to plasma membrane	magnesium ion binding	0.0012008	2.42547
ZNF555	transcription	intracellular	DNA binding	0.000857385	2.42594
TPSD1	proteolysis	extracellular region	serine-type endopeptidase	0.000288331	2.44257

			activity		
C11orf9	regulation of transcription	membrane	transcription factor activity	0.000107408	2.46749
NF-E4	transcription	nucleus		0.000328869	2.46976
TBXA2R	signal transduction	plasma membrane	receptor activity	2.22E-05	2.51593
ST6GALNAC2	protein amino acid glycosylation	Golgi membrane	sialyltransferase activity	7.42E-06	2.52625
HS3ST3B1	heparan sulfate proteoglycan biosynthetic process, en	Golgi membrane	sulfotransferase activity	0.00105888	2.54211
ALS2	endosome organization	ruffle	guanyl-nucleotide exchange factor activity	0.000879132	2.56021
RPSAP58	translational elongation	intracellular	structural constituent of ribosome	5.41E-05	2.58299
TLR2	response to molecule of fungal origin	plasma membrane	lipopolysaccharide receptor activity	0.000747629	2.60807
TBL1X	transcription	nucleus		5.34E-05	2.61725
UNQ6490		extracellular region		0.000127724	2.62828
SMAGP		plasma membrane		2.02E-07	2.65315
DNAH8	ciliary or flagellar motility	cytoplasm	nucleotide binding	0.000392471	2.66548
PLCG1	intracellular signaling cascade	cellular_component	phosphoinositide phospholipase C activity	9.50E-05	2.73251
GRK5	protein amino acid phosphorylation	soluble fraction	nucleotide binding	6.30E-05	2.74553
FLNB	cytoskeletal anchoring at plasma membrane	nucleus	actin binding	0.000207042	2.74964
HOMER2	metabotropic glutamate receptor signaling pathway	cytoplasm	actin binding	5.86E-08	2.77109
CDCA7	transcription	nucleus		0.000144064	2.84763
CORO2A	intracellular signaling cascade		actin binding	4.93E-05	2.88715
UBASH3B		nucleus		6.87E-05	2.89231
TMEM136		membrane		0.00070043	2.92113
FGFR1	MAPKKK cascade	membrane fraction	nucleotide binding	1.46E-08	2.96314
CAV2	vesicle fusion	Golgi membrane	protein binding	0.00101907	2.98291
FBXL13	modification-dependent protein catabolic process			0.000822359	3.0095

SERPINB10		nucleus	serine-type endopeptidase inhibitor activity	0.00108965	3.01633
SCCPDH	oxidation reduction		binding	7.46E-05	3.01881
AADAT	biosynthetic process	mitochondrion	kynurenine-oxoglutarate transaminase activity	2.10E-05	3.05775
UHRF1	DNA repair	nucleus	transcription factor activity	0.000391989	3.08633
SLC45A3	transport	membrane		0.00118312	3.17098
TNFRSF21	apoptosis	cytoplasm	receptor activity	2.68E-07	3.21
MFAP4	cell adhesion	microfibril	molecular_function	0.000422814	3.2451
KIT	myeloid leukocyte differentiation	extracellular space	nucleotide binding	0.000507242	3.28286
STK32B	protein amino acid phosphorylation		nucleotide binding	3.14E-06	3.42717
GNAI1	signal transduction	intracellular	nucleotide binding	3.71E-06	3.50149
ITM2C		Golgi apparatus		0.000571724	3.5228
TM6SF1	biological_process	membrane	molecular_function	1.68E-05	3.52511
SERPINB9	anti-apoptosis	cytoplasm	serine-type endopeptidase inhibitor activity	2.69E-06	3.53143
NT5DC3			magnesium ion binding	3.70E-05	3.57121
GATM	creatine biosynthetic process	cytoplasm	glycine amidinotransferase activity	4.18E-05	3.58058
NCALD	vesicle-mediated transport	intracellular	actin binding	5.90E-05	3.64514
MTSS1	cell motion	ruffle	actin monomer binding	6.99E-05	3.70334
CACNA2D2	ion transport	membrane	voltage-gated ion channel activity	2.54E-05	3.70803
PRTN3	proteolysis	cellular_component	serine-type endopeptidase activity	0.000201436	3.72313
PTGR1	leukotriene metabolic process	cytoplasm	alcohol dehydrogenase activity	2.40E-05	3.74584
HIST1H2BM	nucleosome assembly	nucleosome	DNA binding	0.000262673	4.04628
PALM	cell motion	plasma membrane		2.11E-10	4.13068
PSD3	regulation of ARF protein signal transduction	intracellular	ARF guanyl-nucleotide exchange factor activity	8.30E-05	4.19546
ITGA9	cell adhesion	integrin	receptor activity	0.00019521	4.19728

		complex		2	
HAL	histidine catabolic process	soluble fraction	histidine ammonia-lyase activity	0.000412807	4.2152
EPCAM		plasma membrane		0.000141482	4.40028
VEGFA	angiogenesis	extracellular region	fibronectin binding	5.76E-07	4.51368
FBLN5	cell adhesion	extracellular region	integrin binding	0.00123704	4.63839
PSD3	regulation of ARF protein signal transduction	intracellular	ARF guanyl-nucleotide exchange factor activity	4.92E-06	4.69641
MAN1A1	metabolic process	Golgi membrane	mannosyl-oligosaccharide 1,2-alpha-mannosidase activi	4.53E-07	4.82998
HPGD	lipid metabolic process	nucleus	prostaglandin E receptor activity	0.000607085	4.97399
GPM6B	multicellular organismal development	nucleus	molecular_function	3.99E-05	5.69264
SLC18A2	response to amphetamine	membrane fraction	monoamine transmembrane transporter activity	0.000183659	5.71725
CEACAM4		membrane		0.00016687	6.22054
C21orf109				0.000189887	6.35642
CAV1	inactivation of MAPK activity	Golgi membrane	structural molecule activity	2.84E-06	7.22331
SIPA1L2	regulation of small GTPase mediated signal transducti	intracellular	GTPase activator activity	2.24E-07	7.2642
BAALC		cytoplasm		1.17E-06	7.54201
PGDS	prostaglandin biosynthetic process	cytoplasm	prostaglandin-D synthase activity	0.00112342	7.61551
SLCO5A1	transport	membrane	transporter activity	7.45E-08	7.67319
ELANE	proteolysis	extracellular region	serine-type endopeptidase activity	3.64E-05	9.34343
CLEC5A	immune response	integral to plasma membrane	binding	6.95E-07	9.50729
IL5RA	signal transduction	extracellular space	receptor activity	1.56E-06	13.5298
TRH	signal transduction	extracellular region	neuropeptide hormone activity	3.44E-06	13.7637
ROBO1	chemotaxis	integral to plasma membrane	receptor activity	1.37E-08	15.0037

CD200		plasma membrane		3.72E-07	15.3419
POU4F1	suckling behavior	nucleus	transcription factor activity	6.23E-07	17.1116
RUNX1T1	generation of precursor metabolites and energy	nucleus	transcription factor activity	8.66E-15	17.166
MPO	anti-apoptosis	extracellular space	chromatin binding	1.75E-05	18.4828
PRAME			protein binding	7.77E-05	18.5729
TSPAN7	protein amino acid N-linked glycosylation	plasma membrane		7.91E-06	20.0498

Appendix B

Alternative Transcripts of AML1-ETO

Cloning of pMSCV-AML1-ETO6a-GFP

A pMSCV-AML1-ETO-GFP (pMiG-A-E) vector was obtained (courtesy of D.Gascoyne). The vector was digested with enzymes MfeI and BamH1 (NEB), which only cut the AML1-ETO sequence once each, on either side of exon 6 (Figure 65). This gives rise to a fragment of approximately 800 bp, which is discarded (Figure 66). The linearised vector, now containing sticky ends, was obtained while the approximate 700bp insert was discarded.

Figure 64 Part of Sequence of ETO highlighting exon 6 and the digestion sites BamH1 and Mfe1
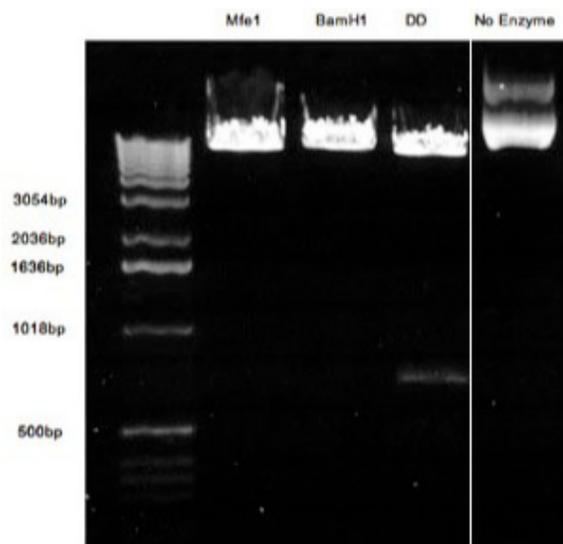
Figure 65 Agarose Gel of a Restriction Digest of pMiG -A-E with enzymes BamH1, Mfe1 and a double digest (DD). Buffer 4 (5ul) (NEB), DNA (3ug) enzyme 1.5ul each enzyme and water to make upto 50ul was mixed and incubated in a water bath at 37C for 3 hours. There is linearisation of vector with the Mfe1 and BamH1 enzymes. The double digest results in a smaller linearised vector, which is retained, and the appearance of an approximately 800bp insert, which is discarded.

Using the same enzymes, the previously cloned TA TOPO vector containing the ETO exon 6a transcript was digested. On this occasion the resulting insert was kept as this contains the 6a exon. However, the TA TOPO vector has an extra BamH1 digest site leading to 2 inserts with the double digest. The larger insert contains the new exon and this was cut out from the gel and the smaller insert and vector discarded (Figures 67 & 68).
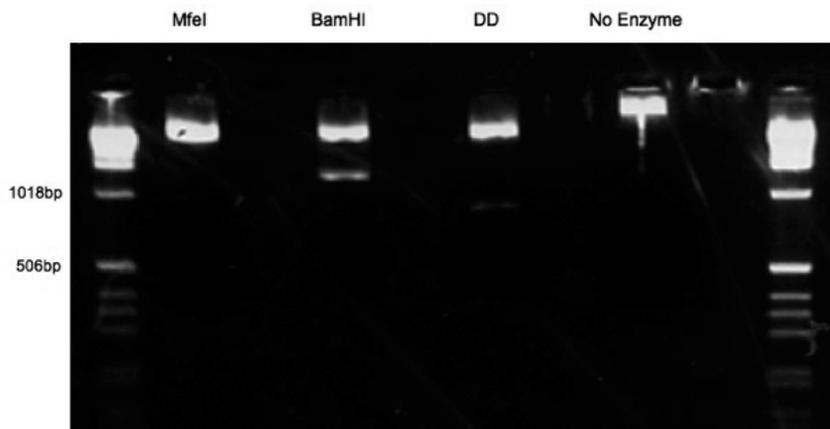
Figure 66 Agarose Gel of Restriction Digestion of TaTOPO vector containing exon 6a with BamH1 and Mfe1 and a double digest (DD). Digestion was performed in same conditions as above. There is an extra BamHI digest site in this vector leading to an extra fragment. Consequently the DD has two inserts in addition to the vector, which are seen more clearly on the figure below.
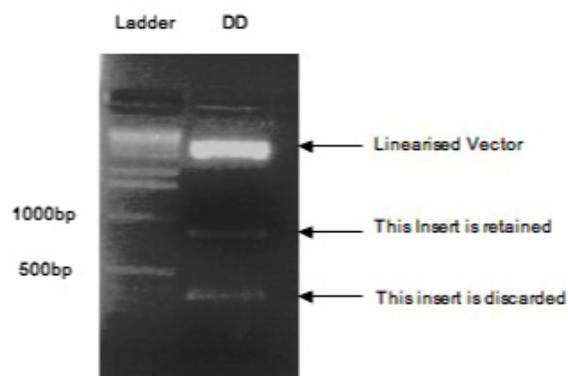
Figure 67 Agarose Gel of a Restriction Digestion of TaTOPO vector containing exon 6a with a double digest, highlighting the vector and both inserts. The larger insert at approximately 850bp was excised from the gel.

The pMiG-A-E vector minus the 700bp fragment was ligated to the insert from the second digest containing the exon 6a to obtain a pMiG-A-E6a plasmid. Ligation was performed with vector to insert ratios of 1:1, 1:2 and vector alone. The products were cloned, 12 colonies selected and minipreps made. The plasmids were checked by digesting with BamH1 and Mfe1 before use for functional assays.

Cloning of pMSCV AML1-ETO-ires hCD2

The pMiG A-E was obtained and digested with restriction enzymes ECoR1 and Sal1 (NEB). This results in a vector and 2 inserts, one containing the ires-GFP and the other a small part of the AML1-ETO coding region. (Figure 69 & 70) The vector is retained.
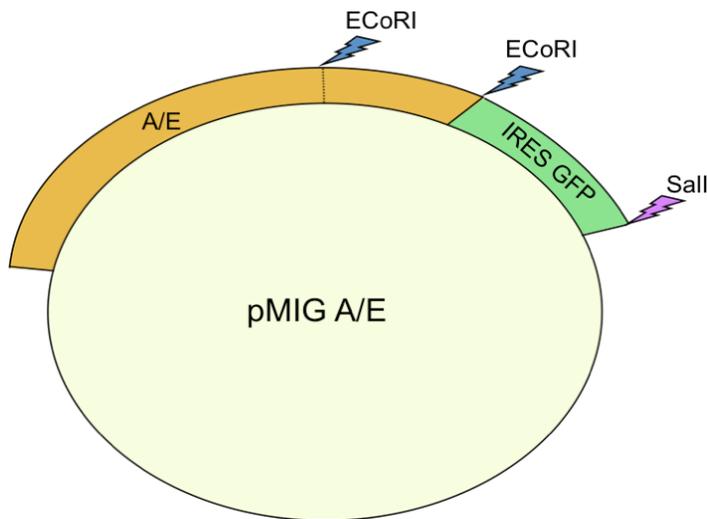
Figure 68 Schematic representation of the pMiGA/E vector with coding region and ires region and highlighting the digestion points.
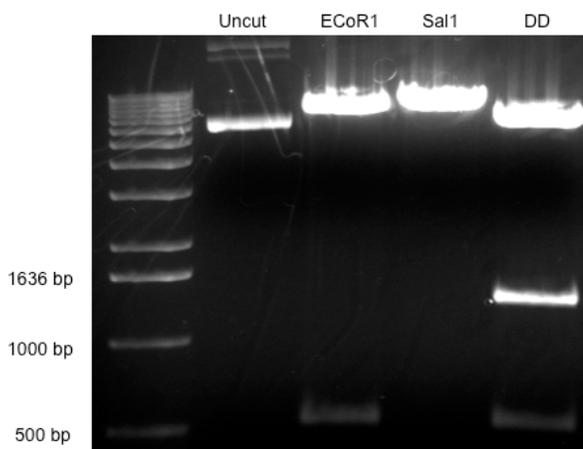
Figure 69 Agarose Gel of Restriction Digestion of pMiGA/E vector with ECoR1 and Sal1 and a double digest (DD). Digestion was performed in Buffer D (4ul) (NEB), DNA (1ul), enzymes (1ul each) and water to make up to 40ul were mixed and digested at 37c for 3 hours. The double digest has two inserts in addition to the vector.

Next a pMSCV-ireshCD2 vector was obtained. The vector was digested with ECoR1 followed by a partial digest with Sal1 to obtain the vector and the ireshCD2 insert (Figure 71& 72). On this occasion the vector is discarded but the insert containing the ires is retained.
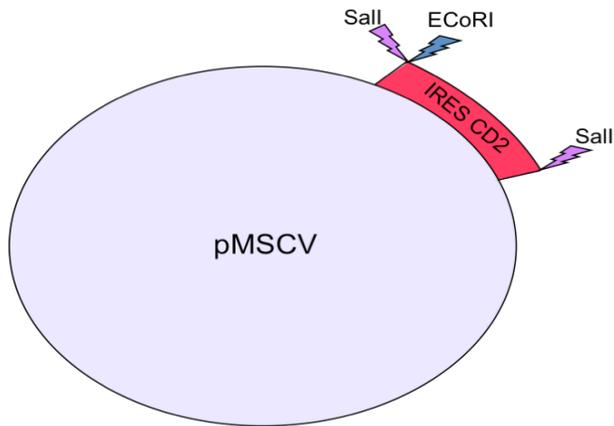
Figure 70 Schematic representation of the pMSCV-ireshCD2 vector highlighting the digestion points.
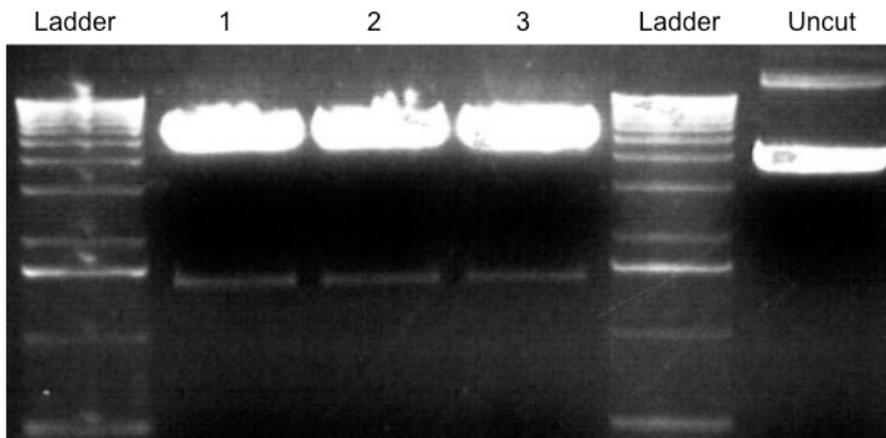
Figure 71 Buffer H (NEB) (4ul), DNA (2ug), ECoR1 (1ul) and water to make to 40ul were mixed and incubated at 37C for a total of 2 hours. After 60, 90 and 105 minutes 0.5ul of Sal1 was added to produce a partial digests 1,2 and 3 respectively.

The vector from the first digest and the insert from the second digest (sample 3) were ligated. A pMihC-A-E vector that is incomplete, lacking a small piece of approximately 500bp of AML1-ETO coding region, is obtained. Plasmids were checked with further digests for verification and the correct plasmids used (Figure 73).
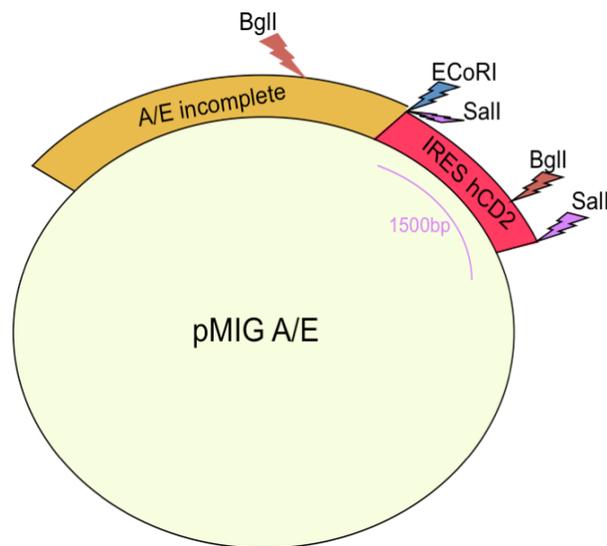
Figure 72 Schematic of newly formed vector and the digestion break points. Digestion with Sal1 should give an approximately 1500bp band and vector, with Bgl1 a 3000bp band and vector and with ECoR1 a linearised vector.

The next aim was to re-insert the missing approximately 500bp insert into this new vector to complete the AML1-ETO coding region. The new vector was digested with ECoR1 to linearise the vector and treated with alkaline phosphotase to stop re-annealing (Figure 74). Then the original pMiG-A-E was digested with ECoR1 and the 500bp fragment was obtained as previously described. The vector and the 500bp fragment were ligated.
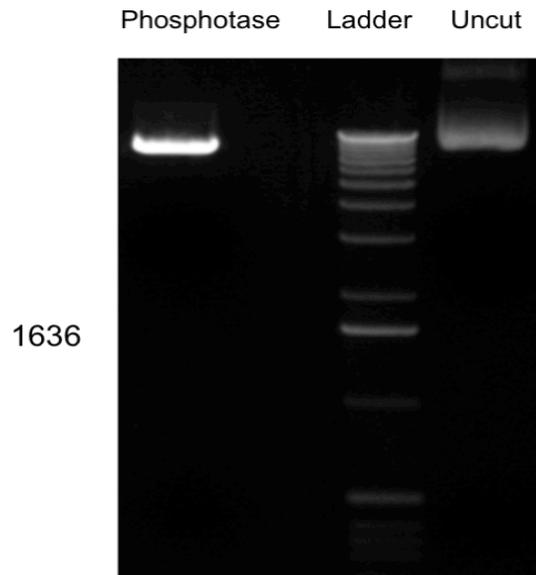
Figure 73 Buffer ECor1 (4ul), DNA (1ug), ECoR1 (5ul) and water to make upto 40ul were mixed and incubated for 2 hours at 37C, heat inactivated at 65C for 5 minutes and incubated at 37C for a further 15 minutes after adding 1ul (1unit/ul) of alkaline phosphotase.

To check ligation and correct orientation of the 500bp insert a series of digests were performed and the appropriate colonies were selected (Figure 75).
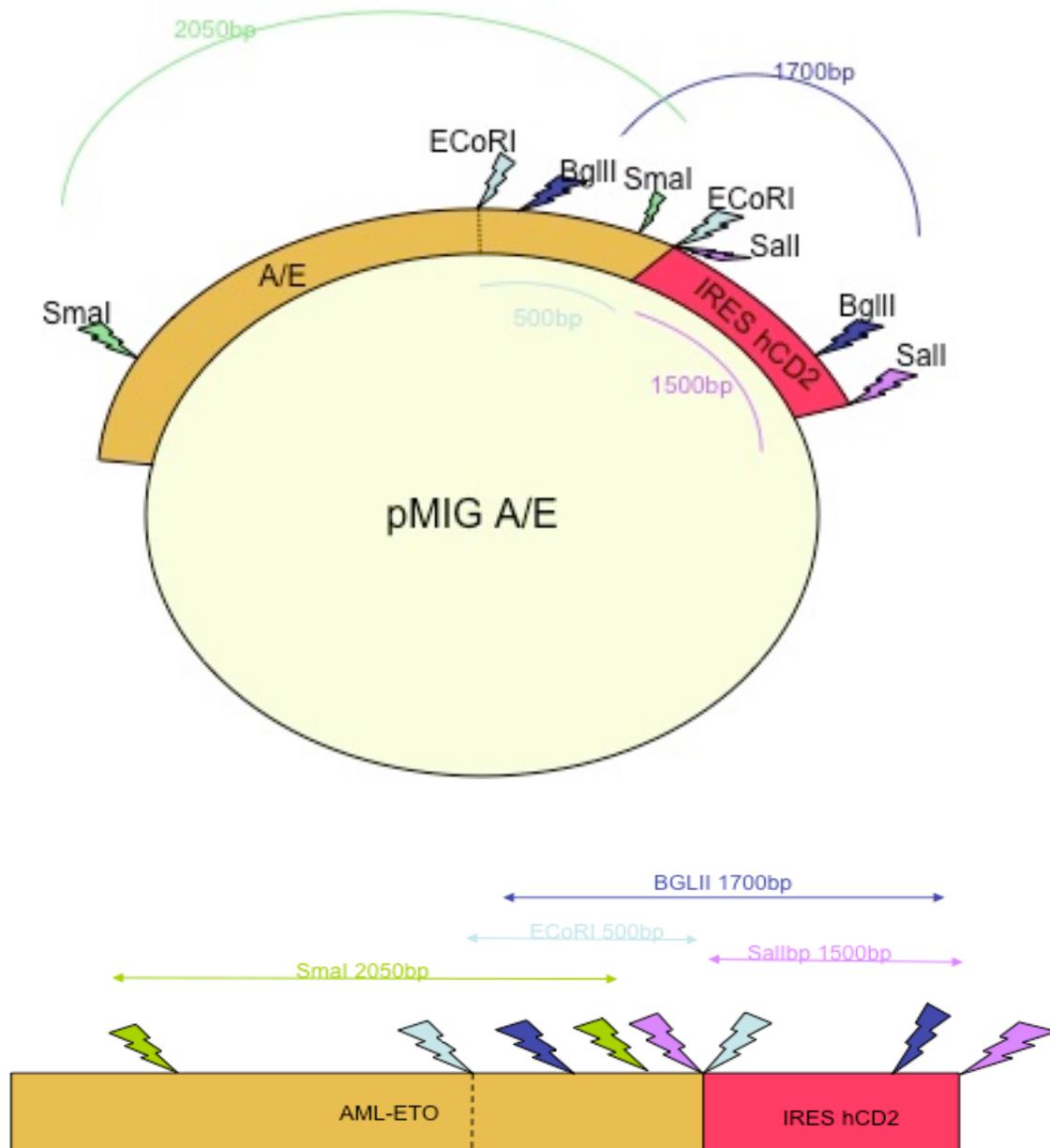
Figure 74 Schematic representation of the size of digests expected. BGLII should give approximately 1700bp and not 1200bp digest; with SmaI 2050bp and not 1550bp if the insert is in the correct orientation.

Appendix C

Gene Lists derived from ChIp-Seq Experiments

Table 31 Gene List of 338 genes which bind AML1-ETO showing the gene name; chromosome location; the total number of reads located in the binding site for the input sample and the test sample; and the p value derived from the comparison of input and test sample.

GENE	CHROMOSOME	INPUT	CHIP	P-VALUE
MX2	21	121	776	0
CDH23	10	74	651	0
TRIM8	10	125	714	0
CCND3	6	141	1123	0
PTCH1	9	77	576	0
STAB1	3	99	974	0
PRKCD	3	129	709	0
HCCA2	11	71	636	0
RPS6KA1	1	126	1011	0
INPP5A	10	74	617	0
GFOD1	6	110	693	0
ZBTB16	11	56	459	0
NEK8	17	88	602	0
MAP2K2	19	29	420	0
BAHCC1	17	117	936	0
USP20	9	30	356	9.81E-45
SH3BP5	3	70	482	7.26E-43
KIAA0427	18	83	519	1.89E-42
LOC374443	12	29	338	3.63E-42
FAM129B	9	22	310	3.92E-42
PTTG1IP	21	80	507	5.35E-42
GALNT6	12	104	567	1.50E-40
BCL3	19	86	512	4.44E-40
ADCY4	14	72	470	4.70E-40
HIP1	7	69	459	8.93E-40
LMNA	1	108	564	1.48E-38
SULF2	20	71	452	8.99E-38
LAPTM5	1	108	558	1.01E-37
C6orf103	6	64	417	1.13E-35
HNRNPM	19	99	517	1.74E-35
CXXC5	5	177	711	2.07E-34
PPP4R1L	20	72	431	3.54E-34
CYP51A1	7	88	474	9.30E-34
GLIPR2	9	70	419	2.86E-33
KIAA0652	11	56	376	3.93E-33
VAV1	19	95	485	1.38E-32
JMJD6	17	84	452	3.09E-32
GRK7	3	165	661	5.52E-32
GNB5	15	110	520	7.40E-32
TADA3	3	98	485	1.80E-31
ACSL1	4	71	407	4.40E-31
RASSF2	20	73	411	7.87E-31
TPM4	19	63	374	1.19E-29
RAI1	17	77	414	1.21E-29
EEPD1	7	54	347	1.29E-29

RARA	17	173	659	1.72E-29
LAT2	7	43	306	1.60E-28
HSH2D	19	32	269	2.47E-28
MSI2	17	105	473	1.90E-27
SLC2A1	1	120	507	6.61E-27
SH3BP2	4	67	364	1.53E-26
TSPAN32	11	103	460	1.90E-26
ZFYVE28	4	98	446	2.59E-26
KIAA0182	16	75	384	3.13E-26
NHSL2	X	14	190	5.31E-26
ALDH16A1	19	50	310	7.87E-26
RASSF5	1	119	494	1.27E-25
C5orf56	5	121	497	2.21E-25
ABCG1	21	33	254	2.53E-25
GPR114	16	99	437	7.84E-25
TOX2	20	77	377	1.55E-24
CD276	15	75	371	1.87E-24
TBC1D14	4	120	486	2.33E-24
SPNS3	17	38	263	3.20E-24
SMG6	17	154	561	1.43E-23
MBP	18	171	599	2.15E-23
STARD9	15	20	198	2.35E-23
PKIB	6	80	376	2.44E-23
MAX	14	66	337	3.74E-23
FNTB	14	66	337	3.74E-23
IL21R	16	55	305	5.82E-23
CCDC135	16	48	283	1.09E-22
TK2	16	134	504	1.68E-22
LOC100130987	11	84	380	1.69E-22
IL17RA	22	101	423	1.86E-22
SORBS1	10	68	335	3.87E-22
ADORA2A	22	92	396	6.34E-22
C22orf45	22	92	396	6.34E-22
MGAT1	5	198	646	6.38E-22
COL23A1	5	52	287	1.33E-21
SIK3	11	152	538	1.41E-21
CNIH3	1	82	365	3.07E-21
SIPA1L3	19	82	362	7.39E-21
LST1	6	100	407	8.11E-21
CD82	11	109	429	8.32E-21
EP400NL	12	77	348	1.02E-20
SPN	16	79	353	1.08E-20
RALGDS	9	115	442	1.26E-20
LHX6	9	104	414	1.77E-20
PI16	6	73	334	2.98E-20
FAM82A2	15	81	353	4.81E-20
RNF166	16	92	379	7.80E-20
SERINC5	5	46	256	1.36E-19
FBRSL1	12	135	480	1.38E-19
HSPG2	1	70	319	2.54E-19
FUT7	9	93	376	3.56E-19

C9orf139	9	93	376	3.56E-19
MAD1L1	7	23	183	4.42E-19
IL1RN	2	36	221	1.05E-18
LMAN2	5	44	244	1.10E-18
ZMIZ1	10	46	248	1.88E-18
PTPRE	10	72	317	2.10E-18
UNC13D	17	90	362	2.19E-18
GAS7	17	47	250	2.43E-18
FOXN3	14	56	274	2.92E-18
CUGBP2	10	14	147	3.03E-18
CDC42BPB	14	73	317	4.41E-18
ADCY8	8	6	114	4.63E-18
XYLT1	16	41	231	4.85E-18
LCP2	5	75	320	8.01E-18
BPI	20	43	235	8.41E-18
MYO1D	17	12	137	8.67E-18
DCC	18	9	125	9.86E-18
MTMR4	17	202	609	1.47E-17
PLD3	19	64	288	2.49E-17
GRID1	10	19	159	2.61E-17
TBPL1	6	38	217	2.96E-17
RUNX1	21	108	395	3.33E-17
LOC100128822	7	40	222	3.74E-17
AGTRAP	1	80	327	3.78E-17
ERICH1	8	51	252	4.19E-17
CRTC1	19	128	440	4.73E-17
TEC	4	56	264	6.32E-17
XKR6	8	16	145	1.14E-16
TSEN54	17	119	415	1.36E-16
AHI1	6	38	212	1.57E-16
AKAP13	15	200	593	1.68E-16
KIAA0513	16	81	324	1.70E-16
SLC43A2	17	73	304	1.80E-16
PLCB1	20	15	139	3.07E-16
ZC3H3	8	82	323	4.35E-16
RIN2	20	68	288	4.81E-16
FOXK2	17	84	327	5.48E-16
PNPLA7	9	20	154	6.05E-16
EFHD2	1	87	333	7.62E-16
SAPS3	11	78	308	1.80E-15
ZNRF1	16	13	126	3.28E-15
CHST11	12	128	420	5.31E-15
PDE4DIP	1	24	159	9.55E-15
LOC153684	5	127	414	1.26E-14
CDA	1	29	173	1.30E-14
SMOX	20	31	178	1.70E-14
TRAP1	16	16	132	1.74E-14
PCSK6	15	118	392	1.93E-14
HIVEP3	1	6	95	1.97E-14
FAM155A	13	8	103	2.05E-14
EMILIN1	2	65	267	2.17E-14

SYT2	1	27	163	5.41E-14
DLG2	11	9	104	6.68E-14
FGGY	1	7	96	7.43E-14
PTPRC	1	54	235	7.54E-14
MTSS1	8	4	83	8.35E-14
UBE2V1	20	120	389	1.08E-13
TMEM189-UBE2V1	20	120	389	1.08E-13
TTC7A	2	112	371	1.11E-13
PFKM	12	10	104	3.02E-13
DFFB	1	89	314	3.33E-13
RERE	1	9	100	3.47E-13
ZSWIM6	5	37	185	3.99E-13
LMTK2	7	27	157	4.19E-13
CCDC109B	4	19	133	4.36E-13
ERGIC1	5	22	142	4.63E-13
EFHC2	X	4	79	5.00E-13
SYNJ2	6	49	215	6.24E-13
DEDD2	19	68	262	6.28E-13
TRPM2	21	111	361	6.68E-13
CLMN	14	50	217	7.31E-13
OSBPL5	11	113	365	7.40E-13
TSPAN16	19	65	254	7.68E-13
RAB3D	19	65	254	7.68E-13
CABLES1	18	11	105	8.44E-13
NTNG2	9	43	198	9.80E-13
CACNA1C	12	28	157	1.09E-12
GRAP2	22	165	475	1.19E-12
PRR5	22	23	142	1.31E-12
HTT	4	45	202	1.40E-12
TNS3	7	3	72	1.48E-12
ARAP1	11	135	410	1.48E-12
TFEB	6	100	332	1.84E-12
AGPAT3	21	21	135	1.90E-12
C18orf45	18	15	116	2.16E-12
SYN3	22	97	323	3.11E-12
ADARB2	10	10	98	3.38E-12
BDNF	11	35	172	4.35E-12
C9orf89	9	99	326	4.41E-12
PPAP2B	1	44	195	5.16E-12
FAM73B	9	74	268	5.36E-12
KIAA1949	6	76	272	6.35E-12
HHEX	10	93	310	7.93E-12
ZNF718	4	14	109	8.89E-12
PPCDC	15	15	112	9.61E-12
UHRF1	19	18	121	1.09E-11
HVCN1	12	40	182	1.15E-11
DNAJB12	10	92	306	1.20E-11
RAB11FIP4	17	7	83	1.77E-11
ALK	2	17	115	3.14E-11
SH3BP5L	1	8	85	3.56E-11
PIP5K1C	19	71	252	5.18E-11

LARS2	3	5	72	6.89E-11
RAPGEF1	9	53	208	7.39E-11
MAPK1	22	81	273	8.24E-11
MED26	19	17	112	9.19E-11
ZNF444	19	15	105	1.26E-10
OR9Q1	11	20	120	1.26E-10
ATP2B4	1	3	62	1.36E-10
NCOR2	12	57	215	1.48E-10
ACVR1B	12	79	266	1.48E-10
PDE3A	12	2	57	1.58E-10
EFNB2	13	73	251	2.06E-10
TNFRSF4	1	64	230	2.18E-10
STAT3	17	31	148	2.74E-10
NFAM1	22	67	236	2.77E-10
AKNA	9	7	76	3.17E-10
ARHGEF2	1	12	93	3.22E-10
INPP5B	1	46	185	3.74E-10
C20orf26	20	1	50	3.76E-10
FAM125B	9	119	349	3.93E-10
ARHGAP26	5	15	101	5.36E-10
SMAD3	15	60	217	5.52E-10
MAP1LC3B2	12	71	242	6.02E-10
ZFHX3	16	21	118	6.55E-10
PI4KA	22	20	115	6.91E-10
WFS1	4	39	165	7.89E-10
PGCP	8	8	77	8.93E-10
TNFRSF1A	12	9	80	1.06E-09
EVL	14	7	73	1.08E-09
BDNFOS	11	7	73	1.08E-09
FLJ43663	7	17	105	1.08E-09
LOC651250	17	15	99	1.10E-09
TRIP4	15	40	166	1.21E-09
UBASH3B	11	38	161	1.21E-09
LOC285045	2	6	69	1.24E-09
FLJ45079	17	101	305	1.25E-09
PCSK5	9	1	47	1.56E-09
RIN3	14	21	115	1.78E-09
C2orf18	2	69	232	2.23E-09
PID1	2	2	51	2.49E-09
TTYH3	7	87	271	2.76E-09
SYK	9	13	89	4.49E-09
TMEM104	17	52	190	4.56E-09
LUZP1	1	40	161	4.89E-09
PACSIN2	22	23	117	5.55E-09
SDC3	1	58	203	5.77E-09
BCL2	18	35	148	5.82E-09
GPR149	3	8	72	6.45E-09
CDC27	17	13	88	6.45E-09
TSNAX-DISC1	1	24	119	6.92E-09
ZNF804B	7	0	38	6.92E-09
HFM1	1	2	48	9.79E-09

DUSP10	1	168	429	1.72E-08
SCFD2	4	4	55	1.83E-08
WDR81	17	15	91	1.83E-08
TRPM1	15	14	88	1.87E-08
OGG1	3	124	339	2.15E-08
ANKRD33B	5	73	231	2.20E-08
ATP9B	18	3	50	2.80E-08
VIM	10	142	373	3.42E-08
C12orf5	12	52	182	3.43E-08
WASF1	6	15	89	3.65E-08
PARN	16	7	64	3.95E-08
LOC100216545	7	136	360	4.01E-08
SASH3	X	22	108	4.09E-08
FGD3	9	4	53	4.28E-08
KIAA1797	9	3	49	4.34E-08
CLCN7	16	13	82	5.47E-08
MYH9	22	13	82	5.47E-08
PI4KB	1	65	208	7.44E-08
SIPA1	11	118	320	7.47E-08
AR	X	0	33	8.20E-08
GSN	9	28	121	8.82E-08
SENP5	3	2	43	9.44E-08
SNAPC3	9	56	186	1.18E-07
RAP1GDS1	4	20	98	1.77E-07
UMODL1	21	137	353	1.93E-07
CARD9	9	76	227	2.16E-07
POLR1A	2	21	100	2.17E-07
ARHGEF6	X	13	78	2.20E-07
MIA	19	79	232	2.89E-07
ETV6	12	83	240	3.17E-07
KCNJ12	17	20	96	3.31E-07
ZMYND10	3	90	254	3.61E-07
RASSF1	3	90	254	3.61E-07
KLF9	9	7	58	4.09E-07
KLF13	15	88	249	4.25E-07
NRD1	1	78	227	5.22E-07
SLC43A3	11	108	289	5.25E-07
FAM53B	10	120	313	5.45E-07
MAN1A1	6	297	650	5.97E-07
NFATC1	18	32	124	6.03E-07
LIMS1	2	86	243	6.10E-07
ZCCHC14	16	20	94	6.15E-07
AXIN1	16	62	191	7.64E-07
CD9	12	7	56	8.81E-07
SH3GL1	19	5	49	1.01E-06
TTLL4	2	13	73	1.21E-06
MOBKL2A	19	58	180	1.27E-06
INPP5D	2	7	55	1.29E-06
CTNND1	11	74	214	1.32E-06
ARID5A	2	89	245	1.38E-06
STAT5B	17	5	48	1.51E-06

ATP2A3	17	19	88	1.74E-06
THRAP3	1	1	32	1.81E-06
DOK7	4	4	44	1.85E-06
PCBD2	5	30	115	1.88E-06
DNAJC15	13	10	63	1.97E-06
SPI1	11	10	63	1.97E-06
DYNC1LI2	16	9	60	1.99E-06
MADD	11	23	97	2.45E-06
CDKL4	2	60	181	2.62E-06
CCDC113	16	2	35	3.37E-06
CNN2	19	65	190	3.70E-06
CASK	X	1	30	4.59E-06
NBPF1	1	7	51	5.79E-06
DUSP3	17	62	181	6.41E-06
PWP2	21	47	147	9.54E-06
FSTL3	19	50	153	1.07E-05
TBC1D1	4	5	43	1.09E-05
DUS2L	16	0	23	1.15E-05
TMEM120B	12	1	28	1.16E-05
LSM4	19	8	52	1.21E-05
PRODH2	19	64	182	1.23E-05
DHCR24	1	4	39	1.42E-05
ZNF445	3	5	42	1.61E-05
DHRS12	13	9	54	1.66E-05
WIPF1	2	11	59	2.04E-05
DNMT3L	21	5	41	2.36E-05
RNF44	5	50	149	2.53E-05
LOC285456	4	3	34	2.64E-05
CEP63	3	1	25	4.60E-05
KIAA1539	9	94	235	4.83E-05
FAM117A	17	26	90	0.000110188
ZNF595	4	105	244	0.000318889
SOLH	16	5	33	0.000474549
TSNAX	1	4	29	0.000697853
C19orf45	19	12	50	0.000755817
PDPR	16	85	200	0.000769034
KIAA1012	18	0	12	0.00264936
PMF1	1	45	114	0.00311452

Table 32 Gene List of top 186 genes generated from analysis of samples 2 & 3 only in order of p-value.

GENE SYMBOL	CHROMOSOME	PEAKS INPUT	PEAKS CHIP	p-VALUE
RGSL1	1	44	701	0
DOCK6	19	23	346	0
MX2	21	121	776	0
RHBDF2	17	156	783	0
CDH23	10	74	651	0
CALR	19	76	690	0
TRIM8	10	125	714	0
CCND3	6	141	1123	0
MYEOV	11	195	1172	0
PTCH1	9	77	576	0
STAB1	3	99	974	0
PRKCD	3	129	709	0
HCCA2	11	71	636	0
RPS6KA1	1	126	1011	0
BLK	8	42	500	0
INPP5A	10	74	617	0
GFOD1	6	110	693	0
ZBTB16	11	56	459	0
NEK8	17	88	602	0
MAP2K2	19	29	420	0
BAHCC1	17	117	936	0
RPS9	19	84	983	0
USP20	9	30	356	9.81E-45
LMO2	11	141	697	2.10E-44
WNT10A	2	110	605	1.54E-43
SH3BP5	3	70	482	7.26E-43
KIAA0427	18	83	519	1.89E-42
LOC374443	12	29	338	3.63E-42
FAM129B	9	22	310	3.92E-42
PTTG1IP	21	80	507	5.35E-42
GPR97	16	171	755	1.17E-41
GALNT6	12	104	567	1.50E-40
BCL3	19	86	512	4.44E-40
ADCY4	14	72	470	4.70E-40
TESC	12	102	557	6.24E-40
HIP1	7	69	459	8.93E-40
LMNA	1	108	564	1.48E-38
SULF2	20	71	452	8.99E-38
LAPTM5	1	108	558	1.01E-37
MPG	16	35	333	2.12E-37
IQCD	12	75	461	2.52E-37
ADRBK1	11	133	611	7.73E-36
C6orf103	6	64	417	1.13E-35
HNRNPM	19	99	517	1.74E-35
SDK2	17	31	304	7.24E-35
CXXC5	5	177	711	2.07E-34
PPP4R1L	20	72	431	3.54E-34

ZFPM2	8	23	270	4.16E-34
CYP51A1	7	88	474	9.30E-34
GLIPR2	9	70	419	2.86E-33
DIS3L2	2	92	481	3.55E-33
KIAA0652	11	56	376	3.93E-33
VAV1	19	95	485	1.38E-32
MCF2L	13	88	465	1.71E-32
MYO1C	17	57	374	2.37E-32
JMJD6	17	84	452	3.09E-32
GRK7	3	165	661	5.52E-32
GNB5	15	110	520	7.40E-32
TADA3	3	98	485	1.80E-31
ACSL1	4	71	407	4.40E-31
RASSF2	20	73	411	7.87E-31
SLC35F3	1	23	251	1.02E-30
SBF1	22	67	391	1.95E-30
ATP10A	15	66	385	5.54E-30
MNAT1	14	57	358	6.50E-30
DMWD	19	105	492	6.51E-30
TPM4	19	63	374	1.19E-29
RAI1	17	77	414	1.21E-29
EEPD1	7	54	347	1.29E-29
RARA	17	173	659	1.72E-29
GRK5	10	140	572	9.86E-29
LAT2	7	43	306	1.60E-28
HSH2D	19	32	269	2.47E-28
ANXA11	10	61	357	5.24E-28
ANXA6	5	11	187	8.46E-28
GFI1B	9	57	343	1.15E-27
BIN1	2	81	410	1.51E-27
MSI2	17	105	473	1.90E-27
SLC2A1	1	120	507	6.61E-27
LFNG	7	165	617	6.65E-27
DIP2C	10	63	354	9.86E-27
FXYD2	11	19	213	1.22E-26
CBFA2T3	16	45	300	1.31E-26
MIR548H2	12	17	205	1.42E-26
RFX4	12	17	205	1.42E-26
ADC	1	100	453	1.52E-26
SH3BP2	4	67	364	1.53E-26
IQGAP2	5	221	743	1.77E-26
C11orf21	11	103	460	1.90E-26
TSPAN32	11	103	460	1.90E-26
ZFYVE28	4	98	446	2.59E-26
KIAA0182	16	75	384	3.13E-26
GHDC	17	105	462	4.78E-26
NHSL2	X	14	190	5.31E-26
LTC4S	5	66	357	6.01E-26
CCDC88B	11	80	395	7.09E-26
ALDH16A1	19	50	310	7.87E-26
PADI2	1	49	307	7.91E-26

RASSF5	1	119	494	1.27E-25
LOC257358	5	149	567	1.42E-25
C5orf56	5	121	497	2.21E-25
ABCG1	21	33	254	2.53E-25
WDR1	4	249	793	3.22E-25
DLGAP2	8	16	193	4.16E-25
MUC16	19	6	150	4.20E-25
DNMT3A	2	45	290	4.49E-25
KCNMA1	10	47	295	6.41E-25
GPR114	16	99	437	7.84E-25
SYTL4	X	10	167	7.94E-25
KRT80	12	47	294	9.08E-25
KIRREL	1	7	153	9.63E-25
GAB2	11	14	183	1.02E-24
C14orf106	14	32	247	1.02E-24
TOX2	20	77	377	1.55E-24
CD276	15	75	371	1.87E-24
SGMS1	10	22	211	2.28E-24
TBC1D14	4	120	486	2.33E-24
TNFRSF14	1	77	375	2.87E-24
DAB1	1	8	155	3.14E-24
STAC	3	8	155	3.14E-24
SPNS3	17	38	263	3.20E-24
GNAO1	16	12	171	5.40E-24
PPM1L	3	27	226	5.52E-24
CDH4	20	7	149	5.82E-24
SFXN3	10	52	302	8.99E-24
SMG6	17	154	561	1.43E-23
ILDR2	1	24	213	1.62E-23
RAD52	12	51	297	1.83E-23
MBP	18	171	599	2.15E-23
ALS2CR12	2	72	355	2.19E-23
STARD9	15	20	198	2.35E-23
DYM	18	80	376	2.44E-23
PKIB	6	80	376	2.44E-23
C7orf50	7	40	263	2.90E-23
MAX	14	66	337	3.74E-23
FNTB	14	66	337	3.74E-23
PEBP4	8	56	309	3.96E-23
IL21R	16	55	305	5.82E-23
UNC5B	10	14	173	6.81E-23
CCDC135	16	48	283	1.09E-22
CSRNP1	3	152	547	1.49E-22
NAV2	11	12	163	1.63E-22
TK2	16	134	504	1.68E-22
LOC100130987	11	84	380	1.69E-22
IL17RA	22	101	423	1.86E-22
MTUS2	13	2	115	2.14E-22
DMGDH	5	14	170	2.38E-22
IL6R	1	40	257	2.41E-22
MOBKL2B	9	20	191	3.81E-22

SORBS1	10	68	335	3.87E-22
DYSF	2	7	139	5.12E-22
WNK2	9	20	190	5.66E-22
OLFM4	13	6	134	5.96E-22
RELT	11	145	525	6.11E-22
ADORA2A	22	92	396	6.34E-22
C22orf45	22	92	396	6.34E-22
MGAT1	5	198	646	6.38E-22
CSMD2	1	8	143	6.39E-22
INTS1	7	59	308	8.39E-22
C9orf3	9	54	294	8.55E-22
STK32C	10	30	222	1.01E-21
CCDC86	11	16	174	1.04E-21
TGFB2	1	25	205	1.30E-21
COL23A1	5	52	287	1.33E-21
SIK3	11	152	538	1.41E-21
AIM1	6	114	447	1.73E-21
ANK1	8	9	145	1.78E-21
LOC284805	20	42	257	1.93E-21
CUX1	7	83	369	2.02E-21
CLIC5	6	44	262	2.61E-21
CNIH3	1	82	365	3.07E-21
RNF152	18	8	139	3.71E-21
DCLK1	13	20	185	4.06E-21
MIR548F5	13	20	185	4.06E-21
LOC146336	16	52	283	4.97E-21
LNPEP	5	174	583	5.17E-21
SHISA6	17	8	138	5.76E-21
TYROBP	19	184	605	5.85E-21
HCST	19	184	605	5.85E-21
SIPA1L3	19	82	362	7.39E-21
ITGAE	17	55	290	7.78E-21
LST1	6	100	407	8.11E-21
CD82	11	109	429	8.32E-21
PRMT7	16	54	287	8.35E-21
C20orf94	20	22	190	8.50E-21

Table 33 Significant processes as detected by GeneGo generated from the definitive gene list of the above186 genes

	NAME	p-VALUE	NETWORK OBJECTS
1	Reproduction_FSH-beta signaling pathway	1.17E-05	10/159
2	Apoptosis_Anti-Apoptosis mediated by external signals via MAPK and JAK/STAT	6.78E-04	8/170
3	Signal transduction_CREM pathway	8.48E-04	6/98
4	Signal transduction_WNT signaling	8.84E-04	8/177
5	Signal transduction_NOTCH signaling	1.41E-03	9/236
6	Cell cycle_G1-S Growth factor regulation	1.65E-03	8/195
7	Cell cycle_G2-M	2.26E-03	8/205
8	Development_Hedgehog signaling	2.27E-03	9/253
9	Cell adhesion_Platelet aggregation	3.28E-03	7/171

Appendix D

Mixtures, Solutions, Buffers & Primers

I. Common Methods

cDNA Synthesis

Component	Volume x1 Rxn (for 100ng RNA)	Volume x1 Rxn (for 2ug RNA)
DW	10.5μl	Up to 12μl
5x Synthesis Buffer	6μl	6uμl
dNTP, 2.5 μM	10μl	10uμl
Random hexamers, 50 μm	1.5μl	1.5uμl
RT	1μl	2μl
RNA	1μg/μl	100ng

RT-PCR Master Mix

Component	Volume x1 Rxn
10x PCR Buffer	10.5μl
dNTP, 2.5uM	4μl
Forward Primer, 20uM	2μl
Reverse Primer, 20 uM	2uμl
Taq Polymerase	0.3μl
DW	35.7μl
cDNA	1μl

RQ-PCR - SYBR Protocol

Components	Volume x1 Rxn
Universal Mastermix	12.5μl
Gene Assay Mix	1.25μl
DW	10.25μl
cDNA	1μlμ
Total	25ul

RQ-PCR- Protocol for Specific Probes

Components	Volume x1 Rxn
Universal Mastermix	12.5
Forward Primer	2.5μl
Reverse Primer	2.5μl
Probe (FAM)	2.5μl
DW	4μl
cDNA	1μl
Total	25μl

II. Exon Array Protocols

Step C RiboMinus Probe Hybridization

Component	Volume x1Rxn(1µg/µl sample concentration)	Volume x 1Rxn (0.31µg/ul-1µg/µl)
Total RNA/Poly-A RNA Controls Mix	3.0ul	Up to 5.2µl
Ribominus Probe, 100pmol/ul	0.8µl	0.8µl
Hybridization buffer with Betaine	20µl	30µl
Total	23.8µl	36.0µl

First-cycle, First-Strand Master mix

Component	Volume in 1 Rxn
5x 1st strand Buffer	2µl
DTT	1µl
dNTP Mix, 10mM	0.5µl
RNase Inhibitor	0.5µl
SuperScript II	1µl
Total Volume	5µl

First-Cycle, Second-Strand cDNA Synthesis

Component	Volume in 1Rxn
RNase-free Water	4.8µl
MgCl2, 17.5mM	4.0µl
dNTP Mix, mM	0.4µl
DNA Polymerase I	0.6µl
RNase H	0.2µl
Total Volume	10µl

First-Cycle, IVT Master Mix

Component	Volume in 1 Rxn
10x IVT Buffer	5µl
IVT NTP Mix	20µl
IVT Enzyme Mix	5µl
Total Volume	30µl

Second-Cycle, cRNA/Random Primers Mix

Component	Volume in 1 Rxn
cRNA 10ug	Variable
Random Primers (3ug/ul)	1.5µl
RNase-free water	Up to 8µl

Total Volume	8μl

Second-Cycle, First Strand Master Mix

Component	Volume in 1 Rxn
5x 1st strand Buffer	4μl
DTT, 0.1mM	2μl
dNTP+dUTP, 10mM	1.25μl
SuperScript II	4.75μl
Total Volume	12μl

Fragmentation Master Mix

Component	Volume in 1 Rxn
Single-Stranded DNA	5.5gμ
10x cDNA Fragmentation Buffer	4.8lμ
UDG 10U/uL	1lμ
APE 1, 1000 U/uL	1lμ
RNase-free Water	Up to 48lμ
Total Volume	48lμ

Labelling Reaction

Component	Volume in 1 Rxn
Fragmented Single-Stranded DNA	45lμ
5x TdT Buffer	12μl
TdT	2μl
DNA Labelling Reagent, 5mM	1μl
Total Volume	60μl

Hybridisation master mix

Fragmented and Labelled DNA	60μl
Control Oligonucleotide B2	3.7μl
20x Eukaryotic Hybridization Controls	11μl
Herring Sperm DNA (10mg.ml)	2.2μl
Acetylated BSA (50mg/ml)	2.2μl
2x Hybridisation Buffer	110μl
DMSO	15.4μl
RNase free Water	Upto 220μl
Total	220μl

III. Alternative Transcripts

Restriction Digestion

Components	Volumes	Volumes
Distilled Water	5.5μl	Upto 40μl
10x Buffer	1μl	4μl
DNA	3μl	2μg
Enzyme	0.5μl	1μl
Total	10μl	40μl

Ligation Protocol

Components	Volume
Distilled Water	13μl
Buffer x10	2μl
DNA (Insert)	2μl
Vector	2μl
Enzyme	1μl
Total	20μl

Wizard® Plus SV Cell Lysis Solution
0.2M NaOH
1% SDS

Wizard® Plus SV Cell Resuspension Solution
50mM Tris-HCl (pH 7.5)
10mM EDTA
100ug/ml RNase A

Wizard® Plus SV Neutralization Solution
4.09M guanidine hydrochloride
0.759M potassium acetate
2.12M glacial acetic acid
final pH 4.2

Wizard® Plus SV Column Wash Solution
162.8mM potassium acetate
22.6mM Tris-HCL (pH 7.5)
0.109mM EDTA (pH 8.0)

Buffer D (Promega)
6mM Tris-HCl
6mM MgCl2
150mM NaCl

1mM Dithiothreitol
pH 7.9 @37C

Buffer H (Promega)
90mM Tris-HCl
10mM MgCl2
50mM NaCl
pH 7.5 @37C

Buffer 1 (New England Biolabs)
10mM bis-Tris-Propane-HCl
10mM MgCl2
1mM Dithiothreitol
pH 7.0 @ 25C

Buffer 4 (New England Biolabs)
50 mM potassium acetate
20 mM Tris-acetate
10 mM Magnesium Acetate
1 mM Dithiothreitol
pH 7.9 @ 25°C

TBE Buffer (x10)
Tris 107.81 g/l
EDTA 5.8 g/l
Boric acid 55.0 g/L

TAE Buffer (x50)
242g Tris
57.1 ml acetic acid
100ml 0.5M EDTA
made up to 1000ml water

T4 DNA Ligase Buffer (New England Biolabs)
50 mM Tris-HCl
10 mM MgCl2
1 mM ATP
10 mM Dithiothreitol
pH 7.5 @ 25°C

TBS-T
Tris pH 10ml
NaCl 30ml
Triton X-100 0.5 ml
Made up to 500ml

Running Buffer
50ml NuPageR Buffer (x20)
100ml Methanol
made upto 1L

Blocking Buffer
Marvel ^{TM} 4% (ie. 2g in 50ml)
TBS-T

Antibodies for Western Blots
Goat IgG α-ETO (Santa Cruz Biotech)
Rabbit IgG α-AML1 (Calbiochem)
Rabbit IgG GAPDH (Cell Signalling)

Primers

AML EXON 4	CTACCGCAGCCATGAAGAACC
ETO EXON 3	AGAGGAAGGCCCATTGCTGAA
ETO Exon 9a	CCATCGATCCTCATATGACCCAGGACAGC
ETO Exon 6a	ACTGGACTGGGAATCAAGAGACCTG
Real Time 6a Primer Forward	TGATTCCCAGTCCAGTGCTCTT
Real Time 6a Primer Reverse	GAGTGCAAAGGCTCTCTGTCAAA
Primer 6a Probe (Dye FAM, Quencher Tamara)	CACTACACCACAATGCCAGTAAGAACCAAAGAAAA

IV. ChIP-Sequencing

Lysis Buffer

Distilled Water	8.3 (for 10ml)
10%SDS	1ml
0.5M EDTA	200ul
1M Tris pH8.0	500ul
200ul PMSF (1mM)	5ul/ml
Aprotinin A1153 1ug/ml	1ul/ml
Leupetin L2884 (1ug/ml)	1ul/ml

Dilution Buffer

Distilled Water	18.7 (for 20ml)
10%SDS	20ul
Triton X-100	200ul
0.5M EDTA	80ul
5M NaCl	600ul
1M Tris pH 8.0	400ul
200ul PMSF (1mM)	5ul/ml
Aprotinin A1153 1ug/ml	1ul/ml
Leupetin L2884 (1ug/ml)	1ul/ml

Wash Buffer 1

Distilled Water	18.72ml (for 20ml)
10% SDS	200ul
Triton X100	200ul
0.5M EDTA	80ul
5M NaCl	600ul
1MTris pH 8.0	200ul

Wash Buffer 2

Distilled Water	17.32 (for 20ml)
10% SDS	200ul
Triton X-100	200ul
0.5M EDTA	80ul
5M NaCl	2ml
1Mtris pH 8.0	200ul

Wash Buffer 3

Distilled Water	14.56ml (for 20ml)
1M LiCl	5ml

NP40	200ul
Deoxycholic acid	200mg
0.5M EDTA	40ul
1M Tris pH 8.0	200ul

Elution Buffer

1% SDS and 1M $NaHCO_3$ (for 10ml solution 0.084g of $NaHCO_3$)

Antibodies

Goat IgG α-ETO (Santa Cruz Biotech)

Abbreviations

AML	Acute Myeloid Leukaemia
AML1	Acute Myeloid Leukaemia 1
AMP	Adenosine Monophosphate
AP-1	proximal activator protein
APE1	apurinic/apyrimidinic endonuclease
APML	Acute Promyelocytic Leukaemia
BAALC	Brain and acute leukaemia cytoplasmic protein
BCL-2	Break-point Cluster Locus -2
BCR-ABL	Break-Point Cluster-ABL
BSA	Bovine serum albumin
bHLH	Basic Helix-Loop-Helix
CBFα	Core Binding Factor alpha
CBF	Core Binding Factor
CBFβ	Core Binding Factor beta
CBP/ p300	CREB Binding Protein
CDK2	Cyclin Dependent Kinase 2
CDKN1a	Cyclin Dependent Kinase inhibitor 1A
CEBPα	CCAAT/enhancer binding protein
CML	Chronic Myeloid Leukaemia
c-Myb	Cellular- myeloblastosis viral oncogene
CR	Complete Remission
CREB	c-AMP response-element binding
C_T	Cycle Threshold
DABG	Detection Abobe Bacground
DEPC	Diethylpyrocarbonate
dNTPs	Deoxynucleotide trisphosphate
dTTP	De-thymidine trisphosphate
dUTP	De-uracil trisphosphate
DNMT	DNA methyltransferase
DW	Distilled Water
EAP	Epstein Barr encoded RNA associated protein
EAR2	Early Growth Response 2
EDTA	Ethylene diamine tetraacetic acid
ERG	v-ets erythroblastosis virus E26 oncogene like
ETO	Eight Twenty-One
ETS	v-ets erythroblastosis virus E26 oncogene homolog
EVI1	ecotropic viral integration site 1
EZH2	Enhancer of zeste homolog 2
FAB	French-American-British
FCS	Fetal Calf Serum
FDR	False Discovery Rate
FLT3	fms-like tyrosine kinase receptor-3

fos	FBJ murine osteosarcoma viral oncogene homolog
GATA1	GATA binding protein 1
G-CSF	Granulocyte Colony Stimulating Factor
GEP	Gene expression profiling
GFI	Growth factor inhibitor
GM-CSF	Granulocyte-macrophage colony-stimulating factor
GPS	G protein pathway suppressor 1
HAT	Histone acetyle transferase
HDAC	Histone deacetylase
HSC	Haemopoietic stem cell
HEB or TCF12	Transcription factor 12
ICSBP	Interferon Consensus Sequence-binding protein
IgA	Immunoglobulin alpha
IDH	Isocitrate dehydrogenase
IL-3	Interleukin 3
IR10	coronin, actin binding protein, 2A
IRES	Internal ribosme entry site
JAK2	Janus Kinase 2
JNK	C-Jun N-terminal kinases
KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolgue
LB medium	Lysogeny broth
LEF	Lymphoid Enhancer Binding Factor
MAD	Mother against DPP
MAPK	Mitogen activated protein kinase
M-CSF	Macrophage colony-stimulating factor
MDM	murine double minute
MDS	Myelodysplasia
MDS1	Myelodysplasia syndrome 1
MDR-1	Mulit drug resistance 1
MEF	Myeloid ELF-1-like factor
MLL	Mixed-Lineage Leukaemia
miR	Micro RNA
MN1	Meningioma (disrupted in balanced translocation)1
MPD	Myeloproliferative disease
MPO	Myeloperoxidase
MRC	Medical Research Council
MTG	Myeloid transforming gene
MYND	MTG8, Nervy, Deformed
NCoR	Nuclear receptor corepressor
NF1	Neurofibromaosis type1
NGF	Nerve Growth Factor
NHR	Nervy homology region
NMTS	Nuclear matrix attachment signal
NPM	Nucleophosmin
NTRK1	Neurotrophic tyrosine kinase, receptor, type 1

p53	Protein 53
PBS	Phosphate buffered saline
PCA	Principal Component Analysis
PDGFR	Platelet Derived Growth Factor Receptor
PEBP2	Polyoma Enhancer Binding Protein 2
PLIER	Probe Logarithmic Intensity Error
PLZF	Promyelocytic leukaemia zinc finger
RA	Retinoic Acid
RARα	Retinoic Acid Receptor alpha
Rb	Retinoblastoma
RHD	Runt homology domain
RMA	Robust Multi-chip Average
RT	Reverse Transcriptase
RUNX	Runt related transcription factor
SAP	SLAM associated protein
SMAD	mothers against DPP homolog
SMRT	Silencing mediator of retinoid and thyroid hormone receptor
SWOG	South-West Oncology Group
TAE	Tris/Acetate/EDTA
T-ALL	T-Lineage Acute Lymphoblastic Leukaemia
TBE	Tris/Borate/EDTA
TBLI	transducin (beta)-like 1
TBS-T	Tris Buffered Saline with Tween-20
TCR	T-cell receptor
TdT	Terminal Deoxynucleotidyl Transferase
TEL/ETV6	ets variant 6
TEL-PDGFR	TEL-Platelet derived growth factor
TGF	Transforming Growth Factor
TLE	Transducin like enhancer
Tris	tris(hydroxymethyl)aminomethane
TRKA	Tyrosine kinase receptor
TRPS	Trichorhinophylangeal syndrome
UDG	Uracil DNA Glycosylase
WHO	World Health Organisation
Wnt	Wingless-type MMTV integration site family
WT-1	Wilms Tumor 1
YAP	Yes associated protein