From the Department of Medicine III, Grosshadern Hospital and
HelmholtzZentrum München, Clinical Cooperative Group "Leukemia"
Ludwig-Maximilians-University, Munich
Chair: **Prof. Dr. med. Wolfgang Hiddemann**

Early target genes of CALM/AF10 as revealed by gene expression profiling

Thesis Submitted for a Doctoral degree in Human Biology
at the Faculty of Medicine Ludwig-Maximilians-University,
Munich, Germany

Submitted by
Medhanie Assmelash Mulaw

From
Asmara, Eritrea

2009

Aus der Medizinischen Klinik und Poliklinik III am Klinikum Großhadern
und HelmholtzZentrum München, Klinische Kooperations Gruppe ,
"Leukämie"
der Ludwig-Maximilians-Universität München,
Direktor: **Prof. Dr. med. Wolfgang Hiddemann**

Frühe Zielgene von CALM/AF10 identifiziert durch Genexpressionsprofilerstellung

Dissertation zum Erwerb des Doktorgrades der Humanbiologie
an der Medizinischen Fakultät der Ludwig-Maximilians-
Universität zu München, Deutschland

vorgelegt von
Medhanie Assmelash Mulaw

Aus
Asmara, Eritrea

2009

Mit Genehmigung der Medizinischen Fakultät der
Universität München

Berichterstatter:	Prof. Dr. med. Stefan K. Bohlander
Mitberichterstatter:	Priv. Doz. Dr. Michael Albert
	Prof. Dr. Wolfram Dempke
Mitbetreuung durch den promovierten Mitarbeiter:	
Dekan:	Prof. Dr. med. Dr. h.c. M. Reiser, FACR, FRCR
Tag der mündlichen Prüfung:	27.07.2009

To my beloved wife, Blen Tsegaye Aregawi

Acknowledgement

I would like to extend my heartfelt gratitude to my supervisor Prof. Dr. med. Stefan K. Bohlander for the opportunity and unreserved support throughout my study. His academic and moral guidance was staggering and conspicuous, without which I couldn't have managed to complete this work in time.

I would also like to acknowledge all the members of the clinical cooperative group (CCG) Leukemia, for the lively and vivid discussions that were quite informative and enlightening. Former and current members of the Prof. Bohlander lab (Belay Tizazu, Deepak Bararia, Alexandre Krause, Luciana-Fontanari Krause, Leticia Fröhlich Archangelo, Philipp Greif, Zlatana Pašalić) are duly acknowledged, whose technical and collegial support was instrumental in the completion of my study.

My deepest gratitude goes to my wife, Blen Tesgaye Aregawi, who has always supported my dreams and encouraged me to fulfill them. Her love and support was pivotal for my study as she patiently and understandingly stayed by my side during my PhD work, paying a gigantic sacrifice that I am hugely indebted to.

I would also like to thank the German Academic Exchange Program (DAAD) for providing me with the scholarship that made it easy for me to pursue my studies.

My acknowledgement also goes to my home institution, Addis Ababa University (Department of Biology, Faculty of Science), which provided me with a study leave to carry out my PhD work.

University of Munich and HelmholtzZentrum München are also duly acknowledged.

Contents

List of Tables

I. Introduction

1.1. Genetic basis of cancer

Worldwide, between 100 and 350 of each 100,000 people die of cancer each year (Lodish *et al.*, 2003). Cancer is due to failures of the mechanisms that usually control the growth and proliferation of cells. During normal development and throughout adult life, intricate genetic control systems regulate the balance between cell birth and death in response to growth signals, growth-inhibiting signals, and death signals.

The loss of cellular regulation that gives rise to most or all cases of cancer are due to genetic damage. Mutations in two broad classes of genes have been implicated in the onset of cancer: **proto-oncogenes** and **tumor suppressor genes** (Lodish *et al.*, 2003). Proto-oncogenes are activated to become oncogenes by mutations that cause the gene to be excessively active in growth promotion, which is either due to increased gene expression or production of a hyperactive product. Tumor suppressor genes normally restrain growth, so damage to them allows inappropriate growth. Many of the genes in both classes encode proteins that help regulate cell birth (i.e., entry into and progression through the cell cycle) or cell death by **apoptosis;** others encode proteins that participate in repairing damaged DNA. Conversion, or activation, of a proto-oncogene into an oncogene generally involves a *gain-of-function* mutation. At least four mechanisms can produce oncogenes from the corresponding proto-oncogenes:

Point mutations (i.e., change in a single base pair) in proto-oncogenes that resuls in constitutively active protein products

Chromosomal translocations which fuse two genes together to produce a hybrid gene encoding a chimeric protein whose activity, unlike that of the parent proteins, often is constitutive

Chromosomal translocations which bring a growth regulatory gene under the control of a different promoter that causes inappropriate expression of the gene

Amplification (i.e., abnormal DNA replication) of a DNA segment including a proto-oncogene, so that numerous copies exist, leading to overproduction of the encoded protein

An **oncogene** formed by either of the first two mechanisms encodes an "oncoprotein" that differs from the normal protein encoded by the corresponding proto-oncogene. In contrast, the other two mechanisms generate oncogenes whose protein products are identical with the normal proteins; their oncogenic effect is due to production at higher than normal levels or their presence in cells where they normally are not produced (Lodish *et al.*, 2003).

1.2. Chromosomal Translocations

1.2.1. Definition

Translocation (*trans* = across; *location* = place) is the term used to describe a rearrangement of chromosome material involving two or more chromosomes.

1.2.2. Types of chromosomal translocations

Two types of chromosomal translocations exist:

1. **Reciprocal translocations:** Reciprocal translocations are the most common type of translocation. These translocations can occur between any of the chromosomes and involve pieces of any size. Pieces of each of these chromosomes have

changed places and the pieces have become attached to the other chromosome. In this case, where there does not appear to have been any loss or gain of chromosome material, the translocation is described as balanced**.**

2. **Robertsonian translocations:** Named after the American insect geneticist who described them, W. R. B. Robertson, Robertsonian translocations only involve exchanges between human chromosome numbers 13, 14, 15, 21 and 22. The exchange involves loss of the short arms of two chromosomes and fusion of the remaining two long arms at their centromeres. Robertsonian translocations can also occur between the two chromosomes of the same pair.

1.2.3. Occurrence of chromosomal translocation

Though the mechanism and reason underlying chromosomal translocations is not well understood, there are at least two theories (Levitan *et al.*, 2003):

1. Rearrangements occur at random, but selection acts to eliminate those that do not provide the cell with a proliferative advantage
2. Alternatively, certain changes may occur preferentially and, thus, may be the ones that that are observed

1.2.4. Causes of chromosomal translocation

Though the cause of chromosomal rearrangement is in most cases not known exactly, one of the main causes is believed to be an exposure to radiation for a prolonged time period. Illegitimate V(D)J recombination, class switch recombination, homologous recombination, non-homologous end joining, and fragile sites all have potential roles in the production of non-random chromosomal translocations (Aplan, 2006). In addition, mutations in DNA repair pathways have been implicated in the production of

chromosomal translocations in humans, mice, and yeast (Aplan, 2006). Some breakage events in pre–B or pre–T cells are due to the RAG (recombination activating gene) proteins recognizing heptamer/nonamer-like sequences (Lieber *et al.*, 2008).

1.2.5. Consequences of chromosomal translocations

Based on the potential effects that chromosomal translocations might entail, two consequences are noted:

a. enforced oncogene expression that occurs when an oncogene is subjected to a new chromosomal environment such as a very active promoter or enhancer element, or
b. gene fusion between exons of two genes, whereby the two genes affected by the translocation form a fusion gene which gives rise to a fusion protein.

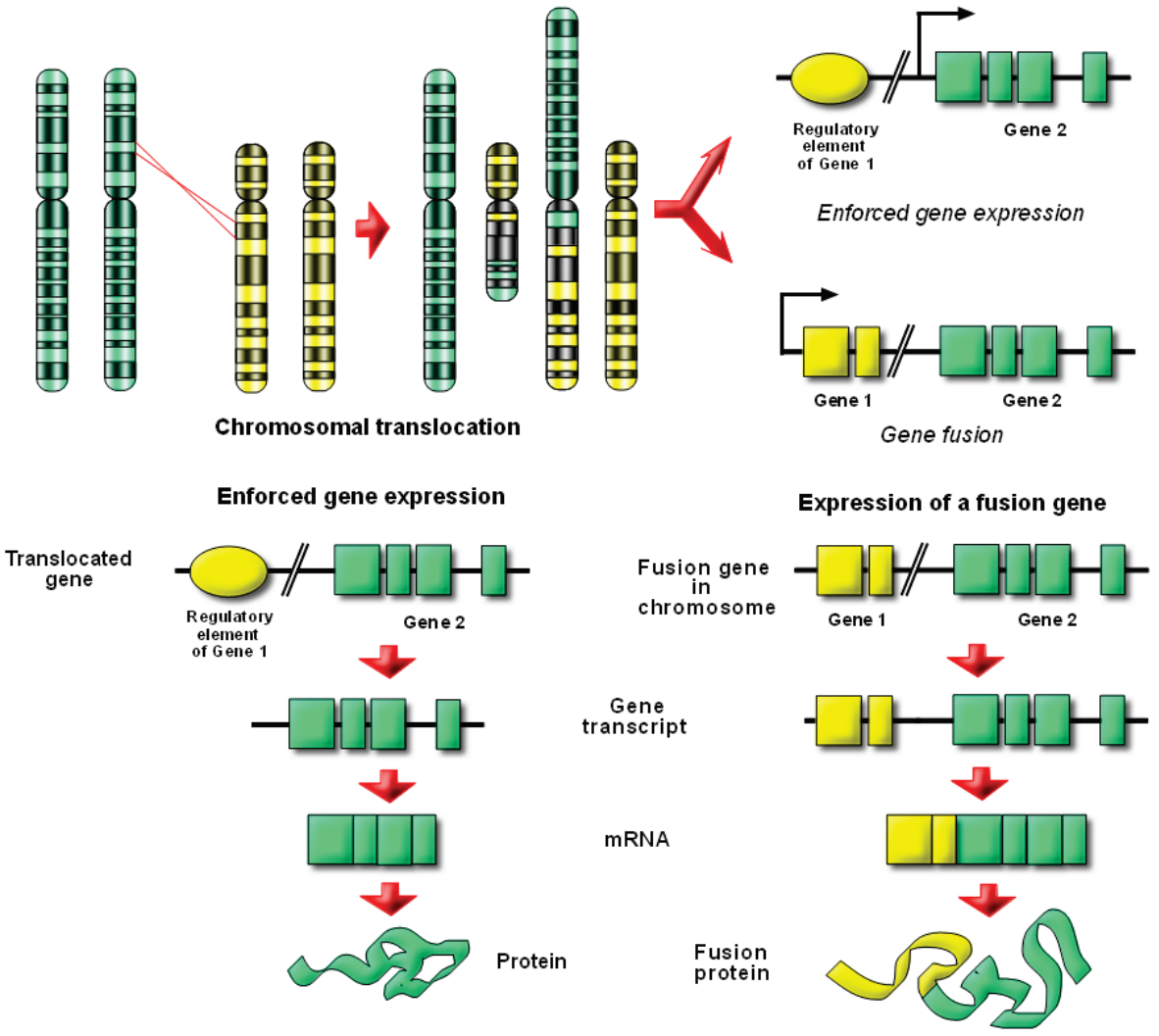


Figure 1.1. Consequences of chromosomal translocations (a) enforced gene expression, (b) fusion gene.

1.3. Chromosomal translocations in hematopoietic malignancies

Chromosomal translocations, especially recurrent chromosomal translocations, are requently found in malignancies and play an important role in malignant transformation. As discussed above, translocations can lead to the juxtaposition of the coding region of a gene near the transcriptionally active promoter/enhancer region of another gene, hence leading to over-expression of the former gene (Nambiar *et al.*, 2008). Examples of such events are the IgH-*BCL2* and IgH-*MYC* rearrangements (Korsmeyer, 1992, Rabbitts,

1991) where *BCL2* and *MYC* show overexpression. Translocations can also result in the formation of a fusion gene. For instance, the t(9;22) translocation results in the fusion of the *ABL1* gene on chromosome 9 with the *BCR* gene on chromosome 22, thereby forming the fusion gene due to the formation of a unique in-frame fusion mRNA and protein (Rowley, 2001, Rowley and Potter, 1976, Rowley, 1973). The BCR/ABL fusion gene is under the control of the BCR promoter. Table 1.1 and table 1.2 give lists of the different translocation events and their consequences.

Table 1.1. Some selected chromosomal translocations involved in hematopoietic malignancies leading to altered gene expression (adapted from Nambiar *et al.* (2008) with some modifications).

Type of translocation	Affected gene	Type of tumor	Mechanism of activation	References
t(8;14)(q24;q32)	c-myc(8q24)	BL, B-ALL	Relocation to IgH locus	Rabbits et al., Adv Immun, 1991. 50:119–146
t(7;19)(q35;p13)	lyl1(19p13)	T-ALL	Relocation to TCR-β locus	Mellentin *et al*., Cell, 1989. 58: 77–83
t(1;14)(p32;q11)	tal1/scl(1p32)	T-ALL	Relocation to TCR-α Locus	Baer *et al*., Sem. Cancer Biol., 1993. 4:341–347
t(11;14)(p15;q11)	rbtn1/ttg1(11p15)	T-ALL	Relocation to TCR-δ Locus	Boehm *et al*., EMBO J, 1988. 7:385–94 McGuire *et al*., Mol. Cell. Biol, 1989. 9:2124–32 Boehm *et al*., PNAS, USA, 1991. 88:4367–71
t(10;14)(q24;q11)	hox11(10q24)	T-ALL	Relocation to TCR-α/β locus	Lu et al., Genes Chr Cancer, 1990. 2:217–22

				Kagan *et al*., PNAS, USA, 1989. 86:4161–5 Zutter *et al*., PNAS, USA, 1990. 87:3161–5
t(11;14)(q13;q32)	Bcl1(11q13)	B-CLL	Relocation to IgH locus	Tsujimoto *et al*., Science, 1984. 224:1403–6 Tsujimoto *et al*., Nature, 1985. 315:340–3 Welzel *et al*., Cancer Res., 2001. 61: 1629–1636
t(5;14)(q31;q32)	IL-3(5q31)	pre-B-ALL	Relocation to IgH locus	Grimaldi *et al*., Blood, 1989. 73:2081–2085 Meeker *et al*., Blood, 1990. 76:285–289
t(1;7)(p34;q34)	lck(1p34)	T-ALL	Relocation to TCR-β Locus	Tycko *et al*., J. Exp. Med., 1991. 174:867–873
t(X;14)(q28;q11)	C6.1B(Xq28)	T-PLL	Relocation to TCR-α locus	Stern *et al*., Oncogene, 1993. 8:2475–2483
t(14;21)(q11;q22)	bhlhb1(21q22)	pre-T-LBL	Relocation to TCRA/D locus	Wang *et al*., PNAS, 2000. 97: 3497–3502

Table 1.2. Some selected chromosomal translocations involved in hematopoietic tumors leading to gene fusions (adapted from Nambiar *et al*. (2008) with some modifications).

Type of translocation	Type of tumor	Genes involved	References
t(8;21)(q22;q22)	AML-M2	AML1/CBFα(21q22) and ETO/MTGβ(8q22)	Ohki M, Sem. Cancer Biol., 1993. **4**:369–376
t(15;17)(q21;q21)	APL	PML (15q21) and RARA (17q21)	Gillard *et al*., Sem. Cancer Biol., 1993. **4**:359–368
t(8;16)(p11;p13)	AML	MOZ(8p11) and CBP(16p13)	Borrow J *et al*., Nat Genet., 1996. **14**:33–41
t(9;22)(q34q11.2)	CML/ALL	BCR(22q11) and c-ABL(9q34)	de Klein *et al*., Nature, 1982. **300**:765–767
t(11;17)(q23;q21.1)	AML-M3	PLZF (11q23) and RARA (17q21)	Chen *et al*., EMBO J., 1993. **12**:1161–1167
t(9;11)(p22;q23)	AML-M4, pre-B-ALL	MLL (11q23) and AF9 (9p22)	Nakamura *et al*., PNAS, 1993. **90**: 4631–4635
t(6;11)(q27;q23)	AML-M5, ALL	MLL (11q23) and AF6 (6q27)	Prasad *et al*., Cancer Res., 1993. **53**:5624–5628
t(6;9)(p23;q34)	AML-M1,M2,M4,M5	DEK (6p23) and CAN (9q34)	von Lindern, Mol. Cell Biol., 1992. **12**:1687–1697
t(16;21)(p11;q22)	AML	FUS(16p11) and ERG(21q22)	Shimizu *et al*., PNAS, 1993. **90**:10280–10284
t(16;21)(q24;q22)	t-AML, MDS	AML1(21q22) and MTG(16q24)	Gamou T *et al*., Blood, 1998. **91**:4028–4037.
t(12;22)(p13;q11–12)	MDS	ETV6(12p13)/TEL and MN1(22q11)	Buijs A *et al*., Oncogene, 1995. **10**:1511–1519

t(12;21)(p12–13;q22)	Pre-B-ALL	TEL(12p12) and AML1(21q22)	Romana *et al*., Blood, 1995. **86**:4263–4269 Golub et al., Proc. Natl. Acad. Sci. USA, 1995. **92**: 4917-4921
t(4;11)(q21;q23)	Pre-B-ALL	MLL(11q23) and AF4(4q21)	Djabali *et al*., Nature Genet., 1992. **2**: 113–118 Gu *et al*., Cell, 1992. **71**:701–708
t(11;19)(q23;p13)	Pre-B-ALL, T-ALL	MLL(11q23) and ENL(19p13)	Tkachuk *et al*., Cell, 1992. **71**:691–700 Yamamoto *et al*., Oncogene, 1993. **8**: 2617–2625
t(2;5)(p23;q35)	ALCL (NHL)	NPM (5q35) and ALK (2p23)	Morris *et al*., Science, 1994. **263**:1281–1284
inv(16)(p13q22)	AML, MDS at times	MYH11	Bernard et al.,Leukemia, 1989. 3 (10) : 740-745. Campbell et al., 1991, Genes, chromosomes & cancer.**3** (1) : 55-61.
t(10;11)(p12;q14)	pre-T-ALL, AML	CALM (11q14) and AF10(10p12)	Dreyling *et al*., Proc. Natl. Acad. Sci. USA. 1996. **93**, 4804-4809.

1.4. CALM/AF10

The t(10;11)(p12;q14) is a recurring chromosomal translocation that is found in acute myeloid and acute lymphoblastic leukemia as well as in malignant lymphoma (Dreyling *et al.*, 1996). This translocation results in the fusion of AF10 (Chaplin *et al.*, 1995), a putative zinc finger transcription factor containing an N-terminal LAP/PHD zinc finger motif, a nuclear localization signal, an AT-hook domain, and a leucine zipper and with the CALM gene (Clathrin assembly protein lymphoid myeloid leukemia gene) that encodes a clathrin assembly protein.

(Linder *et al.*, 2000, Saha *et al.*, 1995) have identified a novel protein-protein interaction motif in the AF10 protein comprising the extended LAP/PHD-finger domain that mediates homo-oligomerisation of recombinant AF10 and is conserved in several proteins, including MLL itself. AF10 binds cruciform DNA via a specific interaction with an AT-hook motif and is localized to the nucleus by a defined bipartite nuclear localization signal in the N-terminal region (Linder *et al.*, 2000). Furthermore, it has recently been shown that PHD domains interact with H3K4me3 (Li *et al.*, 2006, Peña *et al.*, 2006), potentially playing a crucial role in transcription initiation as H3K4me3 modifications have recently been shown to be associated with gene expression (Barski *et al.*, 2007).

CALM (for Clathrin Assembly Lymphoid Myeloid; also known as PICALM) is located on chromosome 11q23, is ubiquitously expressed and encodes a 652 amino acid protein with multiple domains involved in endocytosis (Dreyling *et al.*, 1996). These domains include the epsin N-terminal homology (ENTH) domain (Asp-Pro-Phe), a DPF (ASP-Pro-Phe) motif, an NPF (Asn-Pro-Phe) motif and type I and II clathrin-binding sequences

(CBS I and II) (Anika Meyerholz, 2005, Tebar *et al.*, 1999, Klebig *et al.*, 2003). The predicted CALM protein is similar to the neuronal specific monomeric clathrin assembly protein AP180, which was first identified in coated vesicles of bovine brain (S Ahle and Ungewickell, 1986, Takei *et al.*, 1996). CALM homologues have been identified in rat, mouse and cow.

The CALM/AF10 fusion mRNA codes for the CALM/AF10 fusion protein which contains almost the complete CALM protein fused in frame to about 90% of the AF10 protein without the N terminal two PHD zinc fingers (Dreyling *et al.*, 1996) (See figures 2 and 3). In the monocytic cell line U937 both the CALM/AF10 and the AF10/CALM fusion mRNAs can be identified (Dreyling *et al.*, 1996). CALM/AF10 chimeric transcripts (six isoforms of the CALM/AF10 fusion transcripts and five of AF10/CALM) were detected by RT-PCR in hematological malignancies with t(10;11)(p13;q14) in patient samples including ALL, AML and lymphoblastic lymphoma, and in three monocytic cell lines (P31/Fujioka, KP-Mo-TS and U937) (Ashihara *et al.*, 2007, Narita *et al.*, 1999).

CALM/AF10 is highly Leukemogenic. CALM/AF10 leads to the development of an aggressive acute leukemia with in a murine bone marrow transplant model (Deshpande *et al.*, 2006). According to Deshpande *et al.* (2006), all mice (n = 13) transplanted with highly purified GFP+ CALM/AF10 (murine stem cell virus (MSCV)-based retroviral construct carrying the CALM/AF10 cDNA upstream of an internal ribosomal entry site (IRES)-green fluorescence protein (GFP) cassette) transduced cells (median retroviral transduction efficiency of 5.7, range 1.2%–16.4%) with or without non-transduced helper

cells developed the disease after a median of 110 days post-transplantation (range 46–366).

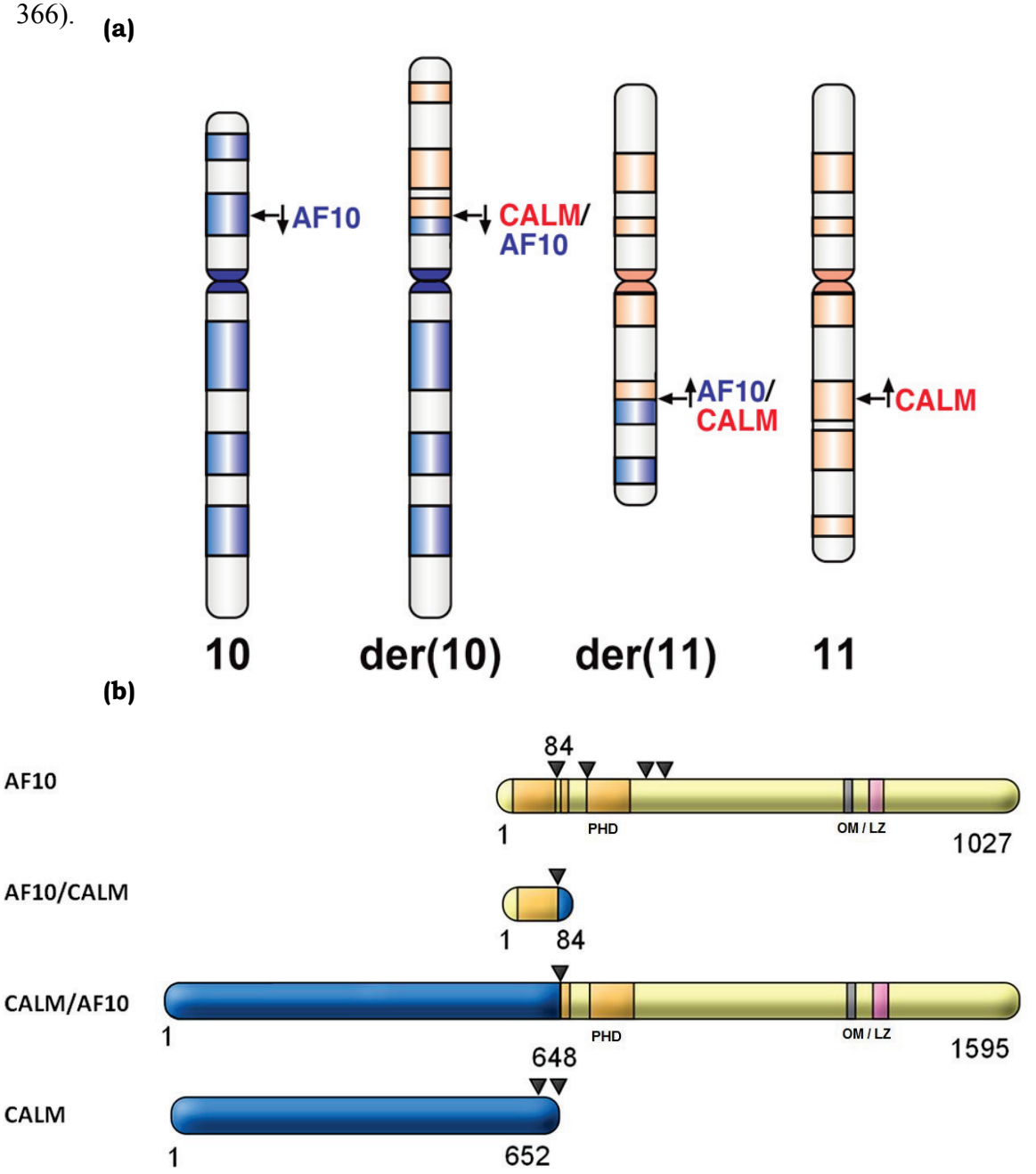


Figure 1. 2. t(10;11)(p12;q14) leading to fusion of CALM and AF10. (a) the translocation event, arrows indicating the approximate position of translocation and gene direction (b) the genes involved (CALM and AF10) and their fusion products as a consequence of t(10;11)(p12;q14). AF10 contains an octapeptide motif and leucine zipper domain (OM/LZ) and a plant homeo domain (PHD) finger. Small triangles indicate the different breaking points in AF10 and in CALM as observed in CALM/AF10 patients. Numbers stand for amino acid counts.

In a recent report, (Caudell *et al.*, 2007) showed that 40% to 50% of the F1 generation of transgenic mice that express CALM/AF10 under the control of the *vav* promoter developed acute leukemia at a median age of 12 months. The transgenic mice were generated by cloning CALM/AF10 into the HS21/45-*vav* vector that uses 5′ and 3′ *vav* regulatory elements to direct expression of cDNA inserts specifically in hematopoietic tissues. Caudell and co-workers also reported, based on real-time RT-PCR experiment, that *Hoxa5, Hoxa7, Hoxa9, Hoxa10*, and *Meis1* were all up-regulated in hematopoietic tissues (bone marrow, spleen, and thymus) from clinically healthy CALM/AF10. These genes were also found to be up-regulated in Myeloid leukemias from CALM/AF10 mice, indicating that up-regulation of these genes occurs in myeloid as well as T-cell tumors associated with CALM/AF10 expression (Caudell *et al.*, 2007). *Hox* genes encode a class of transcription factors called homeobox genes and are found in clusters named A, B, C, and D on four separate chromosomes. Expression of these proteins is spatially and temporally regulated during embryonic development, and they regulate gene expression, morphogenesis, and differentiation. They encode proteins that involved in the placement of hindbrain segments in the proper location along the anterior-posterior axis during development. *Meis1* (myeloid ecotropic viral integration site 1) is also a homeobox gene and encodes a protein belonging to the TALE ('three amino acid loop extension') family of homeodomain-containing proteins.

Caudell *et al*. (2007) assayed the genes *Hoxa5, Hoxa7, Hoxa9, Hoxa10*, and *Meis1* based on the findings of a previous report on patient microarray study that compared CALM/AF10^{+} to CALMAF10^{-} T-ALL (Dik *et al.*, 2005). Dik and co-workers reported that their analysis showed the up-regulation of *HOXA5, HOXA9, HOXA10* and *BMI1* in

the CALM/AF10$^+$ cases, a finding that they validated by quantitative RT-PCR by comparing an independent group of T-ALL to mixed lineage leukemia-translocated acute leukemias (MLL-t AL). The over-expression of *HOXA* genes was associated with overexpression of its cofactor *MEIS1* (myeloid ecotropic viral integration site homeobox 1) in CALM/AF10$^+$ T-ALL, reaching levels of expression similar to those observed in MLL-t AL (Dik *et al.*, 2005). Consequently, Dik *et al.* (2005) reported that CALM/AF10$^+$ T-ALL and MLL-t AL share a specific *HOXA* overexpression, indicating they activate common oncogenic pathways.

Furthermore, *BMI1* (polycomb, ubiquitin ligase, B-lymphoma Mo-MLV integration site), a gene located close to AF10 breakpoint and controls cellular proliferation through suppression of the tumor suppressors encoded by the *CDKN2A* locus, was found to be over-expressed only in CALM/AF10+ T-ALL and not in MLL-t AL (Dik *et al.*, 2005). As *CDKN2A* is often deleted in T-ALL and was found to be conserved in CALM/AF10+ T-ALL, Dik and co-workers suggested that decreased *CDKN2A* activity, as a result of *BMI1* overexpression, contributes to leukemogenesis in CALM/AF10+ T-ALL.

Little is known about the molecular mechanism of leukemogenesis of CALM/AF10 (Caudell and Aplan, 2008). Recently, there has been a report that demonstrated that the H3K79 methyltransferase *hDOT1L* (DOT1-like, histone H3 methyltransferase (*S. cerevisiae*)), a gene that mediates H3K79 methylation and is associated with telomere silencing, meiotic checkpoint control, and DNA damage response) contributes to MLL–AF10-mediated leukemogenesis through its interaction with *AF10* (Okada *et al.*, 2005). The role of *DOT1L* in embryonic development, genome stability and heterochromatin formation has very recently been shown in a report where (Jones *et al.*, 2008) analyzed

mouse embryonic stem (ES) derived from *Dot1L* mutant blastocysts and observed that these cells had global loss of H3K79 methylation as well as reduced levels of heterochromatic marks (H3K9 di-methylation and H4K20 tri-methylation) at their centromeres and telomeres. Jones and co-workers further noted that these changes are accompanied by aneuploidy, telomere elongation, and proliferation defects.

According to Okada *et al*. (2005), mistargeting of *hDOT1L* to *Hoxa9* plays an important role in MLL-AF10-mediated leukemogenesis and suggests that the enzymatic activity of *hDOT1L* may provide a potential target for therapeutic intervention. In another experiment where CALM/AF10 was knocked down in U937 cells using a vector-based RNA interference (RNAi) and later transplanted into NOD/SCID mice. The same group found that *hDOT1L* contributes to CALM–AF10- mediated leukemic transformation by preventing nuclear export of CALM–AF10 and up-regulation of the *Hoxa5* gene through H3K79 methylation (Okada *et al.*, 2006). In contrast to the hypothesis put forward by Dik *et al*. (2005), Okada and co-workers demonstrated that CALM/AF10 fusion is both necessary and sufficient for leukemic transformation.

Even though much is known about the CALM/AF10 fusion from patient samples and the leukemic potential of CALM/AF10 is well established in mouse models, the mechanisms that are required for CALM/AF10-mediated leukemogenesis are still poorly understood. The aim of this study is to identify immediate target genes of CALM/AF10 using gene expression profiling on an Affymetrix microarray platform with an inducible CALM/AF10 cell line system. Identifying early and potentially direct targets of CALM/AF10 will increase our understanding of the molecular mechanism of CALM/AF10-mediated-leukemogenesis.

II. Materials and Methods

2.1. Materials

2.1.1. Chemicals

Table 2.1. List of chemicals

Name of Reagent	IUPAC Name	Molecular formula	Molar Mass (g/mol)	Company
Agarose	(14)-3,6-anhydro-α-L-galactopyranosyl-(13)-β-D-galactopyranan		167.6 ± 7.8 × 10^3	Carl Roth® GmbH, Germany
Aprotinin	Aprotinin	$C_{284}H_{432}N_{84}O_{79}S_7$	6511.51	Sigma-Aldrich®, USA
ATP	5-(6-aminopurin-9-yl) -3,4-dihydroxy-oxolan-2-yl methoxy-hydroxy-phosphoryl oxy-hydroxy-phosphoryl oxyphosphonic acid	$C_{10}H_{16}N_5O_{13}P_3$	507.181	Invitrogen®, USA
Bromophenol blue	2,6-dibromo-4-[3-(3,5-dibromo-4-hydroxyphenyl)-1,1-dioxo-3-benzo[c]oxathiolyl]phenol	$C_{19}H_{10}Br_4O_5S$	669.9607	NEB®, USA
Chloroform	Chloroform	$CHCl_3$	119.38	Carl Roth® GmbH,

				Germany
dATP	Deoxyadenosine triphosphate	$C_{10}H_{16}N_5O_{12}P_3$	491.181623	Invitrogen®, USA
dCTP	Deoxycytidine triphosphate	$C_9H_{16}N_3O_{13}P_3$	467.156923	Invitrogen®, USA
DEPC	Diethylpyrocarbonate	$C_6H_{10}O_5$	162.141	Invitrogen, USA
dGTP	Deoxyguanosine triphosphate	$C_{10}H_{16}N_5O_{13}P_3$	507.181023	Invitrogen®, USA
DMSO	dimethyl sulfoxide	C_2H_6OS	78.13	Carl Roth® GmbH, Germany
DTT	2S,3S)-1,4-Bis-sulfanylbutane-2,3-diol	$C_4H_{10}O_2S_2$	154.26	Invitrogen®, USA
dTTP	Deoxythymidine-triphosphatase			Invitrogen®, USA
EDTA	2-[2-(Bis (carboxymethyl) amino) ethyl-(carboxymethyl) amino]acetic acid	$C_{10}H_{16}N_2O_8$	292.25	Carl Roth® GmbH, Germany
Ethanol	Ethanol	CH_3CH_2OH	46.06844	Carl Roth® GmbH, Germany

Ethidium bromide	3,8-Diamino-5-ethyl-6-phenylphenanthridinium bromide	$C_{21}H_{20}BrN_3$	394.294	Invitrogen®, USA
Ficoll				Carl Roth® GmbH, Germany
Glycerol	Propane-1,2,3-triol	$C_3H_5(OH)_3$	92.09382	Carl Roth® GmbH, Germany
Isopropanol	Propan-2-ol	C_3H_8O	60.10	Carl Roth® GmbH, Germany
Leupeptin	Ac-Leu-Leu-Arg-H; N-acetyl-L-leucyl-L-leucyl-D,L-argininaldehyde	Ac - Leu - Leu - Arg - CHO	426.6	Sigma-Aldrich®, USA
Magnesium acetate	Magnes	$(CH_3COO)_2Mg$	214.4	Carl Roth® GmbH, Germany
Magnesium chloride	Magnesium chloride	$MgCl_2$	95.211	Carl Roth® GmbH, Germany
Magnesium sulphate	Magnesium sulphate	$MgSO_4$	120.415	Carl Roth® GmbH, Germany

Pefabloc	4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride	$C_8H_{10}FNO_2S$	203.235	Sigma-Aldrich®, USA
PEG	poly(oxyethylene)	$C_{2n+2}H_{4n+6}O_{n+2}$	44n+62	Sigma-Aldrich®, USA
Pepstatin	3-hydroxy-4-[2-[3-hydroxy-6-methyl-4-[3-methyl-2-[3-methyl-2-(3-methylbutanoylamino)butanoyl]amino-butanoyl]amino-heptanoyl]-aminopropanoylamino]-6-methyl-heptanoic acid	$C_{34}H_{63}N_5O_9$	685.892	Sigma-Aldrich®, USA
Potassium acetate	Potassium acetate	CH_3COOK	98.15	Carl Roth® GmbH, Germany
SDS	Sodium dodecyl sulfate	$NaC_{12}H_{25}SO_4$	288.38	Carl Roth® GmbH, Germany
Sodium chloride	Sodium chloride	NaCl	58.44277	Carl Roth® GmbH, Germany
Sodium hydroxide	Sodium hydroxide	NaOH	39.9971	Carl Roth® GmbH, Germany

Sucrose	Sucrose	$C_{12}H_{22}O_{11}$	342.29648	Carl Roth® GmbH, Germany
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol	$C_4H_{11}NO_3$	121.14	Carl Roth® GmbH, Germany
Tween 20	Polyoxyethylene (20) sorbitan monolaurate	$C_{58}H_{114}O_{26}$	1227.54	Carl Roth® GmbH, Germany
Xylene cyanol	Xylene cyanol	$C_{25}H_{27}N_2O_6S_2Na$	538.61	Carl Roth® GmbH, Germany
Zinc chloride	Zinc chloride	$ZnCl_2$	136.315	Carl Roth® GmbH, Germany
β-mercaptoethanol	2-Hydroxy-1-ethanethiol	C_2H_6OS	78.13	Sigma-Aldrich®, USA

2.1.2. Enzymes

Table 2.2. List of enzymes

Name of Reagent	Purpose	Unit Definition	Company
Benzonase	Removal of nucleic acids from protein samples	Amount of enzyme required to completely digest 37 µg of DNA in 30 minutes under standard assay conditions	Sigma-Aldrich®, USA
Calf Intestine Alkaline Phosphatase (CIAP)	Dephosphorylation of DNA ends	1 unit hydrolyzes 1 µmol of p-nitrophenyl phosphate in 1 minute at 37°C	Invitrogen®, USA
Polynucloetide Kinase (PNK)	Catalyzes the transfer of the terminal phosphate of ATP to 5' hydroxyl termini of polynucleotides such as DNA and RNA, oligonucleotides and 3' mononucleotides	One unit is the amount of enzyme required to incorporate 1 nmol of radiolabeled ATP into DNA substrate in 30 min at 37°C.	Usb® (United States Biochemical), USA
Proteinase K	Degradation or removal of proteins from samples	1 unit liberates folin-positive amino acids and peptides corresponding to 1 µmol of tyrosine in 1 minute at 37°C in a total reaction volume of 250 µl (1)	NEB®, USA
RNase H	specifically hydrolyzes the phosphodiester bonds of RNA which is	One unit is defined as the amount of enzyme that will hydrolyze 1 nmol of the RNA in [3H]-labeled	Invitrogen®, USA

	hybridized to DNA	poly(rA).poly(dT), to acid-soluble ribonucleotides in a total reaction volume of 50 µl in 20 minutes at 37°C in 1X RNase H Reaction Buffer with 10 nmol [3H]-labeled poly(rA) and 12.5 µg poly(dT).	
RNase I	Degradation or removal of RNA from samples	One unit of RNase I degrades 100 ng of *E. coli* ribosomal RNA per second into acid-soluble nucleotides at 37°C in 10 mM Tris-HCl (pH 7.5), 100 mM NaCl, and 1 mM EDTA	QIAGEN®, Germany
T4 DNA Ligase	Joins the 5' phosphate and the 3'-hydroxyl groups of DOUBLE stranded DNA molecules	amount of enzyme required to give 50% ligation of HindIII fragments of λ DNA (5´ DNA termini concentration of 0.12 µM, 300- µg/ml) in a total reaction volume of 20 µl in 30 minutes at 16°C in 1X T4 DNA Ligase Reaction Buffer	NEB®, USA
T4 DNA Polymerase	Converts DNA fragments with 3′- and 5′-protruding ends to DNA fragments with blunt ends	One unit converts 10 nmoles of dNTPs into acid-insoluble material in 30 minutes at 37^{o}C.	NEB®, USA
Thermoscript Reverse Transcriptase	synthesizes DNA from an RNA template	One unit incorporates 1 nmol of dTTP into acid-precipitable material in 10 min at 37°C using	Invitrogen®, USA

	poly(A).oligo(dT)25 as template-primer

2.1.3. Primers

Table 2.3. List of Primers

Name	Sequence	Purpose
VP16TopNot	5'-GATCGCGGCCGCCCTCCTGAAAGATGAAGCTA-3'	Amplify VP16 transactivation domain
VP16BottomXho	5'-GATCTCGAGGCCGCGGATCCCGGACCCGG-3	
CALM2192for	5'-GCAATCTTGGCATCGGAAAT-3	Primer pairs that amplify the CALM/AF10 break point
AF10254rev	5'-GCCTGTCGACATCCATGTTT-3	

2.1.4. Plasmids

Table 2.4. List of Plasmids

Name	Type
pEYFP C1	Mammalian expression vector
pINCO NGFR	NGF expressing mammalian expression vector
pRTS-1	Mammalian expression vector
pUC19 with SfiI cassette	Shuttle vector for subcloning

2.1.5. Buffers and Solutions

Table 2.5. List of Buffers and Solutions

Name of Buffer/Solution	Purpose	Composition
Blocking Buffer	Blocking protein membranes	5 g Milk powder 0.1% Tween 20 in TBS
cDNA Synthesis Buffer	Reverse transcriptase reaction Buffer	250 mM Tris acetate, pH 8.4 375 mM potassium acetate 40 mM magnesium acetate 20 µg/ml BSA
CIAP dilution Buffer	Dilution buffer for CIAP	25mM Tris.Cl, pH 7.6 1mM $MgCl_2$ 0.1 mM $ZnCl_2$ 50% Glycerol (v/v)
Dephosphorylation Buffer	Dephosphorylation Buffer	50mM Tris.Cl, pH 8.5 0.1 mM EDTA
EB Buffer	DNA elution buffer	10 mM Tris·Cl, pH 8.5
ER		Confidential
Fetal Calf Serum (FCS)	Supplement to culture media	GIBCO®, USA
FT LYSIS Buffer	Cell lysis for protein extraction	600 mM KCl 20 mM

		Tris.Cl, pH 7.8 20% Glycerol
Laemmli Buffer	Protein sample preparation	4% SDS 20% glycerol 10% β-mercaptoethanol 0.004% bromphenol blue 0.125 M Tris.Cl, pH 6.8
Ligation Buffer	Buffer for ligation reaction	132 mM Tris-HCl, pH 7.6 20 mM $MgCl_2$, 2 mM dithiothreitol 2 mM ATP 15% Polyethylene glycol (PEG 6000)
Loading dye	Dye used for loading DNA and RNA samples on agarose gels	Bromophenol blue (0.25%) Xylene xyanol (0.25%) Sucrose (40%) H_2O
Lysis Buffer	Mammalian Genomic DNA extraction	100 mM Tris.Cl pH 8.5 5 mM EDTA 0.2% SDS 200 mM NaCl

		100 μg Proteinase K/ml
MACS Buffer	Magnetic Activated Cell Sorting Buffer	2.5 g BSA/500 ml approx. 0.5%; 2 ml EDTA 0.5 M
Non Essential Amino Acids (NEAA)	Supplement to culture media	GIBCO®, USA
OPTIMEM	Electroporation medium for mammalian cells	GIBCO®, USA
P_1 Buffer	Resuspension buffer for bacterial pellets	50 mM Tris.Cl pH 8.0 10 mM EDTA 100 µg/ml Rnase A
P_2 Buffer	Lysis Buffer	200 mM NaOH 1% SDS (w/v)
P_3 Buffer	Neutralization Buffer	3.0 M Potassium Acetate pH 5.5
PBS	Phosphate Buffered Saline	GIBCO®, USA
PCR Buffer	Traditional PCR Buffer	500 mM KCl 100 mM Tris.Cl ,pH 9.0, 1.0% Triton X 100)
Protein inhibitor cocktail	Solution for freeze thaw protein exptraction procedure	0.4 mg/ml Pefabloc

		10 μg/ml Leupeptin 10 μg/ml Pepstatin 5 μg/ml Aprotinin
Pyruvic acid	Supplement to culture media	GIBCO®, USA
QBT Buffer	Equilibration Buffer	750 mM NaCl 50 mM MOPS, pH 7.0 15% Isopropanol (v/v) 0.15% Triton® X-100 (v/v)
QC Buffer	Wash Buffer	1.0 M NaCl 50 mM MOPS, pH 7.0 15% Isopropanol (v/v)
QG Buffer	Gel solubilization buffer	Confidential
QN Buffer	Elution Buffer	1.6 M NaCl 50 mM MOPS, pH 7.0 15% Isopropanol (v/v)
T4 DNA Polymerase Buffer	T4 DNA Polymerase Buffer	330 mM Tris-acetate, pH 7.8 660 mM potassium acetate 100 mM magnesium acetate

		5 mM DTT
T4 PNK Buffer	T4 PNK Buffer	0.5 M Tris.Cl, pH 7.6 100 mM $MgCl_2$ 100 mM β-mercaptoethanol
TAE Buffer	Agarose electrophoresis Buffer	Tris base Glacial acetic acid EDTA, pH 8.0
TBE Buffer	Agarose electrophoresis Buffer	Tris base Boric acid EDTA, pH 8.0
TBS	Transfer Buffer for SDS-PAGE transfer setup	192mM Glycine 25mM Tris 20%Methanol
TE Buffer	Resuspension of air dried DNA	10 mM Tris.Cl pH 7.5 0.1 mM EDTA
TRIZOL®	RNA extraction solution	Guanidinium thiocyanate-phenol-chloroform
Trypan Blue	Counting cells	GIBCO®, Germany

2.1.6. Media

2.1.6.1. Bacterial culture

Table 2.6. Bacterial culture medium and agar

Name of Media	Purpose	Company
Liquid Broth (LB) Agar	Solid medium for bacteria culture	Carl Roth® GmbH, Germany
Liquid Broth (LB) Medium	Liquid medium for bacteria culture	Carl Roth® GmbH, Germany

2.1.6.2. Mammalian cell culture

Table 2.7. Mammalian cell culture medium

Name of Media	Purpose	Company
DMEM	Culture media for adherent cells	GIBCO®, USA
RPMI 640	Culture media for suspension cells	GIBCO®, USA

2.1.7. Antibodies

Table 2.8. List of Antibodies

Name	Description	Company
Anti AF10	Recognizes the AF10 protein	HelmholtzZentrum Muenchen, Germany
Anti CALM	Recognizes the Clathrin Assembly Lymphoid Myeloid (CALM) protein	Santa Cruz Biotechnology® Inc., USA
Anti FLAG	Detects the FLAG tag	Sigma-Aldrich®, USA
Anti GFP	Recognizes the Green Fluorescent Protein (GFP)	Roche®, Switzerland
Anti NGF	Recognizes the cell membrane protein NGF	Miltenyi®, Germany

2.1.8. Cells

2.1.8.1. Bacterial Cells

Table 2.9. List of chemo/electro-competent bacteria used for transformation

Name	Description	Company
DH5α	Chemically competent cells for transformation purposes	Invitrogen®, Germany
INV 110	Electrocompetent/chemically competent bacterial cells that are *dam*$^-$	Invitrogen®, Germany

XL1Blue	Electrocompentent bacteria cells	Invitrogen®, Germany

2.1.8.1. Mammalian Cells

Table 2.10. List of Mammalian cell lines used

Name	Description	Source
293T	Human Embryonic Kidney cells	HelmholtzZentrum Muenchen, Germany
DG75		HelmholtzZentrum Muenchen, Germany
H1299	human lung cancer *cell line*	HelmholtzZentrum Muenchen, Germany

2.1.9. Computer Operating System and Application Software

Table 2.11. List of software and applications

Software	Purpose	Company/Source
Adobe Photoshop CS2	Image analysis application	Adobe® Inc., USA
dChip	Microarray analysis tool	Harvard (Li Wong), USA
Endnote X1	Reference manager application	Thompson® Company, USA
Gene Pattern 3.1.1	Microarray analysis suite	Broad MIT, USA
Image J 1.410	Image analysis software	National Institute of Health, USA
Linux Fedora Core 8 and 9	Core Operating System	Open Source/Fedora Project
Macromedia Freehand MX	Drawing and image processing application	Adobe® Inc., USA

Microsoft Office 2007	Office applications	Microsoft® Company, USA
R Statistical Package 2.3 – 2.8	Freeware statistical package	CRAN, International
SDS 2.1	Sequence Detection System for Taqman Real Time PCR	Applied Biosystems®, USA
STATISTICA 7	Multipurpose statistical software	Statsoft® Inc., USA
Vector NTI Suite 10	Molecular Biology and related fields application Suite	Invitrogen® Company, USA
Windows Vista	Core Operating System	Microsoft® Company, USA
Windows Xp	Core Operating System	Microsoft® Company, USA

2.1.10. Equipments and Utensils

Table 2.12. List of equipments and utensils

Name of Item	Purpose	Company
7900HT system Real Time PCR System	Station for running Taqman® Real Time PCR reactions	Applied Biosystems®, USA
Bacteria culture dish	Petri-dish for culturing bacteria	Corning Life Sciences, USA
Mammalian cell culture dish	Plates for growing mammalian cells	Corning Life Sciences, USA
Mammalian cell culture flask	Flask for culturing mammalian cells	Sarstedt®, Germany
Conical Flask	Multipurpose plasticware for routine lab activities	Falcon®, Germany
EasyjecT Prima	Bacteria electroporation device	EquiBio® Co., UK
Graduated Cylinder	Measuring volumes	VIT-LAB® GmbH, Germany

High speed Centrifuge with thermoregulator	For cooled centrifugation purposes (max. centrifugation 21160 rcf)	Eppendorf®, USA
High volume Centrifuge with cooling system	For centrifugation of large volumes (max speed 4,500 rpm)	Sigma®, Germany
Incubator Shaker	For incubating samples at specific temperature and rotations, mainly bacterial cultures	Innova®, USA
Micro Fluidic Card Sample Block (Also referred to as the thermal cycler module or thermal block)	Metal heating block required for running low density array cards	Applied Biosystems®, USA
Micro Fluidic Card Sealer	To seal LDA cards	Applied Biosystems®, USA
Microfuge Tubes (1.5 ml)	Multipurpose plastic tubes for routine lab activities	Eppendorf®, Germany
Pipette tips	Plastic tips fitting to pipettes	Carl Roth® GmbH, Germany
Micropipettes	Tool for measuring, adding, removing, and mixing small volumes of solutions	Gilson® Co., USA
QIAfilter Cartridge with cap and Plunger	Filtration column during maxiprep (filters debries and proteins)	QIAGEN®, Germany
QIAGEN-tip 500	Column for filtering and eluting DNA	QIAGEN®, Germany

QIAquick column	Column for filtering solubilized agarose gel from DNA	QIAGEN®, Germany
Sorvall/Heraeus Custom Buckets and Adaptors (4 pcs.)	Centrifugation buckets for LDA cards	Thermo Fisher® Scientific, USA
Table-top Centrifuge	Routine centrifugation needs (max. speed 14,000 rpm)	Eppendorf®, Germany

2.2. Methods

2.2.1. Extraction

2.2.1.1. Nucleic Acids Extraction

2.2.1.1.1. DNA Extraction

2.2.1.1.1.1. Mammalian Genomic DNA extraction

This protocol was adopted from (Laird, 1991). Cells were grown in a 6 well plate (approximately 4 x 10^6 cells) were used as a starting material. The culture media was removed and cells were washed twice with PBS by centrifugation (1100 rpm at 4^{o}C). Lysis buffer, which has been adjusted to allow enzymatic manipulation of the DNA without prior organic solvent extractions, was added to the washed cells. The procedure involved just three manipulations:

1. Addition of lysis buffer to the tissue or cells
2. Addition of isopropanol
3. Transfer of precipitate to TE

1. Lysis: The lysis buffer (0.5 ml) was added to the cells. Digestion was complete within 4 hours at 37°C with agitation.

2. Isopropanol precipitation: One volume of isopropanol was added to the lysate and the samples were mixed or swirled until precipitation was complete (viscosity completely gone).

3. Recovery of precipitate: The DNA was recovered by lifting the aggregated precipitate from the solution using a disposable yellow tip. Excess liquid was dabbed off and the DNA was dispersed in a pre-labeled Eppendorf tube containing, depending on the size of the precipitate, 20 to 500 μl of TE Buffer. Complete dissolution of the DNA may require several hours of agitation at 37°C. It is important that the DNA is completely dissolved to ensure the reproducible removal of aliquots for analysis.

2.2.1.1.1.2. Plasmid DNA Extraction

Miniprep

Approximately 1.5 ml of a bacterial overnight culture was centrifuged at 11,000 rpm for approximately 1 minute. The supernatant was decanted and the pellet was resuspended with 300 μl Buffer P_1. 300 μl P_2 was immediately added and was the mix was left at room temperature for 5 minutes. Following the 5 minute incubation, 300μl P_3 was added and samples were centrifuged at 14,000 rpm for 10 minutes. 900 μl of the supernatant was taken and transferred to a fresh tube. Plasmid DNA was precipitated by adding 0.7 of volume isopropanol. Tubes were kept at -20°C for 10 minutes and were later centrifuged for 30 minutes at 4°C. The supernatant was discarded and the pellet was washed with 70% Ethanol by centrifuging at 14,000 rpm for 15 minutes. The supernatant was discarded and the pellet was dried at room temperature for 10 minutes. The pellet was resuspended with 20 μl of 1x TE buffer and kept at -20^{o}C for future use.

Maxiprep

The whole protocol has been adopted from QIAGEN EndoFree® Plasmid Purification Handbook (http://www1.qiagen.com/HB/CompactPrep_EN). A single bacterial colony was picked from a freshly streaked selective plate and a starter culture of 2–5 ml LB medium containing the appropriate selective antibiotic was inoculated. The starter culture was incubated overnight at 37°C with vigorous shaking (approx. 300 rpm). A tube or flask was used with a volume of at least 4 times the volume of the culture. The starter culture was diluted 1/500 to 1/1000 into selective LB medium. For high-copy plasmids, 100 ml medium was inoculated with 100–200 μl of starter culture. For low-copy plasmids, 250 ml medium was inoculated with 250–500 μl of starter culture. The culture was grown at 37°C for 12–16 h with vigorous shaking (approx. 300 rpm). A flask with a volume of at least 4 times the volume of the culture was used. The culture usually reaches a density of approximately 3–4 x 10^9 cells/ml, which typically corresponds to a pellet wet weight of approximately 3 g/liter medium. Bacterial cells were harvested by centrifugation at 6000 x *g* for 15 min at 4°C. If there was a need to stop the procedure, cells were frozen at -20^{o}C. The bacterial pellet was resuspended in 10 ml Buffer P1. For efficient lysis it is important to use a vessel that is large enough to allow complete mixing of the lysis buffers. The pellet was resuspended completely by vortexing or pipetting up and down until no cell clumps remained. Following Resuspension, 10 ml Buffer P2 was added and mixed thoroughly by vigorously inverting the sealed tube 4–6 times, and incubated at room temperature (15–25°C) for 5 min. During the incubation the QIAfilter Cartridge was prepared. 10 ml pre-chilled Buffer P3 was added to the lysate and mixed immediately and thoroughly by vigorously inverting 4–6 times. Precipitation is enhanced by using chilled Buffer P3. The lysate was poured into the barrel of the QIAfilter

Cartridge and incubated at room temperature (15–25°C) for 10 min. The cap was removed from the QIAfilter Cartridge outlet nozzle. The plunger was gently inserted into the QIAfilter Maxi Cartridge and the cell lysate was filtered into a 50 ml tube. Approximately 25 ml of the lysate is generally recovered after filtration. 2.5 ml Buffer ER was added to the filtered lysate and was mixed by inverting the tube approximately 10 times and incubating on ice for 30 min. QIAGEN-tip 500 was equilibrated by applying 10 ml Buffer QBT and the column was allowed to empty by gravity flow. The filtered lysate was applied to the QIAGEN-tip and allowed to enter the resin by gravity flow. The QIAGEN-tip was washed with 2 x 30 ml Buffer QC. Buffer QC was allowed to move through the QIAGEN-tip by gravity flow. DNA was eluted with 15 ml Buffer QN. DNA was precipitated by adding 10.5 ml (0.7 volumes) room-temperature isopropanol to the eluted DNA. The mix was centrifuged immediately at ≥15,000 x *g* for 30 min at 4°C. The supernatant was carefully decanted. DNA pellet was washed with 5 ml of endotoxin-free room-temperature 70% ethanol and centrifuged at ≥15,000 x *g* for 10 minutes. Again, the supernatant was carefully decanted without disturbing the pellet. The pellet was Air-dried for 5–10 min, and re-dissolved in a suitable volume of endotoxin-free Buffer TE.

2.2.1.1.1.3. Organic precipitation of DNA

This method is useful for precipitation of DNA from enzymatic reactions. 0.6 volume of isopropanol is added to the sample. After mixing the sample, 2 volumes of 95% cold ethanol were added and sample was placed at -80^{o}C for 30 minutes. Following the incubation, the sample was centrifuged at 12,000 rpm at 4^{o}C for 25 minutes. The supernatant was decanted and the pellet was drained by inverting the eppendorf tube. 2

volumes of 80% room temperature ethanol were added to the sample and incubated at room temperature for 10 minutes. The sample was centrifuged for at 12,000 rpm at 4°C for 7 minutes. The supernatant was decanted and pellet air dried for 5 – 10 minutes at room temperature and resuspended in suitable volume of TE buffer or water.

2.2.1.1.2. RNA Extraction

Cells Grown in Suspension were pelleted by centrifugation. Cells were lysed in TRIZOL® Reagent by repetitive pipetting. 1 ml of TRIZOL® reagent per 5-10 × 10^6 of cells was used. Washing cells before addition of TRIZOL® Reagent was avoided as this increases the possibility of mRNA degradation. Insoluble material was removed from the homogenate by centrifugation at 12,000 × *g* for 10 minutes at 2 to 8°C. The resulting pellet contains extracellular membranes, polysaccharides, and high molecular weight DNA, while the supernatant contains RNA. The cleared homogenate solution was transferred to a fresh tube. The homogenized samples were incubated for 5 minutes at 15 to 30°C to permit the complete dissociation of nucleoprotein complexes. 0.2 ml chloroform per 1 ml of TRIZOL® reagent was added. Tubes were shaken vigorously by hand for 15 seconds and incubated at 15 to 30°C for 2 to 3 minutes. The samples were centrifuged at no more than 12,000 × *g* for 15 minutes at 2 to 8°C. Following centrifugation, the mixture separates into a lower red, phenol-chloroform phase, an interphase, and a colorless upper aqueous phase. RNA remains exclusively in the aqueous phase. The volume of the aqueous phase is about 60% of the volume of TRIZOL® reagent used for homogenization. Transfer the aqueous phase to a fresh tube, and save the organic phase if isolation of DNA or protein is desired. RNA was precipitated from the aqueous phase by mixing with isopropanol. 0.5 ml of isopropanol per 1 ml of TRIZOL®

Reagent used for the initial homogenization was used. Samples were incubated at 15 to 30°C for 10 minutes and centrifuged at 12,000 × *g* for 10 minutes at 2 to 8°C. The RNA precipitate forms a gel-like pellet on the side and bottom of the tube. The supernatant was removed. The RNA pellet was washed once with 75% ethanol, adding at least 1 ml of 75% ethanol per 1 ml of TRIZOL® Reagent used for the initial homogenization. The sample was mixed by vortexing and centrifuged at 7,500 × *g* for 5 minutes at 2 to 8°C. The RNA pellet was air-dried for 5 minutes and was dissolved in RNase-free water by passing the solution a few times through a pipette tip, and incubating for 10 minutes at 55 to 60°C.

2.2.1.1.3. Protein Extraction

2.2.1.1.3.1. Freeze thaw method

The culture medium was removed from the cells, and the cells werer washed in ice-cold PBS. The sample was transferred to 1.5 ml microfuge tube. Cells were pelleted at top speed in a microfuge (45s in cold-room). The pellet was resuspended in FT LYSIS Buffer (25-40 μl per confluent 6 cm dish, 50-80 μl per confluent 10 cm dish). The tubes were dropped into liquid nitrogen for shock-freezeing. The samples were taken out of the liquid nitrogen (once they are frozen) and thawed on ice and were vortexed briefly. The freeze-thaw cycle was repeated two more times. Finally, 250 U of Benzonase was added to the samples to digest DNA and incubated at room temperature for 10 minutes. Protein concentration was measured by Bradford method. Equal volume of 2xLaemmli Buffer was added prior to analysis by SDSPAGE and heated at 95°C for 10 minutes. Samples were loaded on the gel immediately (Note that these samples must be loaded while still

warm, because the salt in the FT Buffer will cause the SDS in the Laemmli buffer to precipitate when cold).

2.2.1.1.3.2. RIPA lysis method

Cells were spun at 1000 rpm for 10 minutes at 4°C. The supernatant was removed and pellet was resuspended with cold PBS (approximately 1 ml per 10 cm dish). Cells were further centrifuged using the same settings and washed once more with cold PBS. The supernatant was removed and pellet was resuspended with 80 µl RIPA Buffer with protein inhibitors. The samples were gently mixed in a radial rotor at 15 rpm for 30 minutes at 4°C. Following the incubation, the samples were centrifuged for 10 minutes at 14,000 rpm at 4°C. Supernatant was collected in new tubes and the protein amount was quantified using the Bradford method.

2.2.2. DNA digestions and modifications

2.2.2.1. Restriction Endonuclease digestion

DNA restriction endonuclease digestion is usually carried out with a 2-fold to 10-fold excess of restriction enzyme in the total volume of 20 µl using 0.2-1.5 µg of DNA. A typical restriction enzyme digestion protocol is presented below. The components were added in the following order for a 20 µl reaction:

Table 2.13. Typical reaction mix setup for restriction digestion.

Component	Volume
Water, nuclease-free	16 µl
10X recommended buffer for restriction enzyme	2 µl
Substrate DNA	1 µl (~1 µg)

Restriction enzyme	0.5-1 μl (5-10 u)

Samples were mixed gently and spun down briefly. Incubation was carried out at the optimum temperature for 1-16 hours. The digestion reaction may be scaled up or down.

2.2.2.2. Dephosphorylation

The following protocol was adapted from Invitrogen with some modifications. It was used to dephosphorylate vector backbones digested with one restriction enzyme that were subsequently used in cloning experiments.

The mass of DNA required for 1 pmol 5end was determined by applying the following relations:

$$\text{pmol ends} = \text{pmol DNA} \times [(\text{number of cuts} \times 2) + 2]$$

1 μg of 1000 bp DNA = 3.04 pmol ends

Once the required amount of DNA was determined (amount of DNA having the required 1pmol of DNA ends), the following components were added to a microfuge tube:

4 μl of 10X dephosphorylation Buffer

1 pmol of DNA ends

Distilled water to 39 μl

The calf intestinal phosphatase (CIAP) was diluted in Dilution Buffer such that 1 μl contains the amount of enzyme required for the appropriate 5end (i.e. 1 unit for 5′ -recessed and blunt ends and 0.01 units for 5′ overhang). For 5′ -recessed and blunt-ended DNA, samples were incubated at 50°C for 60 minutes and samples with 5

overhang were incubated at 37°C for 30 minutes. Samples were later run on Agarose gel and extracted or precipitated by organic extraction.

Occasionally, a simplified protocol was also employed. This protocol allows for the dephosphorylation of DNA directly in restriction endonuclease buffer in the presence of the restriction endonuclease by adding 1 unit of CIAP and incubating samples at 50°C for 5 minutes blunt-ended DNA and 37°C for 5 minutes for DNA with 5′ overhang.

2.2.2.3. Phosphorylation

This protocol is written for 10 pmol of oligonucleotide. It can be scaled up or down accordingly. The oligonucleotide can be in any volume up to 25 μl of water or buffer, pH 7.6.

The reaction mixture was as follows:

Table 2.14. Phosphorylation reaction mix

Component	**Volume**
Oligonucleotide (10 pmol)	__ μl
ATP	__ μl
T4 Polynucleotide Kinase (PNK) buffer (10X)	5 μl
T4 Polynucleotide Kinase (PNK) (diluted)	2-5 units
Water	__ μl
Total	50 μl

The contents were mixed well and centrifuged briefly. Incubate was carried out at 37°C for 30 minutes and reaction was terminated by heating at 65°C for 10 minutes or adding EDTA to 5mM.

2.2.2.4. Blunting

This protocol allows the conversion of DNA fragments with 3′- and 5′-protruding ends to DNA fragments with blunt ends. Conversion of the two types of protruding ends can be accomplished simultaneously by the 3'->5' exonuclease and 5'->3' polymerase activities of the T4 DNA Polymerase.

The reaction mix was:

Table 2.15. Reaction mix for blunting procedure

Component	Volume
DNA fragment (with protruding ends)	> 0.1 pmol
10X Buffer	1 µl
Distilled sterile water	up to 9 µl
Total Volume	9 µl

Samples were incubated at 70°C for 5 minutes. 1 µl of T4 DNA Polymerase was added and mixed gently by pipetting. Incubation was carried out at 37°C for 5 minutes. For DNA samples with low G+C content, incubation was done 25°C instead of 37°C. DNA Dilution Buffer was added to bring the DNA concentration to 1 µg/50 µl and mixed thoroughly. Samples were later placed on ice to ensure deactivation of T4 DNA

Polymerase. If the solution was to be stored, DNA extraction was done immediately by organic extraction.

2.2.2.5. Ligation

Ligation experiments are usually carried out in 3:1 proportion of insert to vector backbone ratio. The required amount of insert is calculated by using the following formula:

$$I_a = \frac{V_a \times I_s}{V_s} \times \frac{3}{1}$$

Where I_a is the concentration of insert required, V_a is the amount of vector to be used, and I_s and V_s are insert size and vector size in base pairs, respectively. Reaction mix was:

Table 2.16. Ligation reaction setup

Component	Main reaction	Control 1 (no insert)	Control 2 (no vector)
Vector backbone	__ µl	__ µl	x
Insert	__ µl	x	__ µl
T4 DNA Ligase	1 µl	1 µl	1 µl
Ligation Buffer	2 µl	2 µl	2 µl
Nuclease Free Water	to 20 µl	to 20 µl	to 20 µl

For blunt-end ligation PEG (up to 15%) was added to increase the efficiency.

Incubation was carried out according to the guidelines below:

- Sticky end ligation reactions at room temperature for 1-2 hours (or at 4°C in an ice-water mix that slowly thaws to room temperature overnight).
- Blunt-end ligation reactions at 4°C in an ice-water mix that slowly thaws to room temperature overnight.

2.2.3. Transformation

2.2.3. 1. Preparation of electro-competent bacteria Cells

500 ml LB Medium was inoculated with a starter (a single bacterial colony that had been cultured in 3 ml LB medium overnight) culture and was mixed well. OD_{600} measurement was carried out every 45 min. Once the OD_{600} reading was in the range 0.40-0.60, the flask was taken out of the shaker and put in an ice-water bath for 15 min. The culture was then distributed in 8 (50 ml each) conical bottom Falcon tubes and cells were pelleted at 4500 x g for 5 min at 4°C. The supernatant was discarded and cells were resuspended in 40 ml ice cold water by pooling pellet from all tubes into one tube and subsequently centrifuged at 4500 x g for 5 min at 4 °C. Supernatant was again discarded and cells were resuspended in 40 ml ice cold water. Further centrifugation was carried out at 4500 x g for 5 min at 4 °C. After discarding the supernatant, cells were washed in 40 ml 10% glycerol (prepared my mixing with ice cold water) and centrifuged at 4500 x g for 5 min at 4 °C. Supernatant was discarded and the pellet was resuspended again in 40 ml 10% glycerol and centrifuged at 4500 x g for 5 min at 4 °C. Finally, 50 μl aliquots were prepared in 1.5 ml eppendorf tubes. The aliquots were immediately frozen in liquid nitrogen and later stored at -80 °C.

2.2.4. Electroporation

Electroporation is a procedure by which foreign DNA is introduced into cells by electrical pulse. The concise protocol is as follows. Frozen electro-competent bacterial cells were thawed on ice and mixed with the required amount of DNA, usually higher than 20 ng, and incubated further on ice for 1 minute. The mix was pipetted into electroporation cuvette and put on EasyjecT Prima electroporation device. The sample was pulsed with high voltage (1800-2500 volts) which lasts for few seconds. After a signal that the pulse was successful, sample was resuspended in 1 ml antibiotic free room temperature LB medium and incubated for approximately 45 minutes at 37^{o}C in an incubator with a shaker set to 220 rpm. Sample was later plated on an LB agar plate with the appropriate antibiotic selection.

2.2.5. Electrophoresis

2.2.5.1. DNA electrophoresis

DNA electrophoresis is a technique by which DNA fragments are separated by application of electric current by letting the fragments move through a matrix. The separation, as DNA is negatively charged, is based on size whereby larger fragments move slower and are more cathodal while smaller fragments move faster and are anodal. There are two buffer systems employed for standard DNA electrophoresis procedures: Tris Acetate EDTA (TAE) and Tris Borate EDTA (TBE). Tris-borate buffer (TBE) was advantageous for obtaining a higher resolution of smaller DNA fragments on agarose gels, when compared to the conventional tris-acetate buffer (TAE)(Yutaka MIURA, 1999). Critical DNA sizes and gel concentrations for a clear separation were about 2-kb for the 0.8% agarose and 300-bp for the 2.0% agarose. DNA fragments larger than the

critical size (>2-kb on 0.8% agarose gel) migrate faster in TAE, and the smaller fragments (<300-bp on 2% agarose gel) migrate faster in TBE showing the better resolution in a mini chamber system. Routine preparations were 0.8-1% agarose, which were prepared by mixing 400 – 500 mg agarose with 50 ml 0.5x or 1x TAE/TBE. 3 μl from 10 mg/ml Ethidium bromide was added after the temperature of the gel was below 55^{o}C. The gel was left to solidify at room temperature for 15 to 20 minutes. Samples to be loaded were mixed with Loading dye (1x final concentration) and loaded on the gel after submerging the gel in the same buffer used for preparing it. Runs usually took 45 to 90 minutes at 90 Volts, depending on the size of DNA size to be analyzed.

2.2.5.2. RNA electrophoresis

Gel preparation for whole RNA electrophoresis was 1.2% agarose in 0.5xTBE. Samples were mixed with loading dye (1x final concentration) and loaded on the gel. Electrophoresis was conducted at 90 Volts for 60 minutes.

2.2.5. Gel Purification

This protocol, adapted from QIAGEN® QIAquick Gel Extraction Kit Protocol, is designed to extract and purify DNA from standard or low-melt agarose gels in TAE or TBE buffer. Up to 400 mg agarose can be processed per spin column. This kit can also be used for DNA cleanup from enzymatic reactions.

DNA fragment was excised from the agarose gel with a clean, sharp scalpel. Gel slice was weighed in a colorless tube. 3 volumes of Buffer QG were added to 1 volume of gel (100 mg ~ 100 μl). The maximum amount of gel slice per QIAquick column is 400 mg. Sample was incubated at 50°C for 10 min. To help dissolve gel, sample was mixed by vortexing the tube every 2–3 min during the incubation. 1 gel volume of isopropanol was

added to the sample and mixed thoroughly. To bind DNA, the sample was applied to the QIAquick column, and centrifuged for 1 min at 13,000 rpm. Flow-through was discarded and QIAquick column was placed back in the same collection tube. 0.75 ml of Buffer PE was added to QIAquick column and centrifuged for 1 min at 13,000 rpm. The column was left to stand 2–5 min after addition of Buffer PE before centrifuging. Flow-through was again discarded and the QIAquick column was centrifuged for an additional 1 min at 17,900 x g (13,000 rpm). QIAquick column was placed into a clean 1.5 ml microcentrifuge tube. To elute DNA, 30 μl of Buffer EB (10 mM Tris·Cl, pH 8.5) or water (pH 7.0–8.5) was added to the center of the QIAquick column membrane and was left to stand for 1 minute before centrifuging it for 1 minute.

2.2.6. Cell Culture

The following protocols have been adapted, with some modifications, from Sigma Aldrich® cell culture handbook and SAFC® biosciences general cell culture techniques manual.

2.2.6.1. Serum Preparation

Animal serum is an excellent natural source of nutrients for cells in culture because it contains proteins, lipids, salts, vitamins, minerals, amino acids and other components necessary for growth. When stored and handled correctly, the performance characteristics of serum can be maintained for many years. Serum was stored frozen (-10 to -40 °C) and protected from light.

2.2.6.1.2. Thawing Procedure

The serum bottle was taken out from the freezer and allowed to acclimate to room temperature for approximately 10 minutes. The container was placed in a 30 - 37°C water

bath or incubator and swirled every 10 - 15 minutes until the serum was completely thawed.

2.2.6.1.3. Heat Inactivating Serum

Heat inactivation of serum, which degrades complement proteins by raising the temperature of the serum to 56 °C for 30 minutes, is a required step before the serum could be used. To achieve this, thawed serum bottles were put in a water bath and were periodically agitated. The serum bottles were removed after 30 minutes at 56 °C. The serum was aliquoted in 50 ml falcon tubes and was allowed to cool before using it on cells. Unused portions were refrozen for later use.

2.2.6.2. Media preparation

RPMI640 was used for culturing suspension cells, DG75. The medium was supplemented with 10% Calf Feotal Serum (CFS), 1% NEAA and 1% Pyruvic acid. Complete medium was stored at 4^{o}C until further use.

2.2.6.3. Growing cells

The following protocol has been adapted, with some modifications, from Sigma Aldrich® cell culture handbook.

DG-75 Cells are round to polygonal, single or clustered cells in suspension (pre-warmed 90% RPMI 1640 + 10% FBS). They are maintained at 0.1-0.5 x 106 cells/ml and subcultured 1:5 to 1:10 every 24-48 hours. The doubling time is 40-50 hours and saturation density is 0.75-1.0 x 106 cells/ml (DSMZ, 2008).

Suspension cells Flasks (T25, T75 and T175) were used to culture the cells. Cultures were viewed using an inverted phase contrast microscope. Cells growing in exponential

growth phase are bright, round and refractive. Occasionally, cells were very sticky and required a gentle knock to the flask to detach them. Overgrowth of cells was monitored by checking the change in pH of the medium, indicated by change in color of the phenol in the medium from red to yellow. For sub-culturing, the cells were allowed to settle to the bottom of the flask for 10 – 20 minutes. Once they have settled, the supernatant was removed by leaving a thin layer of cells at the bottom of the flask. Desired number of cells (after calculating cells/ml) was re-seeded into freshly prepared flasks without centrifugation just by diluting the cells. This was repeated every 2-3 days.

2.2.6.4. Counting cells

Trypan blue is the stain most commonly used to distinguish viable from nonviable cells. Viable cells exclude the dye, while nonviable cells absorb the dye and appear blue when viewed with a microscope. Cells should be in suspension as single cells in medium or a buffered saline before counting. Trypan blue has a higher affinity for serum protein than for cellular protein, so suspending cells in medium containing serum generates a dark background.

Trypan blue was first diluted with Phosphate Buffered Saline (PBS) to a working concentration of 0.02% - 0.04%. Cells were prepared in duplicate for counting in 1:100 dilutions in trypan blue solution. After being stained with trypan blue, the cells were counted within three minutes to avoid staining of viable cells as they will begin to take up the dye. A small amount of the stained cell suspension was placed onto the slot of a clean hemacytometer with coverslip. The cell suspension passed under the coverslip by capillary action. The opposite chamber was filled with the second diluted sample. The

hemacytometer was placed on the stage of an inverted microscope and the focus and power was adjusted until a single counting square fills the field. If the counts differed by more than 20%, a third sample was prepared to verify the count. The viable cell number was calculated using the formula:

$$\text{Viable Cell Number/ml} = \frac{\text{Number of viable cells Counted}}{\text{Number of Squares Counted}} \times 10^4 \times \text{Dilution Factor}$$

2.2.6.5. Freezing cells

Cryopreservation is an efficient method by which cells are stored in liquid nitrogen (-196^{o}C) without significantly losing their viability. The following procedure describes a quick method by which cells can be frozen.

Freezing medium was prepared by mixing 90% fetal calf serum and 10% dimethyl sulfoxide (DMSO). Cells were harvested in mid-logarithmic growth using standard procedures. The cell count was obtained by pelleting the cells by centrifugation. Cells were put in the freezing medium at a density of approximately 5×10^6 cells/ml. The cells were aliquoted into sterile vials or ampules. Cells were frozen in a rate-controlled freezer unit overnight. Cells were frozen in the vapor phase of liquid nitrogen.

2.2.6.6. Thawing cells

A vial of cells was removed from the vapor phase of liquid nitrogen. The vial contents were allowed to thaw by continuously swirling the vial in a 37°C water bath until only a small ice pellet remained. The vial was sprayed down with a 70% alcohol solution and was allowed to evaporate. 10 - 20 ml cold PBS was added to the cells and resuspended.

Cells were centrifuged at 1000 rpm for 10 minutes at 4°C. Supernatant was discarded and 10 ml fresh medium was added to the cells. Cells were cultured in an incubator at 37°C with 5% CO_2.

2.2.6.7. Kill Curve Analysis and stably transfected cells

Cells are often transfected with expression vectors containing antibiotic (e.g. puromycin or hygromycin) resistance markers that play key role in the enrichment of transfected cells by selectively killing non transfected ones and thereby contributing to the establishment of stably transfected cell lines. To meet such an end, one needs to determine the exact concentration of the selection agent required to achieve a predetermined level of selection. This is done by initially determining the LD_x (Latent Dose) at which a certain percentage of cells would be alive or dead after being treated by the agent. In the case of establishment of stable cell lines, it is important to determine LD at which there is more than 90% mortality.

2.2.6.7.1. Experimental setup

To predict the concentration of puromycin required to kill ~100% DG75 cells, a Completely Randomized Design (CRD) was used an experimental model. CRD is a type of statistical experimental design that is equivalent to a t-test when only two treatments are examined. In this model, replications of treatments (concentration of puromycin) are assigned completely at random to independent experimental subjects (cells). Adjacent subjects could potentially have the same treatment.

Cells were plated in triplicates in a 6 well plate at a density of 5×10^4 cells per well. Except for control group, four different concentrations of puromycin were used: 500 ng/μl, 750 ng/μl, 1000 ng/μl, 1250 ng/μl.

2.2.6.7.2. Probit Analysis

Data obtained was analyzed by SPSS® 15.0. Analysis was done by selecting Probit from the subcategories of Regression Analysis (RA) under the Analysis Toolbar. Probit is a specialized regression model of binomial response variables. The model tries to estimate a linear function that can approximate the sigmoid distribution.

2.2.7. Transfection

DG75 cells are normally hard to transfect with chemical tranfection agents. Method of choice is thus to use electric current to transfect the cells.

2.2.8. Electroporation

1×10^7 cells (in 50 ml flask) at mid log phase were used for the electroporation procedure. The cells were washed with 1× OPTIMEM. Cells were centrifuged at 500×g for 5 minutes. Supernatant was discarded and pellet was washed and mixed gently with 1× OPTIMEM (4 – 6 times). Cells were further centrifuged at 500×g for 5 minutes. After discarding the supernatant, cells were diluted in 350 μl OPTIMEM. 5 μg of pINCO NGFR and 15 μg pRTS-1 constructs were co-electroporated (230 Volts and 1mF). 5% FCS (from 15% pre-warmed FCS) was added immediately after electroporation.

2.2.8.1. Ficoll procedure

After culturing the cells at 37°C (5% CO_2) overnight, 30 ml of the supernatant out of 50 ml was discarded. Cells were resuspended with the remaining media. 20 ml Ficoll was

very gently applied to the bottom of the 50 ml conical bottom tube without disturbing the overlay and gently moving up along the border of the Ficoll/medium boundary. Cells were spun in a swinging bucket rotor at 2000 rpm for 30 minutes with deceleration rate of 1 rpm per second. Live cells form a ring at the interface between Ficoll and medium while dead cells settle to the bottom of the flask. The ring was pipetted very gently and mixed with 5 ml fresh medium and spun at 1500 rpm for 5 minutes in a swinging bucket rotor with deceleration rate of 1 rpm per second. Finally, the pellet was resuspended in 50 ml fresh medium.

2.2.8.2. Sorting NGF+ cells by MACS

There is an optimal NGF expression in the transfected cells 2-3 days after transfection. To enrich for transfected cells, Magnetic-Activated Cell Sorting targeting the NGF was used. Approximately 40 ml of the 50 ml old medium was discarded (flasks were kept in standing position). Cells were resuspended with the remaining medium (approx. 10 ml). 12.5 ml PBS plus MACS Buffer (2.5 g BSA/500 ml approx. 0.5%; 2 ml EDTA 0.5 M final concentration 2 mM) was added. Cells were spun at 500 × g for 5 minutes at 4°C (from here onwards, everything was carried out at 4°C). The following cocktails were prepared before the experiment started:

Table 2.18. MACS cocktails

Cocktail	Volume	Description
anti NGF antibody mix	800 µl	720 µl of MACS Buffer and 80 µl of anti NGF antibody
Goat anti mouse IgG microbeads	400 µl	360 µl Buffer and 40 µl beads

The supernatant was discarded and the pellet was resuspended in 400 μl **anti NGF antibody mix** and put at 4°C for 20 to 30 minutes. The sample was washed once with ice cold MACS buffer and kept at 4°C. It was later centrifuged at 500 × g for 5 minutes at 4°C. Supernatant was discarded and 200 μl of **Goat anti mouse IgG micro beads** was added and mixed. The sample was left on ice for 20 minutes. 15 ml of MACS Buffer was added after the incubation and centrifuged at 500 × g at 4°C for 5 minutes. Following centrifugation, 500 µl ice cold MACS buffer was added to soften the beads. In the mean time, MACS columns were equilibrated with 3 ml MACS buffer. The 500 μl micro beads solution was added to the pre-equilibrated column and was filled up three times with 3 ml MACS buffer. The column was removed from magnet stand and put on a fresh 15 ml conical bottom falcon tube. 7 ml of MACS buffer was added and the column was plunged with strong force 4 times and the content was collected in the fresh 15 ml tube. This was further centrifuged at 500 × g for 5 minutes at 4^{o}C. Finally, after discarding the supernatant, the pellet was solved in approximately 5 ml medium.

2.2.9. Polymerase Chain Reaction (PCR)

2.2.9.1. Traditional PCR

Polymerase Chain Reaction is widely held as one of the most important inventions of the 20th century in molecular biology. Small amounts of the genetic material can now be amplified to be able to a identify, manipulate DNA, detect infectious organisms, including the viruses that cause AIDS, hepatitis, tuberculosis, detect genetic variations, including mutations, in human genes and numerous other tasks.

To perform several parallel reactions, master mix was prepared. It contains water, buffer, dNTPs, primers and *Taq* DNA Polymerase in a single tube, which was then be aliquoted

into individual tubes. Template DNA was finally added to individual tubes. Master mix preparation is described below.

All solutions were gently vortexed and briefly centrifuged after thawing. In a thin-walled PCR tube put on ice, the following components were added:

Table 2.19. Traditional PCR reaction mix

Reagent	Final concentration	Quantity for 20 µl of reaction mixture
Sterile deionized water	-	variable
10X Taq thermopol buffer	1X	2 µl
10 mM dNTP mix	0.2 mM of each	0.4 µl
Primer Top (10 µM)	0.025 µM	0.5 µl
Primer Bottom (10 µM)	0.025 µM	0.5 µl
Taq DNA Polymerase	0.5 u / 20 µl	0.125 µl
Template DNA	10pg-1 µg	variable

The sample was gently vortexed and briefly centrifuged to collect all drops from walls of tube. Samples were placed in a thermo-cycler with the appropriate program and were run.

2.2.9.2. Reverse Transcriptase PCR (RT-PCR)

RT-PCR is a sensitive and reproducible method for the detection of RNA molecules. It is a two step process (see figure 1):

a. Reverse transcription, a process by which DNA copies are made from RNA template (**cDNA synthesis**)

b. PCR/**Amplification** aims at amplifying sequences of interest from the library using gene/sequence specific primers

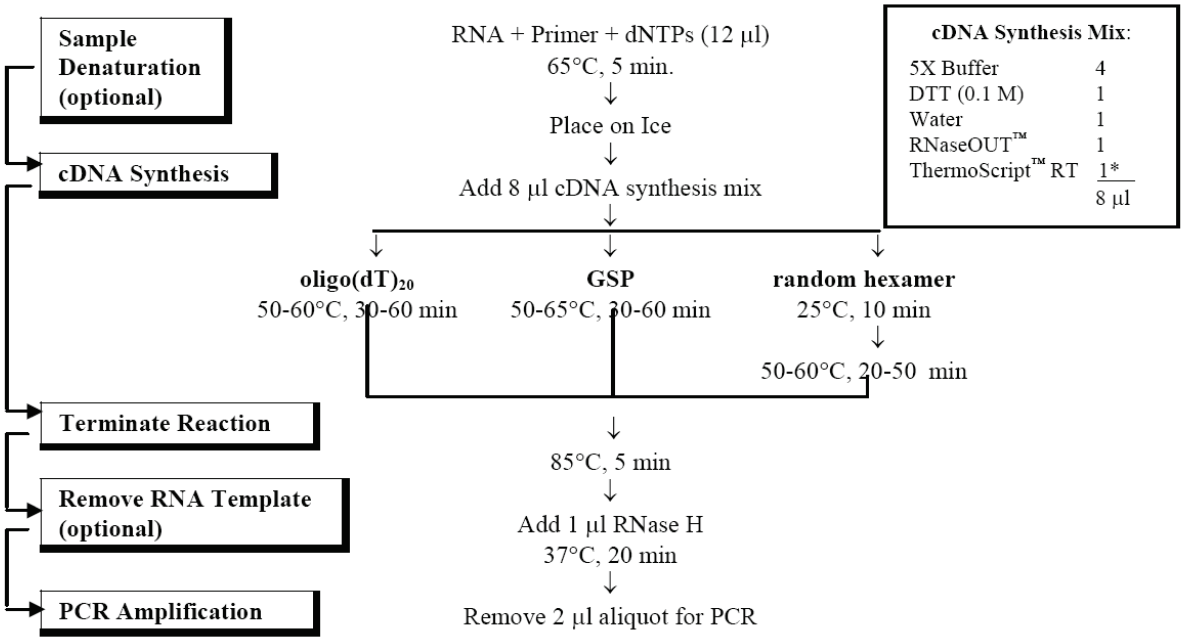


Figure 2.1. Schematic representation of the RT-PCR procedure (adapted from Invitrogen Inc.)

cDNA synthesis: The following components were mixed in 1.5 ml microfuge:

Table 2.20. RT-PCR first reaction mix (Primers, dNTPs and RNA)

Component	Amount
Primer (50 ng/µl)	1 µl
RNA (10 pg -5 µg)	x µl
10 mM dNTP Mix	2 µl
DEPC-treated water	to 12 µl

The mix was denatured by incubating at 65°C for 5 min and later placed on ice. 5X cDNA Synthesis Buffer was vortexed for 5 s just prior to use. Master reaction mix was prepared, as per the following composition, on ice and vortexed gently.

Table 2.21. RT-PCR first strand synthesis mix

Component	Amount
5x cDNA Synthesis Buffer	4 µl
0.1 M DTT	1 µl
RNaseOUT. (40 U/µl)	1 µl
DEPC-treated water	1 µl
ThermoScript. RT (15 units/µl)	1 µl

8 µl of master reaction mix was pipetted into each reaction tube on ice and transferred to a preheated thermal cycler with the following program: 25°C for 10 min, followed by 20-50 min at 50°C (or 50-65°C). Reaction was terminated by incubating at 85°C for 5 min. Finally, 1 µl of RNase H was added and sample was incubated at 37°C for 20 min. Sample was stored at -20°C or used for PCR immediately.

Amplification: PCR was carried out using Platinum® *Taq* DNA Polymerase High Fidelity. 10% of the cDNA synthesis reaction (2 µl) was used for PCR. The following components were mixed in a 0.2- or 0.5-ml, thin-walled, PCR tube:

Table 2.21. PCR mix for gene specific amplification from cDNA

Component	Amount
10X High Fidelity PCR Buffer	5 µl
50 mM $MgSO_4$	2 µl
10 mM dNTP Mix	1 µl
10 µM sense primer	1 µl
10 µM antisense primer	1 µl
Platinum® *Taq* High Fidelity	0.2 µl
cDNA (from cDNA synthesis reaction)	2 µl
DEPC-treated water	37.8 µl
Final volume	50 µl

Samples were mixed gently and put in a preheated thermocycler with the following settings:

Table 2.22. PCR program setup for detection of CALM/AF10 from cDNA preparation

Temperature	Time	Remark
94°C	2 min	Hold
94°C	30 seconds	
54.7°C (temp depends on primers)	30 seconds	30 cycles
68°C	1 min/kb extension	
68°C	7 minutes	Final Extension
4°C	∞	Hold

10 μl of the amplified sample was analyzed by agarose gel electrophoresis.

2.2.10. Colony PCR

This protocol is designed to quickly screen for plasmid inserts directly from *E. coli* colonies. This technique can be used to determine insert size and/or orientation in the vector. Colonies from successful ligation/transformation reaction were picked and cultured in LB medium with the appropriate selection agent overnight at 37^{o}C in a shaker set to 230 rpm.

The next day, the following ingredients were mixed on ice:

Table 2.23. Reaction mix for colony PCR

Component	Amount
sterile distilled water	38 μl
10X PCR buffer	5 μl
25 mM $MgCl_2$	3 μl
10 mM dNTPs (10 mM each dATP, dTTP, dGTP, dCTP)	1 μl
20 µM forward primer	1 μl
20 µM reverse primer	1 μl
Taq polymerase	0.2-1 μl
Total volume	50 μl

To each cold PCR tube containing the PCR reaction, 1 μl from the overnight culture was added and mixed thoroughly.

Table 2.24. Program setup for colony PCR

Temperature	Time	Remark
95°C	5 min	Lysis and denaturation
95°C	1 min	DNA denatures into single strands
54°C (temp depends on primers)	1.5 min	30-40 cycles
72°C	1 min (1 min/kb)	primers are extended from 3'-end by *Taq*
72°C	5 min	final extension
4°C	∞	Hold

2.2.11. Real Time PCR (Applied Biosystems 7900HT Micro Fluidic Card)

The Micro Fluidic Card allows you to use the 7900HT system for profiling gene expression using the Comparative C_T Method of relative quantification. The card evaluates from one to eight cDNA samples or controls generated from total RNA in a two-step RT-PCR. The following protocol describes, adopted with some modifications from Applied Biosystems 7900HT Micro Fluidic Card Getting Started Guide(Biosystems, 2003), the steps following post cDNA synthesis.

2.2.11.1. Amount of cDNA in the PCR Mix

Each 100 μl PCR mix, enough to fill one reservoir, contained 33.75 ng of total RNA converted to cDNA. The cDNA sample volume, with added water, was 50 μl. cDNA sample was thawed by rolling the tube between fingers. Sample was gently vortexed and centrifuged. The following components were added to a labeled 1.5-ml microcentrifuge tube:

Table 2.25. Sample reaction mix for Low Density Array real-time PCR

Component	Volume Per Fill Reservoir (μl)
cDNA	0.75
H_2O	49.25
TaqMan® Universal PCR Master Mix (2X)	50
Total	100

The micro-centrifuge tube was capped and thoroughly mixed by gentle vortexing. The tube was centrifuged to eliminate air bubbles from the mixture. The Micro Fluidic Card was loaded with samples as per the recommended guideline.

Table 2.26. List of different LDA cards available from Applied Biosystems®

TaqMan Low Density Array	Number of Targets and Controls	Number of Assay Replicates	Number of Samples per Card	Minimum Order Size	Part Number
Format 12	11 + 1	4	8	20	4342247
Format 16	15 + 1	3	8	20	4346798
Format 24	23 + 1	2 (4)	8 (4)	20	4342249
Format 32	31 + 1	3	4	20	4346799
Format 48	47 + 1	2 (4)	4 (2)	10	4342253
Format 64	63 + 1	3	2	20	4346800
Format 96a*	95 + 1	2 (4)	2 (1)	10	4342259
Format 96b	95 + 1	2 (4)	2 (1)	20	4342261
Format 384	380 + 1	1	1	50	4342265

*** Type of card used**

Applied Biosystems provides 9 formats of Low Density Array (LDA) cards. Format 96a was used in this experiment. 95 genes that had been found to be differentially regulated were selected for the custom setup. The table above gives details about the types of cards available.

Once samples were loaded, the fluidics card was centrifuged in a Heraeus centrifuge with bucket settings. The following table lists the operational parameters for Micro Fluidic Cards.

Table 2.27. Centrifugation generic setup for LDA cards

Parameter	QUIKSet	EASYSet
Up ramp rate	3	9
Down ramp rate	N/A	9
Rotational speed	1200 rpm	1200 rpm
Centrifugation time	2 × 1 min (0.01 on display)	2 × 1 min (0.01 on display)

Following centrifugation based on the above settings, the fluidics card was sealed and sample loader was trimmed.

2.2.11.1. Program setup and run

Sequence Detection Software version 2.1 (SDS 2.1) was used to run the fluidics cards and acquire data. Further data analysis was also performed using SDS 2.1. New template file was created using the options Relative Quantification ($\Delta\Delta C_t$) and 384 well cards.

After a new document template has been created, the default thermal cycling conditions for Micro Fluidics Cards were set (See table below).

Table 2.28. Applied Biosystems® 7900HT real-time PCR reaction instrument setup for LDA

AmpErase UNG Activation	AmpliTaq Gold DNA Polymerase Activation	Each of 40 Cycles	
		Melt	Anneal/Extend
50 °C	94.5 °C	97 °C	59.7 °C
2 min	10 min	30 sec	1 min
100% ramp	100% ramp	50% ramp	100% ramp

2.2.12. Microarray Analysis

This protocol is a summary of the extensive protocol provided by Affymetrix®(Affymetrix, 2005). Further details are found in the complete standard operation manual (Figure 2.2 summarizes the workflow). The procedure requires a minimum of 5 μg of purified total RNA as starting material.

2.2.12.1. First-Strand cDNA Synthesis

5 μg of total RNA was added to a 1.5-ml RNase-free tube and mixed with 1 μl of 100 pmol/μl T7-(dT)24 primer. Volume was adjusted to 12 μl /reaction with nuclease-free water. The mix was incubated at 65 °C for 10 minutes and centrifuged briefly at about 10,000 rpm for 10 sec and placed on ice. 4 µl of 5X first-strand cDNA buffer, 2 µl of 0.1 M dithiothreitol (DTT), and 1 µl of 10 mM dNTP mix were added and mixed well. The sample was incubated at 42 °C for 2 min. Following the addition of 1 μl of 200 U/μl SuperScript II RT, the mix was further incubated at 42 °C for 1 hr and finally centrifuged briefly at about 10,000 rpm for 10 sec and placed on ice.

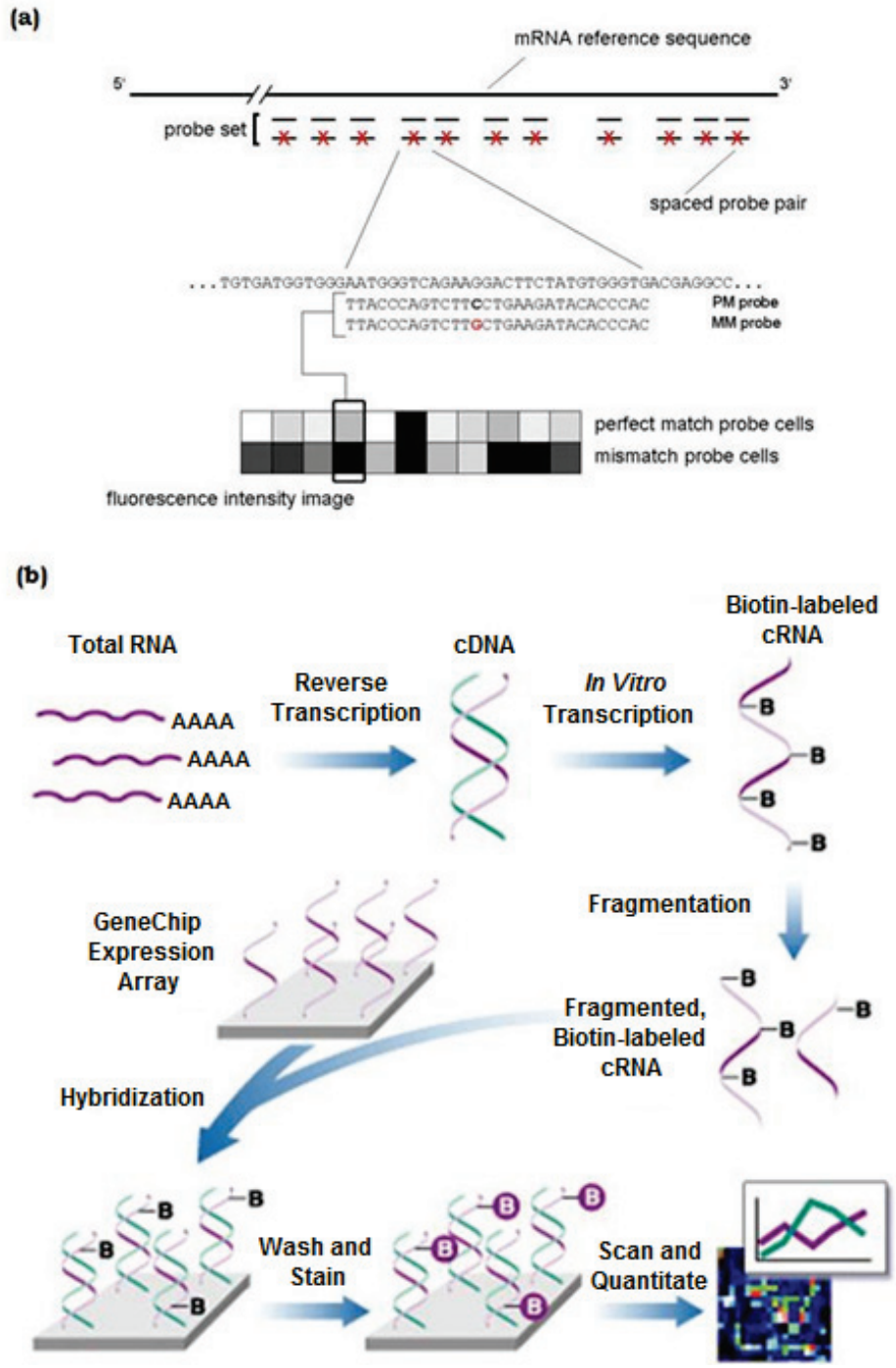


Figure 2.2. (a) A set of short oligonucleotides matching to different parts of an mRNA reference sequence is designed and is referred to as a probeset. A probeset contains pairs of oligonucleotides that either perfectly match (Perfect Match probe, PM) or contain a single mismatch (MisMatch probe, MM, serving as negative control) to the reference mRNA. (b) Affymetrix microarray workflow (both images adapted from http://www.dkfz.de/gpcf/24.html).

2.2.12.2. Second-Strand cDNA Synthesis

30 μl of 5x second-strand reaction buffer, 3 μl of 10 mM dNTP mix, 1 μl of 10 U/μl *E. coli* DNA ligase, 4 μl of 10 U/μl DNA polymerase I, 1 μl of 2 U/μl RNase H, and 91 μl of nuclease-free water were added to the first strand cDNA synthesis reaction and incubated at 16 °C for 2 hrs. 2 µl of 5 U/μl T4 DNA polymerase was added and the mix was further incubated at 16 °C for 5 min. 10 μl of 0.5 M EDTA was added prior to proceeding to cleanup and precipitation.

2.2.12.3. cDNA Cleanup and Precipitation

600 μl of cDNA binding buffer (from GeneChip Sample Cleanup Module) was added to the synthesized cDNA. The sample was applied to a cDNA cleanup spin column sitting in a 2-ml collection tube and centrifuged at 8000 x g for 1 min. The flow-through was discarded. The collection tube was replaced with another 2-ml collection tube. 750 μl of cDNA wash buffer was added to the spin column and centrifuged at 8000 x g for 1 min. The flow-through was again discarded.21. The spin column was centrifuged at maximum speed for 5 min with the column cap opened. The column was transferred to a fresh 1.5-ml collection tube and 14 μl cDNA elution buffer was pipetted directly onto the column membrane. Centrifuge at 8000 x g for 1 min to elute DNA.

2.2.12.4. Biotin-Labeled cRNA Synthesis

10 μl of sample cDNA, 4 μl of 10x HighYield (HY) reaction buffer, 4 μl of biotin-labeled ribonucleotides, 4 μl of DTT, 4 μl of RNase inhibitor mix, and 2 μl of 20x T7 RNA polymerase were mixed with nuclease-free water to provide a final reaction volume of 40 μl. The mix was incubated at 37^{o}C for 5 hr. For cleanup and quantification, the sample volume was adjusted to 100 μl with nuclease-free water and 350 μl of IVT cRNA

binding buffer was added and mixed well by pipetting. 250 μl of 100% ethanol was added and mixed well by pipetting gently 4 to 5 times. Sample was applied to an IVT cRNA cleanup spin column sitting in a 2-ml collection tube and centrifuged at 8000 x g for 15 sec. The column was transferred to a fresh collection tube and 500 μl IVT cRNA wash buffer was added. Further centrifugation was carried out at 8000 x g for 15 sec. The flow-through was discarded and the column returned to same collection tube. 500 μl of 80% ethanol was added and centrifuged at 8000 x g for 15 sec. The spin column was centrifuged at maximum speed for 5 min with the column cap opened to dry the column membrane. The column was transferred to a fresh 1.5-ml collection tube and 25 μl of nuclease-free water was pipetted directly onto the column membrane and centrifuged at 25,000 x g for 1 min to elute RNA. The RNA yield was determined by spectrophotometric analysis, applying the convention that 1 OD at 260 nm equals 40 μg/ml RNA.

2.2.12.5. cRNA Fragmentation

8 μl of 5x fragmentation buffer was added to 20 μg of cRNA and the total volume was brought to 40 μl with nuclease-free water. The mix was incubated at 94°C for 35 min and placed on ice.

2.2.12.6. Array Hybridization

10 μg of fragmented cRNA was mixed with 3.3 μl of 3 nM control oligonucleotide B2, 10 μl of 20X eukaryotic hybridization control, 2 μl of 10 mg/ml herring sperm DNA, 2 μl of 50 mg/ml acetylated BSA, 100 μl of 2X hybridization buffer, and nuclease-free water to a final volume of 200 μl. The probe array was equilibrated to room temperature before use. The hybridization mixture was incubated at 99°C for 5 min. In the meantime, the

probe array was filled with 160 μl of 1X hybridization buffer and incubated at 45°C for 15 min. The buffer was removed from the probe array and the array was filled with 200 μl of hybridization mixture. The filled array was placed in a rotisserie box heated to 45°C for 16 hrs.

2.2.12.7. Array Post-Hybridization Washing

All required buffers and staining reagents, including nonstringent wash buffer, stringent wash buffer, streptavidin-phycoerythrin (SAPE) stain solution, and antibody solution mix were prepared prior to washing procedure. The Fluidics Station was turned on. From the Microarray Suite, the *Fluidics* option was run. The Fluidics Station was primed by selecting *Protocol* → *Prime* from the drop-down menu.

The nonstringent wash buffer was placed in reservoir A while the stringent wash buffer was placed in reservoir B. The nuclease-free was poured into the water reservoir. *Run* was executed to begin priming. Experiment file for each probe array was created by clicking *Run* → *Experiments Info*. The hybridization mixture was transferred from the probe array to a tube. The probe array was fully loaded with 160 μl of nonstringent wash buffer. From the *Fluidics* window, the drop-down menu below *Experiment* was clicked to select the correct experiment name; the drop-down menu below *Protocol* was also clicked to select the appropriate protocol. Once the setup was done, *Run* command was executed to begin washing and staining. The instructions on the monitor were followed and the probe array was inserted when prompted. The microcentrifuge tube was removed from the sample holder of the Fluidics Station. A microcentrifuge tube containing 600 μl of SAPE stain solution was placed in the sample holder. Further instructions on the monitor were followed and when prompted, the microcentrifuge with the staining

solution was replaced with a microcentrifuge tube containing 600 μl of antibody solution mix. Furthermore, another microcentrifuge tube containing 600 μl of SAPE stain solution was put in place after being prompted for it. The Fluidics Station was allowed to continue until the *EJECT CARTRIDGE* message appeared on the monitor, indicating the completion of the protocol. The probe array was removed from the GeneChip Fluidics Station for scanning protocol.

III. Results

3.1. Cloning experiments

3.1.1. Cloning of FLAG-CALM/AF10 into the pUC19 Shuttle Vector

The CALM/AF10 fusion gene that was used for this work was a FLAG-CALM/AF10 open-reading frame (ORF) that was cloned in the pEYFP vector using *Xba*I and *Xho*I (Figure 3.1b). pEYFP vector doesn't contain the appropriate restriction site (*Sfi*I site) required to clone CALM/AF10 into the desired conditional mammalian expression vector pRTS-1 (Figure 3.3a and 3.3b). To facilitate the cloning procedure, FLAG-CALM/AF10 was first cloned into a shuttle vector that contains *Sfi*I cassette (pUC19 vector).

The *Xba*I and *Xho*I sites were used to clone FLAG-CALM/AF10 ORF into the plasmid pUC19 *Sfi*I Shuttle Vector (Figure 1a). CALM/AF10 was cut out from the pEYFP vector backbone by using *Xho*I, *Xba*I, and *Nsi*I (Figure 3.1c, lanes 7 to 10). The pUC19 Shuttle Vector was digested with *Xho*I and *Xba*I (Figure 3.1c, lanes 1 to 5).

3.1.2. Cloning of VP16 upstream of FLAG-CALM/AF10

The VP16 transactivating domain, a strong transactivating domain which is able to transactivate almost all promoter types, was cloned up-stream and in-frame to the CALM/AF10 fusion protein in order to convert CALM/AF10 from a transcriptional repressor into a strong transcriptional activator of these promoters thus leading to a clear detection of genes, in the microarray analysis, that are normally down-regulated by CALM/AF10. It should be noted that this is an additional clone. The transactivation domain of VP16 (tVP16) cloning was carried out by first amplifying the tVP16 from the pVP-FLAG vector by adding *Not*I and *Xho*I sites to the forward and reverse primers,

respectively. These restriction sites were later used for cloning the tVP16 into the pUC19-FLAG-CALM/AF10 construct (Fig 3.1a and 3.1b).

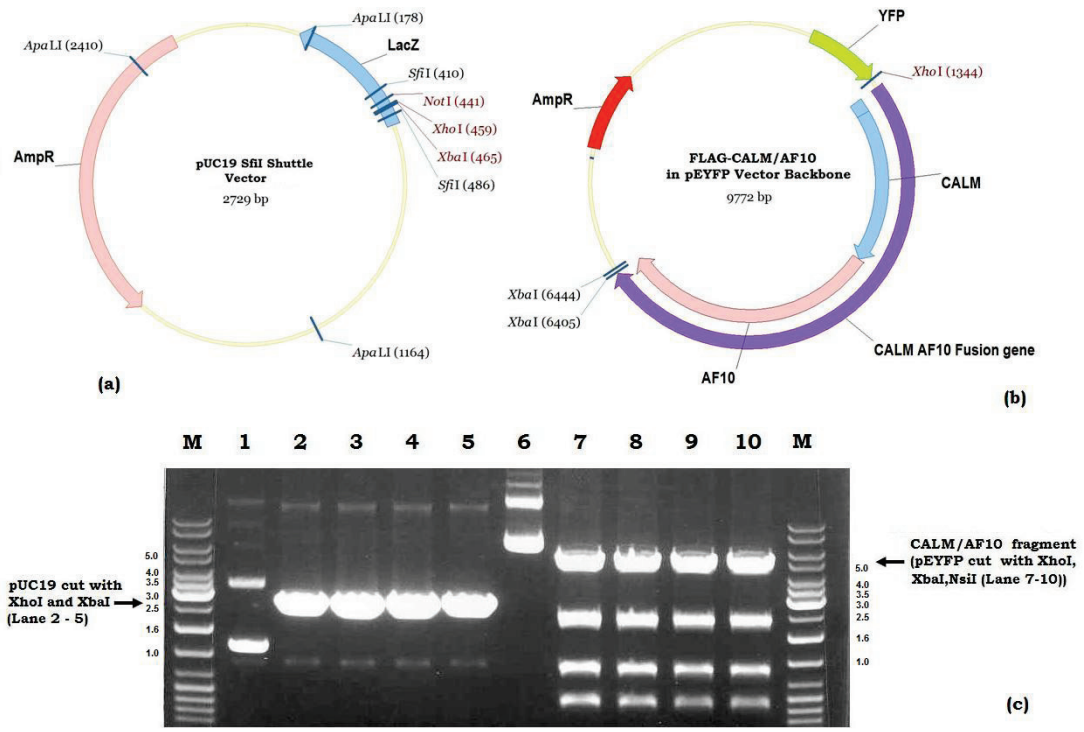


Figure 3.1. Cloning of FLAG-CALM/AF10 into the pUC19 shuttle vector. (a) the pUC19 shuttle vector (b) pEYFP-FLAG-CALM/AF10 (c) Agarose gel showing the fragments produced by the restriction digestion, arrows indicating desired fragments. M is the DNA ladder used and fragment sizes are given on the left and right hand side of the gel (in kbp).

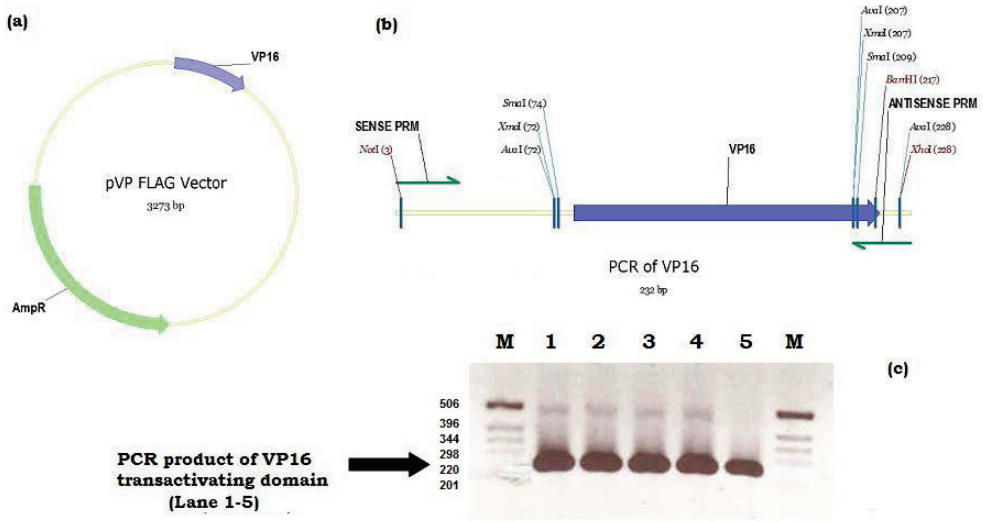


Figure 3.2. Schematic diagram of the pVP-FLAG vector containing the VP16 transactivating domain (a). Diagram (b) and gel (c) showing the PCR product and size. M is the DNA ladder used and fragment sizes are given on the left hand side of the gel (in bp).

3.1.3. Cloning of FLAG-CALM/AF10 and FLAG-VP16-CALM/AF10 into pRTS-1

pRTS-1 is an EBV-derived episomally replicating plasmid that carries all the elements for conditional expression of a gene of interest via Tet regulation. The vector is characterized by

(i) low background activity,

(ii) high inducibility in the presence of doxycycline (Dox) and

(iii) graded response to increasing concentrations of the inducer (Bornkamm *et al.*, 2005).

The gene of interest is expressed from the bidirectional promoter P_{tet}bi-1 that allows simultaneous expression of two genes, of which one may be used as surrogate marker for the expression of the gene of interest. Tight down regulation is achieved through binding of the silencer tTSKRAB to P_{tet}bi-1 in the absence of the inducer (e.g. Doxycycline, Dox). Addition of Dox releases repression and via binding of rtTA2^{S}-M2 activates the promoter P_{tet}bi-1. pRTS-1 has the GFP gene on one side of the promoter and the firefly luciferase gene on the other. Two *Sfi*I sites flank the luciferase gene.

VP16-FLAG-CALM/AF10 and FLAG-CALM/AF10 (without VP16) were cloned into pRTS-1 using *Sfi*I restriction sites (Figure 3.3c). *Sfi*I has a bipartite recognition sequence (two 4 nucleotide sequences) flanking a core of 5 not-defined nucleotides. The enzyme creates a 3 nucleotide long′3overhang in this non -defined internal pentanucleodtide. DNA fragments cut with *Sfi*I will only religate if the 3 nucleotides of this 3′ overhang are complementary, which is the case in for the *Sfi*I sites in the pUC19-SfiI shuttle vector and pRTS-1. This allows the easy directional cloning of fragments from the pUC19-SfiI shuttle into the pRTS-1 vector.

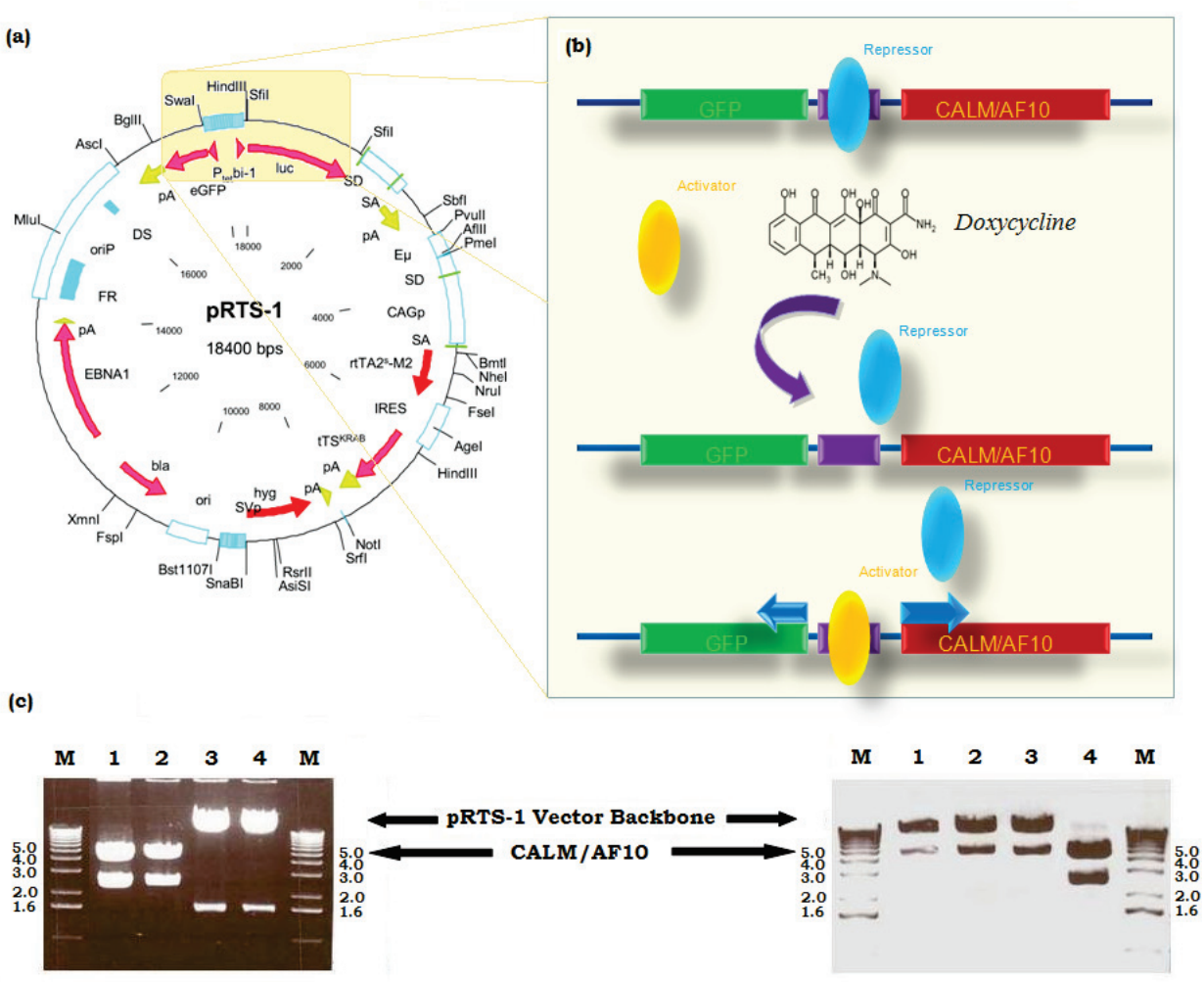


Figure 3.3. ***(a)*** *Schematic diagram of pRTS-1 expression vector,* ***(b)*** *illustration showing how the bidirectional promoter works,* ***(c)*** *Agrose gels showing the pRTS-1 backbone (upper bands) and Luciferase (lower bands) after* SfiI *digestion (Left gel, lanes 3 and 4). Lanes 1 and 2 in the same gel show* SfiI *digested FLAG-CALM/AF10 and VP16-FLAG-CALM/AF10 (upper bands), respectively, the lower bands are the vector backbone of pUC19 Shuttle Vector. Right gel (lanes 1 to 3) shows* SfiI *digestion upon successful cloning of CALM/AF10 (lower bands) into the pRTS-1 vector (upper bands). M is the DNA ladder used and fragment sizes are given on the left and right hand sides of the gels (in kbp).*

3.2. Establishing a stably transfected cell line

There are many different approaches for establishing stable cell lines, depending on the type of expression one is interested in (inducible vs. constitutive) and the construct to be used. The desired end result is a population of cells in which > 95% of the cells are

expressing the protein of interest at approximately the same level. This allows large-scale biochemical analyses of the protein and monitoring of localization throughout the cell cycle, and greatly simplifies many other microscopy techniques as well (Lamond, 2008). DG75 is a human Burkitt lymphoma cell line that was originally established from the pleural effusion of a 10-year-old boy with Burkitt lymphoma in 1975 (Ben-Bassat *et al.*, 1977). It has a near diploid karyotype with a t(8;14) with 6% polyploidy - 46XY, t(8;14)(q24;q32).

3.2.1. Electroporation/transfection

Electroporation was carried out as per the protocol described in the materials and methods section. DG75 cells were transfected with the following vectors and constructs:

Table 3.3. Type of constructs/vectors and amount of DNA used for the electroporation procedure

Construct/Vector	Amount used per 1×10^7 cells	Remark
pRTS-1-FLAG-CALM/AF10	15 μg	Co-transfected with 5 μg of NGFR* expressing vector (pINCO)
pRTS-1-VP16-FLAG-CALM/AF10	15 μg	Co-transfected with 5 μg of NGFR expressing vector (pINCO)
pRTS-1 vector backbone	15 μg	Co-transfected with 5 μg of NGFR expressing vector (pINCO)

* Nerve Growth Factor Receptor

The electroporation resulted in about 40% of cell death. After collecting the live cells using Ficoll and culturing them for 48 hours, Magnet Assisted Cell Sorting (MACS) was carried out using antibodies against the NGFR, a membrane protein that is expressed by

the co-electroporated expression vector pINCO. It takes approximately 48 hours for sufficient NGFR to be present on the membrane of the cells for MACS.

Cells were cultured for an additional 48 hours and were then induced with doxycycline (1 μg/ml) and then the percentage of GFP positive cells was analyzed by flow cytometry. The following table shows the proportion of GFP positive cells for the different transfections:

Table. 3.2. Percentage of GFP positive cells after transfection.

Construct/Vector	% GFP Positive	Remark
pRTS-1-FLAG-CALM/AF10	20	Result after NGFR sorting
pRTS-1-VP16-FLAG-CALM/AF10	18	Result after NGFR sorting
pRTS-1 vector backbone	17	Result after NGFR sorting

3.2.2. Selection/enrichment of transfected cells

Puromycin has been used as a selection agent to enrich transfected DG-75 cells. pRTS-1 contains a gene that encodes puromycin resistance enzyme which is constitutively expressed independent of the inducible bi-directional promoter that controls the GFP and CALM/AF10. As a result, the enrichment of transfected cells can be performed without inducing the cells.

Puromycin dihydrochloride is an aminonuclease antibiotic that inhibits protein synthesis. Puromycin is used for selection and maintenance of cell lines expressing a transfected *Pac* gene, whose product, puromycin acetyltransferase, inactivates puromycin via acetylation. The working puromycin concentration for mammalian cell lines ranges from 1-10 µg/ml. The selection agent needs to be titrated prior to use to determine the optimal

concentration for the target cell line. Normally, the lowest concentration that kills 100% of non-transfected cells in 3-5 days from the start of puromycin selection at an average cell density of 6×10^4 cells/ml is used.

Depending on individual cell type and doubling rate, the selection of stable transfectants will take between 7 and 28 days. The expansion and characterization of single cell clones will take several weeks in addition. The media should be changed every 2-3 days. This eliminates potentially toxic substances produced by dying cells and it keeps the concentration of the antibiotic at a constant level.

To calculate the concentration of puromycin to use for the enrichment of transfected DG-75 cells, non-transfected DG-75 cells were seeded on a six well plate and each well was treated with one specific concentration of puromycin by following Completely Randomized Design (CRD) strategy. The position of replicates (well number) and the concentration of puromycin (treatment) applied was randomized in each plate.

The Mortality rate was estimated 5 days after treatment by counting dead/live cells after staining for Trypan blue. Appendix A1 details the result obtained, adjusting the natural mortality rate observed in control cells to 0 percent mortality (100% survival).

Based on these results (see Appendix A5), we used a concentration of 1.24 μg/ml puromycin to achieve a close to 100% kill of non transfected cells within 5 days at a cell density of 6×10^4 cells/ml.

3.2.3. Conditional expression

Once the required dose of puromycin had been determined, the cells were put under selection at a density of 6×10^4 cells/ml and monitored every week for selection

efficiency. In the first 10 days, a strong enrichment of transfected cells was observed. Table 3.3 shows results obtained after 10 days selection. Cell samples were analyzed by flow cytometry 24 hours after GFP expression had been induced with doxycycline (1 μg/ml).

Table 3.3. Percentage of GFP positive cells after 10 days of puromycin selection

Construct/Vector	% GFP Positive	Remark
pRTS-1-FLAG-CALM/AF10	60	Result after 10 days puromycin selection
pRTS-1-VP16-FLAG-CALM/AF10	65	Result after 10 days puromycin selection
pRTS-1 vector backbone	57	Result after 10 days puromycin selection

Fluorescent microscopic observation was also carried out for the expression of GFP and visual assessment of induction. Figure 3.4 shows these observations for both the FITC (channel showing only GFP expressing cells) and phase contrast (all cells) modes.

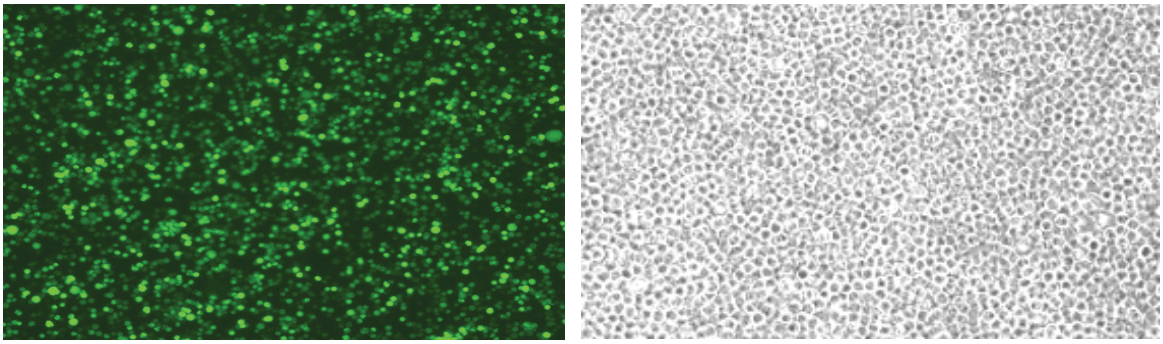


Figure 3.4. GFP expression after induction by Doxycycline. DG-75 cells expressing FLAG-CALM/AF10 (from the pRTS-1-FLAG-CALM/AF10 construct) (24 hours after induction, 1 µg/ml doxycycline). Left image: FITC, 20 X magnifications; Right image: No fluorescence filter, 20 X magnifications.

3.3. RT PCR

Once the desired percentage of GFP positive cells was obtained (> 90%, after 4 weeks of puromycin selection), the expression of CALM/AF10 was confirmed by RT-PCR.

RNA extracted from induced DG75 cells was used as a starting material. 3μg of total RNA was treated with DNaseI to remove DNA contamination. Approximately 900 ng of DNaseI treated total RNA was used for cDNA synthesis using random primers and the protocol provided by Invitrogen (Materials and Methods section 2.2.10.2) for the Thermoscript RT-PCR kit.

A CALM/AF10 specific primer set was used to confirm the expression of CALM/AF10 after induction. The CALM/AF10 primers used were: Forward ′5-GCAATCTTGGCATCGGAAAT-3′ (Forward) and 5′-GCCTGTCGACATCCATGTTT-3′ (Reverse) (a schematic representation of the binding site of the primers in CALM/AF10 and the amplified fragment is shown in Figure 3.5a). The primers were designed to amplify across the breakpoint to ensure that the primers are amplifying the

fusion gene and not the endogenous forms of either CALM or AF10. The RT-PCR amplified the expected products (Figure 3.5b).

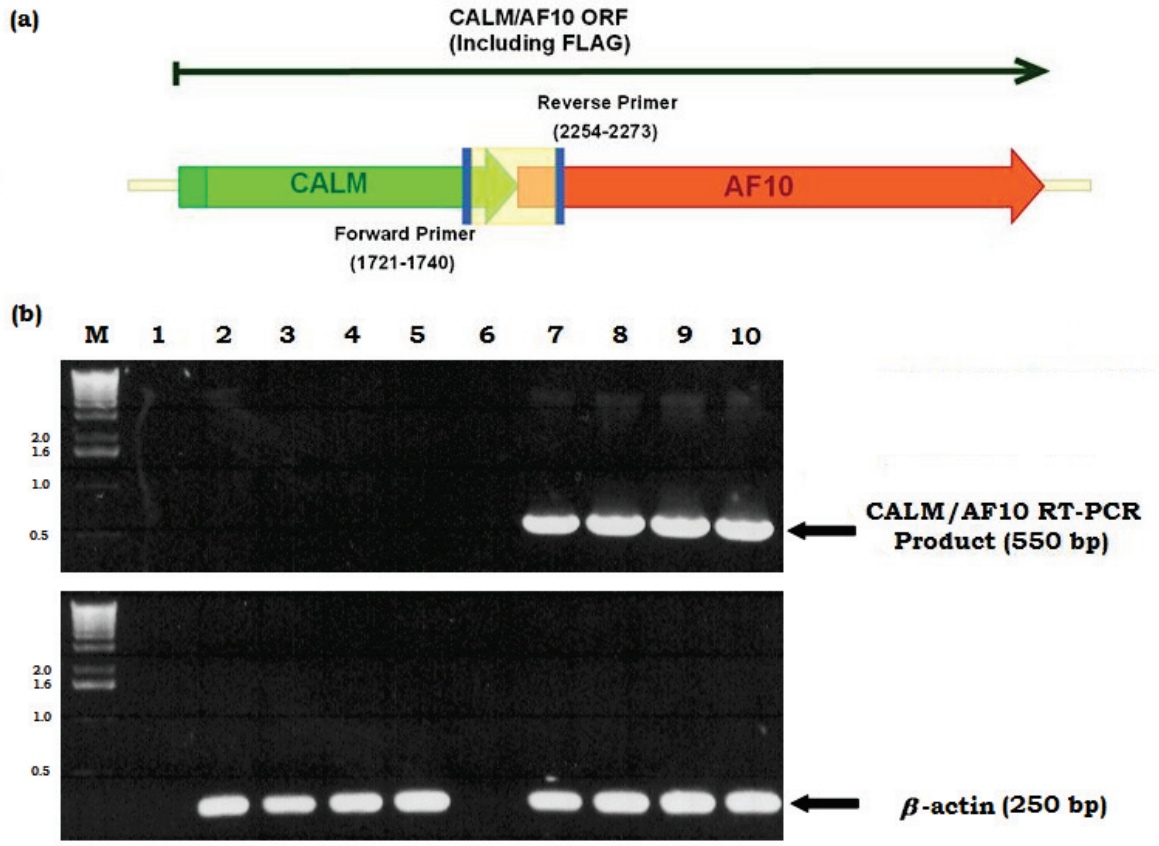


Figure 3.5. ***(a)*** *Schematic representation of the binding sites of the primers used and the part of CALM/AF10 that they amplify (yellow-shaded area)* ***(b)*** *RT-PCR result. Lanes 1 and 6 are negative control (H_2O); Lane 2 = pRTS-1 vector only control (non-induced); Lane 3 = pRTS-1 vector only control (24 hours after induction); Lane 4 = pRTS-1-FLAG-CALM/AF10 (non-induced); Lane 5 = pRTS-1-VP16-FLAG-CALM/AF10 (non induced); Lanes 7 and 8 are pRTS-1-FLAG-CALM/AF10 (24 hours after induction); Lanes 9 and 10 are pRTS-1-VP16-FLAG-CALM/AF10 (24 hours after induction). M is the DNA ladder used and size markers are indicated by numbers on the left side of the gel (in kbp).*

3.4. Microarray

3.4.1. Experimental Setup

RNA samples, that were obtained from induced and non-induced cells and were confirmed for the expression of CALM/AF10 by RT-PCR, were used as a starting material for the microarray experiment.

Each sample obtained was applied to one HGU133 Plus 2.0 Affymetrix gene chip array by employing the protocol detailed in the materials and methods section. Scanned array images were finally stored as .CEL files.

3.4.2. Low-level Analysis of Microarray Data

3.4.2.1. Preprocessing and Normalization

Since scanned images may have different overall brightness, generally the normalization is needed to adjust the brightness of the arrays to comparable level. Currently, there are at least 6 popular algorithms that can normalize microarray data using different mathematical approaches (Millenaar *et al.*, 2006). Recently, more methods have become available for calculating gene expression values, such as Model-Based Expression Indexes (MBEI) from Li and Wong (Li and Wong, 2003) that are implemented in the freeware program dChip. Also available is the Robust Multiarray Average (RMA) software from Irizarry *et al*. (Irizarry *et al.*, 2003). In contrast to MAS 5.0 or GCOS, in which information from only one microarray is used, model-based algorithms incorporate information from multiple microarrays to calculate the expression of a gene. The probe response pattern is fitted over multiple arrays with a multiplicative model in dChip and an additive model in the RMA software. These algorithms use the fitted models to detect abnormally behaving probes, which are subsequently excluded for calculating gene

expression. Therefore, gene expression from these model-based algorithms can be expected to provide more reproducible results. Both dChip and RMA use a stochastic model to estimate gene expression.

3.4.2.1. Median Probe Intensity (MPI) and background Adjustment

dChip (Li and Wong, 2003), a freely available software (http://biosun1.harvard.edu/complab/dchip/), was used to perform the normalization of the microarray data obtained. dChip normalizes arrays at the Perfect Match (probes matching perfectly to a certain part of the reference mRNA, PM) and Mismatch (probes containing a single mismatch, MM) (see figure 2.2a) probe level before computing model-based expression levels. Model-based method assumes that the arrays are already at comparable brightness (Li and Wong, 2003). However, there is also a danger of introducing artifacts if normalization is not done reasonably. To avoid the introduction of artifacts, one can examine the normalization curve and, if the amplification and scanning steps are controlled well and if the median intensities of arrays are close (± 5 Median Probe Intensity (MPI) from the median array), one may skip the normalization step and perform a model-based expression analysis directly (Li and Wong, 2003). It is very critical that a visual inspection of the array images is performed before further analysis is carried out to rule out artifacts.

Table 3.4 summarizes the names of the arrays and the median intensities obtained for the arrays in this study. After the calculation of the Median Probe Intensities (MPI), it was observed that the array with median intensity (taking all the nine arrays together) was pRTS-1-FLAG-CALM/AF10 24hrs (MPI = 80). This array was then used as a baseline for the normalization of the remaining 8 arrays, a process by which the brightness of each

array is adjusted to approximate the MPI of the baseline array. The adjusted MPI values are given in Table 3.4.

Table 3.4. Arrays and their Median Probe Intensity (MPI)

Name of Array	Median Probe Intensity (Brightness)		Remark
	Before Adjustment	After Adjustment	
pRTS-1 0hrs	78	80	Sample from DG-75 transfected with the empty pRTS-1 vector (non-induced)
pRTS-1 24hrs	78	81	Sample from DG-75 transfected with the empty pRTS-1 vector (24 hours after induction)
pRTS-1 72hrs	98	80	Sample from DG-75 transfected with the empty pRTS-1 vector (72 hours after induction)
pRTS-1-VP16-FLAG-CALM/AF10 0hrs	95	80	Sample from DG-75 transfected with the VP16-FLAG-CALM/AF10-pRTS-1 construct (non-induced)
pRTS-1-VP16-FLAG-CALM/AF10 24 hrs	70	80	Sample from DG-75 transfected with the VP16-FLAG-CALM/AF10-pRTS-1 construct (24 hours after induction)
pRTS-1-VP16-FLAG-CALM/AF10 72hrs	85	80	Sample from DG-75 transfected with the VP16-FLAG-CALM/AF10-pRTS-1 construct (72 hours after induction)
pRTS-1- FLAG-CALM/AF10 0hrs	82	81	Sample from DG-75 transfected with the FLAG-CALM/AF10-pRTS-1 construct (non-induced)

pRTS-1- FLAG-CALM/AF10 24hrs	80	80	Sample from DG-75 transfected with the FLAG-CALM/AF10-pRTS-1 construct (24 hours after induction) **NB:** median Brightness across arrays
pRTS-1- FLAG-CALM/AF10 72hrs	77	79	Sample from DG-75 transfected with the FLAG-CALM/AF10-pRTS-1 construct (24 hours after induction)

3.4.2.2. The Invariant Set Normalization

The Invariant Set Normalization method chooses a subset of PM probes with small within-subset rank difference in the two arrays, to serve as the basis for fitting a normalization curve (Li and Wong, 2001). The image of outliers (array and single outliers) identified through model-fitting can be used to assess the quality of an experiment and to identify unexpected problems such as misaligned corner of a DAT file (Li and Wong, 2001).

The fitted curve is the running median curve in the scatterplot of probe intensities of two arrays (with the baseline array on the Y-axis and the array to be normalized on the X-axis). When fitting the running median curve at the two ends, 5% of the "invariant" points are used to fit a ray at one end fixed; this makes the high-end normalization relationship more smooth and robust. The final running median curve is a piece-wise linear curve (Li and Wong, 2001). Then the normalization transformation is performed for all the points (probes) in the array on the X-axis (the Y-axis is the baseline array and is not changed). To obtain the normalized value of a point which has a particular intensity value on the X-axis, we drop an imagined vertical line passing this point, and use the Y-

axis value of the intersection point of this line and the fitted curve as the normalized value. The normalized probe values beyond the range [0, 65535] are truncated at the boundaries. The total effect of the normalization is a rotation and straightening of the scatter plot, so it better centers around the diagonal line y = x (Li and Wong, 2001). Figure 3.6 gives an example of an invariant set normalization carried out on pRTS-1 0hrs by using the baseline array pRTS-1-FLAG-CALM/AF10 24hrs.

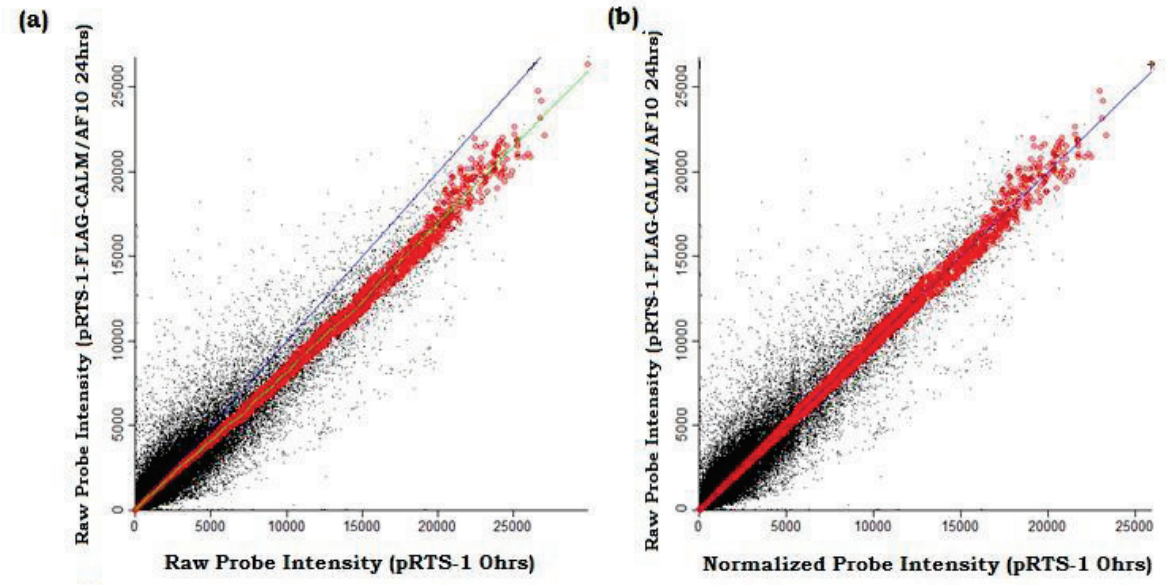


Figure 3.6. Example of Invariant set normalization (pRTS-1 0hrs versus pRTS-1-FLAG-CALM/AF10 24hrs). Y axes represent the base line array (pRTS-1-FLAG-CALM/AF10 24hrs) and X axes represent the array to be normalized (pRTS-1 0hrs in this case). Each black point indicates a probe intensity value. ***(a)*** *Scatterplot of raw intensities as read by the Affymetrix array scanner. The invariant set (5 % of the points that are used to fit a ray that is fixed at one end) are indicated in red circles. Blue indicated fitted function y=x (y equals to the raw intensity of the baseline array, pRTS-1-FLAG-CALM/AF10 24hrs). The green line represents the line that best describes the correlation of raw probe intensities of the base line array (pRTS-1-FLAG-CALM/AF10 24hrs)and the array to be normalized(pRTS-1 0hrs) (correlation coefficient r = 0.957).* ***(b)*** *Normalized probe intensities. The probe intensities of the array to be normalized (pRTS-1 0hrs) are rotated to fit into the function y=x (by rotating the green line to fit the blue line) without changing the values of the base line (pRTS-1-FLAG-CALM/AF10 24hrs) probe intensities.*

All the other arrays were normalized in a similar fashion by taking the pRTS-1-FLAG-CALM/AF10 24hrs array as a base line array.

3.4.3. High-level analysis

3.4.3.1. Sample comparison (t-test)

Given two samples or two groups of samples, we want to identify genes that are reliably differentially expressed between the two groups. This is done for every gene in the array and is based on calculating a t-statistic. The t-statistic tries to account for the level of difference in

expression of a given probe set in two different samples and determines if this difference is statistically significant (Li and Wong, 2003).

Based on the t-test comparison of the different samples, various comparisons were carried out to find genes that are significantly differentially regulated at a confidence interval of 95%. The comparison was done by comparing the expression level of a probeset of a given gene in the experimental setting versus the expression level of the same probeset in the control setting. For instance, to find the genes that were at least 2 fold significantly differentially regulated (up or down) in the sample pRTS-1-FLAG-CALM/AF10 24 hrs (genes that were deregulated 24 hours after CALM/AF10 expression), the pRTS-1-FLAG-CALM/AF10 24 hrs array was compared to the pRTS-1-FLAG-CALM/AF10 0hrs (no CALM/AF10 expression) control. This list of differentially regulated genes was further subjected to other controls; genes that were also found to be 2 fold differentially regulated in the following comparisons were excluded from the final list,

a. pRTS-1-FLAG-CALM/AF10 0hrs to pRTS-1 0hrs (to exclude genes that were already 2 fold up or down-regulated without induction)
b. pRTS-1 24 hrs to pRTS-1 0hrs (to exclude genes that were deregulated because of confounding factors like GFP expression or addition of doxycycline)

Table 3.5 gives the number of genes that were found to be at least 2 fold significantly differentially regulated (up or down regulated) for different comparisons upon the expression of CALM/AF10. The genes that have fulfilled the selection criteria and test

statistics for were also evaluated by plotting them in an M (Intensity Ratio)-A (Average Intensity) plot (Figure 3.7).

Table 3.5. Number of differentially regulated genes in different comparisons.

Comparison number	Experiment	Baseline	Number of genes
1	pRTS-1-FLAG-CALM/AF10 24 hrs	pRTS-1 0hrs, pRTS-1-FLAG-CALM/AF10 0hrs, pRTS-1 24 hrs	1237
2	pRTS-1-FLAG-CALM/AF10 72 hrs	pRTS-1 0hrs, pRTS-1-FLAG-CALM/AF10 0hrs, pRTS-1 72 hrs	564
3	pRTS-1-VP16-FLAG-CALM/AF10 24 hrs	pRTS-1 0hrs, pRTS-1-VP16-FLAG-CALM/AF10 0hrs, pRTS-1 24 hrs	990
4	pRTS-1-VP16-FLAG-CALM/AF10 72 hrs	pRTS-1 0hrs, pRTS-1-VP16-FLAG-CALM/AF10 0hrs, pRTS-1 72 hrs	164
5	pRTS-1-FLAG-CALM/AF10 24hrs and pRTS-1-FLAG-CALM/AF10 72 hrs	pRTS-1 0hrs, pRTS-1-FLAG-CALM/AF10 0hrs, pRTS-1 24 hrs, pRTS-1 72 hrs	477
6	pRTS-1-VP16-FLAG-CALM/AF10 24hrs and pRTS-1-VP16-FLAG-CALM/AF10 72 hrs)	pRTS-1 0hrs, pRTS-1-VP16-FLAG-CALM/AF10 0hrs, pRTS-1 24 hrs, pRTS-1 72 hrs	130

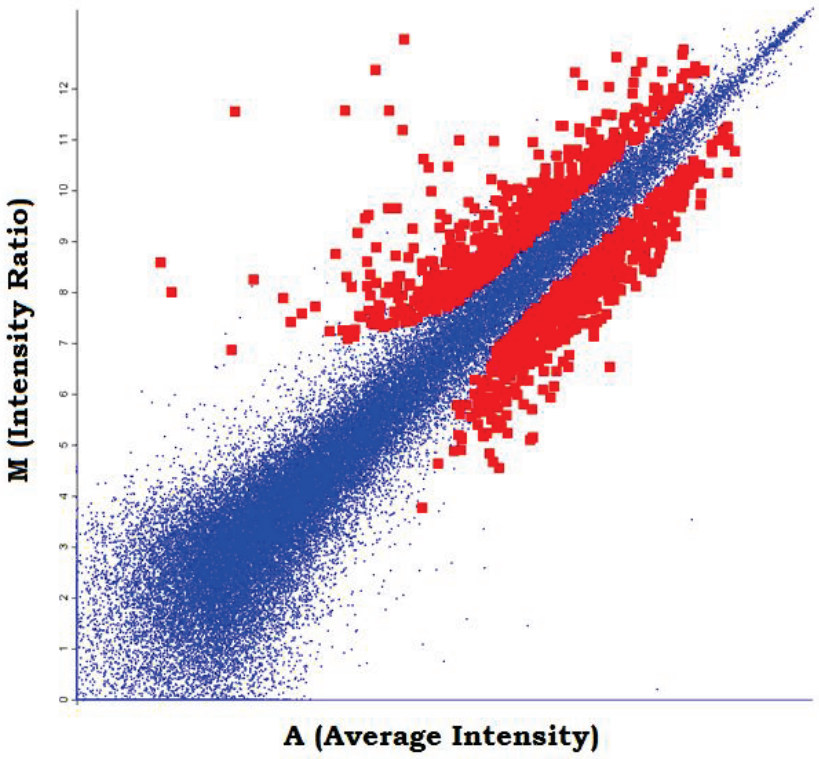


Figure3.7. M-A plot for identification of genes that are 2 fold differentially regulated in pRTS-1-FLAG-CALM/AF10 24hrs. M (y axis) is the intensity ratio given by the log2 difference of the experimental sample (pRTS-1-FLAG-CALM/AF10 24hrs) and control (pRTS-1 0hrs, pRTS-1-FLAG-CALM/AF10 0hrs, and pRTS-1 24hrs)(see text for explanation on how the comparison criteria is setup), and it is of the form $M = \log_2 E - \log_2 C$, where E is experimental sample and C is control. A (y=x line) is the average intensity for a probeset in the plot and is calculated using the formula $A = ½ (\log_2 E + \log_2 C)$, where E and C are experimental expression level and control expression level, respectively. Red squares are probeset that satisfied the selection criteria (up and down-regulated) and blue dots are probesets that are not differentially regulated.

3.4.3.2. The effect of the VP16 transactivating domain

The VP16 transactivating domain is a strong transactivating domain which is able to transactivate almost all promoter types. We speculated that CALM/AF10 might act as a transcriptional repressor on certain promoters. However, due to limitations inherent in our experimental set-up (e.g. not 100% of cells transfected), weakly down-regulated genes would be more difficult to identify. To circumvent this problem, we fused the VP16 transactivation domain to the CALM/AF10 fusion protein in order to convert CALM/AF10 from a transcriptional repressor into a strong transcriptional activator of

these promoters thus leading to a clear detection of genes that would be normally down-regulated by CALM/AF10.

To assess whether VP16 had the expected effect, the list of genes that were found to be differentially regulated in pRTS-1-FLAG-CALM/AF10 24hrs and pRTS-1-FLAG CALM/AF10 72hrs (Table 3.5, 5th comparison, 477 genes) were compared to the list of differentially regulated genes in pRTS-1-VP16-FLAG-CALM/AF10 24hrs and pRTS-1-VP16-FLAG CALM/AF10 72hrs (Table 3.5, 6th comparison, 130 genes). It was observed that 95 of the 130 genes that were differentially regulated in pRTS-1-VP16-FLAG-CALM/AF10 24hrs and pRTS-1-VP16-FLAG-CALM/AF10 72hrs (VP16-CALM/AF10 expressing cells) were also differentially regulated in pRTS-1-FLAG-CALM/AF10 24hrs and pRTS-1-FLAG CALM/AF10 72hrs (CALM/AF10 expressing cells but without VP16). A Pearson correlation was used to evaluate whether the differential expression levels of these genes in these settings were comparable. The correlation analysis showed that there was a statistically significant correlation between the differential expression levels of genes in VP16-CALM/AF10 expressing cells and CALM/AF10 expressing cells but without VP16 ($r=0.959$, $p \approx 0.00$) (Table 3.6 and figure 3.8). In conclusion, the VP16 transactivating domain did not enhance genes that were found to be down-regulated in the pRTS-1-FLAG-CALM/AF10 (pRTS-1-FLAG-CALM/AF10 24hrs and pRTS-1-FLAG CALM/AF10 72hrs) to control (pRTS-1 0hrs, pRTS-1-FLAG-CALM/AF10 0hrs, pRTS-1 24hrs, and pRTS-1 72hrs) comparison.

Table 3.6. Correlation table comparing differential expression levels of VP16-CALM/AF10 expressing cells (pRTS-1-VP16-FLAG-CALM/AF10 24hrs and pRTS-1-VP16-FLAG-CALM/AF10 72hrs) and CALM/AF10 expressing cells but without VP16 (pRTS-1-FLAG-CALM/AF10 24hrs and pRTS-1-FLAG CALM/AF10 72hrs).

All time points taken together		**CALM/AF10 expressing cells but without VP16**
VP16-CALM/AF10 expressing cells	Pearson Correlation	.959(**)
	Sig. (2-tailed)	.000
	N	95

** Correlation is significant at the 0.01 level (2-tailed).

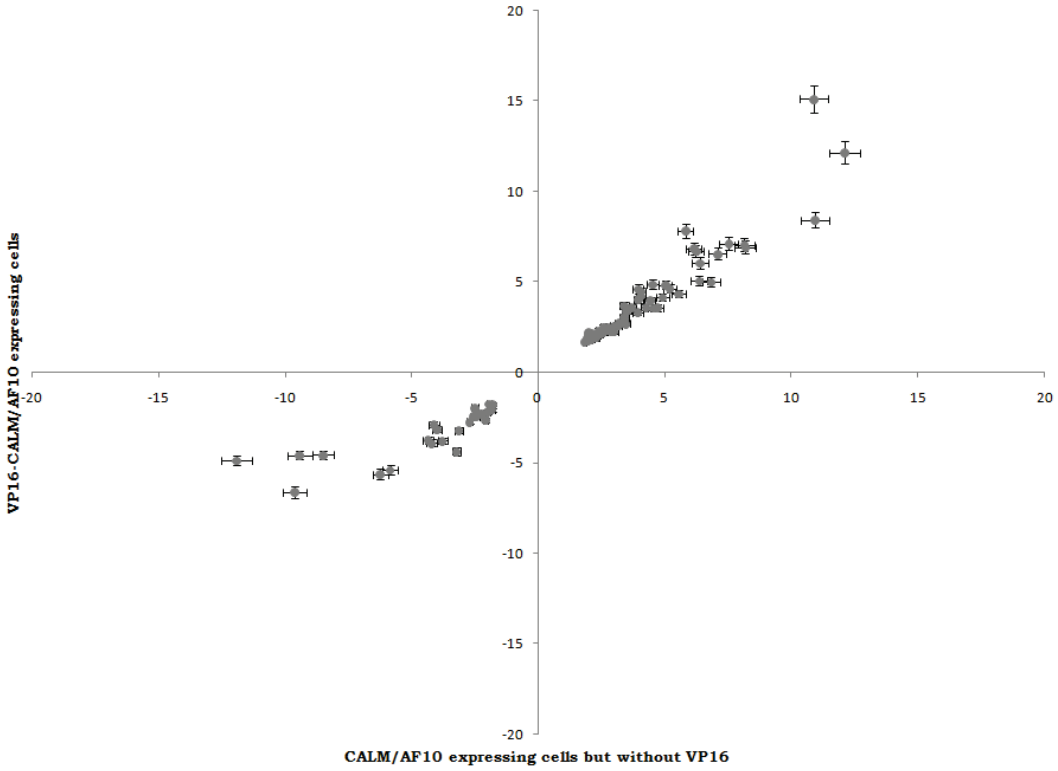


Figure 3.8. Scatterplot showing fold expression levels of the 95 genes that were found to be differentially regulated in both VP16-CALM/AF10 expressing cells (pRTS-1-VP16-FLAG-CALM/AF10 24hrs and pRTS-1-VP16-FLAG-CALM/AF10 72hrs, y axis) and CALM/AF10 expressing cells but without VP16 (pRTS-1-FLAG-CALM/AF10 24hrs and pRTS-1-FLAG CALM/AF10 72hrs, x axis). Error bars indicate the 95% confidence interval.

3.4.3.3. Hierarchical clustering

Given a set of items to be clustered (items can be either genes or chips/experiments), agglomerative hierarchical clustering (HC) recursively merges items with similar items, or with the result of previous merges, according to their pair-wise distance (with the closest item pairs being merged first) (Gould, 2005). The purpose of this method is to group together genes that behave in a similar fashion in a certain set of experiment. As a result, it produces a tree structure, referred to as dendrogram, whose nodes correspond to:

i) The original items (these are the leaves of the tree); and

ii) The merging of other nodes (these are the internal nodes of the tree)

Different distance measures can give different clustering results. The most common types of distance measurement depend on either correlation (parametric or non-parametric) or Euclidean distance (parametric) measurement. Once the distance between different genes is computed depending on how they behave in different samples, there are various ways of building the hierarchy/linkage:

Single linkage - minimum distance between points in different clusters

Complete linkage - maximum distance

Average linkage - mean of all distances between points in different clusters

Centroid distance - represent each cluster by its centroid and measures distances between them.

Based on the results obtained from the differential gene expression by comparing different samples, hierarchical clustering was performed using Euclidean distance as a measure of similarity/dissimilarity of genes in their expression pattern and Average linkage as a hierarchical amalgamation rule.

Figure 3.9 shows heatmaps of some selected genes based on the list of differentially regulated genes obtained by comparing CALM/AF10 expressing cells (pRTS-1-FLAG-CALM/AF10 24hrs and pRTS-1-FLAG-CALM/AF10 72 hrs) to the controls (pRTS-1 0hrs, pRTS-1-FLAG-CALM/AF10 0hrs, pRTS-1 24 hrs, pRTS-1 72 hrs) (comparison 5 in table 3.5, 477 genes). The genes in figure 3.9 have been selected because the pathways that they are involved in were later found to be some of the strongly deregulated cellular processes in this study and are relevant for subsequent discussion. It can be observed that the selected genes shown here are mostly down-regulated after CALM/AF10 expression. Genes close to each other based on the hierarchy show a similar expression pattern indicating that they might be co-regulated by an upstream factor or that they are regulating each other.

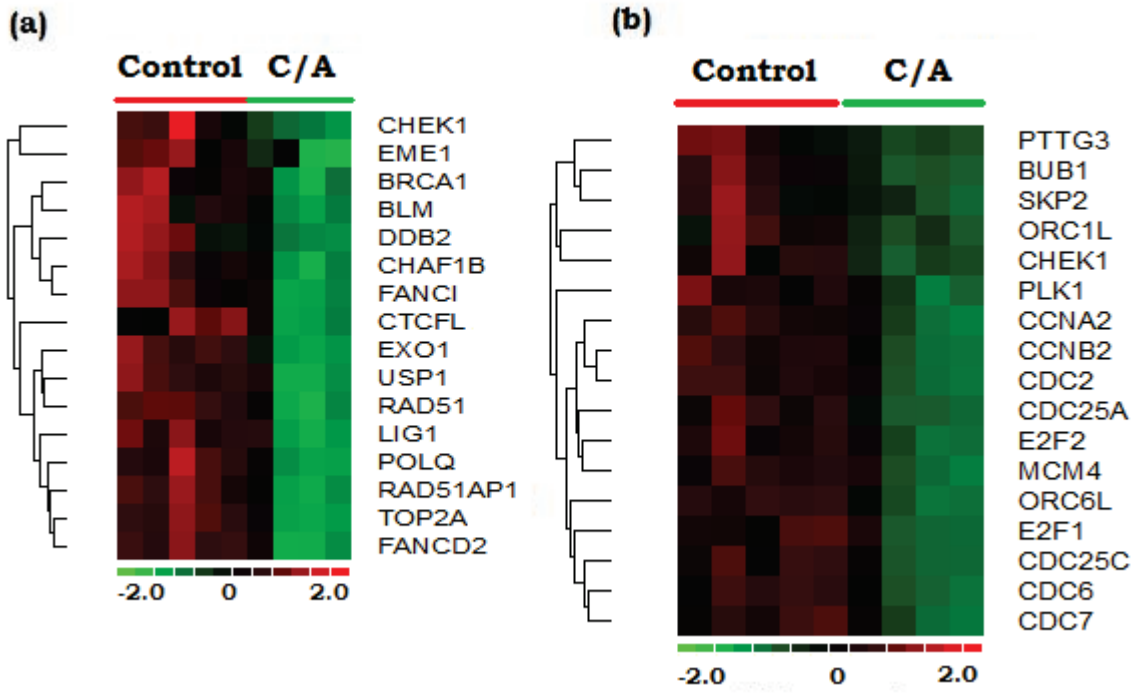


Figure 3.9. Hierarchical clustering of differentially regulated genes obtained by comparing CALM/AF10 expressing cells (pRTS-1-FLAG-CALM/AF10 24hrs and pRTS-1-FLAG-CALM/AF10 72 hrs) to the controls (pRTS-1 0hrs, pRTS-1-FLAG-CALM/AF10 0hrs, pRTS-1 24 hrs, pRTS-1 72 hrs). **C/A** *stands for cells expressing DG-75 (pRTS-1-FLAG-CALM/AF10 24hrs, pRTS-1-FLAG-CALM/AF10 72 hrs, pRTS-1-VP16-FLAG-CALM/AF10 24hrs, and pRTS-1-VP16-FLAG-CALM/AF10 72 hrs) and* **Control** *stands for non-induced DG-75 cells containing CALM/AF10 constructs (pRTS-1-FLAG-CALM/AF10 0hrs and pRTS-1-VP16-FLAG-CALM/AF10 0hrs) and vector only controls (pRTS-1 0hrs, pRTS-1 24hrs, and pRTS-1 72hrs). These genes are involved in DNA repair (a) and cell cycle (b). Red squares indicate up-regulation and green squares show down-regulated (color scale is shown under the heatmaps). The hierarchy is shown on the left side of the heatmaps and gene names are given on the right side.*

3.4.3.3. Principal Component Analysis

In Affymetrix® based microarray studies each sample is represented by one chip. Each chip will thus have a unique expression pattern (snapshot of the expression status of every gene within the cell at the time of sample preparation) if *all the genes on the chip* are simultaneously considered. This is of biological interest as the physiology of a cell at a given time is the result of the interplay of the expression levels of the various genes in the cell. The more similar two samples are biologically, the higher is the similarity in

their expression pattern. As a single microarray chip covers the entire human genome (more than 20,000 genes interrogated by more than 54,000 probesets in the HGU133plus 2.0 chip), it is usually impractical to assess every gene independently and see how its expression contributes to the different expression patterns that are displayed by the different samples. To assess this biological variability among the different samples and find a statistical method that would summarize this variation has always been one of the key challenges in the analyses of microarray data. Towards this end, the principal component analysis has been used extensively in numerous microarray studies.

The principal component analysis (PCA) is a statistical method that tries to find few functions (usually 3, referred to as Eigenvectors or principal components) that best explain the variation that is observed in a dataset by reducing the number of variables (gene expression levels in this case) based on how strong their correlation or co-variation is. This is achieved by grouping genes that vary in their expression in a similar fashion across different samples into one function that captures their expression pattern. The power of a PCA is thus dependent on how highly correlated or co-varied the genes are: the lower the correlation the harder it is to find a small number of principal components that could explain the variation observed in the data set.

In the current study, the R implementation of PCA (pcaMethods) was used to conduct a principal component analysis on the 9 normalized microarrays (R version 2.8.1 (2008-12-22), Copyright© 2008, The R Foundation for Statistical Computing). 3 principal components were extracted that explained cumulatively 98.42% of the variation among the 9 different samples (see table 3.7 for further details). Based on these 3 principal components (PCs), new values are calculated for the expression level of each gene (factor

loading). Based on the loadings, a 3D scatterplot of the samples was generated that showed which of the samples are more similar/different depending on the position of the samples in the 3D space, by giving most weight to the principal component 1 (PC1) axis, followed by PC2 and PC3 based on their contribution to the explanation of the variation (Figure 3.10).

Table 3.7. PCA result. The second (% total variance) and the last column (Cumulative % of variance) are of particular interest in this table.

PCA	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	6.163022	68.47803	6.163022	68.47803
2	2.406050	26.73389	8.569073	95.21192
3	0.289115	3.21239	8.858188	98.42431

It can be observed (figure 3.10) that the CALM/AF10 and VP16-CALM/AF10 expressing samples (pRTS-1-VP16-FLAG-CALM/AF10 24hrs, pRTS-1-VP16-FLAG CALM/AF10 72hrs, pRTS-1-FLAG-CALM/AF10 24hrs, and pRTS-1-FLAG-CALM/AF10 72hrs) cluster together, with slight deviation for the VP16-CALM/AF10 expressing sample at 72 hrs (pRTS-1-VP16-FLAG-CALM/AF10 72hrs). Samples with CALM/AF10 or VP16-CALM/AF10 expression plasmids which are not induced (pRTS-1-VP16-FLAG-CALM/AF10 0hrs and pRTS-1-FLAG-CALM/AF10 0hrs) and the vector only controls (pRTS-1 0hrs, pRTS-1 24hrs, and pRTS-1 72hrs) form a cluster as well. This is as expected because cells that are not induced (i.e. not expressing CALM/AF10 or

VP16-CALM/AF10) should have a gene expression pattern which is similar to the vector only controls (pRTS-1 0hrs, pRTS-1 24hrs and pRTS-1 72hrs).

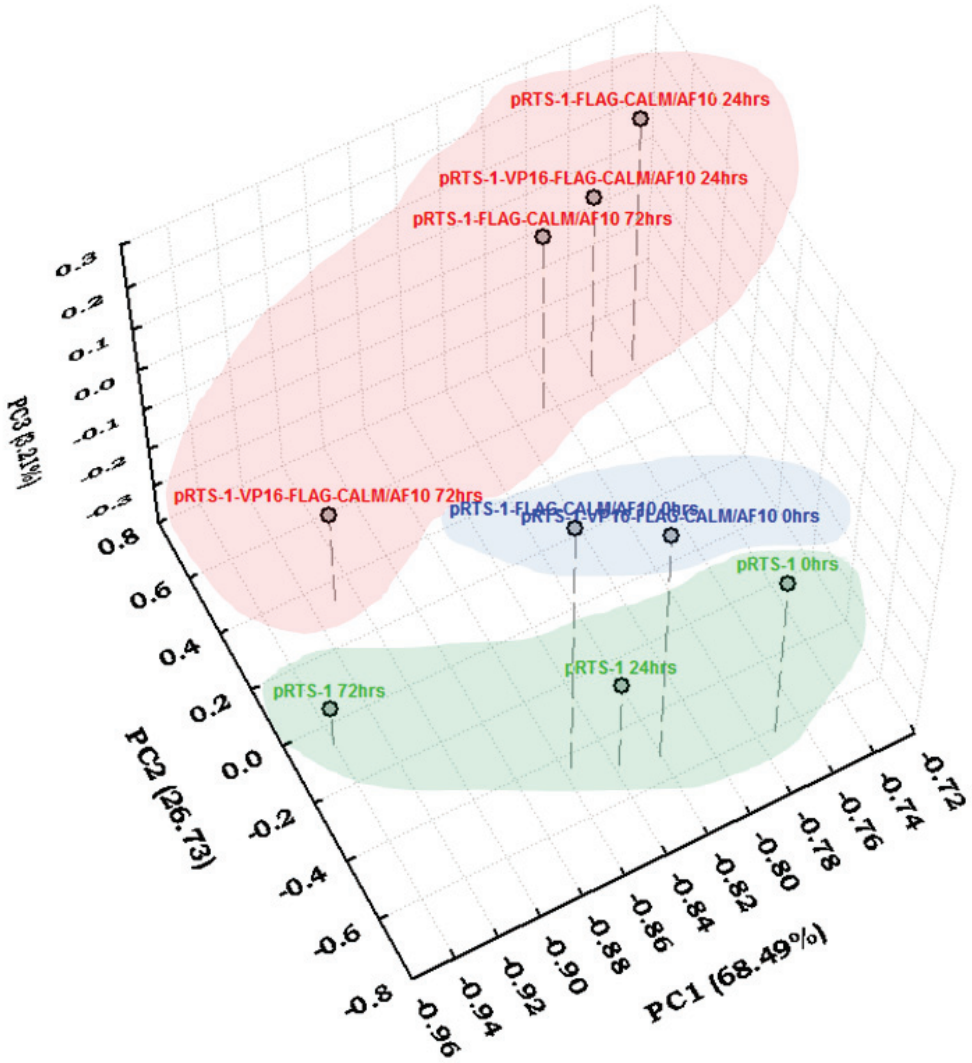


Figure 3.10. 3D scatterplot of the samples with respect to the 3 principal components. Dots represent samples and spikes show the position of the samples on a 2D surface of PC1 and PC2 (PC1 and PC2 together explained 95% of the variance). Red labels (and red-shaded area) are all samples that were induced to express CALM/AF10 (pRTS-1-VP16-FLAG-CALM/AF10 24hrs, pRTS-1-VP16-FLAG CALM/AF10 72, pRTS-1-FLAG-CALM/AF10 24hrs, and pRTS-1-FLAG-CALM/AF10 72hrs). Blue labels and shading represent non-induced experimental samples (pRTS-1-VP16-FLAG-CALM/AF10 0hrs and pRTS-1-FLAG-CALM/AF10 0hrs). Green labeling and shading show vector only controls (pRTS-1 0hrs, pRTS-1 24hrs, and pRTS-1 72hrs). ***NB:*** *shades only show sample classification and not necessarily clustering.*

3.4.4. Chromosomal distribution of differentially regulated genes

One of the interesting ways to look at microarray data is to examine the global transcriptional consequence by plotting the differentially expressed genes according to their chromosomal location. The main objective would be to look at interesting chromosomal locations or positions that might be significantly affected. This might be an indication of positional co-regulation of groups of genes.

3.4.4.1. dChip based analysis of chromosomal distribution of differentially regulated genes

The dChip software can plot the differentially regulated genes according to their location in the genome. The genes are assigned to their location on the chromosome and are indicated by red (up-regulated) or green (down-regulated) bars. The position of a bar with respect to the middle line indicates the direction of transcription: if the bar is above the middle line it indicates that the gene is transcribed from the sense strand and if it is below the middle line it shows that the gene is transcribed from the anti-sense strand. The genes on each chromosome are placed proportionally from chromosomal position 0 to the gene with the maximal chromosomal position in the Genome information file. The test is carried out at a cut point of 500 kbp of a stretch of a chromosome on either side of a differentially regulated gene with respect to its transcription start site. The p-values are calculated for all stretches containing less or equal to 20 differentially regulated genes to assess the significance of "gene proximity". It is an evaluation of the "tightness" (the rank distance of the genes on the two ends of the stretch, in Mbp) of a stretch of *n* genes (real gene distribution) against that of *n* genes randomly put on the chromosome (relative position of genes along the chromosome if they were randomly distributed). This way,

gene density along a chromosome will be taken into account. For example, on chromosome 6, the stretch from gene 1 to 9 has a p-value 0.021799, indicating that this stretch of chromosome is gene rich as a p-value of 0.021799 is significant and rejects the hypothesis that genes are randomly distributed along a chromosome. These significant gene stretches are outlined in blue boxes. If a significant longer stretch contains a significant shorter stretch, only the longer one is reported and outlined. Since there is no correction for multiple hypotheses testing, the p-values here are used for drawing attention to specific genes.

In the current study, the list of differentially regulated genes obtained by comparing pRTS-1-FLAG-CALM/AF10 expressing cells (pRTS-1-FLAG-CALM/AF10 24hrs and pRTS-1-FLAG-CALM/AF10 72hrs) versus controls (pRTS-1 0hrs, pRTS-1-FLAG-CALM/AF10 0hrs, pRTS-1 24 hrs, pRTS-1 72 hrs, comparison criterion 5 in table 3.5) was used to generate figure 3.11. It can be observed that there are a number of significant cluster of differentially regulated genes, as indicated by the blue bars/boxes. Each region can be further investigated by examining the differentially regulated genes. For instance, chromosome 10 is interesting as it is:

a. The chromosome that contains one of the fusion partners of CALM, namely AF10 and
b. Previous work by our group on patient microarray data that compared 10 CALM/AF10^{+} AML and ALL patients to 119 cases of other leukemias (CML, ALL with 1;19 or 4;11, ALL with MLL rearrangement, ALL, ALL with Ph+, AML with MLL rearrangement, AML normal karyotype, AML normal karyotype FLT3+, AML with inv(16), AML with t(8;21), AML with t(15;17), AML with

complex karyotype; all with 10 samples each except ALL with MLL rearrangement (9)) and to 10 normal bone marrow samples showed that genes located close to the breakpoint of the t(10;11)(p12;q14) translocation on 10p12 (COMMD3, SPAG6, BMI and DNAJC1) were found to be at least 1.5 fold differentially regulated (unpublished data). To see if such an effect is also observable in our model, we examined chromosome 10 more closely in the next section.

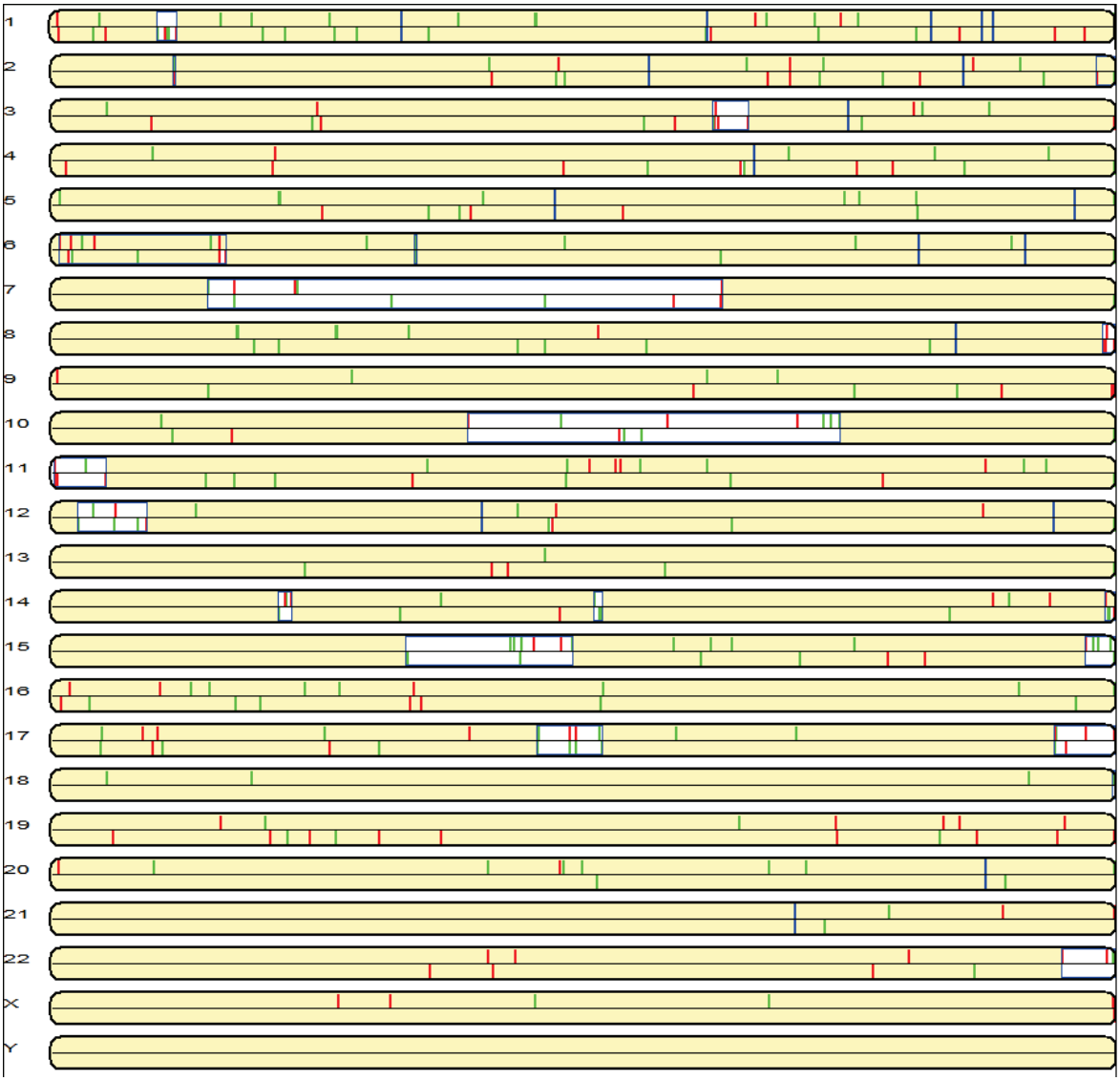


Figure 3.11. Genomic distribution of differentially expressed genes (chromosomes are not drawn to scale) based on comparison of pRTS-1-FLAG-CALM/AF10 expressing cells (pRTS-1-FLAG-CALM/AF10 24hrs and pRTS-1-FLAG-CALM/AF10 72hrs) versus controls (pRTS-1 0hrs, pRTS-1-FLAG-CALM/AF10 0hrs, pRTS-1 24 hrs, pRTS-1 72 hrs, comparison criterion 5 in table 3.5.).Yellow shading indicates the chromosome and non-shaded areas within the chromosome indicate clusters of differentially regulated that are composed of smaller significant clusters.

3.4.4.2. MicroArray Chromosome Analysis Tool (MACAT)

The MicroArray Chromosome Analysis Tool (MACAT) aims at linking the term differential gene expression to the chromosomal localization of genes. It is a statistical approach for identifying regions on the chromosomes that contain significantly differentially expressed regions on the chromosomes based on a regularized t-statistic (Toedling *et al.*, 2004).

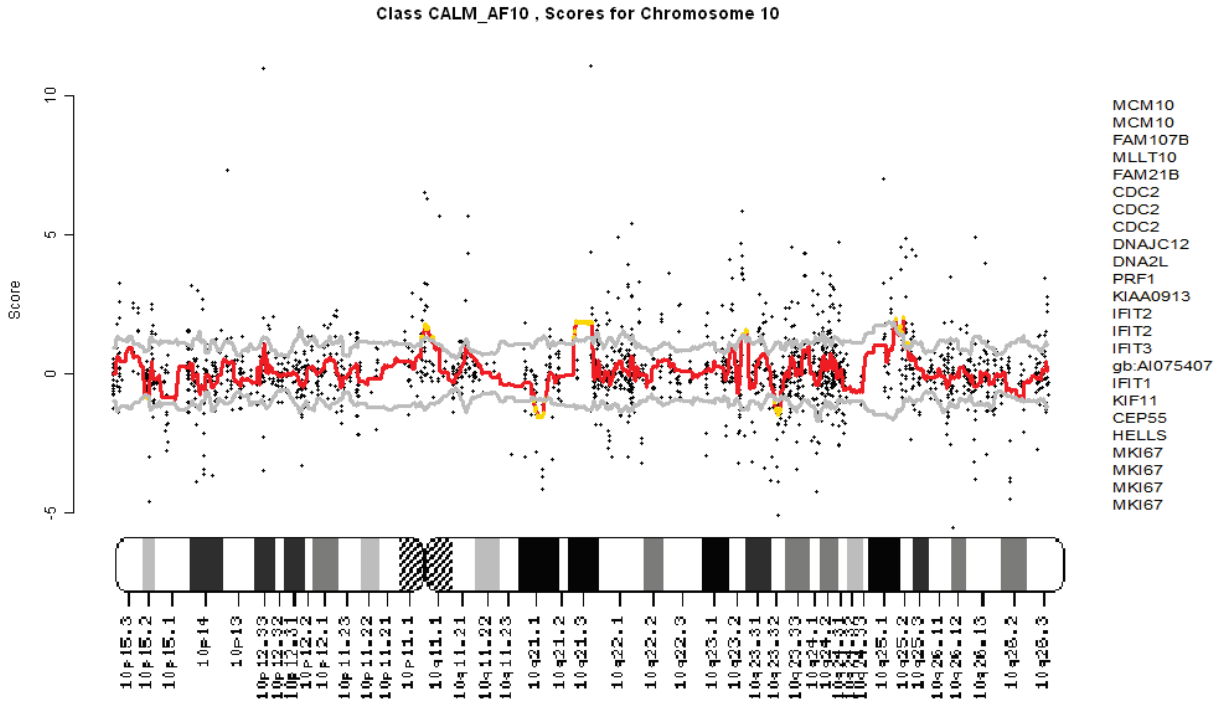


Figure 3.12. MACAT analysis for chromosome 10. Y axis show level of differential regulation for the comparison pRTS-1-FLAG-CALM/AF10 expressing cells (pRTS-1-FLAG-CALM/AF10 24hrs and pRTS-1-FLAG-CALM/AF10 72hrs) versus controls (pRTS-1 0hrs, pRTS-1-FLAG-CALM/AF10 0hrs, pRTS-1 24 hrs, pRTS-1 72 hrs, comparison criterion 5 in table 3.5). Black points show probesets,the red line indicates the mean fold up or down regulation of probesets of a given gene and grey lines indicate cut offs for differential regulation (1.5 fold in this case). Yellow spots on the red line indicate a gene that satisfies the cut point for differential regulation. The list of probesets that significantly deregulated is given on the right-side of the plot.

The MACAT analysis was performed using the Bioconductor© implementation of the MACAT package in R (R version 2.8.1 (2008-12-22), Copyright© 2008, The R Foundation for Statistical Computing) (Toedling *et al.*, 2004). The analysis revealed that there are some closely spaced genes on chromosome 10 that are differentially regulated. The yellow spots indicate genes that are statistically differentially regulated with respect to the calculated t-statistic. The distribution of the genes along chromosome 10 and their probesets are given in figure 3.12. It should be noted that there were no clusters of genes located close enough to each other to suggest co-regulation that could be explained by our current knowledge of chromatin architecture (i.e. genes within a smaller than 1 Mbp region). The MACAT analysis was performed for the entire genome (for all chromosomes) and there were no clusters of genes with very close proximity (<= 100 kbp) that were significantly differentially regulated.

3.4.5. Ontology Analysis (GOSurfer)

One of the major challenges in microarray based experiments is to analyze the data in a way that gives biological insight into the overwhelming amount of data generated. Even though looking at the top up and down-regulated genes in a list of differentially regulated genes seems to be a plausible approach, it is known that biologically relevant gene expression changes might not always be so dramatic. Slight changes in the expression levels in a number of genes belonging to a certain biological processes might be biologically more relevant than a large change in a single gene. As a result, pathway and biological process analysis of microarray data has become one of the important analyses of gene expression profiling studies. Various scientific databases (predominantly Kyoto Encyclopedia of Genes and Genomes (KEGG) and GenMAPP) are striving to arrive at a

consistent and proper annotation of genes with respect to the processes that they are involved in, the functions that they perform, and where exactly they are found within the cell.

The Gene Ontology (GO) project is a collaborative effort to address the need for consistent description of the function of a gene product in different databases (http://www.geneontology.org/GO.doc.shtml). The project began as collaboration between three model organism databases, FlyBase (*Drosophila*), the *Saccharomyces* Genome Database (SGD) and the Mouse Genome Database (MGD), in 1998. The three organizing principles of GO are: **cellular component**, **biological process** and **molecular function**. A gene product might be associated with or located in one or more cellular components; it is active in one or more biological processes, during which it performs one or more molecular functions.

There are many software packages currently available that can build GO hierarchies. The GoSurfer package (Zhong *et al*., 2004) was used to analyze the microarray data in this study. GoSurfer uses Gene Ontology (GO) information to analyze gene sets obtained from genome-wide computations or microarray analyses (Zhong *et al*., 2004). It associates user defined gene lists with GO terms and visualizes such GO terms as a hierarchical tree. The tree output can be manipulated by various means, like setting heuristic thresholds or using statistical tests. Significantly important GO terms, which were found to be significant, can be highlighted.

3.4.5.1. Molecular Function

The molecular function category describes activities such as catalysis or binding activities that occur at the molecular level. The GO analysis was based on the comparison of pRTS-FLAG-CALM/AF10 samples (pRTS-1-FLAG-CALM/AF10 24hrs and pRTS-1-FLAG-CALM/AF10 72 hrs) to the controls pRTS-1 0hrs, pRTS-1-FLAG-CALM/AF10 0hrs, pRTS-1 24 hrs and pRTS-1 72 hrs (comparison criterion 5 in table 3.5, 477 genes). The GO analysis of the microarray data revealed that binding, transcription regulator activity (subcategory transcriptional repressor activity) and catalytic activity were the three GO molecular function classes in which the differentially regulated genes were statistically overrepresented. Subcategories within the binding class that contained most of the differentially regulated genes were “damaged DNA binding”, “chromatin binding”, and “tubulin binding” (Figure 3.13).

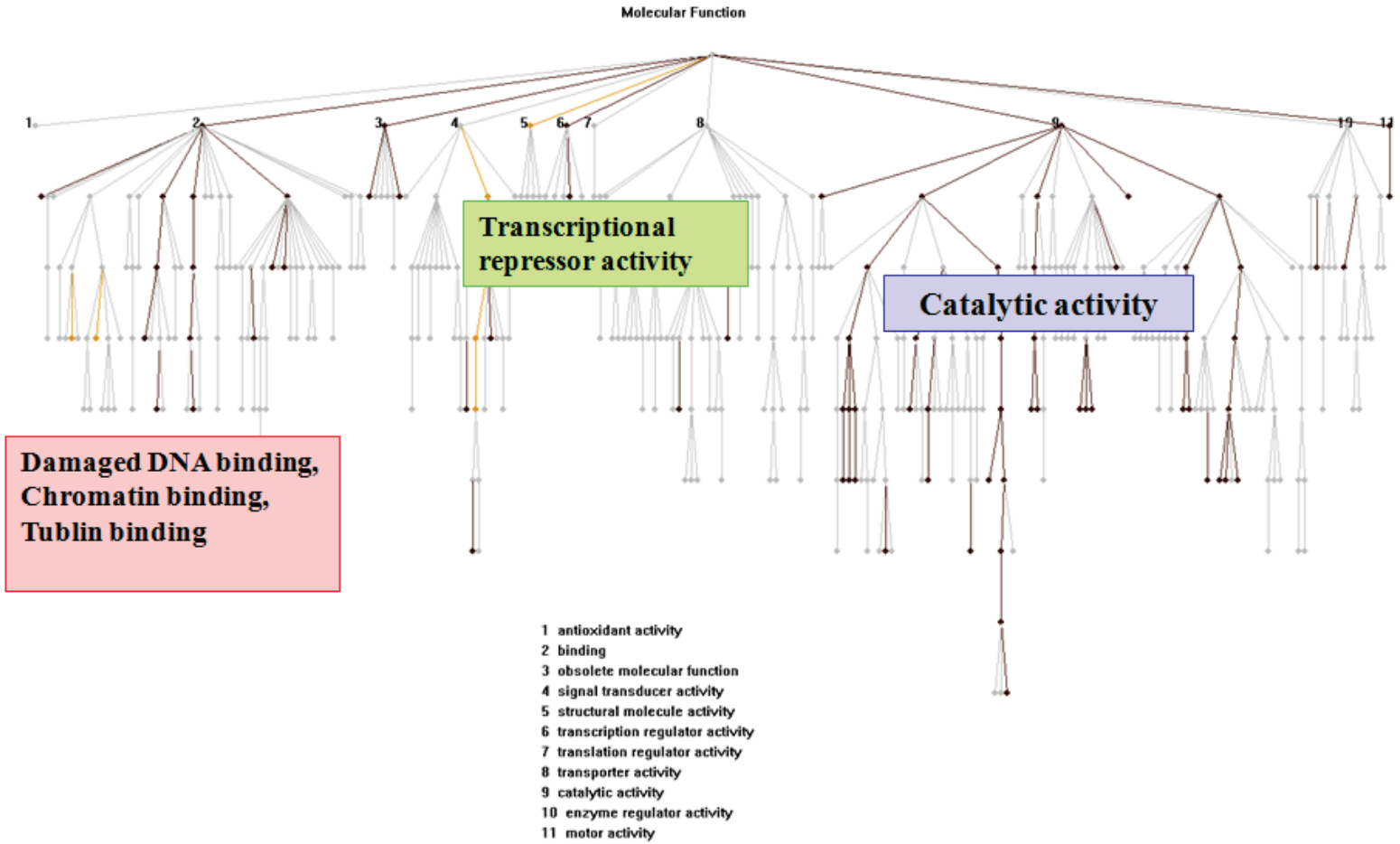


Figure 3.13. Molecular Function. The GO molecular function is classified into 11 functional classes (list shown at the bottom of the figure). Functional classes that showed strong enrichment of the differentially regulated genes based on the comparison of pRTS-FLAG-CALM/AF10 samples (pRTS-1-FLAG-CALM/AF10 24hrs and pRTS-1-FLAG-CALM/AF10 72 hrs) to the controls (pRTS-1 0hrs, pRTS-1-FLAG-CALM/AF10 0hrs, pRTS-1 24 hrs and pRTS-1 72 hrs, comparison criterion 5 in table 3.5, 477 genes) are highlighted. Dark brown and orange branches were found to be statistically significant for the enrichment of the genes in those sub-categories of molecular functions (boxes with different colors and labels), whereas grey branches are statistically not significant.

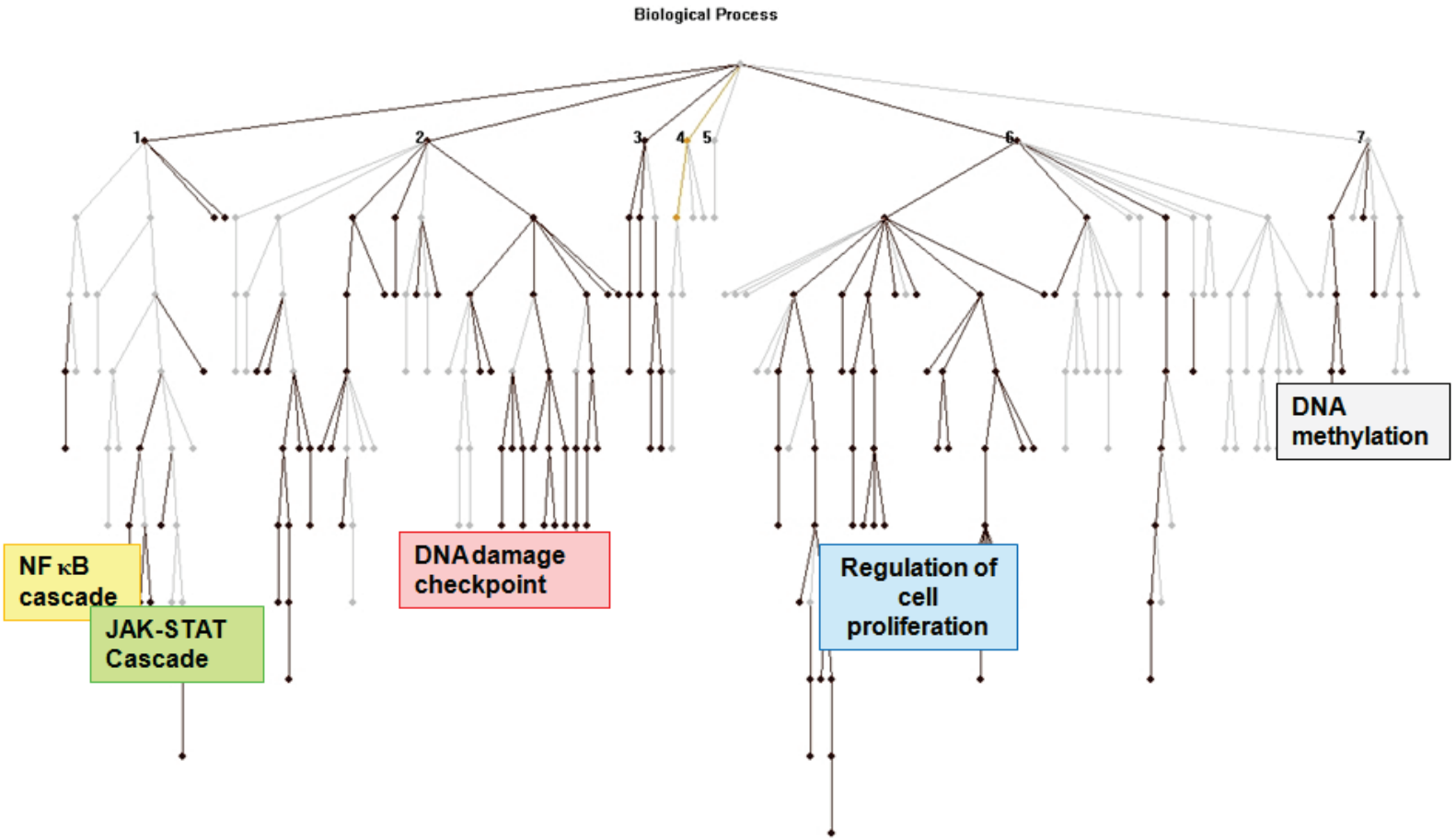


Figure 3.14. Biological Process. The GO molecular function is classified into 7 biological processes. Biological processes that showed strong enrichment of the differentially regulated genes based on the comparison of pRTS-FLAG-CALM/AF10 samples (pRTS-1-FLAG-CALM/AF10 24hrs and pRTS-1-FLAG-CALM/AF10 72 hrs) to the controls (pRTS-1 0hrs, pRTS-1-FLAG-CALM/AF10 0hrs, pRTS-1 24 hrs and pRTS-1 72 hrs, comparison criterion 5 in table 3.5, 477 genes) are highlighted. Dark brown and orange branches were found to be statistically significant for the enrichment of the genes in those sub-categories of biological processes (boxes with different colors and labels) whereas grey branches are statistically not significant.

3.4.5.2. Biological Process

A biological process is series of events accomplished by one or more ordered assemblies of molecular functions. Examples of broad biological process terms are cellular physiological process or signal transduction. Examples of more specific terms are pyrimidine metabolic process or alpha-glucoside transport. It can be difficult to distinguish between a biological process and a molecular function, but the general rule is that a process must have more than one distinct steps. A biological process is not equivalent to a pathway.

Genes from several important biological processes were were significantly overrepresented in the set of differentially regulated genes comparing pRTS-FLAG-CALM/AF10 samples (pRTS-1-FLAG-CALM/AF10 24hrs and pRTS-1-FLAG-CALM/AF10 72 hrs) to the controls pRTS-1 0hrs, pRTS-1-FLAG-CALM/AF10 0hrs, pRTS-1 24 hrs and pRTS-1 72 hrs (comparison criterion 5 in table 3.5, 477 genes). The biological processes that showed statistically significant enrichment among the differentially regulated genes were: NFκB cascade, JAK-STAT cascade, DNA damage checkpoint, and regulation of cell proliferation (Figure 3.14). To further understand which genes in these pathways are affected and how this would affect the whole process, a pathway analysis was carried out.

3.4.6. Pathway Analysis (KegArray)

The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database for uncovering higher-order systemic behaviors of the cell and the organism from genomic and molecular information (Kanehisa and Goto, 2000). The KEGG consortium developed a

number of algorithms that assist the analysis of data sets (e.g. microarray, Real-time PCR or sequencing) by exploiting the information available in the KEGG database. KegArray, for instance, is a tool developed to schematically display the distribution of differentially regulated genes in a given pathway.

Since our data set revealed, based on the hierarchical clustering and GO biological process findings, that cell cycle and DNA damage and repair are some of the significantly affected pathways/processes, the KegArray tool was used to investigate this in further detail.

3.4.6. 1. Cell cycle

Except for three genes (*SMAD2/3, Ink 4a-d,* and *ATM/ATR*), all the differentially regulated genes involved in cell cycle regulation were downregulated (Figure 3.15), based on comparison of pRTS-1-FLAG-CALM/AF10 samples (pRTS-1-FLAG-CALM/AF10 24hrs and pRTS-1-FLAG-CALM/AF10 72 hrs) to the controls (pRTS-1 0hrs, pRTS-1-FLAG-CALM/AF10 0hrs, pRTS-1 24 hrs and pRTS-1 72 hrs; comparison criterion 5 in table 3.5). Two of the six origin recognition complex (ORC) genes and four of the six mini-chromosome maintenance (MCM) genes were down-regulated. How the down-regulated and up-regulated genes interact and how CALM/AF10 might affect the cell cycle will be described in more detail in the discussion section.

3.4.6. 2. DNA damage checkpoint and DNA repair

The list of differentially regulated genes includes a number of genes that have been implicated in various types of cancer including *BRCA1*, *DDB2, CHEK1* and *RAD5*. All four genes were down-regulated (comparison criterion 5 in table 3.5). Most of the other

genes in this category were also down regulated, indicating that DNA damage checkpoint is one of the important pathways affected early after the expression of CALM/AF10. Figure 3.16 shows the nucleotide excision repair process and some of the genes (*DDB2, PCNA,* and *LIG1*) that were found to be differentially regulated in our comparison stated above.

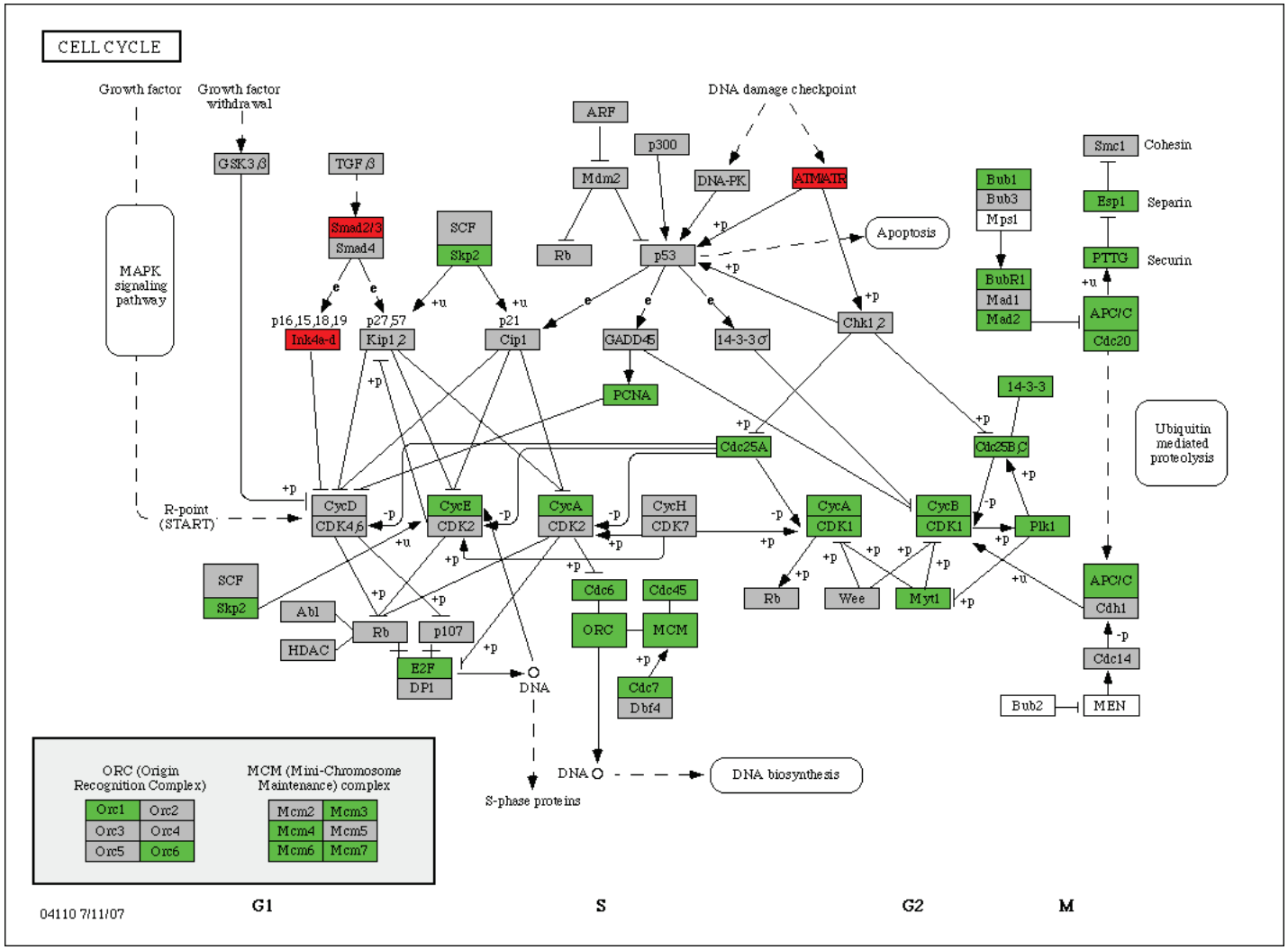
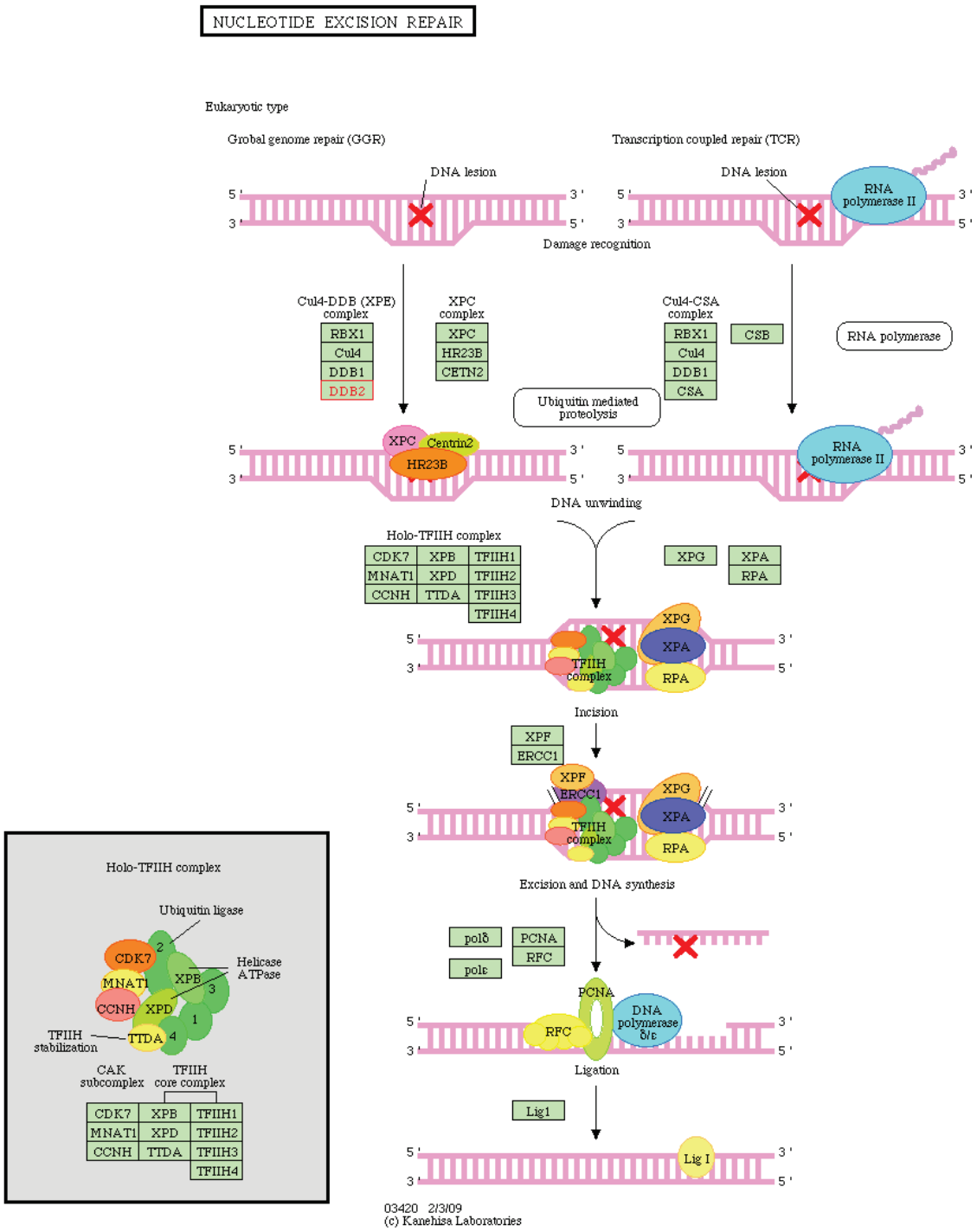


Figure 3.15. Cell cycle based on KEGG. Green boxes are down-regulated genes and red are up-regulated. Grey indicate genes that are not differentially regulated (based on comparison of pRTS-1-FLAG-CALM/AF10 samples (pRTS-1-FLAG-CALM/AF10 24hrs and pRTS-1-FLAG-CALM/AF10 72 hrs) to the controls (pRTS-1 0hrs, pRTS-1-FLAG-CALM/AF10 0hrs, pRTS-1 24 hrs and pRTS-1 72 hrs (comparison criterion 5 in table 3.5)).



3.16. Nucleotide excision repair based on KEGG. A number of genes involved in Nucleotide excision repair (part of the DNA damage repair) including LIG1, DDB2 and PCNA were down0regulated in the comparison of the pRTS-1-FLAG-CALM/AF10 samples (pRTS-1-FLAG-CALM/AF10 24hrs and pRTS-1-FLAG-CALM/AF10 72 hrs) with the controls (pRTS-1 0hrs, pRTS-1-FLAG-CALM/AF10 0hrs, pRTS-1 24 hrs and pRTS-1 72 hrs) (comparison criterion 5 in table 3.5).

3.4.7. Comparison with patient data (Gene Set Enrichment Analysis, GSEA)

The Gene Set Enrichment Analysis is another powerful pathway/biological process orientated analytical method that derives its power by focusing on gene sets, which are defined as groups of genes that share common biological function, chromosomal location, or regulation (Subramanian *et al.*, 2005). Notably, where single-gene analysis finds little similarity between two independent studies (like patient data and cell line data), GSEA might reveal many biological pathways which are affected in both studies. A false discovery rate (FDR) adjustment can be made for multiple testing by setting a certain threshold.

In the current study, it was attempted to employ the power of this tool to find pathways that might be affected by CALM/AF10 both in the patient data set and the cell line data set of this study.

The patient data set used here is derived from microarray based (Affymetrix platform) expression profiling study using mononuclear cells from bone marrow aspirates collected from patients with CALM/AF10 and other types of leukemia. 10 CALM/AF10^{+} AML and ALL patients were compared to 119 patients with various types of leukemia (CML, ALL with 1;19 or 4;11, ALL with MLL rearrangement, ALL, ALL with Ph+, AML with MLL rearrangement, AML normal karyotype, AML normal karyotype FLT3+, AML with inv(16), AML with t(8;21), AML with t(15;17), AML with complex karyotype; all with 10 samples each except the group of ALL with MLL rearrangements (9)). 10 normal bone marrow samples were also included in the study. The purpose of the study was to identify genes which are up and down-regulated in CALM/AF10 patients as compared to other classes of leukemia and normal bone marrow controls. As the nature of the samples

and the comparison method in the patient data set and the cell line data set are not suited for direct comparison, the aim of GSEA is to assess if there are some pathways affected by CALM/AF10 both in the cell line data (immediate target genes) and in the patient data (differentially regulated genes compared to other leukemia subtypes).

The list of genes that were found to be differentially regulated in the cell line data (comparison of pRTS-1-FLAG-CALM/AF10 samples (pRTS-1-FLAG-CALM/AF10 24hrs and pRTS-1-FLAG-CALM/AF10 72 hrs) to the controls (pRTS-1 0hrs, pRTS-1-FLAG-CALM/AF10 0hrs, pRTS-1 24 hrs and pRTS-1 72 hrs) were used as a reference gene set for the GSEA. The reference gene set is the set of genes that was used to build the enrichment analysis when comparing the CALM/AF10 patient data to all the other classes of leukemia (a t-test was used as a metric for finding and ranking differentially regulated genes when comparing CALM/AF10 patient microarray data to other patient classes). In the process, for every match in the reference gene list with the comparison of patient samples for differentially regulated genes in CALM/AF10 patients is encountered, a positive score is entered, named as Enrichment Score (ES). The ES increase steeply if more genes in the reference list are matching with the most differentially regulated genes in the patient microarray data comparisons. The GSEA was run with 1000 iterations based on the reference gene set from the cell line data, there were no statistically enriched genes in the patient comparison. This implies that the set of genes that were differentially regulated in the cell line data are different from the set of differentially regulated genes in the patient data.

To assess if biological processes or molecular functions that were enriched in the cell line data from the GO and KEGG analysis would also be implicated in the patient data, the

GSEA was carried out on the patient data using GO biological process and molecular function as a reference gene set. Using this approach, we discovered that the pathway "DNA repair" was significantly enriched both in the cell line data and the patient data, even though the list of differentially regulated genes in the two different settings (cell line data and patient data) did not overlap. Figure 3.17 summarizes the result of the GSEA of the patient and cell line data. The involvement of DNA repair in both data sets indicates that a compromised DNA repair pathway with its ensuing genomic instability might be one of the oncogenic mechanisms of CALM/AF10-mediated leukemogenesis.

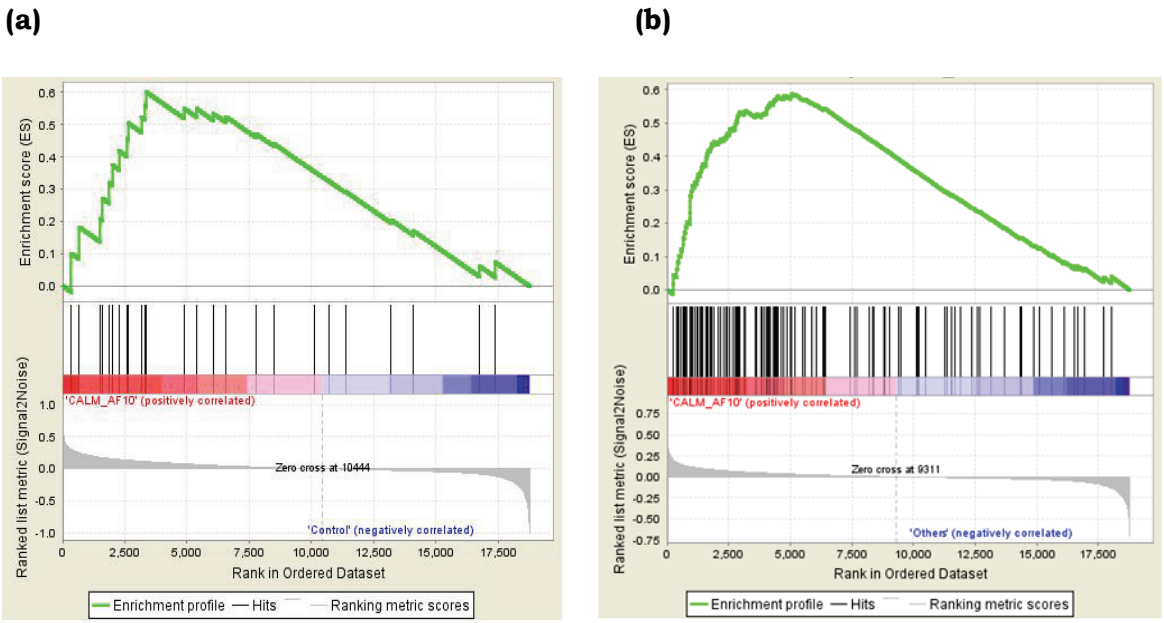


Figure 3.17. GSEA based on cell line and patient data described above (section 3.4.7.).(a) Enrichment of genes involved in DNA repair in the list of differentially regulated genes obtained by comapring pRTS-1-FLAG-CALM/AF10 24hrs and pRTS-1-FLAG-CALM/AF10 72 hrs to the controls (pRTS-1 0hrs, pRTS-1-FLAG-CALM/AF10 0hrs, pRTS-1 24 hrs and pRTS-1 72 hrs) (b) 10 CALM/AF10+ AML and ALL patients were compared to CML, ALL with 1;19 or 4;11, ALL with MLL rearrangement, ALL, ALL with Ph+, AML with MLL rearrangement, AML normal karyotype, AML normal karyotype FLT3+, AML with inv(16), AML with t(8;21), AML with t(15;17), AML with complex karyotype, and Normal bone marrow samples; all with 10 samples each except ALL with MLL rearrangement (9)). After evaluation of differentially regulated genes based on a t-statistic, the genes were arranged in ascending order. Based on this ordered list, the GSEA gives a positive score (y axis) for every gene encountered that belongs to the DNA repair process (green curve). Each hit is represented by a vertical bar on the x axis. The stronger the enrichment of genes belonging to DNA repair, the higher will be the density of bars on the left most side of the x axis.

3.5. Verification of Potential Target Genes by Real Time PCR (Taqman® Low Density Array)

The differential expression of genes found in microarray study is often validated using real-time reverse transcription PCR (RT-PCR) assays.

Real-time reverse-transcriptase (RT) PCR is able to quantify the amount of the template very specifically, sensitively and reproducibly. Real-time PCR monitors the fluorescence

emitted during the reaction as an indicator of amplicon production during each PCR cycle (i.e. in real time) as opposed to endpoint detection (Dorak, 2006). In comparison to conventional RT-PCR, real-time PCR also offers a much wider dynamic range of up to 10^7-fold (compared to 1000-fold in conventional RT-PCR). The dynamic range of any assay determines how much the target concentration can vary and still be quantified. A wide dynamic range ensures that a wide range of ratios of target and normalizer can be assayed with equal sensitivity and specificity (Dorak, 2006). It follows that the broader the dynamic range is, the more accurate the quantitation is. In real-time RT-PCR assays, expression is normalized to a control or housekeeping gene.

An important parameter in real-time PCR analysis is the C_t value, a threshold cycle that should be placed above any baseline activity and within the exponential phase (which looks linear in the log transformation) (Dorak, 2006) of PCR amplification. The Ct is thus the cycle at which the PCR reaction has reach mid-phase of the exponential increase phase, and it serves as a reference cycle for subsequent analysis and comparison between different samples in a given experiment.

TaqMan® Low-Density Arrays based real-time PCR is a medium-throughput method using microfluidics to simultaneously assay the expression of many genes (Lynne V. Abruzzo *et al.*, 2005). For the details of experimental setup, please see Materials and methods section on the Taqman Low Density Array.

We selected ninety four genes from the list of significantly differentially regulated genes (comparing pRTS-1-FLAG-CALM/AF10 samples (pRTS-1-FLAG-CALM/AF10 24hrs and pRTS-1-FLAG-CALM/AF10 72 hrs) to the controls (pRTS-1 0hrs, pRTS-1-FLAG-

CALM/AF10 0hrs, pRTS-1 24 hrs and pRTS-1 72 hrs)) for validation of their expression changes by real-time RT-PCR on the LDA (the genes are listed in Appendix A7). Two house-keeping genes (β-actin and GAPDH) were included in the validation real-time PCR experiment. The differentially regulated genes were selected based on the overall assessment of the analysis of the microarray data and their potential and known biological significance in leukemogenesis and hematopoietic development.

The cDNA that was used for the LDA real-time PCR was produced from the same total RNA that was used as a starting material for the microarray study. All samples from all time points (non-induced, 24 hrs after induction, and 72 hrs after induction) were included in the real-time PCR study. The cDNA synthesis was performed 3 times from the RNA that was used as a starting material for the microarray study and each synthesis product was used as a replicate for the LDA experiment. Samples were loaded randomly on an LDA (no systematic order) to avoid bias. The Materials and Methods section on the Taqman Low Density Array describes the LDA protocol and machine setup.

Once the data were acquired successfully, the individual C_t values of the assayed genes (Figure 3.18) were normalized against the geometric mean of the C_t value of the two internal control genes (β-actin and GAPDH). The raw amplification kinetics of the 94 genes on the TaqMan system is shown in figure 3.18. As can be seen in this figure most of the genes had a C_t value in the range of 25 to 37 (80.1% of the genes). Very few genes (3.7%) showed a relatively narrow C_t value range (high expression level), ranging from 20 to 22. After completion of the normalization, the data points were exported to a spreadsheet for further statistical analysis.

The fold expression change for the 94 genes was calculated using the same approach used for the microarray data (Table 3.5), the only difference being in the LDA data that the mean expression value (and standard deviation) for each gene in a given sample was calculated based on the 3 replicates per sample that were included in the LDA experiment.

The list of genes and their fold expression for the comparison pRTS-1-FLAG-CALM/AF10 samples (pRTS-1-FLAG-CALM/AF10 24hrs and pRTS-1-FLAG-CALM/AF10 72 hrs) to the controls pRTS-1 0hrs, pRTS-1-FLAG-CALM/AF10 0hrs, pRTS-1 24 hrs and pRTS-1 72 hrs) in both the real-time PCR analysis and in the microarray experiments is given in Appendix A7.

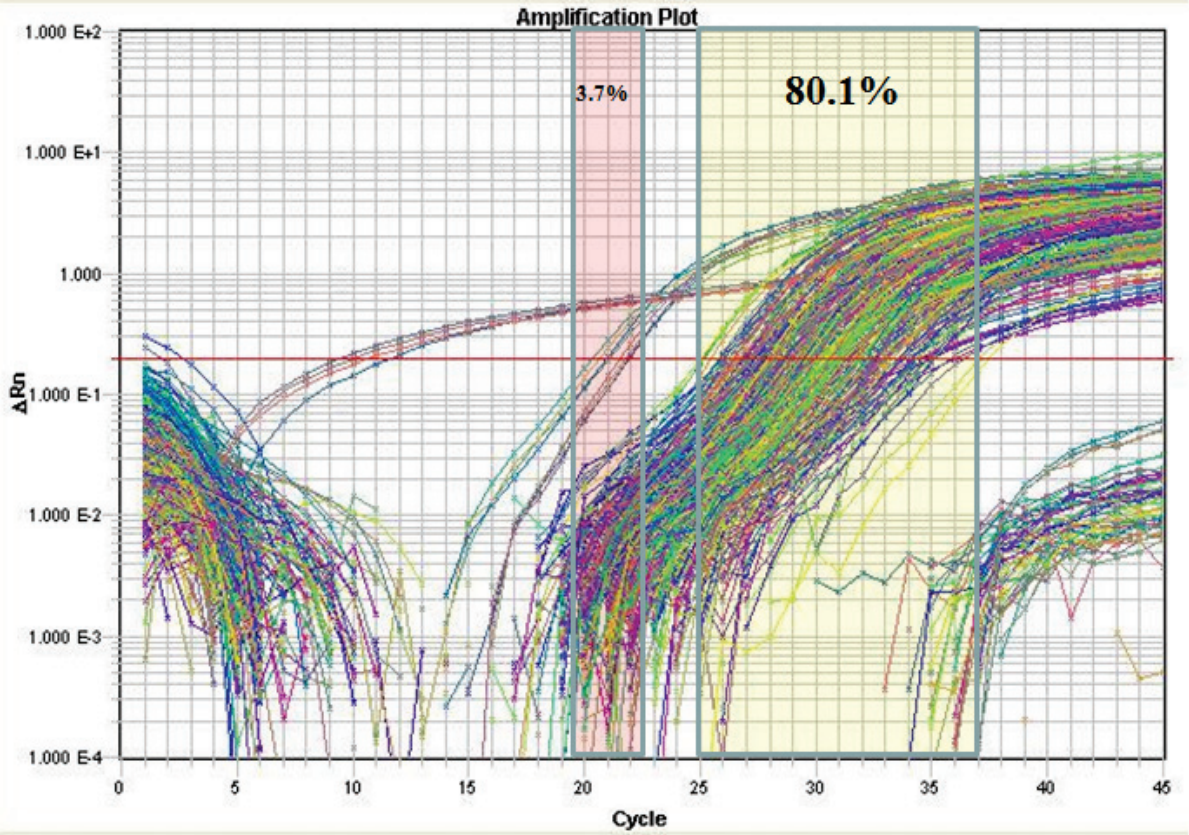


Figure 3.18. Raw real-time PCR data of the 94 genes and all samples in the experiment. Y axis is the magnitude of the signal generated by the given set of PCR condition and is calculated by taking the difference between Rn+ (the reaction containing all components, including the template) and Rn- (the reaction without the template) (Rn is the fluorescence emission intensity of the reporter dye divided by the fluorescence emission intensity of the passive reference dye). Red horizontal line is the C_t value, which is calculated automatically by the software). The X axis represents the cycle number. Different colors indicate readings for different genes. 80.1% of the assayed genes had Ct values ranging from 25 to 37.

3.5.1. Comparison between array and LDA data

The fold expression changes obtained from the LDA real-time PCR analysis were compared to the fold expression changes observed in the microarray data, based on the comparison pRTS-1-FLAG-CALM/AF10 samples (pRTS-1-FLAG-CALM/AF10 24hrs and pRTS-1-FLAG-CALM/AF10 72 hrs) to the controls (pRTS-1 0hrs, pRTS-1-FLAG-CALM/AF10 0hrs, pRTS-1 24 hrs and pRTS-1 72 hrs). A Pearson's correlation analysis

revealed that there was a strong correlation between the two data sets (table 3.8 and figure 3.19). If one considers the absolute value of the fold expression change as shown in figure 3.19., 78% of the genes in the validation list were consistent in the direction of their regulation (most genes up-regulated in the microarray data set upon CALM/AF10 expression were also found to be up-regulated in the LDA analysis; and the same was true for the down-regulated genes). Some gene that did not show matching direction of deregulation include IFI27, USP18, TLX2, RAB8B.

The difference in the magnitude of differential regulation between the real-time PCR data and the microarray data was also evaluated for the pRTS-1-FLAG-CALM/AF10 24hrs and pRTS-1-FLAG-CALM/AF10 72 hrs) to the controls (pRTS-1 0hrs, pRTS-1-FLAG-CALM/AF10 0hrs, pRTS-1 24 hrs and pRTS-1 72 hrs) comparison. It was observed that there is an overall higher magnitude of differential regulation (on average 2.5 fold higher) in the real time PCR data as compared to the microarray data. This difference in the magnitude of fold expression can be accounted for by the greater dynamic range of real-time PCR as compared to the Affymetrix microarray.

Table 3.8. Pearson's correlation of Affymetrix data and real-time PCR LDA data (In bold, significant values (except diagonal) at the level of significance alpha=0.050 (two-tailed test))

	LDA	Affymetrix
LDA	1.000	**0.780**
Affymetrix	**0.780**	1.000

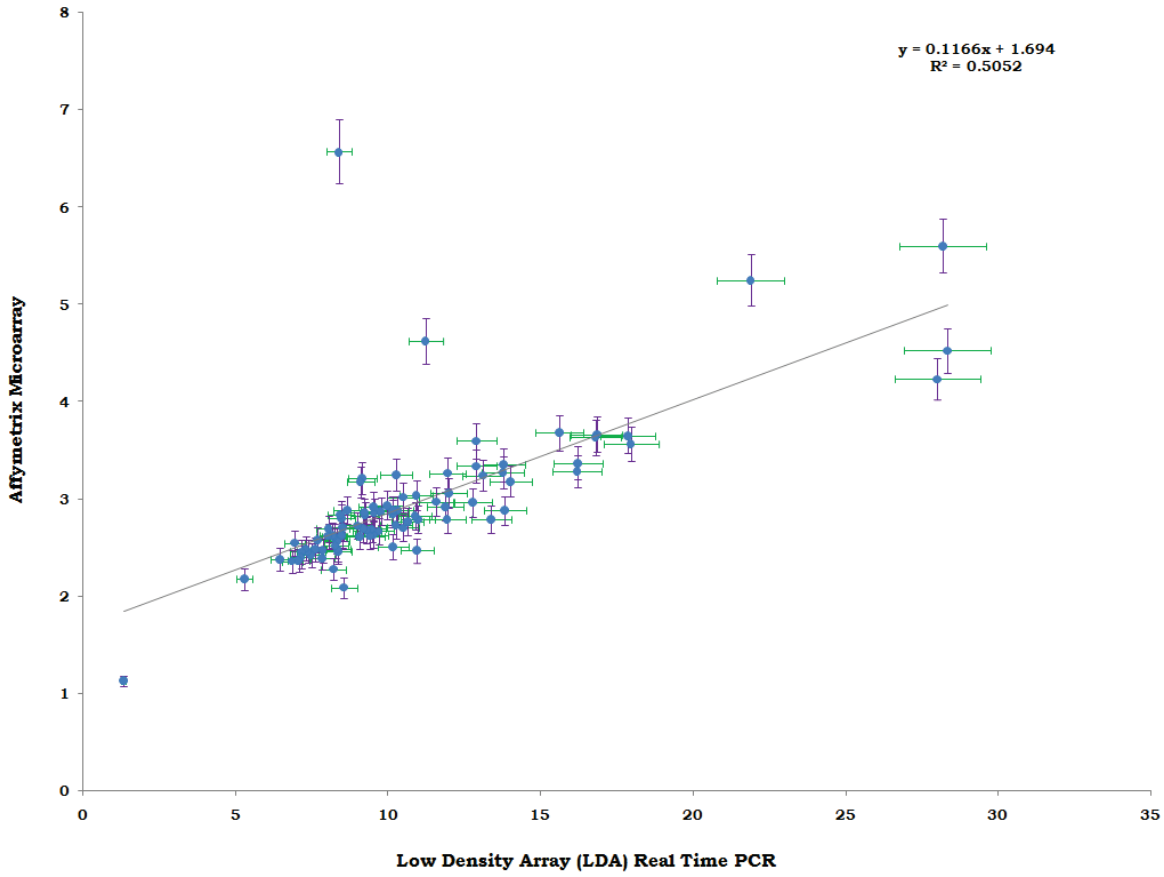


Figure 3.19. Correlation of microarray data and LDA real-time PCR data. The X axis represents the fold difference in regulation of the selected genes in the real-time PCR setting for the comparison of pRTS-1-FLAG-CALM/AF10 24hrs and pRTS-1-FLAG-CALM/AF10 72 hrs to the controls (pRTS-1 0hrs, pRTS-1-FLAG-CALM/AF10 0hrs, pRTS-1 24 hrs and pRTS-1 72 hrs) (log2 scale, green error bars indicating standard deviation for the fold regulation based on the technical replicates). Y axis indicates fold regulation for microarray data for the same comparison criterion, purple error bars indicating 95% confidence intervals.

IV. Discussion

It was the aim of this work to identify the early or immediate targets of CALM/AF10. To achieve this CALM/AF10 was expressed from the inducible mammalian expression vector pRTS-1 in the cell line DG-75, and the changes in the expression levels of ~20,000 genes was measured by expression profiling on the Affymetrix platform. The conditional expression of a gene of interest gives control over the time and dose of expression of the gene to be induced. As clean and appealing as a conditional expression system is there are some inherent limitations which have to be considered. The choice of the cell line plays a critical role in an experiment like ours, as the appropriateness of the cell line for the problem to be addressed and the technical feasibility of using the cell line have to be reconciled. For instance, out of the different cell lines screened for this work, the DG-75 cell line performed best with regard to inducibility and absence of leaky expression of CALM/AF10 prior to induction. DG-75 is a human memory pro B cell line (Burkitt's lymphoma cell line) and is certainly appropriate for this work as it is a hematopoietic. However, this cell line is a partially differentiated cell line and has accumulated certain epigenetic modifications that might mask the effect of CALM/AF10 on certain target genes. Despite the afro mentioned limitations and other intrinsic setbacks, gene expression profiling using Affymetrix® microarray of the different samples collected at different time points after induction of expression of CALM/AF10 provided us with a good overview of the transcriptional consequence of CALM/AF10 both at the gene and pathway level.

It is quite obvious that the number of differentially regulated genes in this is study is very large and these results cannot be discussed in all the details. Therefore, we chose to focus our discussion on the genes and pathways that appeared most relevant to us and that were validated in our real time PCR experiments (see results section and figure 4.1).

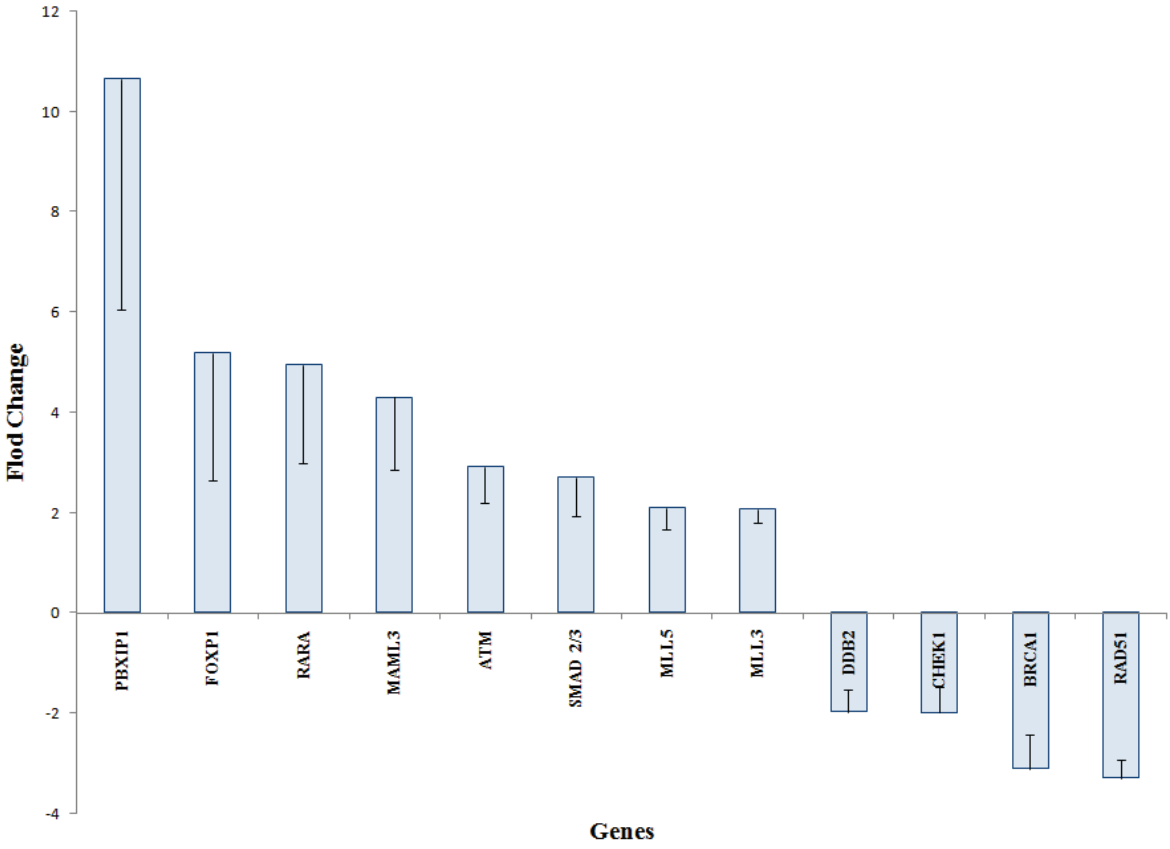


Figure 4.1. Some selected genes differentially regulated after CALM/AF10 expression in the DG-75 cell line.

4.1. Some genes implicated in cancer are immediate targets of CALM/AF10

BRCA1

Several known oncogenes and tumor suppressor genes were found to be differentially regulated after CALM/AF10 expression. *BRCA1* (breast cancer 1, early onset), is one of these genes. *BRCA1* was down-regulated in DG-75 cells expressing CALM/AF10.

BRCA1 encodes a nuclear phosphoprotein that plays a role in maintaining genomic stability and acts as a tumor suppressor. It was initially identified as the gene mutated in families with early onset breast cancer (Futreal *et al.*, 1994, Chen *et al.*, 1996). The BRCA1 protein associates with other tumor suppressor proteins, DNA damage sensors, and signal transducers to form a large multi-subunit protein complex known as BASC for BRCA1-associated genome surveillance complex (Wang *et al.*, 2000). BRCA1 muations are responsible for approximately 40% of familial breast cancers and for more than 80% of familial breast and ovarian cancers and it has recently been implicated as a predictive marker of poorer overall survival in sporadic ovarian cancer (Weberpals *et al.*, 2009). It is tempting to speculate that the down-regulation of BRCA1 might play a role in the early stages of CALM/AF10-mediated oncogenesis by facilitating the accumulation of additional mutations which are necessary for the development of the full leukemic phenotype.

RARA

RARA (retinoic acid receptor, alpha) was another gene that was found to be up-regulated in this experiment. Translocations which involve the *RARA* gene are a cardinal feature of acute promyelocytic leukemia (APL) (Goddard *et al.*, 1992), the most common of these translocations is the t(15;17)(q21;q22), which fuses the *RARA* gene with the *PML* gene (Vitoux *et al.*, 2007). The result of the different *RARA* fusions is a loss of wild type *RARA* function. Therefore, the significance of a higher expression of *RARA* in the context of leukemogenesis is not clear. It has been shown in gene profiling of colonic serrated adenomas by using oligonucleotide microarray that *RARA* is one of the three significantly up-regulated genes (Kim *et al.*, 2008). In contrast, it has recently been reported that

diminished *RARA* expression contributes to leukemogenesis in acute myeloid leukemia cell lines, though the regulation of expression of this gene was very complex and varies at different stages of the disease (Glasow *et al.*, 2008).

FOXP1

Forkhead box transcription factors play important roles in the regulation of tissue- and cell type-specific gene transcription during both development and adulthood. One of these genes, *FOXP1* (forkhead box P1), was found to be one of the immediate targets of CALM/AF10. The up-regulation of *FOXP1* has also been associated with invasive breast carcinomas (Bates *et al.*, 2008). It has also been reported that deregulation of *FOXP1* expression (over-expression) plays an important role in lymphoma development (Goatly *et al.*, 2008). Similarly, up-regulated mRNA expression of the *FOXP1* forkhead transcription factor in response to normal B-cell activation and high expression in a poor prognosis subtype of diffuse large B-cell lymphoma (Banham *et al.*, 2005) is an indication that this gene is one of the more interesting targets of CALM/AF10.

PBXIP1

PBXIP1 (pre-B-cell leukemia homeobox 1 interacting protein 1), also known as *HPIP* (hematopoietic PBX-interacting protein), was up-regulated upon the expression of CALM/AF10. *PBX1* is a transcription factor originally discovered as the chromosome 1 participant in the translocation t(1;19) which results in the E2A/PBX1 fusion protein, seen frequently in pediatric acute lymphoblastic leukemia (Kamps *et al.*, 1990). *HPIP* interacts with *PBX1*, which is a homeodomain protein that functions in complexes with other homeodomain-containing proteins to regulate gene expression during

developmental and/or differentiation processes (Abramovich *et al.*, 2000). Abramovich *et al.* (2000) showed that *HPIP* inhibits the ability of PBX1-HOX heterodimers to bind to target sequences and suggested that *HPIP* is a new regulator of *PBX1* function (Abramovich *et al.*, 2002).

4.2. The *MLL* gene family

MLL (myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila)) is a frequent target for recurrent translocations in acute leukemias that may be characterized as acute myeloid leukemia, acute lymphoblastic leukemia (ALL), or mixed lineage (biphenotypic) leukemia (MLL) (Krivtsov and Armstrong, 2007). Leukemias with translocations involving *MLL* possess unique clinical and biologic characteristics and are often associated with poor prognosis (Krivtsov and Armstrong, 2007). *MLL* rearrangements are found in more than 70% of infant leukemias, where the immunophenotype is more consistent with ALL or AML-M6, but are less frequent in leukemias from older children. *MLL* translocations are also found in approximately 10% of AMLs in adults (Matthew *et al.*, 2008), as well as in therapy-related leukemias, that develop in patients previously treated with chemotherapeutic agents (mainly topoisomerase II inhhibitors) for other malignancies (Millot *et al.*, 2005). A key feature of MLL fusion proteins is their high efficiency in transforming hematopoietic cells (Krivtsov and Armstrong, 2007). The *MLL* gene encodes a DNA-binding protein that methylates histone H3 lys4 (H3K4) and positively regulates expression of target genes, notably multiple *HOX* genes (Ansari *et al.*, 2008). MLL-like genes are part of the trithorax-polycomb system which is crucial in maintaining gene expression patterns established during developmental processes (epigenetic memory).

In the current experiment, two members of the *MLL* gene family (MLL3 and MLL5) were found to be up-regulated. *MLL3* encodes a nuclear protein and has not yet been associated with malignancies while *MLL5* has been shown to upregulated in certain subsets of myeloid leukemias (Emerling *et al.*, 2002). Though the role of *MLL5* overexpression in cell cycle regulation has been reported (Deng *et al.*, 2004, Cheng *et al.*, 2008), additional functional studies are required to assess the role of *MLL3* or *MLL5* up-regulation in CALM/AF10-mediated leukemogenesis.

4.3. Important Signalling Pathways are Affected by CALM/AF10

4.3.1. Notch signaling

Notch signaling is one of the most important pathways in the cell and controls differentiation and proliferation in various cell types and is associated with several diseases (Karlsson *et al.*, 2008). The *Notch* family of evolutionarily conserved proteins regulates a broad spectrum of cell-fate decisions and differentiation processes during fetal and post-natal development. The best characterized role of *Notch* signaling during mammalian hematopoiesis and lymphopoiesis is the essential function of the *Notch1* receptor in T-cell lineage commitment. More recent studies have addressed the roles of other *Notch* receptors and ligands, as well as their downstream targets, revealing additional novel functions of *Notch* signaling in intra-thymic T-cell development, B-cell development and peripheral T-cell function. It has been shown that more than 50% of human T-ALLs, including tumors from all major molecular oncogenic subtypes, have activating mutations that involve the extracellular heterodimerization domain and/or the C-terminal PEST domain of *NOTCH1* (Weng *et al.*, 2004). Importantly, the CALM/AF10 fusion gene is most frequent fusion gene in T-ALL with T-cell receptor γ/δ

rearrangements, being present in 20% of the cases. Aberrant *Notch* signaling has also been found in other neoplastic and non-neoplastic conditions like osteoarthritis (Karlsson *et al.*, 2008), melanoma (Bedogni et al., 2008) and mesothelioma (Graziani et al., 2008).

In this study, it was observed that one of the positive regulators of *Notch* signaling, *MAML3* (mastermind-like 3 (Drosophila)), is significantly up-regulated. *Mastermind* (*Mam*) is one of the evolutionarily conserved elements of the *Notch* signaling pathway. The *Drosophila* gene *mastermind* is an important positive regulator of the pathway (Lin *et al.*, 2002). Lin *et al.* (2002) discussed that both MAML2 and MAML3 stabilize and are part of the DNA-binding complex containing RBP-J/CBF-1 and the Notch intracellular domain, which assembles at the promoter of *Notch* target genes. It is very likely that *MAML3*-mediated *Notch* activation plays a role in CALM/AF10-mediated leukemogenesis since all CALM/AF10 positive patients (both AML and ALL) have a TCR γ/δ rearrangement. In addition, we observed that *NOTCH1* itself is up-regulated in CALM/AF10 patients using microarray analysis (unpublished data).

4.4. DNA damage check-point and DNA repair

Interestingly, several genes involved in DNA damage sensing, DNA damage check-point and DNA repair were found deregulated after expression of CALM/AF10. Most of these genes were found to be down-regulated.

One of the very interesting genes in the DNA damage check-point pathway is *ATM* (ataxia telangiectasia mutated), a gene that belongs to the PI3/PI4-kinase family and plays an important role in cell cycle control after DNA damage. This gene was found to be significantly up-regulated. Mutations in the *ATM* gene lead to Ataxia teleangiectasia, a

recessive autosomal disease characterized by cerebellar ataxia (progressive difficulty in walking and other movements), teleangiectasia (dilatation of small arteries) and immunodeficiency. It is also characterized by increasing sensitivity to radiations and by a high probability of developing leukemias or other tumours.

ATM functions as a regulator of a wide variety of downstream proteins, including tumor suppressor proteins like TP53 and BRCA1, the checkpoint kinase CHK2, the checkpoint proteins RAD17 and RAD9, and the DNA repair protein NBS1 (Antoccia *et al*., 2008). *ATM* is crucial for the maintenance of chromosomal integrity. The absence of *ATM* kinase activity may lead to the rapid accumulation of irreversible chromosome damage (White *et al.*, 2008). Mutations and deletions of *ATM* are frequently found in chronic lymphocytic leukema (CLL) (Wei *et al.*, 2008) and childhood malignancies (Fabienne *et al.*, 2008). Some findings also suggest that overexpression of *ATM* may be one of the early events in the carcinogenesis of oral squamous cell carcinoma (He *et al.*, 2008). It has been reported that there is an increased need for ATM signaling for double strand break and repair as the heterochromatic component of a genome expands (Goodarzi *et al.*, 2008). Choo and co-workers showed that *ATM* can signal to cyclin D1, a key molecule for transition between the G1 and S phases of the cell-cycle with pivotal roles in the development of several human cancers, may be required for maintenance of genomic integrity achieved by rapid arrest of the cell-cycle and that disruption of this crosstalk may increase susceptibility to cancer (Choo *et al.*, 2009).

Down-regulation of *DDB2* (damage-specific DNA binding protein 2), a gene involved in cellular response to DNA damage with a direct DNA binding activity (Kulaksiz *et al.*, 2005), is also interesting as it further signifies the impact that CALM/AF10 might have in

DNA damage and repair. Interestingly, *BRCA1* upregulates *DDB2*, with some evidence of P53 involvement in its regulation (El-Deiry, 2002), via interactions with the RAD51 protein family. Interestingly, *RAD51* (RAD51 homolog (RecA homolog, E. coli) (*S. cerevisiae*)) was found to be down-regulated in the current data. The interaction between *RAD51* and *BRCA1/BRCA2* was found to be important for the cellular response to DNA damage results indicate that alterations in *RAD51* activity may contribute to the disturbances of DNA repair involving *RAD51* and/or *BRCA2* penetration and thus enhance the risk of breast cancer development (Maria *et al.*, 2008). The association of *RAD51* and the BRCAs has been credited to *CHEK1* (another down-regulated gene in this study) and *CHEK2* (Bahassi *et al.*, 2008).

Up-regulation of *ATM* and differential regulation of various genes involved in genomic stability and DNA damage repair in the current study indicate that CALM/AF10 expression might induces genomic instability and thereby affecting DNA break and repair pathways which in turn might lead to cell cycle arrest.

4.5. Cell cycle is one of the most dramatically affected pathways by CALM/AF10

The connection between cell cycle control and cancer is quite obvious. The cell cycle machinery controls cell proliferation, and cancer is a disease of inappropriate cell proliferation (Collins *et al.*, 1997). Fundamentally, all cancers permit the existence of too many cells. However, this excess number of cells is linked in a vicious cycle with a reduction in sensitivity to signals that normally tell a cell to adhere, differentiate, or die (Collins *et al.*, 1997). At least two types of cell cycle control mechanisms are recognized:

a cascade of protein phosphorylations (involving a highly regulated kinase family) that are important for the progression from one stage of the cell cycle to the next, and a set of checkpoints that monitor completion of critical events and delay progression to the next stage if necessary (Collins *et al.*, 1997).

In this study, cell cycle was found to be one of the most dramatically affected pathways by CALM/AF10 which may be as a result of the significant impact of the fusion gene on DNA damage checkpoint genes (*ATM* being the checkpoint that mediates DNA damage checkpoint and cell cycle as discussed in the previous section). As can be noted from the pathway analysis (see results section on Pathway analysis sub-section Cell cycle), a significant proportion of the genes in cell cycle (approx. 50%) are deregulated.

Another entry point into the cell cycle from other pathways is via *SMAD2/3*, which mediate the link between TGFβ signaling and cell cycle, playing a role in cellular proliferation, apoptosis, actin cytoskeleton regulation, cell motility, transcription, and Ras or insulin signaling (Brown *et al.*, 2008). The *SMAD* genes *SMAD2 and SMAD3* were significantly up-regulated in the current study. Aberrant expression of *SMAD2* plays a role in hepatocellular carcinoma (Wu *et al.*, 2007) and its up-regulation has been associated with Glioma (Bruna *et al.*, 2007). It has also been reported that breast cancer patients with high expression of *SMAD2* tended to have a better prognosis (Liapis *et al.*, 2007).

The combined impact of DNA damage checkpoint and TGFβ signaling via *ATM/ATR* and *SMAD2/3*, respectively, seems to indirectly lead to cell cycle arrest as most of the critical

positive regulators of cell cycle are down-regulated in CALM/AF10 expressing DG-75 cells.

4.6. Genes close to the breakpoint of either AF10 or CALM are not immediate targets of CALM/AF10

In a microarray-based comparison of CALM-AF10+ T-ALL to CALM-AF10- T-ALL, it was reported that the genes located close to AF10 breakpoint of t(10;11)(p13;q14-21) (*DNAJC1, COMMD3, BMI1* and *SPAG6*) were found to be significantly up-regulated (Dik *et al.*, 2005). In our CALM/AF10 patient microarray data set (section 3.4.7.), we were able to confirm that these genes were up-regulated (unpublished data). Interestingly, these genes were also found to be up-regulated in a CALM/AF10 mouse bone marrow transplantation model (except Spag6) despite the fact that these mice do not have the t(10;11)(p13;q14-21) translocation (unpublished data), implying that the up-regulation of these genes might be due to the direct/indirect impact of the expression of CALM/AF10 and not as a consequence of the translocation event.

We analyzed the expression levels of these genes (*DNAJC1, COMMD3, BMI1* and *SPAG6*) in the current data set. None of these genes was found to be statistically significantly deregulated, indicating that they might be intermediate-late targets of CALM/AF10 or that strong repressive chromatin modifications are present at the these genes in the DG75 cell line.

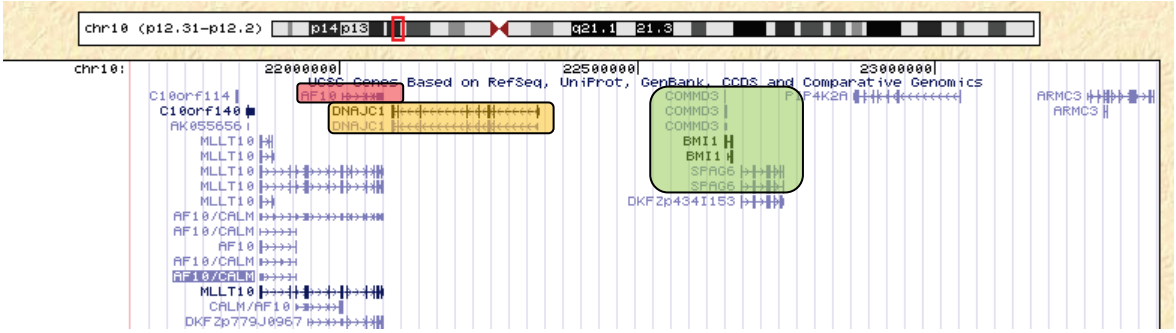


Figure 4.2. Location of AF10 and proximal genes that were analyzed (± 500 kbp)

4.7. Genomic instability and DNA repair might be the oncogenic pathway of CALM/AF10 as revealed by comparative pathway analysis of patient and cell line data

It has been widely reported that genomic instability is one of the key features of malignancies. For instance, cases of ETP (Early T-cell Precursors)-ALL showed increased genomic instability, in terms of number and size of gene lesions, compared with those with typical T-ALL (Coustan-Smith *et al.*, 2009). Similar results were also reported in proB acute lymphoblastic leukemia (Bumbea *et al.*, 2007). The possible involvment of DNA damage and aberrant repair associated with internal tandem duplication of FLT3 (FLT3/ITD) for poor prognosis in AML (Sallmyr *et al.*, 2008) indicates that genomic instability may play a pivotal role in disease progression of AML. The role of genomic instability and the probable causes and consequences has recently been reviewed (Annahita *et al.*, 2008). Annahita and co-workers stated that disease progression in myeloid malignancies results from the accumulation of "mutations" in genes that control cellular growth and differentiation but also pointed out that the mechanism(s) by which these cells acquire genetic alterations or genomic instability is

less well understood. In reference to DNA-microarray and real-time PCR based expression profiling studies, a recent review also indicated that deviant gene expression patterns in acute myeloid leukemias should be assessed on the background of gene expression levels of housekeeping functions, particular differentiation stages and epiphenomena due to genomic instability (Oyan *et al.*, 2007).

Using the Gene Set Enrichment Analysis (GSEA) we could clearly show that DNA repair and genomic stability related pathways are significantly affected in both the CALM/AF10 patients' microarray data and in the microarray data of the DG-75 cell line after CALM/AF10 expression. It was also observed that the genomic stability and DNA repair genes that are differentially regulated in the two different settings were not the same, indicating that early targets (cell line data) might be different from those found in patients. This might be explained by the difference in the nature of the samples in that what is observed in the cell line we see the immediate impact of CALM/AF10 in a very controlled setting while in the CALM/AF10 positive patient samples the microarray study would detect late or terminal targets. It should also be noted that in the patient microarray study, the samples are compared with other types of leukemia and normal bone marrow, which might not be the ideal controls. Despite the huge differences in the systems used (controlled cell line system versus "messy" patient samples) and the different comparisons made (induced versus uninduced, and CALM/AF10-posistive versus other leukemia) , it was still observed that CALM/AF10 affects genomic stability. This indicates that alterations of genomic stability might be one of the very important pathways in CALM/AF10 mediated leukemogenesis.

V. Summary

The t(10;11)(p12;q14) is a recurring chromosomal translocation that is found in acute myeloid and acute lymphoblastic leukemia as well as in malignant lymphoma. This translocation results in the fusion of *AF10,* encoding a putative zinc finger transcription factor containing an N-terminal LAP/PHD zinc finger motif, a nuclear localization signal, an AT-hook domain, and a leucine zipper and with the CALM gene (Clathrin assembly protein lymphoid myeloid leukemia gene) that encodes a clathrin assembly protein. In the monocytic cell line U937 both the CALM/AF10 and the AF10/CALM fusion mRNAs can be identified. The CALM/AF10 fusion mRNA codes for the CALM/AF10 fusion protein which contains almost the complete CALM protein fused in frame to about 90% of the AF10 protein without the N terminal plant homeo domain (PHD) domain. CALM/AF10 is highly leukemogenic with its expression in primary bone marrow cells leading to the development of an aggressive acute leukemia with a short latency of 10 weeks in a murine bone marrow transplant model. We set out to identify immediate target genes of CALM/AF10. The fusion gene was cloned into pRTS-1, an episomally replicating tetracycline/doxycycline inducible (tet-on) expression vector. The construct was stably transfected into the Burkitt's lymphoma B cell line DG75. The expression of CALM/AF10 after induction was confirmed by RT-PCR. Gene expression profiling experiments were conducted using the Affymetrix® Human Genome U133 Plus 2.0 Array analyzing non-induced and induced (24 and 72 hours after induction) samples. As controls, DG75 cells transfected with the pRTS-1 vector without the CALM/AF10 fusion gene were used. The results were analyzed using the R statistical package and dChip (DNA Chip Analyzer) software. The expression of 1237 genes was found to be changed at least 2 fold 24 hours after the induction of CALM/AF10. 594 (48%) genes were downregulated, while 643 (52%) were upregulated. Downregulated genes included genes involved in DNA repair (*DDB2*, *TOP2A*, *BRCA1*), cell cycle check point control (*CCNE2*, *CHEK1*, *CDC2*), and chromosome maintenance (*MCM3*, *MCM7*, *MCM10*). Upregulated genes included signal transduction molecules like RAB8B (a member of the *RAS* oncogene family) and *STAT* family members (*STAT1* and *STAT2*), chromatin

remodeling factors like *BAZ2A* (bromodomain adjacent to zinc finger domain, 2A), and *MAML3* (master mind like 3), a positive regulator of *Notch* signaling. A pathway analysis using KegArray showed an enrichment of differentially regulated genes in processes like "cell cycle regulation" and "DNA replication and repair". Interestingly, no significant upregulation of *Hox* genes was observed, as was previously reported in a study of CALM/AF10 positive patient samples (Dik *et al.*, 2005). The changes in the expression levels of some selected genes were confirmed using a Taqman® Low Density Array (LDA). This analysis showed a statistically significant correlation ($r = 0.78$, $p = 0.001$) between the real time PCR results and the expression levels obtained from the Affymetrix arrays. A gene set enrichment analysis, comparing the data obtained after inducible expression of CALM/AF10 in DG75 cells and data obtained from gene expression profiling in CALM/AF10-positive patient samples, showed a significant deregulation of genes involved in DNA repair and genomic stability both in the differentially regulated genes in the cell line experiments and in the patient data. Our results demonstrate that the expression of the leukemogenic CALM/AF10 fusion protein leads to a severe deregulation of critical cellular processes, and we could identify candidate genes as direct target of the CALM/AF10 fusion protein.

Zusammenfassung

Die t(10;11)(p12;q14) ist eine wiederkehrende chromosomale Translokation, die man bei der akuten myeloischen und akuten lymphoblastischen Leukämie (AML/ALL) und auch bei malignen Lymphomen findet. Diese Translokation führt zu einer eine Fusion des *AF10*-Gens, mit dem CALM Gen. AF10 kodiert einen putativen Zinkfinger-Transkriptionsfaktor, welcher ein N-terminales LAP/PHD Zink-Fingermotiv, ein nukleäres Lokalisationssignal, ein AT-Hook und ein Leucinzipper enthält. *CALM* hingegen kodiert das *Clathrin Assembly Lymphoid Myeloid Leukemia* Protein und spielt bei der rezeptorvermittelten Endozytose eine wichtige Rolle. In der monozytischen Zelllinie U937 konnten die *mRNAs*, die für CALM/AF10- und AF10/CALM-Fusionsproteine kodieren, identifiziert werden. Die CALM/AF10 *mRNA* beinhaltet fast das komplette *CALM*-Gen, an das fast 90% des offenen Leserahmens des *AF10*-Gens anschließen. Lediglich die N-terminale PHD-Domäne des *AF10*-Gens fehlt in dieser Fusions-*mRNA*.

Die Expression eines CALM/AF10 Fusionproteins in primären Knochenmarkszellen ist hochgradig leukämogen im Maus-Transplantationsmodell, da nach einer sehr kurzen Latenzzeit von nur 10 Wochen eine akuten Leukämie in den Mäusen diagnostizierbar ist.

Ziel dieser Arbeit war es, bisher unbekannte, früher Zielgene des CALM/AF10-Fusionsproteins zu identifizieren. Dazu wurde das Fusions-Gen in den episomalen pRTS-1 Vektor einkloniert, dessen Promotor mit Hilfe des Tetrazyklin-Derivats Doxyzyklin conditional aktivierbar ist (Tet-on System). Dieses Kontrukt wurde dann in die Burkitt-Lymphomzelllinie DG75 stabil transfiziert und die Expression von CALM/AF10 nach Induktion des Vektors mittels RT-PCR überprüft. Als Negativkontrolle dienten DG75 Zellen, die mit einem pRTS-1 Vektor ohne Fusionsgen stabil transfiziert worden waren. Mit Hilfe des Genchips Affymetrix® hgU133 plus 2.0 wurden Genexpressionsprofile erstellt. Um zwischen frühen und späten effekten der CALM/AF10 Expression unterscheiden zu können, wurden Genchipanalysen zu unterschiedlichen Zeitpunkten (24h und 72h nach Vektorinduktion) durchgeführt.

Die statistische Datenanalyse des Genexpressionsprofil wurde mit Hilfe der Software "R" und der dCHip Software durchgeführt und ergab folgende Ergebnisse: 24 Stunden nach der Induktion von CALM/AF10 hatte sich das Expressionsniveau von 1237 Genen um mindestens das 2 fache des Ausgangswertes verändert. Bei 594 Genen (=48%) erniedrigte sich das Expressionniveau, im Gegensatz zu 643 Genen (=52%), deren Expressionsniveau sich erhöhte. Unter den Genen, deren Expressionniveau sich erniedrigte, fanden sich Schlüsselgene der DNA-Reparatur (*DDB2*, *TOP2A*, *BRCA1*), der Zellzykluskontrolle (*CCNE2*, *CHEK1*, *CDC2*) und der Chromosomen-Stabilität (*MCM3*, *MCM7*, *MCM10*). Hochregulierte Gene kodierten für Signaltransduktionsmoleküle wie RAB8B (ein Mitglied der *RAS* Onkogen-Familie), Chromatin-Remodelling Faktoren wie *BAZ2A* (bromodomain adjacent to zinc finger domain, 2A), (*STAT*-Familienmitglieder (*STAT1* und *STAT2*) und *MAML3* (mastermind like 3), das den *Notch*-Signalweg positiv

reguliert. Die Einordnung der differenziert regulierten Gene in Signalwegskaskaden mittels KegArray zeigte eine Anreicherung der differentiell regulierten Gene im Kategorien wie "Zellzyklusregulation" und "DNA-Replikation und Reparatur". Einen Anstieg des Expressionsniveaus der *Hox* Gene konnte jedoch nicht beobachtet werden, obwohl dies zuvor in einer Patientenstudie gefunden wurde (Dik *et al*., 2005). Zur Verifizierung der gefundenen Ergebnisse wurden die Veränderungen im Expressionsniveau einiger zuvor ausgewählter Gene mittels Taqman® Low Density Array (LDA) bestätigt. Dabei zeigte sich eine statistisch-signifikante Korrelation ($r = 0.78$, $p = 0.001$) zwischen den Realtime-PCR-Ergebnissen und den mittels Affymetrix-Array ermittelten Veränderungen im Expressionsniveaus der einzelnen Gene. Eine Gene Set Enrichment Analyse zeigte zudem, dass Gene, die bei der DNA-Reparatur und der genomischen Stabilität eine Rolle spielen, sowohl bei bei den deregulierten Genen in unserem Zelllinіensystem als auch bei den differentiell regulierten Genen bei CALM/AF10-positiven Patienten signfikant überrepräsentiert sind. Zusammenfassend zeigen unsere Untersuchungen, dass die Expression des leukämogenen Fusionsproteins CALM/AF10 zu einer starken Deregulierung von Schlüsselgenen in kritischen, zellulären Prozessen führt, und wir konnten somit erste Kandidaten-Gene für frühe Zielgene des CALM/AF10 Fusionsproteins ermitteln.

VI. References

Abramovich, C., Shen, W.-F., Pineault, N., Imren, S., Montpetit, B., Largman, C. and Humphries, R. K. (2000). Functional Cloning and Characterization of a Novel Nonhomeodomain Protein That Inhibits the Binding of PBX1-HOX Complexes to DNA. *J. Biol. Chem.*, **275**, 26172-26177.

Abramovich, C., Chavez, E. A., Lansdorp, P. M. and Humphries, R. K. (2002). Functional characterization of multiple domains involved in the subcellular localization of the hematopoietic Pbx interacting protein (HPIP). *Oncogene*, **21**, 6766-6771.

Affymetrix (2005). Eukaryotic Sample and Array Processing.

Anika Meyerholz, L. H. S. G. P.-C. E. G. B. E. J. U. (2005). Effect of Clathrin Assembly Lymphoid Myeloid Leukemia Protein Depletion on Clathrin Coat Formation. *Traffic*, **6**, 1225-1234.

Annahita, S., Jinshui, F. and Feyruz Virgilia, R. (2008). Genomic instability in myeloid malignancies: Increased reactive oxygen species (ROS), DNA double strand breaks (DSBs) and error-prone repair. *Cancer Letters*, **270**, 1-9.

Ansari, K. I., Mishra, B. P. and Mandal, S. S. (2008). Human CpG binding protein interacts with MLL1, MLL2 and hSet1 and regulates Hox gene expression. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, **1779**, 66-73.

Antoccia, A., Sakamoto, S., Matsuura, S., Tauchi, H. and Komatsu, K. (2008). NBS1 Prevents Chromatid-Type Aberrations through ATM-Dependent Interactions with SMC1. *Radiation Research*, **170**, 345-352.

Aplan, P. D. (2006). Causes of oncogenic chromosomal translocation. *Trends Genet*, **22**, 46-55.

Ashihara, E., Nakamura, S., Inaba, T., Taki, T., Hayashi, Y. and Shimazaki, C. (2007). A novel AF10-CALM fusion transcript in g/d-T cell type lymphoblastic lymphoma. *American Journal of Hematology*, **82**, 859 - 860.

Bahassi, E. M., Ovesen, J. L., Riesenberg, A. L., Bernstein, W. Z., Hasty, P. E. and Stambrook, P. J. (2008). The checkpoint kinases Chk1 and Chk2 regulate the

functional associations between hBRCA2 and Rad51 in response to DNA damage. *Oncogene*, **27**, 3977-3985.

Banham, A. H., Connors, J. M., Brown, P. J., Cordell, J. L., Ott, G., Sreenivasan, G., Farinha, P., Horsman, D. E. and Gascoyne, R. D. (2005). Expression of the FOXP1 Transcription Factor Is Strongly Associated with Inferior Survival in Patients with Diffuse Large B-Cell Lymphoma. *Clin Cancer Res*, **11**, 1065-1072.

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., Zhao, K. and 1Laboratory of Molecular Immunology, N. H., Lung (2007). High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell*, **129**, 823-837.

Bates, G., Fox, S., Han, C., Launchbury, R., Leek, R., Harris, A. and Banham, A. (2008). Expression of the forkhead transcription factor FOXP1 is associated with that of estrogen receptorβ in primary invasive breast carcinomas. *Breast Cancer Research and Treatment*, **111**, 453-459.

Bedogni, B., Warneke, J. A., Nickoloff, B. J., Giaccia, A. J. and Powell, M. B. (2008). Notch1 is an effector of Akt and hypoxia in melanoma development. *J. Clin. Invest.*, **118**, 3660-3670.

Ben-Bassat, H., Goldblum, N., Mitrani, S., Goldblum, T., Yoffey, J. M., Cohen, M. M., Bentwich, Z., Ramot, B., Klein, E. and Klein, G. (1977). Establishment in continuous culture of a new type of lymphocyte from a "Burkitt like" malignant lymphoma (line D.G.-75). *Int J Cancer*, **19**, 27-33.

Biosystems, A. (2003). Applied Biosystems 7900HT Micro Fluidic Card Getting Started Guide

Bornkamm, G. W., Berens, C., Kuklik-Roos, C., Bechet, J., Laux, G., Bachl, J., Korndoerfer, M., Schlee, M., Hölzel, M., Malamoussi, A., Chapman, R. D., Nimmerjahn, F., Mautner, J., Hillen, W., Bujard, H. and Feuillard, J. (2005). Stringent doxycycline-dependent control of gene activities using an episomal one-vector system. *Nucleic Acids Res.*, **33**, e137.

Brown, K. A., Ham, A. J., Clark, C. N., Meller, N., Law, B. K., Chytil, A., Cheng, N., Pietenpol, J. A. and Moses, H. L. (2008). Identification of novel Smad2 and

Smad3 associated proteins in response to TGF-beta1. *Journal of Cellular Biochemistry*, **105**, 596-611.

Bruna, A., Darken, R. S., Rojo, F., Ocaña, A., Peñuelas, S., Arias, A., Paris, R., Tortosa, A., Mora, J., Baselga, J. and Seoane, J. (2007). High TGF²-Smad Activity Confers Poor Prognosis in Glioma Patients and Promotes Cell Proliferation Depending on the Methylation of the PDGF-B Gene. **11**, 147-160.

Bumbea, H., Serghei, L., Vlădăreanu, A. M., Stefan, L. M., Casleanu, D., Voican, I. and Begu, M. (2007). Cryptic genomic abnormalities associated with coexpression of KOR-SA3544 and NG.2 in proB acute lymphoblastic leukemia. *Rom J Intern Med.*, **45**, 387-391.

Caudell, D., Zhang, Z., Chung, Y. J. and Aplan, P. D. (2007). Expression of a CALM-AF10 Fusion Gene Leads to Hoxa Cluster Overexpression and Acute Leukemia in Transgenic Mice. *Cancer Res*, **67**, 8022-8031.

Caudell, D. and Aplan, P. D. (2008). The role of CALM-AF10 gene fusion in acute leukemia. *Leukemia*, **22**, 678-685.

Chaplin, T., Ayton, P., Bernard, O. A., Saha, V., Della Valle, V., Hillion, J., Gregorini, A., Lillington, D., Berger, R. and Young, B. D. (1995). A novel class of zinc finger/leucine zipper genes identified from the molecular cloning of the t(10;11) translocation in acute leukemia. *Blood*, **85**, 1435-1441.

Chen, Y., Farmer, A. A., Chen, C.-F., Jones, D. C., Chen, P.-L. and Lee, W.-H. (1996). BRCA1 Is a 220-kDa Nuclear Phosphoprotein That Is Expressed and Phosphorylated in a Cell Cycle-dependent Manner. *Cancer Res*, **56**, 3168-3172.

Cheng, F., Liu, J., Zhou, S. H., Wang, X. N., Chew, J. F. and Deng, L.-W. (2008). RNA interference against mixed lineage leukemia 5 resulted in cell cycle arrest. *The International Journal of Biochemistry & Cell Biology*, **40**, 2472-2481.

Choo, D. W., Baek, H. J., Motoyama, N., Cho, K. H., Kim, H. S. and Kim, S. S. (2009). ATM is required for rapid degradation of cyclin D1 in response to [gamma]-irradiation. *Biochemical and Biophysical Research Communications*, **378**, 847-850.

Collins, K., Jacks, T. and Pavletich, N. P. (1997). The cell cycle and cancer. *Proc. Natl.Acad. Sci. USA*, **94**, 2776-2778.

Coustan-Smith, E., Mullighan, C. G., Onciu, M., Behm, F. G., Raimondi, S. C., Pei, D., Cheng, C., Su, X., Rubnitz, J. E., Basso, G., Biondi, A., Ching-Hon Pui, C.-H., Downing, J. R. and Campana, D. (2009). Early T-cell precursor leukaemia: a subtype of very high-risk acute lymphoblastic leukaemia. *Lancet Oncology*, **10**, 147-156.

Deng, L.-W., Chiu, I. and Strominger, J. L. (2004). MLL 5 protein forms intranuclear foci, and overexpression inhibits cell cycle progression. *Proc. Natl.Acad. Sci. USA*, **101**, 757-762.

Deshpande, A. J., Cusan, M., Rawat, V. P. S., Reuter, H., Krause, A., Pott, C., Quintanilla-Martinez, L., Kakadia, P., Kuchenbauer, F., Ahmed, F., Delabesse, E., Hahn, M., Lichter, P., Kneba, M., Hiddemann, W., Macintyre, E., Mecucci, C., Ludwig, W.-D., Humphries, R. K., Bohlander, S. K., Feuring-Buske, M. and Buske, C. (2006). Acute myeloid leukemia is propagated by a leukemic stem cell with lymphoid characteristics in a mouse model of CALM/AF10-positive leukemia. *Cancer Cell*, **10**, 363-374.

Dik, W. A., Brahim, W., Braun, C., Asnafi, V., Dastugue, N., Bernard, O. A., van Dongen, J. J. M., Langerak, A. W., Macintyre, E. A. and Delabesse, E. (2005). CALM-AF10+ T-ALL expression profiles are characterized by overexpression of HOXA and BMI1 oncogenes. *Leukemia*, **19**, 1948-1957.

Dreyling, M. H., Martinez-Climent, J. A., Zheng, M., Mao, J., Rowley, J. D. and Bohlander, S. K. (1996). The t(10;11)(p13;q14) in the U937 cell line results in the fusion of the AF10 gene and CALM, encoding a new member of the AP-3 clathrin assembly protein family. *Proc. Natl.Acad. Sci. USA*, **93**, 4804-4809.

DSMZ (2008). DG-75.

El-Deiry, W. S. (2002). Transactivation of repair genes by BRCA1. *Cancer Biol Ther.*, **1**, 490-491.

Emerling, B. M., Bonifas, J., Kratz, C. P., Donovan, S., Taylor, B. R., Green, E. D., Le Beau, M. M. and Shannon, K. M. (2002). MLL5, a homolog of Drosophila trithorax located within a segment of chromosome band 7q22 implicated in myeloid leukemia. *Oncogene*, **21**, 4849-4854.

Fabienne, G.-P., Hulya, O., Mary, K.-B., David, R. B., Philippe, M. and AndrÃ©-Pascal, S. (2008). Detection of ATM gene deletion/duplication by multiplex ligation-dependant probe amplification in childhood lymphoid malignancies: A report from the Children's Oncology Group. *Leukemia research*, **32**, 1207-1213.

Futreal, P. A., Liu, Q., Shattuck-Eidens, D., Cochran, C., Harshman, K., Tavtigian, S., Bennett, L. M., Haugen-Strano, A., Swensen, J. and Miki, Y. (1994). BRCA1 mutations in primary breast and ovarian carcinomas. *Science*, **266**, 120-122.

Glasow, A., Barrett, A., Petrie, K., Gupta, R., Boix-Chornet, M., Zhou, D.-C., Grimwade, D., Gallagher, R., von Lindern, M., Waxman, S., Enver, T., Hildebrandt, G. and Zelent, A. (2008). DNA methylation-independent loss of RARA gene expression in acute myeloid leukemia. *Blood*, **111**, 2374-2377.

Goatly, A., Bacon, C. M., Nakamura, S., Ye, H., Kim, I., Brown, P. J., Ruskone-Fourmestraux, A., Cervera, P., Streubel, B., Banham, A. H. and Du, M.-Q. (2008). FOXP1 abnormalities in lymphoma: translocation breakpoint mapping reveals insights into deregulated transcriptional control. *Mod Pathol*, **21**, 902-911.

Goddard, A. D., Borrow, J. and Solomon, E. (1992). A previously uncharacterized gene, PML, is fused to the retinoic acid receptor alpha gene in acute promyelocytic leukaemia. *Leukemia*, **6 Suppl 3**, 117S-119S.

Goodarzi, A. A., Noon, A. T., Deckbar, D., Ziv, Y., Shiloh, Y., Löbrich, M. and Jeggo, P. A. (2008). ATM Signaling Facilitates Repair of DNA Double-Strand Breaks Associated with Heterochromatin. *Molecular Cell*, **31**, 167-177.

Gould, J. (2005). HierarchicalClustering Documentation. *Broad MIT*,

Graziani, I., Eliasz, S., De Marco, M. A., Chen, Y., Pass, H. I., De May, R. M., Strack, P. R., Miele, L. and Bocchetta, M. (2008). Opposite Effects of Notch-1 and Notch-2 on Mesothelioma Cell Survival under Hypoxia Are Exerted through the Akt Pathway. *Cancer Res*, **68**, 9678-9685.

He, Y., Chen, Q. and Li, B. (2008). ATM in oral carcinogenesis: association with clinicopathological features. *Journal of Cancer Research and Clinical Oncology*, **134**, 1013-1020.

Irizarry, R. A., Bolstad, B. M., Collin, F., Cope, L. M., Hobbs, B. and Speed, T. P. (2003). Summaries of Affymetrix GeneChip probe level data. *Nucl. Acids Res.*, **31**, e15-.

Jones, B., Su, H., Bhat, A., Lei, H., Bajko, J., Hevi, S., Baltus, G. A., Kadam, S., Zhai, H., Valdez, R., Gonzalo, S., Zhang, Y., Li, E. and Chen, T. (2008). The Histone H3K79 Methyltransferase Dot1L Is Essential for Mammalian Development and Heterochromatin Structure. *PLoS Genet.* , **4**, e1000190.

Kamps, M. P., Murre, C., Sun, X.-h. and Baltimore, D. (1990). A new homeobox gene contributes the DNA binding domain of the t(1;19) translocation protein in pre-B all. *Cell*, **60**, 547-555.

Kanehisa, M. and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.*, **28**, 27-30.

Karlsson, C., Brantsing, C., Egell, S. and Lindahl, A. (2008). Notch1, Jagged1, and HES5 Are Abundantly Expressed in Osteoarthritis. *Cells Tissues Organs*, **188**, 287-298.

Kim, K., Park, U., Wang, J., Lee, J., Park, S., Kim, S., Choi, D., Kim, C. and Park, J. (2008). Gene profiling of colonic serrated adenomas by using oligonucleotide microarray. *International Journal of Colorectal Disease*, **23**, 569-580.

Klebig, M. L., Wall, M. D., Potter, M. D., Rowe, E. L., Carpenter, D. A. and Rinchik, E. M. (2003). Mutations in the clathrin-assembly gene Picalm are responsible for the hematopoietic and iron metabolism abnormalities in fit1 mice. *Proc. Natl.Acad. Sci. USA*, **100**, 8360-8365.

Korsmeyer, S. (1992). Chromosomal translocations in lymphoid malignancies reveal novel proto-oncogenes. *Ann Rev Immun*, **10**, 785-807.

Krivtsov, A. V. and Armstrong, S. A. (2007). MLL translocations, histone modifications and leukaemia stem-cell development. *Nat Rev Cancer*, **7**, 823-833.

Kulaksiz, G., Reardon, J. T. and Sancar, A. (2005). Xeroderma Pigmentosum Complementation Group E Protein (XPE/DDB2): Purification of Various Complexes of XPE and Analyses of Their Damaged DNA Binding and Putative DNA Repair Properties. *Mol. Cell. Biol.*, **25**, 9784-9792.

Laird, P. W. Z., A. Linders, K. Rudnicki, M. A. Jaenisch, R. and Berns, A. (1991). Simplified mammalian DNA isolation procedure. *Nucleic Acids Research*, **19**, 4293.

Lamond, A. (2008). Establishing Stable Cell Lines. http://www.lamondlab.com/,

Levitan, D., Nanda, R., Rowley, J. and Olopade, O. (2003). Recurring Chromosome Rearrangements in Human Cancer. *American Cancer Society*, **6**,

Li, C. and Wong, W. H. (2001). Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. *Genome Biology*, **2**, 1-11.

Li, C. and Wong, W. H. (2003). DNA-Chip Analyzer (dChip). *Springer*, 120-141.

Li, H., Ilin, S., Wang, W., Duncan, E. M., Wysocka, J., Allis, C. D. and Patel, D. J. (2006). Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature*, **442**, 91-95.

Liapis, G., Mylona, E., Alexandrou, P., Giannopoulou, I., Nikolaou, I., Markaki, S., Keramopoulos, A. and Nakopoulou, L. (2007). Effect of the different phosphorylated Smad2 protein localizations on the invasive breast carcinoma phenotype. *Apmis*, **115**, 104-114.

Lieber , M., Raghavan , S. and Yu, K. (2008). Mechanistic Aspects of Lymphoid Chromosomal Translocations. *J Natl Cancer Inst Monogr*, **39**, 8-11.

Lin, S.-E., Oyama, T., Nagase, T., Harigaya, K. and Kitagawa, M. (2002). Identification of New Human Mastermind Proteins Defines a Family That Consists of Positive Regulators for Notch Signaling. *J. Biol. Chem.*, **277**, 50612-50620.

Linder, B., Newman, R., Jones, L. K., Debernardi, S., Young, B. D., Freemont, P., Verrijzer, C. P. and Saha, V. (2000). Biochemical analyses of the AF10 protein: the extended LAP/PHD-finger mediates oligomerisation. *Journal of Molecular Biology*, **299**, 369-378.

Lodish, H., Berk, A., Matsudaira, P., Kaiser, C. A., Krieger, M., Scott, M. A., Zipursky, L. and Darnell, J. (2003). Molecular Cell Biology. *W.H. Freeman & Company*,

Lynne V. Abruzzo, L. V., Lee, K. Y., Fuller, A., Silverman, A., Keating, M. J., Medeiros, L. J. and Coombes, K. R. (2005). Validation of oligonucleotide microarray data using microfluidic low-density arrays: a new statistical method to normalize real-time RT-PCR data. *Biotechniques*, **38**, 785-792.

Maria, N.-Z., Magdalena, B., Hanna, R.-M., Andrzej, K. and Wanda, M. K. (2008). Loss of heterozygosity in the RAD51 and BRCA2 regions in breast cancer. *Cancer detection and prevention*, **32**, 144-148.

Matthew, G., Raluca, Y., Constance, A. G. and Denise, A. S. B. (2008). Acute mixed lineage leukemia and a t(6;14)(q25;q32) in two adults. *Cancer genetics and cytogenetics*, **185**, 28-31.

Millenaar, F. F., Okyere, J., May, S. T., van Zanten, M., Voesenek, L. A. C. J. and Peeters, A. J. M. (2006). How to decide? Different methods of calculating gene expression from short oligonucleotide array data will give different results. *BMC Bioinformatics*, **7**, 1-16.

Millot, F., Brizard, F., Sorel, N., Preudhomme, C., Cividin, M., Guilhot, F. and Brizard, A. (2005). Therapy-related acute lymphoblastic leukemia with MLL rearrangement following treatment of Burkitt's leukemia. *Leukemia and Lymphoma*, **46**, 925 - 927.

Nambiar, M., Kari, V. and Raghavan, S. C. (2008). Chromosomal translocations in cancer. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, **In Press, Corrected Proof**,

Narita, M., Shimizu, K., Hayashi, Y., Taki, T., Taniwaki, M., Hosoda, F., Kobayashi, H., Nakamura, H., Sadamori, N., Ohnishi, H., Bessho, F., Yanagisawa, M. and Ohki, M. (1999). Consistent detection of *CALM-AF10* chimaeric transcripts in haematological malignancies with t(10;11)(p13;q14) and identification of novel transcripts. *British Journal of Haematology*, **105**, 928-937.

Okada, Y., Feng, Q., Lin, Y., Jiang, Q., Li, Y., Coffield, V. M., Su, L., Xu, G. and Zhang, Y. (2005). hDOT1L Links Histone Methylation to Leukemogenesis. *Cell*, **121**, 167-178.

Okada, Y., Jiang, Q., Lemieux, M., Jeannotte, L., Su, L. and Zhang, Y. (2006). Leukaemic transformation by CALM-AF10 involves upregulation of Hoxa5 by hDOT1L. *Nat Cell Biol*, **8**, 1017-1024.

Oyan, A. M., Bø, T. H., Jonassen, I., Ulvestad, E., Tore Gjertsen, B., Bruserud, O. and Kalland, K. H. (2007). Global gene expression in classification, pathogenetic

understanding and identification of therapeutic targets in acute myeloid leukemia. *Curr Pharm Biotechnol*, **8**, 344-354.

Peña, P. V., Davrazou, F., Shi, X., Walter, K. L., Verkhusha, V. V., Gozani, O., Zhao, R. and Kutateladze, T. G. (2006). Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature*, **442**, 100-103.

Rabbitts, T. (1991). Translocations, Master Genes, and Differences between the Origins of Acute and Chronic Leukemias. *Cell*, **67**, 641-644.

Rowley, J. and Potter, D. (1976). Chromosomal banding patterns in acute nonlymphocytic leukemia. *Blood*, **47**, 705-721.

Rowley, J. (2001). Chromosome translocations:dangerous liaisons revisited. *NATURE REVIEWS*, **1**, 245-249.

Rowley, J. D. (1973). A new consistent chromosomal abnormality in chronic myelogenous leukemia identified by quinacrine fluorescence and Giemsa staining. Nature 1973. *Nature*, **243**, 290-3.

S Ahle, S. and Ungewickell, E. (1986). Purification and properties of a new clathrin assembly protein. *EMBO J*, **5**, 3143–3149.

Saha, V., Chaplin, T., Gregorini, A., Ayton, P. and Young, B. D. (1995). The leukemia-associated-protein (LAP) domain, a cysteine-rich motif, is present in a wide range of proteins, including MLL, AF10, and MLLT6 proteins. *Proc. Natl.Acad. Sci. USA*, **92**, 9737-9741.

Sallmyr, A., Fan, J., Datta, K., Kim, K.-T., Grosu, D., Shapiro, P., Small, D. and Rassool, F. (2008). Internal tandem duplication of FLT3 (FLT3/ITD) induces increased ROS production, DNA damage, and misrepair: implications for poor prognosis in AML. *Blood*, **111**, 3173-3182.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. and Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl.Acad. Sci. USA*, **102**, 15545-15550.

Takei, K., Mundigl, O., Daniell, L. and De Camilli, P. (1996). The synaptic vesicle cycle: a single vesicle budding step involving clathrin and dynamin. *J. Cell Biol.*, **133**, 1237-1250.

Tebar, F., Bohlander, S. K. and Sorkin, A. (1999). Clathrin Assembly Lymphoid Myeloid Leukemia (CALM) Protein: Localization in Endocytic-coated Pits, Interactions with Clathrin, and the Impact of Overexpression on Clathrin-mediated Traffic. *Mol. Biol. Cell*, **10**, 2687-2702.

Toedling, J., Schmeier, S., Heinig, M., Georgi, B. and Roepcke, S. (2004). MACAT - MicroArray Chromosome Analysis Tool. *Bioconductor package manual*,

Vitoux, D., Nasr, R. and de The, H. (2007). Acute promyelocytic leukemia: new issues on pathogenesis and treatment response. *Int J Biochem Cell Biol*, **39**, 1063-1070.

Wang, Y., Cortez, D., Yazdi, P., Neff, N., Elledge, S. J. and Qin, J. (2000). BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. *Genes & Development*, **14**, 927-939.

Weberpals, J., Garbuio, K., O'Brien, A., Clark-Knowles, K., Doucette, S., Antoniouk, O., Goss, G. and Dimitroulakos, J. (2009). The DNA repair proteins BRCA1 and ERCC1 as predictive markers in sporadic ovarian cancer. *International Journal of Cancer*, **124**, 806-815.

Wei, X., Jian-Yong, L., Yu-Jie, W., Hui, Y., Qiu-Dan, S., Li, L., Lei, F. and Hong-Xia, Q. (2008). Prognostic significance of ATM and TP53 deletions in Chinese patients with chronic lymphocytic leukemia. *Leukemia research*, **32**, 1071-1077.

White, J. S., Choi, S. and Bakkenist, C. J. (2008). Irreversible chromosome damage accumulates rapidly in the absence of ATM kinase activity. *Cell Cycle*, **7**, 1277-1284.

Wu, S. K., Wang, B. J., Yang, Y., Feng, X. H., Zhao, X. P. and Yang, D. L. (2007). Expression of PTEN, PPM1A and P-Smad2 in hepatocellular carcinomas and adjacent liver tissues. *World J Gastroenterol.*, **13**, 4554-4559.

Yutaka MIURA, H. W., and Taiji KATO (1999). TBE, or not TBE; that is the question: Beneficial usage of tris-borate for obtaining a higher resolution of small DNA fragments by agarose gel electrophoresis. *Nagoya Medical Journal*, **31**, 1-6.

VII. Appendix

A1. Data obtained for Kill Curve Analysis (KCA)

Concentration (µg)	Percent Survival after 5 days	Reference total
.50	75.00	100.00
.50	70.00	100.00
.50	73.00	100.00
.60	55.00	100.00
.60	60.00	100.00
.60	58.00	100.00
.70	44.00	100.00
.70	48.00	100.00
.70	45.00	100.00
.80	30.00	100.00
.80	32.00	100.00
.80	34.00	100.00
.90	10.00	100.00

.90	5.00	100.00
.90	5.00	100.00
1.00	3.00	100.00
1.00	4.00	100.00
1.00	4.00	100.00

A2. Test of Normality of the data using K-S (Kolmogorov-Smirnov) test.

One-Sample Kolmogorov-Smirnov Test		**Concentration (µg/ml)**					
Description	**Statistic**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1**
Normal Parameters(a,b)	**Mean**	72.66	57.66	45.66	32	6.66	3.66
	Std. Deviation	2.51	2.51	2.08	2	2.88	0.57
Most Extreme Differences	**Absolute**	0.21	0.21	0.29	0.17	0.38	0.38
	Positive	0.18	0.18	0.29	0.17	0.38	0.28
	Negative	-0.21	-0.21	-0.21	-0.17	-0.28	-0.38
Kolmogorov-Smirnov Z		0.37	0.37	0.5	0.3	0.66	0.66
Asymp. Sig. (2-tailed)		0.99	0.99	0.95	0.99	0.76	0.76

A3. Result of ANOVA of different concentrations of puromycin kill-efficiency

Mortality	Sum of Squares	df	Mean Square	F	Sig.
Between Concentrations	11484.94	5	2296.99	464.56	.000
Within Concentrations	59.33	12	4.94		
Total	11544.28	17			

A4. Independent comparison of different concentrations showing that increment in concentration of puromycin yielded significant increment in percentage of dead cells, except for the increment from 0.9 to 1 µg.

	N	Subset for alpha = .05				
Concentration (µg)	**1**	**2**	**3**	**4**	**5**	**1**
.50	3	27				
.60	3		42			
.70	3			54		
.80	3				68	
.90	3					93
1.00	3					96
Sig.		1.00	1.00	1.00	1.00	.58

Means for groups in homogeneous subsets are displayed.
a Uses Harmonic Mean Sample Size = 3

A5. Calculated/estimated Probit Analysis model based on the mortality data collected

						95% Confidence Interval	
	Parameter	**Estimate**	**Std. Error**	**Z**	**Sig.**	**Lower Bound**	**Upper Bound**
PROBIT(a)	Concentration (ng)	13.17	1.32	9.97	.00	10.58	15.76
	Intercept	-37.86	3.89	-9.74	.00	-41.75	-33.98

PROBIT model: PROBIT(p) = Intercept + BX (Covariates X are transformed using the base 10 logarithm.)

A6. Estimated kill probability (probit) for different puromycin concentration levels based on the mortality data collected.

	Probability	**95% Confidence Limits for Concentration (ng)**			**95% Confidence Limits for log(Concentration (ng))(a)**		
		Estimate	**Lower Bound**	**Upper Bound**	**Estimate**	**Lower Bound**	Upper Bound
PROBIT(b)	.010	500.035	411.157	562.288	2.699	2.614	2.750
	.020	524.449	438.222	584.218	2.720	2.642	2.767
	.030	540.553	456.297	598.592	2.733	2.659	2.777
	.040	552.993	470.375	609.648	2.743	2.672	2.785
	.050	563.323	482.140	618.800	2.751	2.683	2.792
	.060	572.267	492.380	626.705	2.758	2.692	2.797
	.070	580.226	501.533	633.724	2.764	2.700	2.802
	.080	587.446	509.868	640.081	2.769	2.707	2.806

.090	594.091	517.566	645.922	2.774	2.714	2.810
.100	600.274	524.751	651.350	2.778	2.720	2.814
.150	626.564	555.535	674.362	2.797	2.745	2.829
.200	648.277	581.220	693.306	2.812	2.764	2.841
.250	667.504	604.139	710.054	2.824	2.781	2.851
.300	685.255	625.424	725.515	2.836	2.796	2.861
.350	702.126	645.742	740.227	2.846	2.810	2.869
.400	718.519	665.545	754.562	2.856	2.823	2.878
.450	734.743	685.175	768.816	2.866	2.836	2.886
.500	751.068	704.921	783.260	2.876	2.848	2.894
.550	767.755	725.052	798.178	2.885	2.860	2.902
.600	785.091	745.841	813.906	2.895	2.873	2.911
.650	803.420	767.582	830.891	2.905	2.885	2.920
.700	823.200	790.616	849.781	2.916	2.898	2.929
.750	845.093	815.364	871.592	2.927	2.911	2.940
.800	870.157	842.430	898.035	2.940	2.926	2.953
.850	900.312	872.949	932.179	2.954	2.941	2.969
.900	939.742	909.794	980.352	2.973	2.959	2.991
.910	949.522	918.507	992.806	2.978	2.963	2.997
.920	960.262	927.921	1006.674	2.982	2.968	3.003
.930	972.211	938.230	1022.313	2.988	2.972	3.010
.940	985.733	949.713	1040.248	2.994	2.978	3.017
.950	1001.383	962.802	1061.286	3.001	2.984	3.026

.960	1020.089	978.206	1086.773	3.009	2.990	3.036
.970	1043.565	997.240	1119.218	3.019	2.999	3.049
.980	1075.610	1022.798	1164.216	3.032	3.010	3.066
.990	1128.126	1063.886	1239.456	3.052	3.027	3.093

a Logarithm base = 10.
b A heterogeneity factor is used.

A7. Gene list and fold regulation (LDA and affymetrix)

Gene	LDA	Affy
MLF1	28.33257	4.525701
OAS1	28.17326	5.604744
IFI6	27.99181	4.231276
IFI44L	21.87256	5.252087
CDC2	17.95673	3.569363
DHRS2	17.84711	3.650866
UNC13C	16.83723	3.665108
BCL2A1	16.79263	3.633611
CKS1B	16.21683	3.370275
OAS2	16.19548	3.287767
CDKN3	15.61251	3.681336
DNAJB1	14.01363	3.180718
HOXD10	13.82583	2.881668
TREX1	13.78496	3.351746
BIRC5;EPR1	13.76273	3.271797
LY9	13.38118	2.79076

CCNB1	13.10304	3.242981
STAT1	12.89233	3.597469
RAP1A	12.89008	3.33826
ANKRD39	12.76305	2.962647
DDB2	11.99111	2.063714
BIRC4BP	11.95635	3.26541
ETV6	11.93008	2.786431
JMJD1A	11.86804	2.920787
CDCA3	11.5734	2.970424
OTOF	11.22087	4.623846
MZF1	10.96904	2.787606
SEPP1	10.94182	2.468817
CDC6	10.91783	3.035429
IRX5	10.88233	2.824842
SUV39H1	10.63244	2.76463
HBP1	10.50044	2.702757
SPBC24	10.48712	3.017905
MCM10	10.30153	2.881917
BCL2	10.26461	3.25085
JMJD2B	10.25509	2.732236
ETV7	10.15794	2.881668
FHL2	10.15794	2.881668
HOXB3	10.15794	2.881668

HOXD13	10.15794	2.881668
PBX1	10.15794	2.846048
PTN	10.15794	2.881668
RAB8B	10.15794	2.881668
WNT3	10.15794	2.507769
CDC25C	9.973241	2.933275
DNAJB9	9.756624	2.87052
RECQL5	9.693124	2.667386
MAML3	9.544616	2.854931
CCNF	9.529396	2.631359
TSGA10	9.517653	2.920866
METT10D	9.460456	2.690977
JMJD2C	9.414562	2.618284
RAD51	9.293371	2.677388
NEK2	9.236538	2.829051
RAD51AP1	9.216778	2.860128
PBXIP1 (HPIP)	7.59870	10.202341
SMAD2/3	2.43180	9.546690
SKP2	9.189228	2.69633
FHL1	9.15025	3.21343
USP18	9.08962	3.175629
SKIL	9.073743	2.612052
CREM	8.990205	2.726596

KIF23	8.650458	2.885807
IFI27	8.543739	2.089735
MAP4K3	8.524462	2.70185
CHEK1	8.514601	2.706704
MLL3	8.4914	2.621705
BRCA1	8.486075	2.62364
PBX2	8.477455	2.800326
RMI1	8.447185	2.839224
BRIP1	8.38464	6.569919
ATM	8.367658	2.461272
MYBL2	8.344836	2.480506
ESPL1	8.32113	2.565788
CASC5	8.271757	2.521866
STAT2	8.224094	2.613837
TOP2A	8.215572	2.277137
CENPE	8.178639	2.624413
CCNE2	8.048637	2.69551
FOXP1	8.036818	2.616995
RUNX1	7.862116	2.470978
HELLS	7.837011	2.388881
DNMT3B	7.694172	2.575032
RARA	7.615044	2.482983
UHRF1	7.476315	2.4227

MLL5	7.364065	2.461649
CHAF1B	7.299285	2.490731
TLX2	7.157793	2.456056
BAZ2A	7.151682	2.400109
SUV39H2	7.065186	2.363607
MYBL1	6.937232	2.546511
BCL2L11	6.84793	2.359521
CHAF1A	6.461873	2.379459
TCF7	5.285275	2.175603
CDCA8	1.32764	1.128068

Curriculum Vitae

Name: Medhanie Assmelash Mulaw

Date of Birth: September 30, 1976

Nationality: Ethiopian

Marital Status: Married

Formal Academic Background (post secondary school):

2006 – 2009 – PhD in Human Biology from the Department of Medicine III, Faculty of Medicine, University of Munich, Munich, Germany

1998/1999 – 1999/2000 – Masters of Science in Biology (Applied Genetics), from the Department of Biology, Faculty of Science, Addis Ababa University (AAU), Addis Ababa, Ethiopia

1993/94 – 1996/1997 – Bachelor of Science (B.Sc.) in Plant Sciences from the College of Agriculture and Aquatic Sciences of University of Asmara, Asmara, Eritrea

Work Experience:

2000 - : Part time and then full time Lecturer at the Department of Biology, Faculty of Science, Addis Ababa University (AAU). Courses thought include Molecular Biology, Principles of Genetics (Lab), Gene Technology, Biotechnology, Biostatistics, Techniques in Molecular Biology.

Publications:

Bararia D, Trivedi AK, Zada AA, Greif PA, **Mulaw MA**, Christopeit M, Hiddemann W, Bohlander SK, Behre G.(2008). Proteomic identification of the MYST domain histone acetyltransferase TIP60 (HTATIP) as a co-activator of the myeloid transcription factor C/EBPalpha. *Leukemia* **22**(4): 676-677.

Seminars and conferences attended:

Attended: July 2006, Annual meeting of the Department of Medicine III, University of Munich – Herrsching, Germany

Poster Presentation: September 2006, Presented a poster at the SFB Transregio meeting on Chromatin Inheritance-Munich, Germany

Attended: February 2007, Harvard-Munich meeting- Eibsee, Germany

Poster Presentation: July 2007, Presented a poster (and won the poster prize) at the Annual meeting of the Department of Medicine III, University of Munich – Herrsching, Germany

Speaker: February 2008, Speaker at the Harvard-Munich meeting-Garmisch, Germany

Poster Presentation: April 2008, Presented a poster at the Annual Meeting of the society of Human Genetics – Hannover, Germany

Speaker: July 2008, Speaker at the Annual meeting of the Department of Medicine III, University of Munich – Herrsching, Germany

Speaker: November 2008, Nationales Genomforschungsnetz Deutschland (NGFN) meeting, Berlin, Germany

Poster Presentation: December 2008, 50th Annual conference of the American Society of Hematology (ASH), San Franscisco, CA, USA

Course Attended:

March 2007: *Bioinformatics for Molecular Biologists* – Annual course organized by the University of Munich, Technical University of Munich and HelmholtzZentrum Institute of Bioinformatics – Bertinoro, Italy