

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Identification and characterization of genomic target sites for MLL fusion proteins

Michael Lehmbacher

aus

Fürstenfeldbruck

2006

Erklärung

Diese Dissertation wurde im Sinne von § 13 Abs. 3 bzw. 4 der Promotionsordnung vom 29. Januar 1998 von Herrn Prof. Dr. Meisterernst betreut.

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbständig, ohne unerlaubte Hilfe erarbeitet.

München, am 31.07.2006

Michael Lehmbacher

Dissertation eingereicht am 31.07.2006

1. Gutachter Prof. Dr. Meisterernst

2. Gutachter Prof. Dr. Jansen

Mündliche Prüfung am 14.11.2006

Summary

In this study gene regulation by MLL fusion proteins was investigated. Inducible cell lines using the estrogen receptor system were established that allowed for analysis of the early events after activation of an oncogene. A fusion of the MLL aminoterminal and the activation domain of Herpes simplex VP16 was used to imitate leukemic MLL translocation events, the majority of which fuses the aminoterminal part of MLL to a transcriptional activator. It could be demonstrated that expression of biological targets of MLL and MLL fusion constructs increased following activation of the inducible protein MLL–VP16–ER–HA. The regulatory sequence of *Hoxa9*, expression of which is critically depending on MLL, was investigated in detail in order to understand the mechanism of MLL activation. Episomal reporter constructs were used to analyze the activation of different regions of the human *Hoxa9* promoter by MLL–VP16–ER–HA. A fragment comprising nucleotides –118 / +46 was identified as sufficient for the response to induction of the fusion protein. Monoclonal antibodies were raised against MLL and used to monitor the recruitment of MLL–VP16–ER–HA to the episomal as well as to the chromosomal *Hoxa9* promoter. Site–directed mutagenesis of the regulatory sequences of *Hoxa9* led to identification of short sequence motifs that are important for gene activation by MLL–VP16–ER–HA. Three different motifs in the *Hoxa9* promoter could be characterized, some of which are also present in the upstream sequences of other MLL target genes. Furthermore it was demonstrated that core promoter sequences are critical for the effect of MLL–VP16–ER–HA pointing out a global role for MLL in transcription. Comparison of *Hoxa9* with *p21* and *c–myc*, two new target genes that have been identified in this study, provided insights into the mechanism of MLL recruitment to promoters. This analysis showed that the promoter sequences of these target genes share a bipartite structure, in which two modules containing one or more putative binding sites for MLL and MLL fusion proteins cooperate to mediate gene activation. In many cases the binding sites either contain CpG dinucleotides or are located next to CpGs. Mutation of these cytosine residues to guanosine led to increased activation of *Hoxa9* indicating that guanosine residues, but not cytosine residues are important for the effect. Based on this observation a model is proposed in this study for recruitment of MLL and MLL fusion proteins to GC rich regions where

Summary

cytosine residues have a negative regulatory function, which hints at a mechanism involving DNA methylation.

In a genome-wide analysis new candidate target genes were identified. A murine hematopoietic stem cell system was used to study the global effects of induction of MLL-VP16-ER-HA. More than 200 genes were found to be significantly upregulated, some of which have been found to be overexpressed in human leukemias deriving from 11q23 translocation. A host of new candidates, however, has not been described before as MLL target genes. These genes await further analysis that will clarify their role in MLL leukemogenesis.

Table of Contents

1. INTRODUCTION	1
1.1 FROM GENE TO PROTEIN	1
1.2 REGULATORY ELEMENTS OF PROTEIN-CODING GENES	2
1.3 EUKARYOTIC RNA POLYMERASES	3
1.4 GENERAL TRANSCRIPTION FACTORS	4
1.5 MECHANISM OF TRANSCRIPTIONAL INITIATION	5
1.6 ACTIVATORS OF TRANSCRIPTION.....	5
1.6.1 Nuclear hormone receptors.....	8
1.6.2 The viral activator VP16.....	8
1.7 REPRESSORS	9
1.8 TRANSCRIPTIONAL COFACTORS.....	10
1.8.1 TBP-associated factors (TAFs)	10
1.8.2 Cofactors of the USA fraction.....	10
1.8.3 Mediator complexes.....	11
1.9 CHROMATIN	11
1.9.1 Histone modifications.....	12
1.9.1.1 Histone acetylation.....	13
1.9.1.2 Histone methylation.....	14
1.9.1.3 Additional histone modifications	14
1.9.2 Chromatin remodeling	15
1.10 TRANSCRIPTIONAL MAINTENANCE	15
1.10.1 DNA methylation	15
1.10.2 Polycomb and trithorax.....	17
1.10.2.1 MLL: a member of the human trithorax family.....	19
1.10.2.2 MLL-containing complexes	21
1.10.2.3 Leukemic transformation by mutation of MLL	22
1.10.2.4 MLL target genes	23
1.11 OBJECTIVES OF THIS STUDY.....	25
2. MATERIALS AND METHODS	26
2.1 INSTRUMENTS AND ACCESSORIES.....	26
2.2 CHEMICALS AND BIOCHEMICALS.....	26
2.3 ADDITIONAL MATERIAL	29
2.4 ENZYMES.....	29
2.5 ANTIBODIES	30
2.6 GENERAL BUFFERS	30
2.7 OLIGONUCLEOTIDES	31
2.8 PLASMIDS	33
2.9 CLONING.....	37
2.9.1 Polymerase chain reaction (PCR).....	37

Table of Contents

2.9.2 Restriction digests, fragment isolation and ligation of DNA	37
2.9.3 Transformation into <i>E. coli</i> DH5a	38
2.9.4 Plasmid generation	38
2.9.5 Site directed mutagenesis of the human <i>Hoxa9</i> promoter.....	40
2.9.6 Generation of <i>Hoxa9</i> –78 / +46 containing additional binding sites	41
2.10 CELL CULTURE.....	41
2.10.1 Separation of viable and dead cells via density gradient centrifugation	41
2.10.2 Freezing and thawing of cell lines.....	41
2.10.3 Isolation of monoclonal cell lines using semisolid medium.....	42
2.10.4 Preparation and staining of cytopins.....	42
2.10.5 Culture of FDCPmix cells	42
2.10.6 Culture of U937 cells and THP-1 cells.....	43
2.10.7 Production of interleukin 3 – conditioned medium	43
2.10.8 Differentiation of FDCPmix cells into granulocytes and macrophages	44
2.11 TRANSFECTION EXPERIMENTS.....	44
2.11.1 Electroporation of suspension cells	44
2.11.2 Transient transfection of FDCPmix cells.....	44
2.11.3 Stable transfection of FDCPmix cells.....	45
2.11.4 Transient transfection of U937 cells.....	45
2.11.5 Stable transfection of U937 cells.....	46
2.11.6 Transfection of 293T cells	46
2.12 PREPARATION OF PROTEIN EXTRACTS FROM MAMMALIAN CELLS.....	47
2.12.1 Whole cell lysates from U937 cells	47
2.12.2 Isolation of nuclear extract from 293T.....	47
2.12.3 Determination of luciferase reporter gene activity.....	48
2.13 PROTEIN ANALYSIS	49
2.13.1 Recombinant protein expression and purification	49
2.13.2 Purification of GST-tagged proteins.....	49
2.13.3 Preparation of antigens for immunization of rodents.....	50
2.13.4 Immunoprecipitation	50
2.13.5 Sodium–dodecylsulphate polyacrylamide gel electrophoresis (SDS–PAGE).....	51
2.13.6 Coomassie staining.....	52
2.13.7 Western Blot analysis of MLL protein expression in U937 cells.....	52
2.13.8 Chromatin–Immunoprecipitation (IP)	53
2.14 RNA EXPRESSION ANALYSIS.....	57
2.14.1 cDNA microarray analysis	57
2.14.2 RT-PCR.....	58
2.15 BIOINFORMATIC ANALYSIS	59
2.15.1 Identification of transcription factor binding sites using MatInspector	59
2.15.2 Gene2Promoter.....	59

Table of Contents

2.15.3 Bibliosphere.....	60
3. RESULTS	61
3.1 GENERATION AND SELECTION OF MONOCLONAL ANTIBODIES AGAINST MLL	61
3.2 AN INDUCIBLE MLL FUSION MODEL SYSTEM	65
3.2.1 Generations of cell lines stably expressing MLL–VP16–ER–HA or VP16–ER–HA.....	66
3.2.2 Transcription of human Hoxa9 is upregulated following induction of MLL–VP16–ER–HA	67
3.2.3 MLL–VP16–ER–HA is targeted to the Hoxa9 promoter upon induction	70
3.3 IDENTIFICATION OF A PROMOTER REGION CRITICAL FOR REGULATION OF HOXA9 BY MLL–VP16–ER–HA	71
3.4 SITE-DIRECTED MUTAGENESIS OF THE HUMAN HOXA9 PROMOTER REGION.....	73
3.4.1 Hox protein binding sites have a dual role for Hoxa9 regulation	74
3.4.2 A GC rich region in the Hoxa9 promoter is mediating epigenetic regulation	75
3.4.2.1 CpG mutations in the Hoxa9 promoter lead to strong activation of transcription by MLL–VP16–ER–HA.....	76
3.4.2.2 SP1 protein overexpression does not facilitate gene activation by MLL–VP16–ER–HA	80
3.4.3 Introduction of synthetic binding sites into the Hoxa9 minimal promoter	82
3.4.4 A consensus TATA box stimulates activation of Hoxa9 by MLL–VP16–ER–HA.....	86
3.5 IDENTIFICATION OF NEW GENOMIC TARGET SITES FOR MLL FUSION PROTEINS	88
3.5.1 FDCPmix cells as a myeloid progenitor model system.....	89
3.5.2 c–myc promoter sequences as targets for MLL–VP16–ER–HA.....	95
3.5.3 p21 as MLL fusion protein target.....	97
3.5.3.1 MLL–VP16–ER–HA strongly activates a p21–luciferase reporter gene	97
3.5.3.2 p21 promoter sequences are activated by MLL and MLL fusion proteins	101
3.5.3.3 SP1 overexpression does not affect p21 activation by MLL–VP16–ER–HA	102
3.5.3.4 Expression of endogenous p21 mRNA after induction of MLL–VP16–ER–HA activity.....	103
3.5.3.5 MLL–VP16–ER–HA activated transcription of episomally stable p21 reporters.....	105
3.5.3.6 Influence of DNA methylation on activation of p21 transcription	106
3.5.4 Overview: results from reporter gene experiments	108
3.6 MICROARRAY ANALYSIS OF MLL–VP16–ER–HA TARGET GENES	112
4. DISCUSSION	122
4.1 DEFINITION OF CRITICAL REGIONS IN THE HOXA9 PROMOTER.....	122
4.2 MLL–VP16–ER–HA ACTS ON THE CORE PROMOTER REGION OF HOXA9	123
4.3 A MODEL FOR MLL BINDING TO UPSTREAM ACTIVATING SEQUENCES	127
4.3.1 Heterogeneity in the sequence motifs recruiting MLL–VP16–ER–HA	127
4.3.2 MLL–VP16–ER–HA might compete with other transcription factors.....	129
4.3.3 A two–module hypothesis for MLL binding	132
4.3.4 Spatial distribution of binding sites is critical for MLL fusion protein function	134
4.3.5 Epigenetic control of Hoxa9 transcriptional activation by MLL–VP16–ER–HA.....	134
4.4 IDENTIFICATION AND CHARACTERIZATION OF NEW TARGET PROMOTERS	136

Table of Contents

4.4.1 The human p21 promoter is targeted by MLL and MLL fusion proteins.....	136
4.4.2 Activation of p21 by MLL–VP16–ER–HA is dependent from the chromatin environment....	137
4.4.3 MLL–VP16–ER–HA activates transcription from the c–myc promoter.....	138
4.4.4 A hematopoietic stem cell system for MLL fusion target gene analysis.....	139
4.4.5 A genome-wide screen for MLL fusion target genes.....	140
5. REFERENCES	143
6. APPENDIX	158
6.1 ALIGNMENT OF HOXA9 PROMOTER VARIANTS.....	158
6.2 MICROARRAY RESULTS	163
6.3 LIST OF AVAILABLE BINDING MATRICES (GENOMATIX)	175
6.4 PROMOTER SEQUENCES OF CUTL1, P18 AND P27	190
ACKNOWLEDGEMENTS	191
CURRICULUM VITAE MICHAEL LEHMBACHER.....	192

1. Introduction

1.1 From gene to protein

The individual life cycle of any living matter is based on the sum of genetic information, which is inherited from the parental generation. This collection of hereditary material is referred to as the genotype. What defines an organism in the context of its environment, however, is the so-called phenotype, the complete set of characteristics that are observable under certain conditions. Therefore the inherited genetic information has to be translated into biological processes in a way that serves certain purposes in a given environment. This process is called gene expression. The material encoding the genetic information is deoxyribonucleic acid (DNA). A process termed transcription generates a blueprint of the information, which is stored in the genomic DNA. This blueprint, the so-called messenger RNA (mRNA), leaves the nucleus, the cellular compartment where DNA is stored, and travels to the cytoplasm. It is there where the information contained in the mRNA molecule is decoded and translated into proteins.

At a given time point only a fraction of the total amount of genes encoded in the genome of an organism is expressed. Especially in metazoans gene expression profiles vary dramatically over the course of development. The essential function of the coordination of these changes in gene expression becomes obvious in fatal diseases and failures in development that are due to deregulated or wrongly timed expression of certain genes. Nature found multiple ways to control gene expression on different levels. Most protein-coding genes, however, are regulated on the transcriptional level. This is reflected in the genomic structure where protein-coding genes are interspersed with regulatory regions, which do not by themselves contain information that can be translated into protein but are necessary for proper expression of adjacent coding sequences.

1.2 Regulatory elements of protein-coding genes

Generally speaking there are two different classes of regulators that influence the transcription rate: gene-specific elements that control expression of a single gene only and global regulators, which in contrast control groups of genes.

Gene-specific regulators are called promoters, which consist of proximal and distal elements. The DNA region immediately upstream of the transcriptional start point of a given gene is often referred to as UAS (upstream activating sequence) or URS (upstream repressing sequence). These proximal promoter sequences are sufficient for initiation of transcription and determine the basal transcription rate (Smale and Kadonaga 2003). One feature that is frequently found promoters of the start site is the so-called TATA-box, which is usually located 30 nucleotides upstream of the start site. Another common motif is the pyrimidin rich initiator (Inr) sequence, around which the start site is centered. There is considerable variability, however, amongst different promoters: additional elements that are found sometimes but in other cases the promoters lack a well-conserved TATA-box or an Inr sequence. This influences in turn the basal transcription rate. Distal promoter elements can have a stimulatory (enhancers) or an inhibitory effect (silencers) and function over a distance of several kilobases independently from their location (upstream or downstream of a gene) and their orientation. Like proximal promoter elements they are bound by transcription factors in a sequence-specific manner that influence the transcription rate.

Global regulators are widespread in Metazoa. Matrix attachment regions (MARs) and scaffold attachment regions (SARs), respectively, are thought to influence the localization of large chunks of DNA by linking them to the nuclear matrix (Francastel et al. 2000; Bell et al. 2001; Fernandez et al. 2001). The details behind this type of transcriptional regulation are only poorly understood but it seems likely that the specific localization of the chromosome region in the nucleus has an effect on the chromatin structure. Certain sequences termed insulators have been found to separate euchromatic from heterochromatic regions. Locus control regions (LCRs) are aggregates of enhancer sequences and exert an effect on a large chromosomal regional while at the same time they control transcription in a gene-specific manner (Li et al. 2002).

1.3 Eukaryotic RNA polymerases

The vast complexity of eukaryotic genomes illustrates the requirement of enzymes that can process the stored information. Three different types of polymerases are responsible for transcription of different classes of RNA (Roeder and Rutter 1969). While RNA polymerase I transcribes ribosomal RNA and RNA polymerase III mainly synthesizes tRNA, it is RNA polymerase II that is responsible for transcription of all mRNAs generated in a living cell. It is possible to distinguish these enzymes biochemically according to their sensitivity to α -amanitin (Table 1).

Table 1: Eukaryotic RNA polymerases. Pol I, Pol II and Pol III stand for the three DNA-dependent polymerases. rRNA: ribosomal RNA; mRNA: messenger RNA; snRNA: small nuclear RNA, tRNA: transfer RNA.

Type	Genes	Transcripts	Localisation	Response to α -amanitin
Pol I	class I	18S-, 5.8S- and 28S-rRNA	nucleoli	none
Pol II	class II	pre-mRNA, snRNA	nucleoplasm	strong, $K_D=10^{-8}$ M
Pol III	class III	tRNA, 5S-rRNA, snRNA	nucleoplasm	weak, $K_D=10^{-6}$ M

RNA polymerase II is a protein complex that contains 12 subunits (Rpb1 - 12), four of which are homologous to subunits from polymerase I and III complexes and four of which are identical to subunits found in these enzymes (Cramer et al. 2000; Asturias 2004; Cramer 2004). The largest subunit of eukaryotic RNA polymerase II contains the so-called CTD (carboxy-terminal domain). This domain is made up by a species-specific number of heptapeptide repeats and contains binding sites for proteins that regulate processes like transcriptional initiation, elongation, termination and mRNA processing (Palancade and Bensaude 2003).

1.4 General transcription factors

Eukaryotic as well as prokaryotic RNA polymerases cannot recognize promoters and transcription start sites by themselves. Furthermore RNA polymerase II needs a helicase activity for the opening of doublestranded DNA around this start site. The proteins required for transcription from model promoters in vitro have been defined as general transcription factors (GTFs, Table 2) (Hampsey 1998; Lee and Young 2000).

Table 2. General transcription factors.

Factor		Function
TFIID	TBP	promoter recognition (TATA-box) recruitment of TFIIB
	TAF _{II} s	promoter recognition; positive and negative regulatory functions
TFIIA		stabilization of TBP–DNA complex; stabilization of TAF _{II} –DNA interactions; antirepression
TFIIB		recruitment of RNA polymerase II and TFIIF; determination of transcriptional start site
TFIIF		recruitment of RNA polymerase II; destabilization of unspecific RNA polymerase II–DNA interactions
TFIIE		recruitment of TFIIH; modulation of kinase, helicase and ATPase activities of TFIIH; facilitation of strand separation
TFIIH	XBP/ ERCC3	3' – 5' helicase; promoter melting, open complex formation
	XPD/ ERCC2	5' – 3' helicase; DNA repair (nucleotide excision repair)
	Cdk7/ MO15	CTD phosphorylation
	CyclinH	regulation of Cdk7

1.5 Mechanism of transcriptional initiation

The process of transcription can be formally divided into the three steps initiation, elongation and termination. Using in vitro transcription systems it was shown that initiation is starting with a process called pre-initiation complex (PIC) formation (Van Dyke et al. 1988; Buratowski et al. 1989). This event could be dissected and it was demonstrated that PIC formation is achieved through step-wise binding of GTFs and RNA polymerase II to the DNA template.

TFIID binds to the promoter region in a sequence specific manner forming a binary complex with DNA, which is stabilized through binding of TFIIA. This ternary complex is the binding site for TFIIIB that in turn is recruiting RNA polymerase II and TFIIIF. Finally TFIIIE and TFIIH bind to the complex. The helicase activity can now convert the so-called closed PIC into an open PIC by creating a single stranded DNA region between -9 and +2. Besides this function as "PIC isomerase" TFIIH is exerting another important effect during initiation and early elongation: the so-called CAK sub-complex of TFIIH (Cdk7, cyclin H, MAT1) phosphorylates the CTD of Rpb1. The phosphorylation status of CTD is critical for interactions with a series of factors. The non-phosphorylated isoform of CTD is bound by the Mediator cofactor complex. During initiation CTD becomes phosphorylated at serine 5 by TFIIH and later during elongation pTEFb phosphorylates CTD at serine 2 (Pinhero et al. 2004). Furthermore CTD is the binding site for factors involved in polyadenylation thus representing a link between transcription and mRNA processing.

1.6 Activators of transcription

Activators are regulatory proteins that by definition meet two different criteria: on one hand they specifically recognize target sequences on the DNA, on the other hand they interact with the transcriptional machinery. This functional dualism is usually reflected in the modular structure of these proteins: a DNA binding domain can usually be separated from one or more activation domains that work independently from each other. This functional independence is also underlined by the existence of activators that lack an intrinsic DNA binding activity but interact with DNA binding

factors, e. g. β -catenin that is recruited to promoters via LEF-1 (Eastman and Grosschedl 1999).

Activators can bind to DNA in a variety of different ways and the molecular structure of DNA binding domains is used for the classification of these factors (Harrison 1991; Pabo and Sauer 1992):

- Zinc finger proteins (e. g. SP1, GAL4) are characterized by DNA binding domains that are stabilized by a metal ion. The Zinc-ion is complexed by two cysteine and two histidine residues. The residues in between loop out and form a finger-like structure, which contacts DNA. Very often this motif is repeated several times creating an array of several zinc fingers that are separated by short α -helical stretches.
- Helix-turn-helix (HTH) motifs were initially identified in DNA binding domains of repressors in bacteriophages. Two α -helices are connected by a short linker (4 aa). One helix contacts the major groove of the DNA while the other helix is taking a position perpendicular to the DNA double strand. This type of DNA binding domain functions as a dimer. A variant of this motif is found in vertebrates: the homeobox contains three α -helices. One binds the major groove of the DNA, while the other two are positioned perpendicular to the DNA molecule. Homeobox proteins can bind DNA both as monomer and as dimer. They are found not only in vertebrates (e. g. Hox proteins) but also in *Drosophila* (e. g. antennapedia, ultrabithorax) and function as important regulators of development. The family of POU proteins (e. g. Pit1, Oct1, Oct2) is characterized by the presence of a homeobox and a POU domain, which is another variant of the HTH motif. In contrast to the homeobox the POU domain contains four helices, however.
- Another group of activators (e. g. MyoD, E12) have in common a basic helix-loop-helix (bHLH) motif. Two amphipathic α -helices are connected by a short loop and a basic region is found in the immediate vicinity that binds to DNA. bHLH factors function as homo- or heterodimers, in which the helical regions mediate dimerization.
- Also the basic leucine zipper (bZIP) motif contains an amphipathic α -helix in combination with a basic region. Every seventh residue of the α -helix is leucine

and like that this residues form an interaction surface with another bZIP protein: the two α -helices bind to each other where the leucines work like a zipper to form a coiled coil. This brings the basic regions into a fork-like position to each other, which bind to the DNA from opposite sides contacting the major groove. Mammalian examples for bZIP proteins are c-jun, c-fos and CREB that can homo- or heterodimerize, respectively.

- bHLH-bZIP proteins like c-myc, Max, Mad or USF resemble the bZIP family. In between the amphipathic α -helix and the basic region, however, a HLH motif is found.
- The T-box motif is an extraordinarily large DNA binding domain (200 aa) and consists of a β -barrel and four α -helices. In contrast to DNA binding by HTH factors α -helices of T-box factors (e. g. Tbx1, Tbx3) contact the minor groove, however. Formation of a dimer leads to further contacts involving the β -sheets and the major groove of the DNA.

The structural variability of activation domains, however, does not allow for this type of classification. Activation domains usually consist of 30 to 100 amino acids and it was observed that very often there is a dominance of certain chemical properties amongst these residues. This led to the definition of so-called glutamine rich (SP1, GAGA), proline rich (Oct2, EKLF), serine / threonine rich (Sox-2, v-Rel) and acidic (VP16, E1A, NFkB) activation domains.

Activation domains can interact with components of the basal transcription machinery or elongation factors to influence the transcription rate. Alternatively they can bind to coactivators, which mediate between activators and the basal machinery, or to cofactors that modify the chromatin state (Lemon and Tjian 2000; Orphanides and Reinberg 2002). These interactions lead to increased recruitment of GTFs, changes in PIC conformation and to a more open chromatin structure around the start site.

1.6.1 Nuclear hormone receptors

There are two different classes of nuclear hormone receptors:

- Steroid hormone receptors (glucocorticoid, androgen, estrogen and corticoid receptors) are retained in the cytoplasm by binding to heat shock proteins that mask their nuclear localization signals. Upon ligand binding they can translocate to the nucleus where they bind as a homodimer to palindromic target sequences on the DNA to regulate gene expression.
- The second class of hormone receptors (e. g. thyroid hormone receptor, vitamin D receptor, retinoic acid receptor) is usually found bound to DNA also in absence of the ligand where these proteins exert a negative effect on transcription. In contrast to steroid hormone receptors this class binds as heterodimer to target sequences, which consist of direct repeats.

Both classes have some characteristics in common, however: the DNA binding domain of these proteins consist of two zinc finger motifs and two α -helices, one of which contacts the hormone response element (HRE) on the DNA. Furthermore they have two activation domains. While the N-terminal domain (AF1) is constitutively active, the C-terminal domain (AF2) is only active in presence of the ligand (Warnmark et al. 2003).

1.6.2 The viral activator VP16

The activation domain (aa 411 – 490) of the Herpes simplex virus protein 16 (VP16) is well-studied model for acidic activators (Triezenberg et al. 1988). It was shown that only upon binding to its interaction partners the unstructured activation domain is converted into an α -helix (Uesugi et al. 1997), a mechanism that has been termed induced fit.

A host of cellular factors have been shown to interact with the VP16 activation domain:TFIIA (Kobayashi et al. 1995), TFIID (Goodrich et al. 1993), TFIIF (Zhu et al. 1994), TFIIH (Xiao et al. 1994), PC4 (Kretzschmar et al. 1994), MED25 (Mittler et al. 2003), CBP / p300 (Ikeda et al. 2002), yeast SAGA and NuA4 complexes (Uitley et al. 1998), yeast Swi / Snf complex (Neely et al. 1999).

1.7 Repressors

Besides the passive repression that is due to a lack of transcriptional activators for example in certain cell types there are also proteins that actively contribute to transcriptional silencing of a gene. Several different mechanisms can be distinguished in eukaryotes:

- Repressors can compete directly with GTFs for binding to the promoter. A prominent example is NC2, which binds to TBP and prevents recruitment of TFIIA and TFIIB to the promoter (Goppelt et al. 1996).
- Another class of repressors competes with activators for the same binding site on DNA. The homeodomain protein Cut11 (CCAAT displacement protein) for example prevents CTF and C/EBF from binding to the CCAAT motif present in many promoters (Ottolenghi et al. 1989; Neufeld et al. 1992).
- Masking the activation domain of a transcription factor is another way to repress transcription. The yeast protein GAL80 for example can bind to the GAL4 activation domain and thereby prevent transcriptional activation.
- The recruitment of corepressors promotes a closed chromatin structure that renders chromosomal regions less accessible for transcription factors (Courey and Jia 2001). mSin3a as well as N-CoR (nuclear receptor corepressor) and SMRT (silencing mediator for retinoid and thyroid hormone receptors) all recruit a complex of histone deacetylases (HDACs) that modify histone tails and lead to a more closed chromatin structure.
- A less direct way of transcriptional repression is represented by factors that sequester activators in the cytoplasm preventing them from entering the nuclear compartment and binding to target sequences. An example is I κ B, which binds to NF κ B (Silverman and Maniatis 2001). Only after phosphorylation and degradation of I κ B can NF κ B migrate to the nucleus and induce transcription of target genes.

1.8 Transcriptional Cofactors

While some activators can directly interact with GTFs others need accessory proteins to establish a contact with the basal transcriptional machinery. These so-called cofactors of transcription usually lack a sequence specific DNA binding activity.

1.8.1 TBP-associated factors (TAFs)

Cloning of TBP facilitated biochemical studies that led to the identification of TBP-associated factors (TAFs) (Dymlacht et al. 1991). Members of this group function as coactivators in activated transcription, facilitate promoter opening and display enzymatic activity. TAF1 for example has three different activities: phosphorylation of RAP47 (TFIIF), HAT activity, ubiquitin-activating / conjugating activity for modification of histone H1 (Dikstein et al. 1996; Mizzen et al. 1996; Wassarman and Sauer 2001). It seems that several TAFs contribute to binding of TFIID to the promoter. Different TAFs inside the TFIID complex can contact different activation domains thereby increasing complexity of regulatory pathways (Verrijzer and Tjian 1996).

1.8.2 Cofactors of the USA fraction

The observation that increased purity of GTFs isolated from nuclear extracts has a negative influence on activator-driven transcription led to discovery of proteins that can mediate between activators and the basal transcription machinery (Flanagan et al. 1991; Meisterernst et al. 1991). USA (upstream stimulatory activity), a protein fraction from nuclear extracts containing this activity was further characterized and a series of cofactors were identified that were initially classified as positive (PC) or negative cofactors (NC) even though later it became clear that some of them play dual roles (Kaiser and Meisterernst 1996). The group of positive cofactors comprises PC1 (poly (ADP-ribose) polymerase, PARP), PC2 (a human mediator complex), PC3 (DNA topoisomerase I), PC4, PC5 and PC6. NC1 and NC2 together with HMG proteins 1 and 2 and Ada1 / Mot1 have been classified as negative cofactors.

1.8.3 Mediator complexes

The search for proteins that could reconstitute activated transcription *in vitro* also led to identification of the yeast mediator complex and finally to its purification (Flanagan et al. 1991; Kim et al. 1994). Despite a low level of sequence homology among different species this protein complex seems to be conserved throughout evolution (Bourbon et al. 2004). Over the last 15 years several different approaches were made to purify the mammalian complex, which led to the isolation of a range of complexes (TRAP, ARC / CRSP, DRIP, PC2, SMCC) that share most of the subunits but also exhibit subtle differences that are probably due to the purification method (Fondell et al. 1996; Ito et al. 1999; Naar et al. 1999; Naar et al. 2001; Sato et al. 2004; Blazek et al. 2005). The complex consists of 22 – 28 subunits and binds to activators as well as to the CTD of RNA polymerase II. Mediator promotes CTD phosphorylation in a two-fold manner: Cdk8 phosphorylates serine 5 of CTD but it also phosphorylates cyclin H, which in turn stimulates the activity of the TFIIH associated kinase Cdk7. The discovery of factors interacting with this huge complex is still ongoing and indicates that the complex plays a major role in modulating transcription in metazoan organisms.

1.9 Chromatin

The genetic material as it is found in the nucleus of eukaryotic cells is usually not present in the form of naked DNA but rather in a nucleoprotein complex that is termed chromatin. One obvious purpose of the formation of this structure is the compaction of about 2 meters of DNA. The proteins that bind to DNA to form chromatin are five types of histones: an octamer consisting of H2A, H2B, H3 and H4 are forming the protein core around which 146 bp of DNA are wrapped twice to form a nucleosome, the basic unit of chromatin (Thomas and Kornberg 1975; Luger et al. 1997). A stretch of about 50 nucleotides connects two adjacent nucleosomes. This linker is bound by histone H1, which is contributing to further compaction of the chromatin by contacting H2A and tethering neighboring nucleosomes together. Finally another level of compaction is achieved through binding of HMG and SIR

proteins that induce formation higher order chromatin loops and ultimately chromosomes.

1.9.1 Histone modifications

Histones fold into globular proteins that form the nucleosome core. An unstructured N-terminal unstructured tail is protruding from this structure, however. These histone tails are the targets for a range of different modifications. A directed link between these modifications and transcriptional regulation was established through the observation that certain co-activators are capable to modify residues in histone tails. Over the past years the efforts of many laboratories world-wide led to identification of a wide spectrum of different modifications and their relevance for the regulation of transcription. The picture emerging from these studies is suggesting a histone code, i. e. a scenario in which a combination of different histone tail modifications render chromatin regions more or less poised for transcription (Fig. 1) (Jenuwein and Allis 2001).

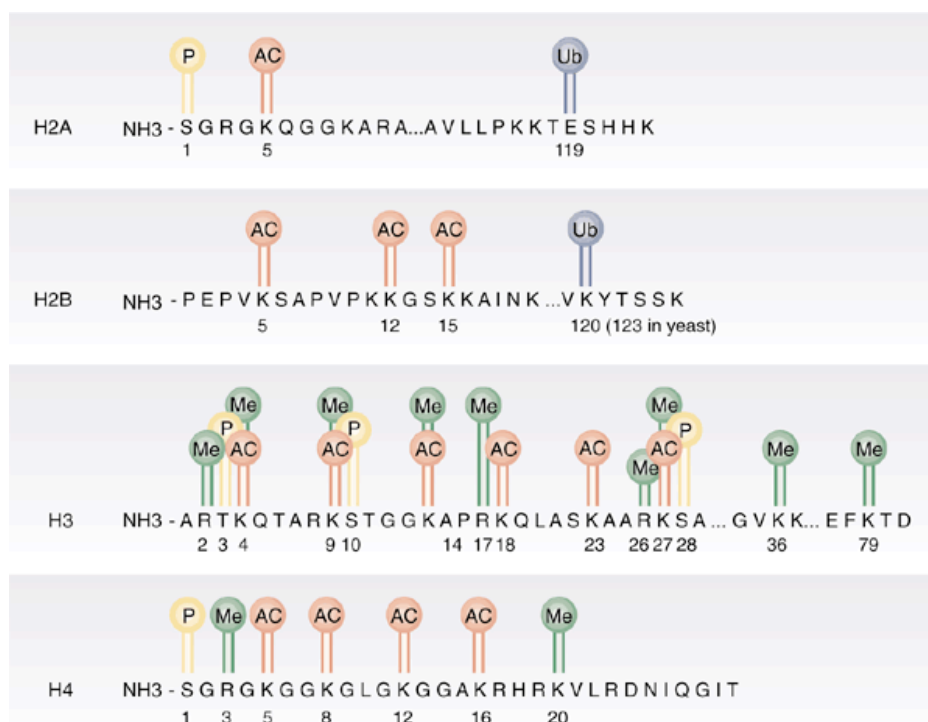


Fig. 1. The histone code. Tails of the nucleosome core histones H2A, H2B, H3 and H4 are shown together with the possible modifications that can take place on these residues (Jaskelioff and Peterson 2003). AC: acetylation, Me: methylation, P: phosphorylation, Ub: ubiquitination.

1.9.1.1 Histone acetylation

Acetylated histone tails are associated with transcriptionally competent chromatin regions (Peterson and Laniel 2004). It is speculated that the acetyl-groups weaken the interaction between DNA and nucleosomes. Many transcription factors can bind acetylated histones through the so-called bromodomain. It has also been shown that chromatin remodeling complexes work more efficiently with acetylated nucleosomes. The enzymes responsible for modifying histones in the nucleus are called histone acetyltransferases (HATs) and can be roughly categorized in three classes (Marmorstein and Roth 2001):

- GNAT (GCN5 related N-acetyltransferases) family members (e. g. GCN5, PCAF) have a C-terminal bromodomain and preferentially acetylate histone H3 at position lysine 14.
- Most of the members of the MYST (MOZ, Ybf2 / Sas3, Sas2, Tip60) family of HATs contain a chromodomain, which binds RNA and methylated histones. With the exception of Ybf2 / Sas3 these enzymes preferentially modify histone H4.
- CBP and p300 are two HATs with a very high degree of sequence homology. They do not only acetylate histones but a wide range of transcription factors (e. g. p53, TFIIE, TFIIIF, E2F1) can function as substrates as well.

Histone deacetylases (HDACs) are the enzymes that remove acetyl-groups from histones thereby counter-acting the activating function of HATs (Khochbin et al. 2001). There are three classes of HDACs:

class I HDACs comprise HDAC1, HDAC2, HDAC3 and HDAC8. HDAC1 and HDAC2 are components of Sin3 / HDAC and NurD / Mi2.

class II HDACs (HDACs 4–7) are poorly characterized.

The function of class III HDACs could be shown to depend on NAD⁺ in vitro providing possibly linking these enzymes to metabolism. The yeast Sir and the mammalian SIRT proteins (SIRT1–SIRT7) are representatives of this group.

1.9.1.2 Histone methylation

Both arginine and lysine residues of histone tails can be methylated (Fig. 1). Therefore two different types of histone methyltransferases (HMTs) are distinguished (Jenuwein 2001): PRMTs (protein arginine methyltransferases) methylate arginine residues in the H3 and H4 histone tails, which usually facilitates transcription from these chromatic regions. Both mono- and dimethylation has been reported. Lysine methyltransferases in contrast lead to mono-, di- or trimethylated residues in histone tail (Zhang et al. 2003). With the exception of Dot1 lysine HMTs have a SET domain, which has first been identified in *Drosophila* and which is also found in yeast. Histone lysine methylation can influence transcription in both directions: while for example H3K4 methylation is usually stimulating transcription, H3K9 methylation is a marker for transcriptionally repressed loci. The complete picture, however seems to be very complex with a multitude of different modified positions and crosstalk amongst them (Jenuwein and Allis 2001).

For a long time histone methylation has been seen as a stable mark that stays imprinted on chromosomes. This view was changed completely with the discovery of the first demethylase (LSD1) that demethylates H3K4 residues (Shi et al. 2004). Recently enzymes could be identified that remove methyl groups from H3K9 and H3K36 (Chen et al. 2006; Tsukada et al. 2006).

1.9.1.3 Additional histone modifications

Besides acetylation and methylation there is a range of other modifications that histone tails can undergo (Peterson and Laniel 2004). Serine and threonine residues seem to be phosphorylated in a cell-cycle dependent manner. H2A, H2B and H3 are subject to ubiquitination and ubiquitinated histones correlate with actively transcribed gene loci. ADP-ribosylation has been reported to be linked to transcription, apoptosis and genomic stability. SUMOylation and biotinylation have also been reported.

1.9.2 Chromatin remodeling

ATP-dependent chromatin remodeling machines do not covalently modify histone tails but rather introduce conformational changes at sites of transcriptional activity that increase the accessibility of DNA for transcription factors (Johnson et al. 2005). The replacement of nucleosomes involves breaking and reformation of histone–DNA contacts. This process requires energy, which is reflected in the presence of ATPases in all known remodeling complexes. Different types of these complexes have been found throughout Eukarya and groups of complexes have been classified according to their ATPases: Swi / Snf, ISWI, Mi2.

1.10 Transcriptional maintenance

In the context of development and cell differentiation it becomes particularly evident that certain transcriptional events affect the fate of daughter cells. Genes that have to be expressed transiently in embryogenesis for example need to be stably silenced. Other gene products are required in certain tissues but not others. If therefore transcription was initiated in a given cell there needs to be a way for the cell to transmit this information to daughter cells. Since this type of inherited information is adding another layer of regulation on top of the information that is contained in the genetic material per se it has been termed epigenetic regulation. While there are many open questions regarding the mechanisms of this process the picture emerging from the work presented in the last years suggests cooperativity of several molecular events.

1.10.1 DNA methylation

Cytosine residues in DNA from multicellular organisms can be methylated and this modification correlates with a repressed chromatin state and inhibition of gene expression (Bird and Wolffe 1999). Targets for methylation are usually found as CpG dinucleotides, which often appear clustered in regulatory regions.

The DNA modifying enzymes are called DNA methyltransferases (DNMTs) and two groups are distinguished based on the preferred DNA substrate: de

novo methylases (DNMT3a, DNMT3b) introduce cytosine methylation at previously non-methylated regions. The maintenance enzyme DNMT1 in contrast methylates newly synthesized DNA strands during replication using the methylation pattern of the parental strand as a template.

There seem to be several mechanisms to target DNMTs to their genomic target sites. It was shown that the PWWP domain of DNMT3b is necessary for binding of the protein to chromatin (Ge et al. 2004). The PWWP domain binds DNA in a sequence-independent fashion (Qiu et al. 2002) and its structure resembles the tudor domain found in 53BP1, a protein that binds to the tail of histone H3 methylated at lysine 79 (H3K79) (Huyen et al. 2004).

Two different mechanisms have been described for gene silencing via DNA methylation:

- Some DNA-binding factors are blocked from binding to methylated target sequences (e. g. USF) (Watt and Molloy 1988).
- Methyl-CpG-binding proteins (MBPs, e. g. MBD2, MeCP2) recruit co-repressors to methylated DNA-regions (Boyes and Bird 1991; Hendrich and Bird 1998). These molecules affect the chromatin structure through histone deacetylation (Jones et al. 1998; Nan et al. 1998; Ng et al. 1999), methylation (Sarraf and Stancheva 2004) or nucleosome remodeling (Wade et al. 1999; Zhang et al. 1999). It is not completely understood, however, what comes first in this process. There is evidence for example for the positive feedback of DNA methylation on H3K9 methylation (Fuks et al. 2003b; Sarraf and Stancheva 2004). On the other hand there have been reports showing the interaction of DNMTs with HP1, which recognizes H3K9 methylation as well as with HMTs and suggesting histone methylation as the primary event that recruits DNMTs in a second step (Fuks et al. 2003a; Lehnertz et al. 2003)

1.10.2 Polycomb and trithorax

Another class of epigenetic regulators comprises members of the Polycomb (Pc) and trithorax (trx) groups of genes. These genes were initially identified in *Drosophila* as proteins required for maintenance of spatial patterns of Hox gene expression, which is initiated during early embryogenesis by segmentation genes (Kennison 2004). Most of the Polycomb group (PcG) and trithorax group (trxG) genes have homologs in mammals that serve similar functions (Brock and Fisher 2005). In many cases, however, gene duplication led to the existence of more than one mammalian homolog for each *Drosophila* gene. Between these closely related proteins there seems to be functional overlap to a certain degree as seen for example with mammalian trithorax homologs (Glaser et al. 2006). PcG and trxG proteins are seen as antagonists in transcription maintenance: while PcG proteins keep loci repressed, trxG proteins are often required to maintain transcription (Hanson et al. 1999). Many PcG and trxG proteins contain conserved domains that have been identified in transcription factors or chromatin-associated proteins, e. g. chromodomains, bromodomains, zinc fingers, WD40 repeats (Brock and Fisher 2005). Furthermore some members display enzymatic activities that closely link them to processes like chromatin remodeling and histone tail modification:

- Some trxG members like for example hBRM contain an ATPase domain and act as the enzymatic subunit of chromatin remodeling complexes (Johnson et al. 2005).
- SET (Su(var)3–9, Enhancer of zeste) domains are found in proteins from both the PcG (e. g. ezh2) and the trxG (e. g. MLL) group (Alvarez-Venegas and Avramova 2002). These SET domains confer HMT activities, through which Polycomb and trithorax-like proteins influence histone tail methylation patterns. While transcriptional repression through PcG is associated with H3K27 methylation (Cao et al. 2002; Czermin et al. 2002; Kuzmichev et al. 2002; Muller et al. 2002; Lee et al. 2006), the activating mark introduced into histone tails by trxG is H3K4 methylation (Beisel et al. 2002; Milne et al. 2002; Nagy et al. 2002; Nakamura et al. 2002; Guenther et al. 2005; Milne et al. 2005c). While methylation of H3K4 leads to an opening of the chromatin structure and facilitates transcription H3K27 methylation is followed by chromatin compaction (Fig. 2).

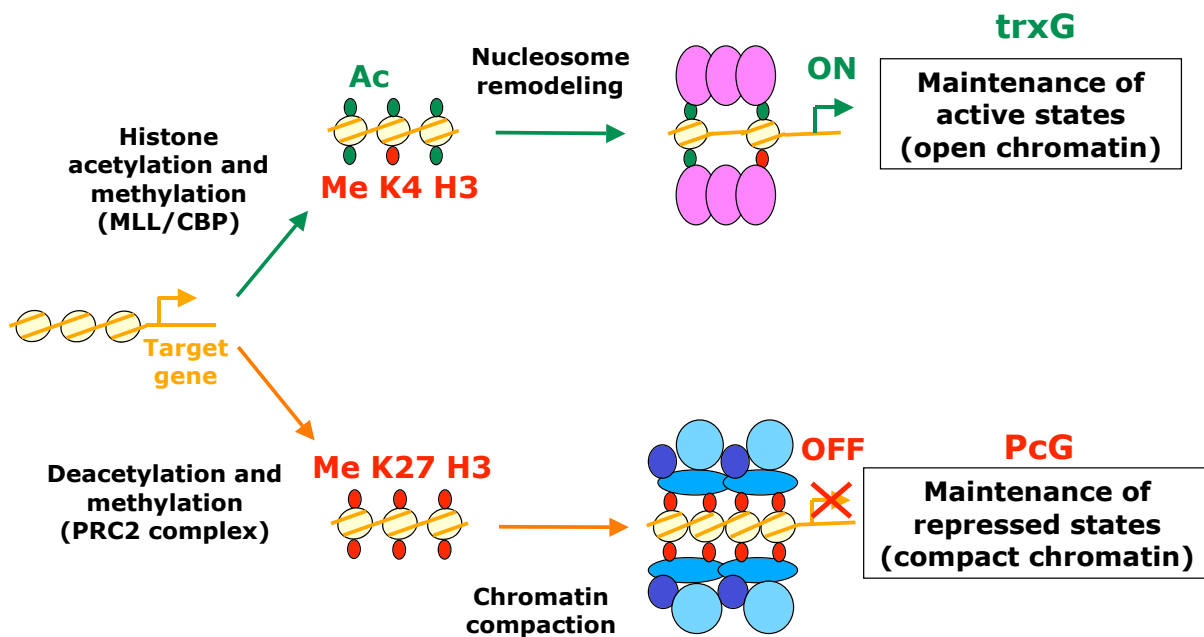


Figure 2. Transcriptional maintenance through trxG and PcG proteins. HMTases of the trithorax group of proteins (e.g. MLL) introduce H3K4 methyl marks in the chromatin and interact with HATs (e.g. CBP) that acetylate the same regions. These modifications lead to opening of the chromatin and recruitment of other activating complexes (e.g. chromatin remodeling complexes). trxG proteins (pink) maintain transcription after initial activation. PcG proteins (blue) antagonize these functions by introduction of H3K27 methylation, recruitment of deacetylation complexes and heterochromatin formation.

The exact mechanism of PcG and trxG action is not yet clear and the sequence of events in particular remains to be clarified. Single members of these gene groups are present in several protein complexes that play key roles in epigenetic regulation of transcription (Brock and Fisher 2005). The emerging picture is reflected by the so-called polarization model, according to which positive feedback loops involving different regulatory pathways stimulate each other leading to establishment of more or less stable chromatin states (Fig. 3) (Jaenisch and Bird 2003).

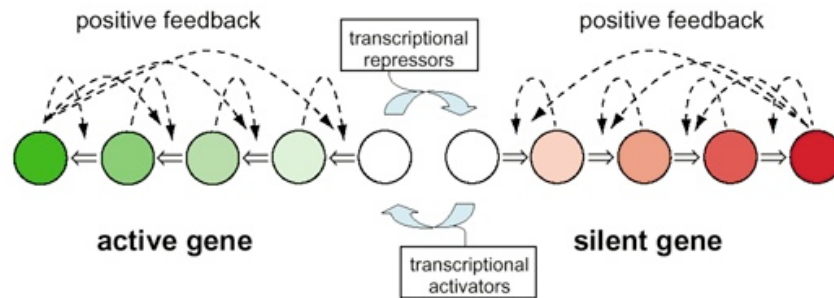


Figure 3. Polarization model. Each circle represents a particular chromatin modification status. The effect of positive feedback is to polarize domains at either the active or the inactive extremity. Broken lines indicate positive feedback loops between products and reactions that interconvert intermediate states. Negative interactions between the silent and active states of the gene occur but are not shown. The likelihood that a chromatin domain will occupy a particular state is indicated by the color intensity, with dark shades representing more probable states. (Jaenisch and Bird 2003)

1.10.2.1 MLL: a member of the human trithorax family

The protein MLL (ALL-1, MLL1, HRX) is the closest human homolog of *Drosophila* trithorax. Even before its role in transcriptional regulation was investigated this gene was already attracting the attention of scientists since it is involved in leukemogenesis. Cloning of the MLL gene (Ziemin-van der Poel et al. 1991; Djabali et al. 1992; Gu et al. 1992; Tkachuk et al. 1992) spurred research since many domains involved in transcription are contained in this large protein (Daser and Rabbitts 2005).

Two different regions have been identified that can mediate binding to DNA: three AT hooks are located near the N-terminus (Fig. 4) and it was reported that they bind the minor groove of AT-rich DNA (Zelevnik-Le et al. 1994). Another region (aa 1053–1119) has been termed MT domain for its homology to DNMT1 (Ma et al. 1993) and binds to unmethylated CpG (Birke et al. 2002). This domain is located inside a region (aa 1032–1395), which has been reported to repress transcription (RD, repression domain) (Slany et al. 1998). The RD region is bound not only by HDACs but also by the Polycomb repressor proteins HPC2 and BMI-1 and the co-repressor CTBP (Xia et al. 2003). In between the AT hooks and the repressive domain two motifs (SNL1 and SNL2) have been identified that are responsible for subnuclear localization of MLL (Yano et al. 1997). Site-specific cleavage of MLL by a threonine protease (taspase 1) (Yokoyama et al. 2002; Hsieh et al. 2003a) is taking place at two positions (CS1, CS2) (Hsieh et al. 2003b). After

this process the two parts of the protein (p180, p300) are still found in one complex and association is mediated through the FYRN and FYRC domains (Hsieh et al. 2003b).

The breakpoint cluster region (BCR) is marking the 3' border of the MLL moiety, which is retained in fusion proteins resulting from translocations at that position. On the 3' side of the breakpoint (and therefore lost in all fusions) are three PHD zinc finger motifs and a bromodomain. The nuclear cyclophilin CYP33 has been reported to bind to the third MLL PHD finger and negatively influence Hox gene transcription (Fair et al. 2001). It has been shown that CYP33 is also binding HDAC1 bound to the repression domain thereby stabilizing the HDAC1–MLL interaction (Xia et al. 2003). The transcriptional activation domain interacts with the co-activator CBP and facilitates its binding to phosphorylated CREB (Ernst et al. 2001). The C-terminal SET domain of MLL has been reported to methylate histone H3 at lysine 4 (Milne et al. 2002; Nakamura et al. 2002) and to bind to Ini1, a component of the Swi / Snf remodeling machinery (Cui et al. 1998), pointing out the importance of the protein as an epigenetic regulator.

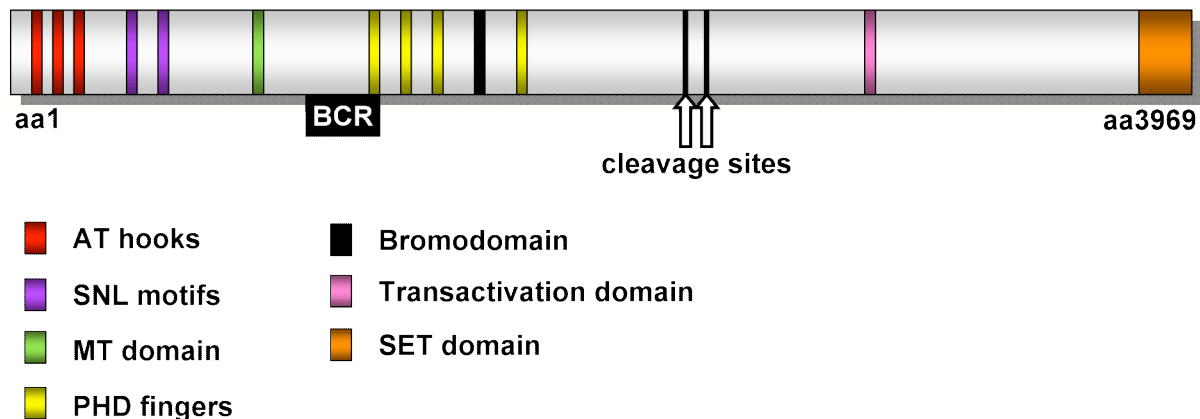


Figure 4. Domain structure of the MLL protein. Arrows are indicating points of cleavage through Taspase. Not to scale. SNL: subnuclear localization, MT: methyltransferase homology, BCR: breakpoint cluster region.

1.10.2.2 MLL-containing complexes

Biochemical approaches using chromatographic techniques to purify MLL from mammalian cell extracts led to the identification of three different complexes (Nakamura et al. 2002; Yokoyama et al. 2004; Dou et al. 2005). The first published report described a "supercomplex" containing proteins from six other complexes that are involved in transcriptional regulation (Nakamura et al. 2002). The HMTase activity associated with these complexes is specific for non- or monomethylated lysines and failed to trimethylate H3K4. The second purified complex is much smaller and there is not much overlap with the previously published supercomplex (Yokoyama et al. 2004). No enzymatic activity was demonstrated for this complex. It contains, however, the tumor suppressor protein Menin, which has been shown to bind directly to the MLL aminoterminal and to be necessary for Hox gene regulation (Yokoyama et al. 2005). Menin binds to both the MLL wild type protein as well as to MLL fusion proteins and loss of Menin leads to impaired Hox gene expression. Therefore it has been proposed that Menin is playing a role in tumorigenesis as well as in normal transcriptional regulation by MLL (Yokoyama et al. 2005).

The smaller MLL complex published by Yokoyama *et al.* seems to represent the "core unit" of the complex found in the cell. This is illustrated by the fact that a third complex that was purified in the Roeder laboratory is containing the same subunits alongside 23 other proteins (Dou et al. 2005). Interestingly most of the proteins initially identified in the MLL "supercomplex" are not present in this preparation. It contains, however, components of TFIID and the MOF, a MYST family HAT. Dou and coworkers demonstrated that the purified complex has HAT as well as HMT activity. In this report it was also shown that H3K4 is trimethylated by the MLL complex. This level of H3K4 methylation is strongly associated with transcribed regions of active genes in yeast as well as in higher eukaryotes (Krogan et al. 2002; Ng et al. 2003; Santos-Rosa et al. 2003; Schneider et al. 2004). SET1, the only H3K4 methyltransferase in yeast and a relative of MLL is introducing this type of modification suggesting that the MLL complex containing this enzymatic activity is the physiologically most relevant one.

1.10.2.3 Leukemic transformation by mutation of MLL

The human MLL gene is involved in acute leukemias, in particular infant leukemias and treatment-related leukemias, which are generally associated with a bad prognosis. More than 30 different fusion proteins have been described that arise from translocations of MLL to more than 50 different gene loci. These events all take place at the major break point region of MLL and result in fusions that consistently contain the MLL aminotermminus including AT hooks and MT domain. PHD fingers, bromodomain, transactivation domain and SET domain are lost on the other hand. Some of the fusion partners are cytoplasmic proteins (e. g. LASP1) (Strehl et al. 2003), the vast majority, however, is represented by nuclear factors involved in transcription (Daser and Rabbitts 2005).

Several models have been proposed for the mechanism of transformation through MLL fusions:

1. The translocation could create a fusion protein, which is acting as a dominant-negative allele and which represents a gain-of-function in comparison to the wild type allele. This model is supported by the fact that both the wild type and MLL fusion proteins target Hox gene promoters (Milne et al. 2002; Yokoyama et al. 2005). Since most of the fusion partners provide a transcriptional activation domain one could speculate that MLL fusions lead to overexpression due to loss of the C-terminal MLL moiety, which could have a regulatory role balancing the transcription rates from Hox gene promoters.
2. One obvious argument against this theory is the activating effect of the SET domain, which is lost in MLL fusion proteins together with a series of other features that might mediate transcriptional activation. The loss-of-function model for MLL leukemogenesis therefore proposes impairment of normal transcription by MLL wild type in the heterozygous situation. To date, however, no target gene for wild type MLL has been reported whose transcription rate is reduced by the presence of MLL fusion protein.
3. Another model for transformation by MLL is claiming that dimerization of MLL molecules might lead to deregulation of target genes. This model is supported by the identification of partial tandem duplications of MLL in some leukemia patients (Caligiuri et al. 1996; Schnittger et al. 2000). To test this hypothesis

MLL aminoterminal constructs were employed that had been genetically manipulated to form dimers or tetramers (Dobson et al. 2000; Martin et al. 2003) and transforming capacity of these constructs was reported to be linked to multimerization. In two MLL fusion partners (GAS7, AF1P) coiled–coil domains have been identified, that are thought to be mediating dimerization. These domains have been shown to be necessary for transformation suggesting that dimerization of MLL fusion proteins might be involved in leukemogenesis (So and Cleary 2003).

1.10.2.4 MLL target genes

The question how MLL fusion proteins transform mammalian cells is closely linked to the identification of target genes for the wild type protein as well as for the leukemic fusions. Early studies showed the importance of MLL as a maintenance factor for Hox gene transcription during mouse development (Yu et al. 1995; Yu et al. 1998). Mouse embryonic fibroblasts (MEFs) lacking MLL (MLL – / –) also revealed impaired Hox gene expression patterns (Schraets et al. 2003). Microarray analysis of leukemias demonstrated overexpression of Hoxa cluster genes indicating that wild type as well as leukemic fusion proteins act (at least in part) through the same target genes (Armstrong et al. 2002; Kohlmann et al. 2005). The overexpression of Hoxa9 has some leukemogenic potential (Kroon et al. 1998; Calvo et al. 2000). However, it has been reported that MLL fusion proteins can transform hematopoietic progenitor cells in absence of Hoxa9 (So et al. 2004) pointing out the existence of more than one pathway for cell immortalization. One step into the direction of identifying additional targets of MLL was the observation that MLL is critical for the transcription of cyclin–dependent kinase inhibitors p27 and p18 (Milne et al. 2005b). The authors proposed a model in this report according to which MLL is recruited to the promoters of these genes by the tumor suppressor Menin. Considering that both of these genes are cell cycle regulators that have been reported to be lacking in tumors this observation is providing some support for a loss–of–function model for MLL leukemogenesis.

Recently a genome–wide location analysis (ChIP–chip) showed colocalization of MLL with the vast majority of actively transcribed genes as assessed by detection of both RNA polymerase II and trimethylated H3K4 (Guenther et al. 2005). The authors

propose a role for MLL as a global transcription factor that binds to most of the actively transcribed gene promoters.

Another recent publication, however, showed that promoter occupancy by RNA polymerase II is not strictly correlated with MLL occupancy (Milne et al. 2005a). On the single gene level it was demonstrated that a MLL knockout mutation leads to decreased RNA polymerase II levels on the promoters of some direct target genes of MLL (e. g. *Hoxa7* and *Meis1*) but not on the promoters of other genes, which are actively transcribed in an MLL-independent manner (e. g. *GAPDH*). These data argue for a more specific role of MLL in transcription. It is becoming clear, however, that the target gene spectrum might well exceed Hox genes. One of the most important aspects of MLL research therefore remains the investigation of direct target genes for MLL and in particular the differences between the target gene populations of the wild type protein and leukemogenic fusions.

1.11 Objectives of this study

In this study the process of gene activation by MLL fusion proteins should be analyzed. In particular the question should be addressed, which binding sites are required to target MLL fusions to a promoter and what are the mechanisms of this recruitment. An inducible MLL fusion variant should be generated as a model system for the investigation of this process in hematopoietic cells. Mutagenesis of the Hoxa9 promoter should be used to identify critical sequences for the recruitment of MLL fusion proteins and for the characterization of the motifs that are involved in this event. Furthermore the inducible system should be employed for identification of new target genes of MLL fusion protein. On the single gene-level some candidates should be investigated using reporter assays as well as RT-PCR. A global screen for new MLL fusion protein target genes should be performed using Affymetrix microarrays.

2. Materials and Methods

2.1 Instruments and Accessories

Acrylamide gel electrophoresis	Amersham / Hoefer / BioRad
Agarose gel electrophoresis	BioRad
Analytical balance	AE 100 and 163, Mettler
Centrifuges	Avanti, Beckman Multifuge 3L-R, Heraeus, 5417 / 5415R, Eppendorf
Developing machine	Curix60, Agfa
Electroblot	semi-dry BioRad
Geigercounter	LB122, Berthold
Heatingblock	Thermomixer compact, Eppendorf
Homogenizer	Douncer, Wheaton
Incubator	WJ311, Forma Scientific Unequip, Unitherm B6200, Heraeus
Light microscope	Axiovert 25, Zeiss
PCR-Thermocycler	GeneAmp 2400, Applied
pH-Meter	Calimatic 760, Knick
Photometer	GeneQuant Pro, Amersham
Rotors	JA10, JA25-50, SW41, SW28, Beckman
Sonifier	W250 and 250-D, Branson
Ultra-centrifuges	L7, L8-M, Heraeus
UV-Illuminator	Bachofer

2.2 Chemicals and Biochemicals

Acetic acid (p.a.)	Roth
Acrylamide/Bisacrylamide 30%	Roth
Acrylamide 30% / 40%	Roth
Agarose	Invitrogen
Ammonium persulfate (APS)	Merck
Ampicillin	Roth
Aprotinin	Sigma
Bacto Agar	Difco
Bacto Trypton	Difco
Bacto Yeast Extract	Difco
Benzamidine	Sigma

Bisacrylamide 2%	Roth
Boric acide	Roth
Bradford reagent	BioRad
5-Bromo-4-Chloro-3-indolyl-phosphate (BCIP)	Peqlab
Bromphenol Blue	Sigma
BSA	Roche
Caesiumchlorid	Sigma
Calciumchloride	Merck
Calciumhydrogenphosphate	Merck
Calciumhydroxide	Merck
CHAPS	Sigma
Chloroform	Merck
Coomassie brilliant blue R-250	Sigma
DAPI	Sigma
Deoxycholot (DOC)	Sigma
Dimethylsulfoxide	Sigma
Dithiothreitol (DTT)	Roth
Doxycycline	Sigma
DMEM medium	Invitrogen
dNTPs	Roche
Ethanol	Sigma
Ethanolamine	Sigma
Ethidium bromide	Sigma
Ethylendiamintetraacetate disodium salt (EDTA)	Merck
Fetal calf serum (FCS)	Invitrogen
Ficoll 1.077	Amersham
Fish DNA	Sigma
G-CSF	Aventis
Giemsa solution	Merck
GM-CSF	Aventis
Glucose	Merck
Glykogen	Sigma
Glycerol	Roth
Glycine	Roth
HEPES	Biomol
4-Hydroxa-tamoxifen	Sigma
Hygromycin	Sigma
IGEPAL CA630 (NP-40)	Sigma
IL3, murine recombinant	Roche
IMDM medium	Invitrogen

IPTG	Roth
Isoamyl alcohol	Merck
Isopropanol	Merck
Leupeptin	Roche
Lithiumchlorid	Sigma
Magnesiumchloride	Merck
May-Gruenwald solution	Merck
b-Mercaptoethanol	Sigma
Methanol	Merck
Milk powder	Heirler, Roth
N-Laurylsarcosin	Sigma
Nitro-blue-tetrazolium (NBT)	Peqlab
Penicillin-Streptomycin	Invitrogen
Phenol/Chloroform	Roth
Phenylmethylsulfonfluoride (PMSF)	Biomol
Phorbolmyristylester (PMA)	Roche
Ponceau S	Sigma
Protein G-Sepharose	Amersham
Protein A-Sepharose	Amersham
Puromycin	Sigma
RPMI 1640 medium	Invitrogen
Sodium azide	Sigma
Sodium borate	Roth
Sodium carbonate	Merck
Sodium chloride	Roth
Sodiumdodecylsulfate (SDS)	Merck
Sodium fluoride	Sigma
Sodium hydroxid	Merck
Sucrose	Sigma
Tetramethylethylendiamin (TEMED)	Sigma
Thymidine	Sigma
Trishydroxidimethyl-aminomethan (Tris; p.a.)	Sigma
Triton X-100	Sigma
Trypsin-EDTA	Invitrogen
Urea	Roth
Xylene cyanole	Fluka

2.3 Additional material

Disposable plastic material	Greiner, Nunc, TPP, Falcon
Dialysing tubes (Viking, MWCO 15 kDa)	Roth
ECL Western Blot Kit	NEN
Film X-OMAT, BioMax	Kodak
Gel Drying Kit	Promega
GFX Gel Band Purification Kit	Amersham
Luciferase Kit	Promega
RNeasy Kit	Qiagen
Nitrocellulose membrane	BioRad
NuPage LDS sample buffer	Invitrogen
Nucleobond AX Plasmid DNA Kit	Machery & Nagel
RT-PCR Kit	Invitrogen
Siliconized Plastic tubes	Sorenson
Silver Staining Kit PlusOne	Amersham
Sterilfilters	Roth
Bottle top filters	Nalgene
Whatman 3MM Paper	Whatman
TNT Kit	Promega

2.4 Enzymes

Calf intestine phosphatase	MBI Fermentas
Klenow Fragment	MBI Fermentas
Lysozym	Sigma
Pfu Polymerase	MBI Fermentas
Restriction enzymes	NEB or MBI Fermentas
RNase A	Roche
T4 DNA ligase	MBI Fermentas
T4 polynucleotide kinase	MBI Fermentas
Taq polymerase	MBI Fermentas

2.5 Antibodies

Table 3. Primary antibodies

antibody	origin	provider	dilution for WB
MLC 5A9	rat	E. Kremmer	1 : 5
MLC 4G10	rat	E. Kremmer	1 : 5
MLC 5C10	rat	E. Kremmer	1 : 5
MLN 7D4	rat	E. Kremmer	1 : 5
MLN 7E12	rat	E. Kremmer	1 : 5
anti-HA 3F10 (ROO1)	rat	E. Kremmer	1 : 5
anti-dog-CD3	rat	E. Kremmer	1 : 5
ER (sc-8002)	mouse	Santa Cruz	1 : 500

Table 4. Secondary antibodies

antibody	provider	dilution for WB
anti-rabbit	Promega	1:5000
anti-mouse	Promega	1:5000
anti-rat	Biomol	1:3000

2.6 General buffers

10 x TBE:	1 M Tris 1 M Boric acid 20 mM EDTA, pH 8.0	20 x PBS:	160 g NaCl 4 g KCl 23 g Na ₂ HPO ₄ x 2 H ₂ O 4 g KH ₂ PO ₄ H ₂ O ad 1000 ml pH 7.2 –7.4
10 x TGS:	250 mM Tris 1.92 M Glycine 1 % (w/v) SDS		
BC-Puffer:	20 mM Tris-HCl, pH 7.3 0.2 mM EDTA, pH 8.0 0 – 2000 mM KCl 20 % (v/v) Glycerin	10 x TBS:	24.2 g Tris 80 g NaCl 2 g KCl H ₂ O ad 1000 ml, pH 7.6

2.7 Oligonucleotides

Table 5. Oligonucleotides

oligo	sequence	purpose
mllab4up	CGAATTCTTGCGAGAACCGACATTTAG	PCR primers for cloning of GST-MLLaa612-935 (pML13)
mllab4down	CCGGATCCCCTTCTTCCGCCCTGTCGT	
mllab5up	CGAATTCACCCAAAATCCAGCAAATGAA	PCR primers for cloning of GST-MLLaa3620-3969 (pML14)
mllab5down	CCGGATCCCTTAGTTGATGAACTTCCGGCA	
HAMCS sense	AATTCTCGAGATCTGCGGCCGCTACCCATACGATGT TCCAGATTACGCTCTTTGAG	multiple cloning site for vector pML3 (pML15)
HAMCS antisense	CTAGCTCAAAGAGCGTAATCTGGAACATCGTATGGG TAGCGGCCGCAGATCTCGAG	
VP16AD XhoIup	CCGCTCGAGCGGTGACGGCCCCCGACCGAT	PCR primers for cloning of VP16 activation domain (pML16)
VP16AD XhoIup	GAAGATCTTCCCACCGTACTCGTCAATT	
ERT2.2up	GCGGCCGCGATCCATCTGCTGGAGACA	PCR primers for cloning of ER ^{T2} (pML27)
ERT2.2 down	GCGGCCGCGAGCTGTGGCAGGGAAACCCT	
HRXN1up	TCGAATTCCGCCACCATGGCG	PCR primers for cloning of MLL aa1-1419 (pML26)
HRXN1 down	CCGCTCGAGCGGCCTCCCATCTCCCA	
pREP4 MCS sense4	GATCTGGTACCGCTAGCGCGGCCGCA	multiple cloning site for vector pML94 (pML73)
pREP4 MCS antisense4	AGATCTGCGGCCGCGCTAGCGGTACCA	
p21 -2325	AGAGGTACCGCTTGGGCAGCAGGCTGTG	PCR primers for cloning of human p21 promoter (pML76, pML78)
p21 -215	AGAGGTACCGCACGCGAGGTTCCGGG	
p21 +8	GTCTGCTAGCACTTCGGCAGCTGCTCACACC	
hoxa9 -1953	AGAGGTACCGACTGCGGGGTATTTAGGACACGGTA CCGACTGCGGGGTATTTAGGACACGGT	PCR primers for cloning of human hoxa9 promoter constructs (pML74, pML82, pML84, pML86, pML93)
hoxa9 -1010	AGAGGTACCCTGGGAATCCTGATTGCCAGCTGATG AGA	
hoxa9 -510	AGAGGTACCCAGCCTGTGTGGCTTCTGAAACAATA AC	
hoxa9 -207	AGAGGTACCCATCGTAGAGCGGCACGATCCCTT	
hoxa9 +46	GTCTGCTAGCCCGACCCACGGAAATTATGAAACTGC AGA	

hoxa9 +98	GTCTGCTAGCCGACCCACGGAAATTATGAAACT	
CpGmut_s	CGGTGGGGAGTGATTTAGGGGTTATTGTTCTGCTG GAGGGGCAC	cloning of pML109
CpGmut_as	GTGCCCTCCAGCAGAACAATAACCCCTAAATCACT CCCCACCGGTAC	
GCCCbox mut_s	CCGTGCGGAGTGATTTACGCGTTATTGTTCTGCTGG ACGTGCAC	cloning of pML111
GCCCbox mut_as	GTGCACGTCCAGCAGAACAATAACGCGTAAATCACT CCGCACGGGTAC	
hoxa9site mut_s	CCGTGCGGAGTGATGCACGCGTTATTGTTCTGCTG GACGGGCAC	cloning of pML113
hoxa9site mut_as	GTGCCCGTCCAGCAGAACAATAACGCGTGCATCAC TCCGCACGGGTAC	
a9_WT_s	CCGTGCGGAGTGATTTACGCGTTATTGTTCTGCTGG ACGGGCAC	cloning of pML115
a9_WT_as	GTGCCCGTCCAGCAGAACAATAACGCGTAAAATCAC TCCGCACGGGTAC	
2xSP1FF_as	GTGACCCCGCCCCGCTGCCAGCACCCCGCCCCG CGGTAC	cloning of pML126
2xSP1FF_s	CGCGGGGCGGGGTGCTGGGCAGCGGGGCGGGGT CAC	
3xMZF1_as	GTGTTCCCCACTGTTCCCCACTGTTCCCCACTGGCG GTAC	cloning of pML128
3xMZF1_s	CGCCAGTGGGGAACAGTGGGGAACAGTGGGGAAC AC	
2xGAL4_as	GTGGCGGAGTACTGTCTCCGAGCGGAGTACTGTC CTCCGAGCGGGGTAC	cloning of pML134
2xGAL4_s	CCCGCTCGGAGGACAGTACTCCGCTCGGAGGACAG TACTCCGCCAC	
3xSP1FFF _as	GTGACCCCGCCCCACCCCGCCCCACCCCGCCCCG CGGTAC	cloning of pML122
3xSP1FFF_s	CGCGGGGCGGGGTGGGGCGGGGTGGGGCGGGGT CAC	
3xSP1FRF _as	GTGACCCCGCCCCGGGGCGGGGTACCCCGCCCCG CGGTAC	cloning of pML124
3xSP1FRF_s	CGCGGGGCGGGGTACCCCGCCCCGGGGCGGGGT CAC	
3xEBOX_as	GTGCCACGTGCTGGCCACGTGGCGCACGTGGCAG CGGTA	cloning of pML132

3xEBOX_s	CGCTGCCACGTGCGCCACGTGGCCAGCACGTGGC AC	
3xUSF_as	GTGGTCACGTGGTGTACGTGGTGTACGTGGTGC GGTA	cloning of pML130
3xUSF_s	CGCACCACGTGACACCACGTGACACCACGTGACCA C	
HA9F	ATCCCAATAACCCAGCAG	RT-PCR primers for human Hoxa9
HA9R	CAGAAACTCTTTCTCCAGTTCC	
B-actinFw	TGCGTTGTTACAGGAAGTCCC	RT-PCR promoters for human beta-actin
B-actinRev	CTATCACCTCCCCTGTGTGGA	
p21-FW	CATGTGTCCTGGTTCCCCTT	RT-PCR primers for human p21
p21-Rev	TCAGCATTGTGGGAGGAGC	
hoxa9F-65	GCCGGCAACTTATTAGGTGACTG	ChIP primers for episomal Hoxa9 promoter
lucR_short	AGCTTACTTAGATCTGCGGCCG	
hoxa9end2	ACGGGAGAGTACAGAGACAAGG	ChIP primers for chromosomal Hoxa9 promoter
hoxa9end_r	GGGGGAAGTACAGTCACCTAATA	

2.8 Plasmids

Table 6. Plasmids used in this work.

plasmid	features	remarks
pML1	pCMV-MLL	gift from R. Slany
pML3	pCAG3SIP	gift from T. Schroeder
pML9	pGEX-B	
pML13	GST-MLL-aa613-935	
pML14	GST-MLL-aa3620-3969	
pML31	MLL-NTD-HA	
pML33	MLL-VP16-HA	
pML37	MLL-VP16-ER-HA	
pML46	ER-VP-16-HA	
pML73	episomal reporter vector (pREP4-luc)	
pML51	MIG-MLL-AF4	gift from C. Buske
pML52	MIG-MLL-AF9	gift from C. Buske
pML56	pMSCV-FLAG-MLL	gift from R. Slany
pML60	c-myc-luc -2332/+513	pCG362 from D. Eick
pML61	c-myc-luc -2332/+513/IgEnh	pCG363 from D. Eick

pML62	c-myc-luc +66/+513	pRF235 from D. Eick
pML63	c-myc-luc +66/+513/IgEnh	pCG508 from D. Eick
pML64	c-myc-luc -101/+66	pRF278-1 from D. Eick
pML65	c-myc-luc -101/+66/IgEnh	pCG5-2 from D. Eick
pML74	episomal Hoxa9 reporter plasmid -1010/+46 luciferase	
pML76	episomal p21 reporter plasmid -2325/+8 luciferase	
pML78	episomal p21 reporter plasmid -215/+8 luciferase	
pML82	episomal Hoxa9 reporter plasmid -207/+46 luciferase	
pML105	episomal Hoxa9 reporter plasmid -78/+46 luciferase	
pML107	episomal Hoxa9 reporter plasmid -207/+98 luciferase	
pML109	episomal Hoxa9 reporter plasmid -118/+46 mutant oligo "CpGmut"	CpG mutant 2
pML111	episomal Hoxa9 reporter plasmid -118/+46 mutant oligo "GCCboxmut"	GGGC mutant
pML113	episomal Hoxa9 reporter plasmid -118/+46 mutant oligo "hoxa9sitemut"	Hox site mutant
pML115	episomal Hoxa9 reporter plasmid -118/+46 mutant oligo "a9_WT"	
pML123	episomal Hoxa9 reporter plasmid -78/+46 with oligo 3xSP1FFF	MAZR1
pML124	episomal Hoxa9 reporter plasmid -78/+46 with oligo 3xSP1FRF	MAZR3
pML126	episomal Hoxa9 reporter plasmid -78/+46 with oligo 2xSP1FF	MAZR2
pML128	episomal Hoxa9 reporter plasmid -78/+46 with oligo 3xMZF1	MZF 1

pML131	episomal Hoxa9 reporter plasmid –78/+46 with oligo 3xUSF	E–BOX1
pML133	episomal Hoxa9 reporter plasmid –78/+46 with oligo 3xEBOX	E–BOX2
pML135	episomal Hoxa9 reporter plasmid –78/+46 with oligo 2xGAL	GAL
pML146	episomal Hoxa9 reporter plasmid –207/+46 Geneart construct 055107	Hox site mutant
pML147	episomal Hoxa9 reporter plasmid –207/+46 Geneart construct 055106	GC mutant
pML151	episomal Hoxa9 reporter plasmid –207/+46 Geneart construct 055113	GGGC mutant
pML152	episomal Hoxa9 reporter plasmid –207/+46 Geneart construct 055114	GGC mutant
pML153	episomal Hoxa9 reporter plasmid –207/+46 Geneart construct 055115	SP1 sites up–mutant
pML154	episomal Hoxa9 reporter plasmid –207/+46 Geneart construct 055112	CpG mutant 1
pML155	episomal Hoxa9 reporter plasmid –207/+46 Geneart construct 055111 nt 9459 deleted from pML160	TATA–less mutant 1
pML156	episomal Hoxa9 reporter plasmid –207/+46 Geneart construct 055110	AdML–TATA box
pML157	episomal Hoxa9 reporter plasmid –207/+46 Geneart construct 055109	AdML–Inr

pML158	episomal Hoxa9 reporter plasmid –207/+46 Geneart construct 055108	core promoter C–stretch mutant
pML160	episomal Hoxa9 reporter plasmid –207/+46 Geneart construct 055111	TATA–less mutant 2
pML161	Gal93–SP1.Q1 SP1 132–243 fused to Gal1–93	P. Halle
pML162	pBXG1 / pPF36 control vector for pML161	P. Halle
pML163	pCMV–SP1 / pPF27 full–length SP1	P. Halle
pML165	Gal–SP1 SP1 75–597 fused to Gal1–147	gift from W. Hammerschmidt
pML166	Gal only vector (control for pML165)	gift from W. Hammerschmidt
pLS23	p21 reporter plasmid –2325 / +8 luciferase	gift from T. Kardassis
pLS24	p21 reporter plasmid –215 / +8 luciferase	gift from T. Kardassis
pLS25	p21 reporter plasmid –143 / +8 luciferase	gift from T. Kardassis
pLS26	p21 reporter plasmid –2325 / +8, delta –122 / –60 luciferase	gift from T. Kardassis

2.9 Cloning

2.9.1 Polymerase chain reaction (PCR)

PCR reactions were performed using Pfu polymerase (MBI, cat.no. 1433) and the following protocol:

10 µl	buffer w/o MgSO ₄
10 µl	DMSO 50%
1 µl	dNTP 25mM
10 µl	MgSO ₄ 25mM
0.5 µl	forward primer (100 pmol / µl)
0.5 µl	reverse primer (100 pmol / µl)
10 µl	template DNA (10 ng / µl)
2 µl	Pfu Polymerase (2.5u / µl)
56 µl	H ₂ O

Duration of elongation cycles were adjusted according to the length of the desired amplicon. Pfu polymerase is thought to have a processivity of 500 bp / min.

2.9.2 Restriction digests, fragment isolation and ligation of DNA

Restriction digests were done using enzymes from MBI Fermentas and the accompanying buffer solutions following the manufacturer's protocol. In general 5 µg of DNA were digested overnight with 10 u of enzyme. After a heat inactivation step (10 min / 70°C) fragments were purified on agarose gels containing Ethidiumbromide. Depending on fragment length gels containing between 0.5 and 1.5% agarose were used. Fragments were cut out from the gel under UV light and DNA was isolated from the gel using the GFX Gel Band Purification Kit (Amersham)

Ligation reactions

Ligations were usually performed overnight using T4 DNA ligase (MBI) in a total volume of 15 µl. The insert was added to the reaction in a five-fold excess in comparison to the vector. 1µl of enzyme was added on ice before incubation at 16°C.

2.9.3 Transformation into *E. coli* DH5 α

Before transformation into bacteria the ligation reaction was stopped and enzymes were inactivated at 70°C for 10 min. A heat shock protocol was used to transform *E. coli* DH5 α : bacteria were thawed on ice, DNA was added and incubated together with the bacteria for 20 min on ice. The heat shock was performed for 90 sec at 42°C, after which 1 ml of LB medium was added to the transformation sample. After an incubation in a shaker at 37°C bacteria were spun down (1.000 g, 3 min). The pellet was resuspended in 150 μ l of LB medium and plated on LB plates containing the appropriate antibiotic for selection.

2.9.4 Plasmid generation

MLL–VP16–ER–HA expression plasmid (pML37)

A multiple cloning site containing a hemagglutinine tag (HAMCS) was introduced into pCAG3–SIP (T. Schroeder), thereby generating pML15. The VP16 activation domain was amplified via PCR using the primer pair "VP16ADXhoI" and pKI14 as a template. The PCR product was digested with XhoI and BglII and cloned into pML15, creating pML16. The ERT2 variant of the human estrogen receptor was amplified via PCR, which was performed using the primer pair "ERT2.2" and pML2 as a template. The product of this reaction was digested with NotI, gel–purified and and subcloned into a pBluescript vector (Stratagene) and the resulting construt was termed pML27. The NotI–fragment was isolated from this vector, purified and ligated into pML16 that had been linearized using NotI. The vector resulting from this cloning step is pML28. Finally MLL1–1419 was amplified via PCR using the primer pair "HRXN1" and pML1 as a template. The product was digested with EcoRI and XhoI, purified on an agarose gel and subcloned into pBluescript. The resulting plasmid was called pML26. The EcoRI / XhoI–fragment was isolated from this vector and ligated with the corresponding sites in pML28. The resulting vector is called pML37.

Cloning of GST fusion proteins (pML13, pML14)

Primer pairs "mllab4" and "mllab5" have been used to amplify the MLL cDNA regions coding for amino acids 612–935 (pML13) and 3620–3969 (pML14), respectively. PCR products were digested with EcoRI and BamHI, purified on an agarose gel and ligated into the GST–expression vector pGEX–B, which had been linearized using the same enzymes.

Generation of an episomal reporter vector (pML73)

An episomal reporter plasmid was generated by transfer of the luciferase gene from pGL2 basic (Promega) into pREP4 (Invitrogen). The RSV promoter was removed from pREP4 by removal of the Sall / XbaI–fragment. The remaining vector backbone was ligated with the luciferase gene isolated from pGL2 basic by restriction digest using NheI and Sall. This vector was termed pML94. For easier cloning a multiple cloning site was introduced ("pREP4MCS4") upstream of the luciferase coding region. This plasmid was called pML73.

p21 episomal reporter constructs (pML76, pML78)

PCR reactions were performed using pLS23 as a template. Specific primers were used to generate p21 –2325 / +8 (pML76) and p21 –215 / +8 (pML78). PCR products were digested with NheI and KpnI, agarose–gel purified and ligated to pML73, which had been linearized using the same enzymes.

Hoxa9 episomal reporter constructs (pML74, pML82, pML84, pML86, pML93)

Hoxa9 promoter constructs comprising Hoxa9 –1953 / +46 (pML84), Hoxa9 –1010 / +46 (pML74), Hoxa9 –510 / +46 (pML86), Hoxa9 –207 / +46 (pML82) and Hoxa9 –207 / +98 (pML93) were generated as described for the p21 promoter constructs. A RZPD clone containing the human genomic Hoxa locus (RPCIP704019170Q2) was used as a template for PCR reactions. Primers are listed below.

Generation of Hoxa9 –78 / +46 (pML105) and Hoxa9 –118 / +46 (pML115)

pML82 was digested with KpnI and PmlI and religated after blunting of restriction sites to generate the construct Hoxa9 –78 / +46 (pML105). Hoxa9 –118 / +46 (pML115) was created by removal of the KpnI / PmlI–fragment from pML82 and insertion of the oligonucleotide "a9_WT".

2.9.5 Site directed mutagenesis of the human Hoxa9 promoter

The Hoxa9 –118 / +46 point mutants were generated in a similar way to pML115: the KpnI / PmlI–fragment was removed from pML82 and oligonucleotides comprising the region –118 to –78 of the Hoxa9 promoter were inserted using the KpnI and PmlI sites. Different point mutants were generated by this procedure:

pML109	Hoxa9 –118 / +46	CpG mutant 2
pML111	Hoxa9 –118 / +46	GGGC mutant
pML113	Hoxa9 –118 / +46	Hox site mutant

The longer Hoxa9 mutant constructs (–207 / +46) were created by transfer of the KpnI / NheI–fragment from a series of vectors synthesized by Geneart that are containing mutant variants of Hoxa9 –207 / +46 into pML73 linearized using KpnI and NheI.

pML146	Hoxa9 –207 / +46	Hox site mutant
pML147	Hoxa9 –207 / +46	GC mutant
pML151	Hoxa9 –207 / +46	GGGC mutant
pML152	Hoxa9 –207 / +46	GGC mutant
pML155	Hoxa9 –207 / +46	TATA–less mutant 1
pML156	Hoxa9 –207 / +46	AdML–TATA box
pML157	Hoxa9 –207 / +46	AdML–Inr
pML158	Hoxa9 –207 / +46	core promoter C–stretch mutant
pML160	Hoxa9 –207 / +46	TATA–less mutant 2

2.9.6 Generation of Hoxa9 –78 / +46 containing additional binding sites

After removal of the KpnI / PmlI-fragment from pML82 oligonucleotides containing binding sites for different transcription factors were inserted to generate pML123 (MAZR1), pML124 (MAZR3), pML126 (MAZR2), pML128 (MZF1), pML131 (E-BOX1), pML133 (E-BOX2), pML135 (GAL).

2.10 Cell culture

All cell lines were grown at 37°C, 5% CO₂ and 95% humidity in an incubator (WJ311, Forma Scientific). A Neubauer chamber was used to determine cell densities. Dead cells were excluded from the count by staining with Trypan Blue.

2.10.1 Separation of viable and dead cells via density gradient centrifugation

Centrifugation on a Ficoll cushion was used to remove dead cells from U937 and FDCP-mix cell suspensions. Ficoll with an appropriate density for myeloid cells (1.077, Amersham) was overlaid with cells suspended in growth medium. and centrifuged for 20 min at 400 g (RT). Viable cells were enriched in a layer separating on top of the lower phase and were transferred to a new tube using a pipette. These cells were then washed and kept in culture.

2.10.2 Freezing and thawing of cell lines

Long term storage of cell lines was performed by keeping frozen aliquots in liquid nitrogen. To generate these aliquots $5 \times 10^6 - 1 \times 10^7$ cells were resuspended in the corresponding growth serum and cooled down slowly using freezing containers (Nalgene) that provide for decrease of temperature at a rate of 1°C per minute when transferred from room temperature to – 80°C.

To thaw cells these aliquots were brought to 37°C and 10 ml of growth medium were added drop-wise. After 10 min of incubation at room temperature cells were centrifuged and the pellet was resuspended in 6 to 8 ml of growth medium and incubated at standard conditions.

2.10.3 Isolation of monoclonal cell lines using semisolid medium

For the generation of monoclonal cell lines and for determination of the colony forming capacity of cell populations U937 and FDCP–mix cells were resuspended in growth medium in a serial dilution. The clonogenicity of a population of undifferentiated FDCPmix cells is in the range of 5 – 10% of the total number of viable cells. For practical reasons it is therefore advisable to seed roughly 300 to 500 cells per 6-well. The cell suspensions were mixed with Bakto–Agar (Difco, Germany) to reach a final agar concentration of 0.33% and 3ml each were transferred to 6-well plates. After 10–12 days of incubation cell clusters of more than 100 cells were counted using a microscope.

For isolation of stable monoclonal puromycin-resistant cell lines puromycin (Sigma, Germany) was added to the growth medium at a concentration of 1–1.5 µg / ml. Single colonies were picked using a 200 µl–Gilson pipette and were transferred for expansion to a 96–well plate containing liquid growth medium.

2.10.4 Preparation and staining of cytopins

Suspension cells can be immobilized on glass slides and stained for morphological analysis in a light-microscope. So–called Cytospins are prepared by centrifugation (400 rpm, 4 min, RT) of suspension cells in a Cytospin centrifuge (Shandon, England). The slides are air-dried after centrifugation, stained with May-Gruenwald solution (Merck, Germany) for 4 min, rinsed with water, stained with 5% Giemsa solution (Merck, Germany) for 16 min, rinsed thoroughly and left for drying. The number of different cell types was determined in blind studies involving microscopical analysis of 100–200 cells per sample.

2.10.5 Culture of FDCPmix cells

FDCPmix cell lines A4 and A7 were grown in Iscove's Modified Dulbecco's Medium (IMDM, Invitrogen, Germany) with an osmolarity of 310–320 mosmol / kg. The medium was containing 20% horse serum (Invitrogen, Germany) and 10 – 15% of interleukin 3 – conditioned medium (IL3–cm). Cell density was checked and kept in a range between 2×10^5 and 1×10^6 / ml.

2.10.6 Culture of U937 cells and THP-1 cells

The cell lines U937 and THP-1 were treated equally and were grown in RPMI medium (Invitrogen, Germany) containing 10% fetal bovine serum. In some cases Penstrep (Invitrogen, Germany) was added to the medium to reduce the risk of contamination.

2.10.7 Production of interleukin 3 – conditioned medium

To provide for IL3 in the FDCPmix growth medium conditioned medium is added. This is basically comprised of a supernatant from a cell line secreting murine IL3. The cell line used for this work was mL3 (T. Schroeder, personal communication), which carries an episomal expression vector for murine IL3, which mediates neomycin-resistance. The cell line can be grown and expanded in RPMI (Invitrogen, Germany) supplemented with 10% fetal bovine serum and 1 mg / ml of neomycin (Sigma, Germany). After expansion a given number of cells was pelleted, washed and seeded in IMDM containing 5% horse serum at a density of 2×10^5 / ml. After 60 hours the supernatant was collected using centrifugation in combination with bottle top-filters with a pore size of 0.44 μm (Nalgene, Germany). Aliquots were prepared and stored at -20°C . IL3cm prepared in this way can be stored for 6 – 12 months.

For determination of the biological activity of a certain batch of IL3cm colony forming assays were performed. IL3cm was serially diluted and added to IMDM containing 20% horse serum. FDCPmix cells were washed with IL3-free medium and added to the different dilutions of IL3cm. The concentration of IL3cm leading to the highest colony numbers was considered the optimal working concentration. This usually corresponded to 10 – 15% depending on the batch of IL3cm. As a control recombinant murine IL3 (Roche, Germany) can be used. The optimum final concentration in this case is 150 u / ml.

2.10.8 Differentiation of FDCPmix cells into granulocytes and macrophages

FDCPmix cells were washed twice in IL3-free growth medium and resuspended at a cell density of $1-2 \times 10^5$ cells per ml in differentiation medium, which is containing 20% pre-tested FCS. 2 units recombinant murine IL3 (Roche) / ml, 250 units GM-CSF (Aventis) / ml and 1000 units recombinant G-CSF (Aventis) were contained as well in the medium for the first three days of differentiation. On day 4 cell density was adjusted again to 2×10^5 per ml by dilution with differentiation medium without IL3. Granulocytes are usually developing in the course of 7 days, macrophage differentiation takes 8–10 days.

2.11 Transfection experiments

2.11.1 Electroporation of suspension cells

For electroporation 4×10^6 cells (FDCPmix) and 1×10^7 cells (U937) were resuspended in 400 μ l of standard growth medium and transferred to 4mm-cuvettes (BioRad, Germany). GenePulser2 (BioRad, Germany) was set to 260 V and 1050 μ F while the capacitance extender (BioRad, Germany) was set to maximum output. The resulting time constants were usually in the range of 30 – 35 ms.

2.11.2 Transient transfection of FDCPmix cells

FDCPmix cells were electroporated as outlined above. In general 10 μ g of reporter plasmid DNA were used in combination with different amounts of expression vector (usually 1 – 15 μ g). After electroporation cells were transferred to a 6-well plate containing 4 ml of growth medium. Luciferase-assays were performed 24 hours after transfection.

For normalization a β -galactosidase assay was performed. 5 μ g of pML48 (CMV- β -Gal) were co-transfected with the constructs of interest. In general 20 μ l of the lysate that was used for luciferase detection were added to 180 μ l of β -Gal substrate solution. The color conversion was measured at 415 nm using a plate reader.

β-Gal substrate solution:

1.1mM MgCl ₂ ,
1mg / ml ONPG,
82mM Na ₂ HPO ₄ ,
18mM NaH ₂ PO ₄ ,
50mM β-Mercaptoethanol.

2.11.3 Stable transfection of FDCPmix cells

For the generation of stable cell lines electroporations were performed as outlined above using 15 µg of linearized plasmid DNA. 24 hours after transfection the medium was replaced by fresh growth medium containing puromycin at a concentration of 0.3 µg / ml. After another two days the majority of cells were non-viable. Dead cells and debris was removed using separation over a Ficoll gradient. The remaining cells were resuspended in fresh growth medium at a density of 2×10^5 / ml. Stably transfected cells usually started proliferating at day 6 or day 7 after electroporation. From that point on puromycin concentration was gradually increased up to 1.5 µg / ml. If necessary medium was replaced and / or cells were separated using Ficoll gradient centrifugation. In general it took about two weeks to obtain a stably proliferating population of puromycin-resistant cells. An aliquot of these was frozen as a back-up, the rest was used for selection of monoclonal cell lines in semisolid medium.

2.11.4 Transient transfection of U937 cells

U937 cells were electroporated following basically the protocol described for the transient transfection of FDCPmix cells. In the case of U937, however, 1×10^7 cells were used for each sample and therefore the medium volume for the following incubation was increased to 8 ml and was performed in tissue culture flasks. Luciferase-assays were performed after 36 hours. A β-Gal assay was used for normalization (see 2.11.2).

2.11.5 Stable transfection of U937 cells

The protocol closely resembles the one outlined for FDCPmix. Like in the transient assays, however, 1×10^7 U937 cells were used for electroporation. The following selection procedure was carried out exactly as described for FDCPmix cells.

2.11.6 Transfection of 293T cells

Calcium phosphate transfection

293T and HepG2 cells were transfected using the classical calcium phosphate transfection method. In brief, one day before transfection cells were seeded in a 10 cm plate, in 10 ml growth medium, at 20% confluency in order to be 30–50% confluent at the moment of transfection. 16 to 24 hours later, a total amount of 20µg DNA at a concentration of 1µg / µl in 0.1 x TE was mixed with 0.5 ml of a 0.25 M CaCl₂ solution followed by addition of 0.5 ml 2xBES–buffered–saline (BBS) solution, mixed and incubated for 10–20 min / RT. Before using them, the two solutions were also pre–equilibrated to a temperature of 20–25°C. The DNA–CaCl₂–BBS mixture was then added dropwise to the cells and the dishes were gently swirled to ensure equal distribution of the solution. After incubation of the cells at 3.4% CO₂ for 18–24 hours at 37°C, the medium was replaced followed by normal growth conditions at 5% CO₂ for another 24 hours. After about 48 hours after transfection cells were rinsed with PBS and harvested after incubation for a minimum of 30 min in 3 ml of cold PBS / 0.5 mM EDTA solution.

2x BBS: 1.07g BES (N,N-bis[2-hydroxyethyl]-2-aminoethanesulfonic acid)
 1.6g NaCl
 0.027g Na₂HPO₄.
 adjust to pH 6.96 with NaOH
 H₂O ad 100ml
 pass through a 0.22µm filter and store aliquots at –20°C.

PolyFect transfection

293T cells were transfected using PolyFect reagent (Qiagen, cat. no. 301105) according to the manufacturers protocol.

2.12 Preparation of protein extracts from mammalian cells

2.12.1 Whole cell lysates from U937 cells

About 5×10^6 U937 cells were pelleted and washed with cold PBS. The cell pellet was resuspended in 50 μ l of extraction buffer, snap-frozen in liquid nitrogen and thawed at 37°C. This procedure was repeated twice before centrifugation at 13.000 rpm in a table top centrifuge for 15 min at 4°C. The supernatant was collected and transferred to a new tube on ice and protein concentrations were determined as follows: 5 μ l of lysate were added to 500 μ l 0.1% SDS. OD was measured at 230 and 260 nm. Protein concentrations were calculated according to the following equation: concentration μ g / ml = $100 \times (183 \times A_{230} - 75.8 \times A_{260})$

Extraction Buffer	20 mM Hepes pH 8.0	4 ml (0.5 M Hepes pH 8.0)
	350 mM NaCl	35 ml (1 M NaCl)
	10% Glycerol	10 ml (100% Glycerol)
	0.1% Tween-20	1 ml (10% T-20)
	2mM EDTA	400 ml (0.5 M EDTA)
		49.5 ml H ₂ O
	sterile filter and add before use:	
	1 mM DTT	
	1 mM PMSF	
	protease inhibitors (Sigma P8340, 10 μ l / ml)	

2.12.2 Isolation of nuclear extract from 293T

48 hours after transfection cells were washed with cold 1x PBS and harvested using a rubber policeman. Cell pellets were then resuspended in cold NEX A buffer and left swelling on ice for 10 minutes. After addition of 0.2% NP40 cytoplasmic membranes were broken mechanically using a douncer and cytoplasm was removed after centrifugation at 1.000 g for 10 min (4°C). Pellets were resuspended in NEX B buffer to a final salt concentration of 0.27 M and rocked at 4°C for a minimum time of 30 min. Finally, samples were centrifuged at 16.000 g and the supernatants (nuclear extracts) were transferred to new tubes and aliquots were frozen in liquid nitrogen.

NEX A:	10mM HEPES pH 7.9 10mM KCl 0.1mM EDTA pH 8.0 0.1mM EGTA pH 8.0 Immediately before use add 1.0mM DTT, 0.5mM PMSF, 2µg/µl Aprotinin, 2µg/µl Leupeptin.
NEX B:	20mM HEPES pH 7.9 0.4M NaCl 1mM EDTA pH 8.0 1mM EGTA pH 8.0 10% Glycerin Immediately before use add 1mM DTT, 1mM PMSF, 2µg/µl Aprotinin, 2µg / µl Leupeptin.

2.12.3 Determination of luciferase reporter gene activity

About 5×10^6 cells were used for the measurement of luciferase activity. Cells were harvested by centrifugation at 200 g for 6 minutes. The cell pellets were kept on ice, resuspended in 6 ml of cold PBS and centrifuged again. Supernatants were aspirated quantitatively and after addition of 100 µl of 1 x Passive Lysis Buffer (Promega) pellets were resuspended by vortexing for 10 sec. The samples were transferred to Eppendorf tubes and incubated at RT for 10 min. Lysates were centrifuged in a table top centrifuge at full speed for 5 min at 4°C. The supernatants were then transferred to new Eppendorf tubes. 50 µl of the lysate were used for measurement of luciferase activity which was performed in a microplate using the Luciferase Assay System from Promega (cat. no. E1501). Measuring time was 10 sec for each well.

In the case of stable cell lines the luciferase values were normalized by calculating the ratio between enzymatic activity and total protein content. Protein amounts were determined doing a standard Bradford assay (Bio-Rad, Protein Assay, cat. no. 500–0006). A minimum of two stable cell lines derived from independent transfection experiments was used to determine the activity of a given reporter construct. Mean induction levels were calculated between four independent luciferase measurements of two different cell lines.

2.13 Protein analysis

2.13.1 Recombinant protein expression and purification

Recombinant proteins were expressed in *E.coli*, using expression plasmids encoding the protein of interest in the BL21 *E.coli* expression strain. An overnight starter culture was diluted to an OD600 of 0.1 into 200 ml LB medium and grown at 37°C to an OD600 of 0.6 - 0.8. At that point the expression of the protein was induced by addition of 0.5 mM IPTG final concentration. To prevent the formation of inclusion bodies and to enhance the expression of full-length protein, the culturing temperature was reduced to 30°C for 2 - 6 hours. The cells were harvested by centrifugation for 15 minutes at 3.500 rpm (4°C). Subsequent purification steps are carried out at 4°C.

2.13.2 Purification of GST-tagged proteins

Bacterial cell pellets were resuspended in 10 ml of lysis buffer and lysed by incubation with 10 mg of lysozyme for 10 minutes and sonification. For sonification the microtip and an output amplitude of 30% for a total time of 2 minutes with a repetitive cycle of 10 seconds on-time and 50 seconds off-time was used. During sonification the samples were cooled in an ice-water bath. The lysate was cleared by centrifugation for 10 minutes at 10.000 g (4°C). In the meantime, 200 µl of Glutathione-Sepharose 4B (Amersham, cat. no. 17-0756-01) were washed and equilibrated in lysis-buffer. The lysate was incubated together with the beads for 90 minutes at 4°C on a rotating wheel to allow binding of the recombinant protein to the matrix. The supernatant was removed and the remaining beads were subsequently washed with 100 column volumes of BC2000 and BC150 buffer. To elute the immobilized fusion proteins, the columns were incubated at 4°C in 500 µl of elution buffer for 10 minutes. For quality control, an aliquot was analyzed in a Coomassie stained SDS-PAGE gel.

Lysis buffer: 20 mM HEPES pH 7.5
 100 mM KCl
 1 mM EDTA
 10% Glycerol
 0.1% NP-40
 5 mM β -Mercaptoethanol
 add protease inhibitors before use

Elution buffer: 25 mM Tris-HCl pH 8.2 RT
 100 mM KCl
 10% Glycerol
 0.1% NP-40
 30 mM reduced Glutathione
 add protease inhibitors before use

2.13.3 Preparation of antigens for immunization of rodents

Recombinant proteins that were used for generation of antibodies were mixed with adjuvants before injection. For this purpose 50 μ g of GST–protein was mixed with 10 μ l of CPG2000 (Sigma) and 500 μ l of Freund's Incomplete Adjuvant (IFA, Sigma). PBS was added to a total volume of 1 ml. Of this 500 μ l were injected subcutaneously, while 500 μ l were injected intraperitoneally. After 4 weeks the animals were boosted with this antigen preparation to ensure a strong antigenic reaction.

2.13.4 Immunoprecipitation

Immunoprecipitation (IP) of MLL from 293T nuclear extracts was performed using monoclonal antibodies against MLL immobilized on Protein G–Sepharose. Columns were prepared by incubation of 500 μ l of Protein G–Sepharose with 50 ml of monoclonal antibody supernatant. The slurry was rocked overnight at RT, washed twice with PBS and stored in PBS / 0.01% NaN_3 . 20 μ l of beads were incubated with 400 μ g of nuclear extracts. The reaction was brought to a total volume of 410 μ l by addition of 350 μ l of BC 150 / 0.1% NP–40. IP was performed at 4°C for four hours. The batch was then divided into four different samples and different washing

buffers (washing buffer 1 – 4) were used. The columns were washed five times with 1 ml of washing buffer. Precipitated protein was eluted from the column using 60 μ l of 2x SDS loading buffer. 24 μ l of this were loaded onto a 5% SDS gel and analyzed via Western Blot.

washing buffer 1:	20 mM Tris-HCl, pH 7.3 0.2 mM EDTA, pH 8.0 500 mM KCl 20 % (v/v) Glycerin 0.1% NP- 40
washing buffer 2:	20 mM Tris-HCl, pH 7.3 0.2 mM EDTA, pH 8.0 150 mM NaCl 20 % (v/v) Glycerin 1% Triton X-100 0.1% SDS
washing buffer 3:	20 mM Tris-HCl, pH 7.3 0.2 mM EDTA, pH 8.0 500 mM NaCl 20 % (v/v) Glycerin 1% Triton X-100 0.1% SDS
washing buffer 4:	20 mM Tris-HCl, pH 7.3 0.2 mM EDTA, pH 8.0 250 mM LiCl 20 % (v/v) Glycerin 1% NP- 40 1% SDS

2.13.5 Sodium-dodecylsulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were separated on an SDS-PAGE using either a maxi-gel system from Hoefer or the minigel system from Bio-Rad. Depending on the size of the protein, either 5, 10, 12 or 15% gels were used. To achieve better resolution and blotting efficiency for large proteins, 170:1 acrylamide / bisacrylamide gels were used. For electrophoresis, proteins were mixed 1:6 with 6x loading buffer, heat denatured at 95°C and loaded onto the gel. Proteins were separated applying a current of 30 mA for minigels and 50 mA for large gels. In both cases TGS was used as electrophoresis buffer. Maxigels were connected to a cooling system. For molecular weight determination, unstained marker was loaded in parallel (Bio-Rad,

SDS-PAGE standards Low Range, cat. no. 161–0304; Bio-Rad, SDS-PAGE standards High Range, cat. no. 161–0303; MBI Fermentas, Protein MW Marker, cat. no. SM0431). Following electrophoresis, proteins were stained with Coomassie Brilliant Blue G250 or transferred to a nitrocellulose membrane by blotting.

6x SDS loading buffer: 0.35 M Tris-HCl (pH 6.8 RT)
 0.12 mg/ml Bromphenol blue
 40 mM β -Mercaptoethanol

2.13.6 Coomassie staining

For Coomassie staining of polyacrylamide gels, the gels were incubated with staining solution for at least 30 minutes on a slowly rocking platform. To visualize protein bands, the gels were destained overnight in destaining solution or in H₂O. After destaining, the gels were scanned and then dried between cellophane film at RT.

Coomassie staining solution:	Methanol	400 ml
	Glacial acetic acid	100 ml
	Coomassie R-250	0.25 g
	H ₂ O	500 ml
Destaining solution:	Methanol	500 ml
	Glacial acetic acid	100 ml
	H ₂ O	400 ml

2.13.7 Western Blot analysis of MLL protein expression in U937 cells

Whole cell lysates were prepared as described earlier (see 2.12.1). Proteins were separated by electrophoresis on a 5% SDS-gel. 20 to 40 μ g of whole cell extract were loaded per lane. Commercial loading buffer was added in this case (NuPage LDS sample buffer 4x, Invitrogen NP0007) and samples were denatured at 70°C for 10 min immediately before loading. After electrophoresis the gel was blotted onto a nitrocellulose membrane using a semi-dry transfer cell (Bio-Rad). The setting was 25 V constant for 1 hour. After blotting the membrane was blocked for a minimum of 1 hour at RT or overnight at 4°C using TBS / 6% milk. For detection of the overexpressed MLL fusion protein MLC5A9 was used as a primary antibody.

Alternatively an antibody against the HA-tag (R001, 3F10) was used as well as an antibody against the ER α ligand binding domain (sc-8002, Santa Cruz Biotechnology). The membrane was incubated with the primary antibody at RT for a minimum of 2 hours. The secondary antibodies used for this work are commercial conjugates of anti-mouse, anti-rat and anti-rabbit immunoglobulines with horseradish peroxidase (Promega). Before addition of the secondary antibody the membrane was washed three times for 5 min with TBS-T. Secondary antibody incubation lasted 45 min and was followed by three washing steps with TBS-T (3 x 15 min at RT). For detection the chemiluminescence kit from Perkin Elmer (Western Lightning, cat. no. NEL105) was used following the manual.

Transfer buffer (pH 8.3):	2.9 g Glycine	39 mM Glycine
	5.8 g Tris base	48 mM Tris base
	1850 μ l 20% SDS	0.037% SDS
	200 ml Methanol	20% Methanol
	H ₂ O ad 1000 ml	

1x TBS-T: For 1L use 100ml 10x TBS, 2ml 10% TWEEN20

2.13.8 Chromatin-Immunoprecipitation (IP)

Crosslinking

U937 cells were induced with 4-hydroxytamoxifen (OHT) at a final concentration of 1 μ M. 1×10^8 – 1×10^9 cells were used while cell density was around 5×10^5 / ml.

After 24 hours cells were harvested by centrifugation in 250 ml buckets (Corning, 300 g, 5 min, RT). Pellets were resuspended in 40 ml of pre-warmed medium containing fetal calf serum (10%) and pooled in a 50 ml Falcon tube. After another centrifugation step the pellet was resuspended in 43.78 ml of serum-free medium (RT). 1.22 ml of formaldehyde (37%, Roth, cat. no. 4979.1) were added and Falcon tubes were incubated on a roller table for 9 min at RT. Glycine was added to a final concentration of 125 mM (3 ml of a 2 M solution) and the samples were mixed quickly and put on ice for 10 min. From that point on all steps of the protocol were carried out on ice. Cells were centrifuged (300 g, 5 min, 4°C) and the pellets were washed thrice in cold PBS and transferred to .15 ml Falcon tubes. At this point the

procedure can be interrupted by shock-freezing the pellet and storage at -80°C . Alternatively one can proceed directly.

Cell lysis and sonication

To isolate nuclei pellets were resuspended in 10 ml of Lysis Buffer 1 (LB1) and rocked gently at 4°C . Nuclei were centrifuged (4.000 g, 10 min, 4°C) and resuspended in 10 ml of Lysis Buffer 2 (LB2). Again the samples were rocked gently at 4°C for 10 min before spinning them down again (4.000 g, 10 min, 4°C). Pellets were resuspended in 5 ml of Lysis Buffer 3 (LB3) without Triton X-100 and glycerol and sonicated in an ice/ethanol bath using a Branson 250 sonifier with the microtip (40% output, 6 min total: 15 sec on / 45 sec off). Resulting fragment lengths were ranging from 300 to 500 bp.

After sonication 1/20th of the total volume of 10% Triton X-100 (0.5% final concentration) was added to the samples, which were then transferred to 1.5 ml tubes. Samples were centrifuged (15.000 g, 15 min, 4°C) and supernatants were transferred to new 15 ml Falcon tubes. Sample volumes were adjusted with LB3 in order to reach a concentration of double-stranded DNA (as obtained through optical measurement) of 1–2 mg/ml. Glycerol was added to a final concentration of 10%. 1 ml aliquots were snap-frozen and stored at -80°C . 10 μl of the sample were digested with 2 μl of Proteinase K (PCR grade, Roche) for 2 h at 65°C . After DNA purification (Qiaquick, Qiagen) fragment lengths were checked on an agarose gel. If necessary, further sonication was carried out.

Lysis Buffers: Complete (Roche, cat. no. 11697498001) protease inhibitor mix was added to all lysis buffers before use.

25x Complete

- Grind 1 tablet in between 2 sheets of balance paper
- Dissolve fine powder in 2 ml of H₂O
- Aliquot and store (labeled 25xC) at -20°C

Lysis Buffer 1 (LB1) final conc.

5,0 ml	1M HEPES-KOH, pH 7.5	50mM
2,8 ml	5M NaCl	140mM
0,2 ml	0,5M EDTA, pH 8.0	1mM
10,0 ml	100% glycerol	10%
5,0 ml	10% NP-40	0,5%
2,5 ml	10% Triton X-100	0,25%
74,5 ml	H ₂ O	

100,0 ml

• store at 4°C

Lysis Buffer 2 (LB2) final conc.

4,0 ml	5M NaCl	200mM
0,2 ml	0,5M EDTA	1mM
0,1 ml	0,5M EGTA	0,5mM
1,0 ml	1M Tris-HCl pH 8	10mM
94,7 ml	H ₂ O	

100 ml

store at 4°C

2x Lysis Buffer 3 (LB3) final conc. (1x)

2,8 ml	5M NaCl	140mM
0,4 ml	0,5M EDTA	1mM
0,2 ml	0,5M EGTA	0,5mM
2,0 ml	1M Tris-HCl pH 8	10mM
10,0 ml	10% N-lauroyl sarcosine	0,5%
84,6 ml	H ₂ O	

100,0 ml

• store at 4°C

• before use, dilute to 1x LB3 and supplement with:

before sonication:

10% Na-deoxycholate 0,1% (freshly prepared) 1x Complete

after sonication:

10% Triton X-100 (at step 2.5) 0,5%

100% glycerol (at step 2.6) 10%

Immunoprecipitation and isolation of DNA fragments

For one IP 100 μ l of chromatin extract (2 mg/ml) were used. For preclearing 10 μ l of blocked beads were used per 100 μ l of extract. Samples were incubated under constant agitation at 4°C for 30–60 min. In parallel blocked beads were pre-incubated with the appropriate antibody. Typically 2–10 μ g of antibody were used per IP. In the case of hybridoma supernatants 100–200 μ l were used. For each IP 15–20 μ l of beads were used. The pre-incubation was done in a total volume of 400 μ l and volumes were adjusted with cold PBS. After incubation at 4°C for 30–60 min tubes were centrifuged (200 g, 2 min, 4°C) and washed once with cold PBS.

100 μ l of pre-cleared chromatin extract was added to 20 μ l of antibody-coated beads. The total volume was adjusted to 400 μ l using LB3 without glycerol. IP reactions were incubated under constant agitation at 4°C overnight.

Beads were washed six times with 1 ml of RIPA buffer and once with 1 ml of TE buffer containing 50 mM NaCl. For elution 100 μ l of pre-warmed (65°C) elution buffer has been added to the beads and samples were rocked for 10 min at 1400 rpm in a thermomixer (65°C). After centrifugation supernatants were transferred to new tubes (RT). Crosslink reversal took place overnight at 65°C.

1 volume of TE buffer containing RNase (0.2 μ g/ μ l final concentration) was added to the samples followed by incubation at 37°C for 1–2 hours. Proteinase K was added to a final concentration of 0.2 μ g/ μ l and samples were incubated at 56°C for 2 hours. DNA was isolated by PCI extraction and ethanol precipitation in the presence of 10 μ g of glycogen. Pellets were resuspended in 50 μ l Tris/HCl. DNA concentrations were determined and 0.5–1 μ g were loaded onto an agarose gel to check for fragment length distribution. 20 ng of this DNA were used as a sample for PCR reactions. For detection of MLL–VP16–ER–HA bound to the chromosomal Hoxa9 promoter primers "hoxa9end2" and "hoxaend_r" were used for PCR, for the episomal Hoxa9 promoter the primers "hoxa9F–65" and "lucR_short".

Wash buffer (RIPA buffer):		final concentration
5ml	1M Hepes (pH 7.6)	50 mM
200µl	1M EDTA	1 mM
10 ml	10% NP-40 (IPGEL)	1%
10 ml	5M LiC	0.5 M
7 ml	10% DOC (Na deoxycholate)	0.7% (freshly prepared)

100 ml		

Elution buffer:	50mM	Tris pH8
	10mM	EDTA
	1%	SDS

2.14 RNA expression analysis

2.14.1 cDNA microarray analysis

FDCP mix cells expressing MLL–VP16–ER–HA were induced with 1 µM 4–hydro–tamoxifen (OHT) and samples of 3×10^7 cells were harvested after 30 min, 3 hours and 12 hours. After washing with PBS RNA was extracted using the Qiagen RNeasy Midikit (Qiagen). The RNase–Free DNase set (Qiagen, cat. no. 79254) was used to eliminate contaminating genomic DNA. RNA concentrations were determined using a spectrophotometer and 5 µg of each RNA sample were sent to the KFB Regensburg, where the microarray analyses were performed. Affymetrix GeneChip 430 2.0 was used, representing more than 39.000 transcripts from the mouse genome thereby covering the complete murine genome. Statistical analysis of the raw data was done by the KFB Regensburg, which yielded lists of genes that were deregulated in comparison to the reference (MLL–VP16–ER–HA cell line before OHT–induction). These target gene lists were then sorted to eliminate those genes with a deregulation that was not statistically relevant. To identify genes with robust expression changes the following procedure was applied to the original target gene lists:

To determine robust increases:

- "absent" calls were eliminated
- "present" calls for increased genes were sorted according to Signal Log Ratio
- probe sets with a Signal Log Ratio < 1.0 were eliminated

To determine robust decreases:

- "absent" calls were eliminated
- "present" calls for decreased genes were sorted according to Signal Log Ratio
- probe sets with a Signal Log Ratio > -1 were eliminated

2.14.2 RT-PCR

Preparation of whole cell RNA

RNA was isolated from 2×10^6 cells (U937, FDCC mix) using the Qiagen RNeasy Minikit (Qiagen, cat. no. 74104). The procedure was carried out as described in the kit protocol. RNA was eluted from the column using 40 μ l of RNase free water. Total yields were usually around 20 μ g of RNA as determined by UV measurement.

DNase treatment of RNA

Isolated RNA was treated with DNase I (Invitrogen, cat. no. 18068-015) in order to eliminate genomic DNA contamination. The reaction volume was 20 μ l. 16 μ l of RNA were incubated at RT for 15 min together with 2 μ l DNase I (1 u / μ l) and 2 μ l of 10 x buffer. The digest was stopped by addition of 2 μ l of EDTA (25 mM) and heat inactivation (10 min / 65°C).

Reverse Transcription

2 μ g of DNase treated RNA were used for reverse transcription (RT) using the ThermoScript™ RT-PCR System (Invitrogen, cat. no. 11146-024). The reaction was carried out following the manual. The oligo dT primer provided in the kit was used for first strand synthesis which was taking place at 50°C for 60 min followed by a denaturation step (5 min, 85°C). The resulting cDNA was incubated with 1 μ l of RNase H in order to degrade complementary RNA molecules (20 min, 37°C).

cDNA was stored at -20°C.

Quantitative Real-Time PCR

Transcript amounts were determined using quantitative Real-Time PCR (qPCR). The reaction was performed using the SYBR Green PCR reagents (Applied Biosystems, cat. no. 4306736) and gene specific primers. The reaction was set up as follows:

2.5	μl	primer mix (5 pmol / μl each)
2.5	μl	SYBR Green buffer
2.5	μl	MgCl ₂ (25 mM MgCl ₂)
2.0	μl	dNTP mix
0.2	μl	AmpliTaq Gold DNA Polymerase (5u / μl)
0.2	μl	AmpErase UNG (1u / μl)
13.1	μl	H ₂ O
2.0	μl	cDNA

Primer pairs were chosen with a T_m of 58°C. 40 cycles of PCR were performed in a Perkin Elmer GeneAmp 5700 instrument and results were normalized to beta-actin RNA levels. As a negative control for qPCR RT reactions done without reverse transcriptase were used as template.

2.15 Bioinformatic analysis

Analyses of promoter sequences for transcription factor binding sites were performed using the programs from Genomatix (www.genomatix.de)

2.15.1 Identification of transcription factor binding sites using MatInspector

The program MatInspector (Quandt et al. 1995; Cartharius et al. 2005) was used for identification of binding sites in promoters and promoter fragments. The analyses presented in this work were performed setting the core similarity to 0.75 and to "optimized" matrix similarity.

2.15.2 Gene2Promoter

Promoters of genes that were identified in the microarray analysis were searched for common transcription factor binding sites using the program Gene2Promoter. 50 genes of each dataset showing the strongest deregulation in expression were used as an input. Transcription factor families are shown that are present in >86% of all input gene promoters.

2.15.3 Bibliosphere

Array data were analyzed using Bibliosphere in order to create a hypothetical network amongst the target genes based on co-citation in literature databases (Scherf et al. 2005). Networks were generated using microarray target genes and transcription factors that were co-cited together with the corresponding gene at the "sentence level", i.e. genes that were found cited in the same sentence in literature databases.

3. Results

3.1 Generation and selection of monoclonal antibodies against MLL

At the time of the beginning of this study MLL-specific antibodies were not commercially available. For protein expression analysis and for ChIP experiments, however, antibodies are precious molecular tools. Therefore a collaboration was initiated with E. Kremmer (GSF, Munich) to generate monoclonal antibodies against MLL. GST-tagged portions of human MLL were injected into rats. The antigens were chosen according to the following criteria:

- a. different regions of the large MLL protein should be targeted therefore one rather N-terminal antigen was chosen and another one that is closer to the C-terminus.
- b. cross-reactivity between human and murine MLL was a desired feature for the monoclonal antibodies. Since there is a high level of homology between the two proteins this seemed feasible and conserved regions were chosen accordingly.

Different fragments of the human MLL cDNA were amplified via PCR and cloned into GST expression vectors. Out of five cloned regions only two could be expressed at high levels in *E. coli* BL21. These two proteins were purified on Glutathion-Sepharose columns and injected into rats.

Both of the antigens show rather high sequence conservation from mouse to human (94% and 98% respectively). One is located in a region of the protein that has been associated with a subnuclear distribution function (C1, amino acids 612 – 935) while the other one is part of the C-terminal SET domain (D1, amino acids 3620 – 3969; Fig. 5). The rat spleens were isolated and B lymphocytes were used for generation of hybridoma cell lines that were singled out in order to obtain monoclonal cell lines. The supernatants of these cell lines were analysed in an ELISA assay and positive clones were tested further for performance in Western Blot analysis and Immunoprecipitation (IP). Whole cell lysates were prepared from 293T cells overexpressing a C-terminally truncated version of MLL (MLL Δ C). Lysates of non-transfected Jurkat cells were used to investigate the function of the MLL antibodies directed against the C-terminus of the protein. Three monoclonals could be identified that precipitate the overexpressed MLL aminotermus (5A9, 4G10, 5C10; Fig. 5). Screening a number of antibodies targeting the carboxyterminus it could be shown

that two monoclonals (7D4, 7E12) were suitable for IP of the C-terminal moiety (p180) of the processed endogenous MLL (Fig. 5).

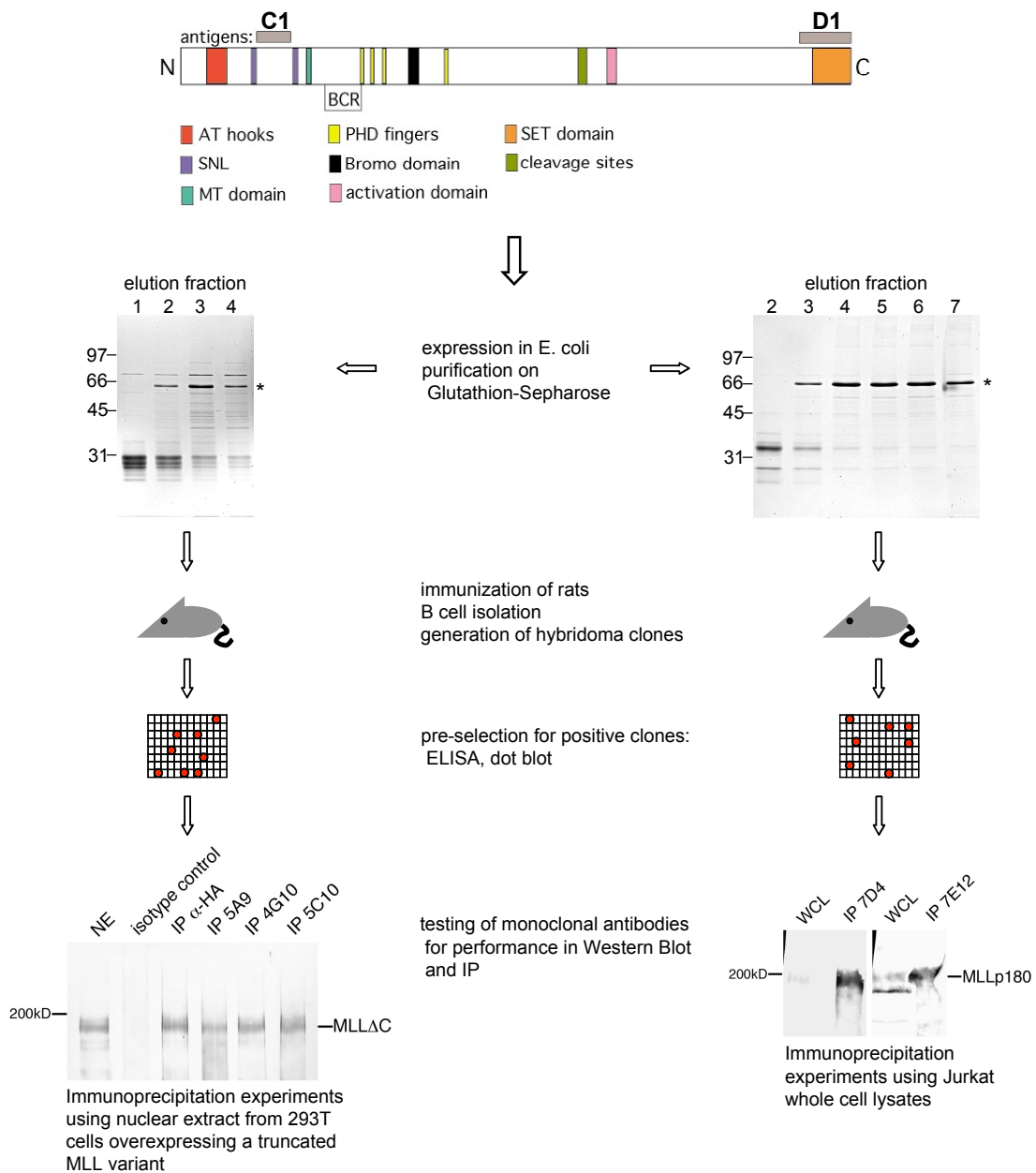


Figure 5. Generation of monoclonal antibodies against MLL. Five monoclonal antibodies were selected: three clones specific for the MLL N-terminus (5A9, 4G10, 5C10), two clones targeting MLL C-terminus (7D4, 7E12). SNL: subnuclear localization domains, MT domain: methyltransferase homology domain.

These initial IP experiments were performed under low stringency conditions (150 mM KCl, 0.1% NP-40). For many biochemical experiments a higher stringency is desirable since this leads to higher specificity of the antibody binding reaction and to increased purity of the bound complexes. The performance of the monoclonal antibodies directed against the MLL aminotermminus was tested using washing buffers that were either containing higher concentrations of salt or detergent or both. Increasing the concentration of KCl to 500 mM led to disruption of the complex formed by MLL Δ C and any of the three antibodies (Fig. 6, lane1). A similar tendency could be seen using NaCl in combination with two detergents: at a concentration of 150 mM interaction was robust, whereas 500 mM NaCl dramatically reduced the IP efficiency (Fig. 6, lanes2 and 3). LiCl at a concentration of 250 mM, however, was tolerated (Fig. 6, lane4). Interestingly, even relatively high concentrations of detergent did not compromise complex stability. MLL Δ C could be detected after washing with 1% Triton X-100 / 0.1% SDS as well as after washing with 1% NP - 40 / 1% SDS.

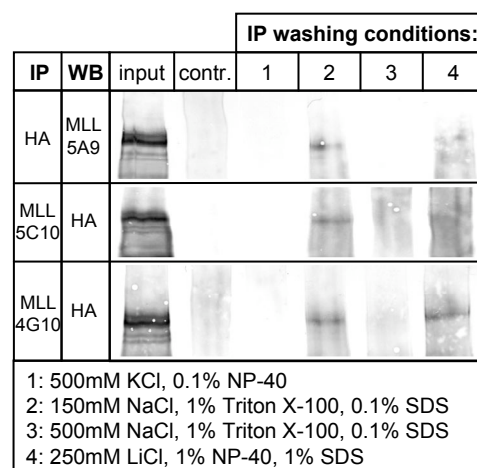


Figure 6. MLL-Immunoprecipitation experiments. After washing of IP samples (HA, MLL 5C10, MLL 4G10) with different buffers (lanes 1-4) a truncated MLL fragment overexpressed in 293T cells was detected using antibodies against MLL (5A9) or HA.

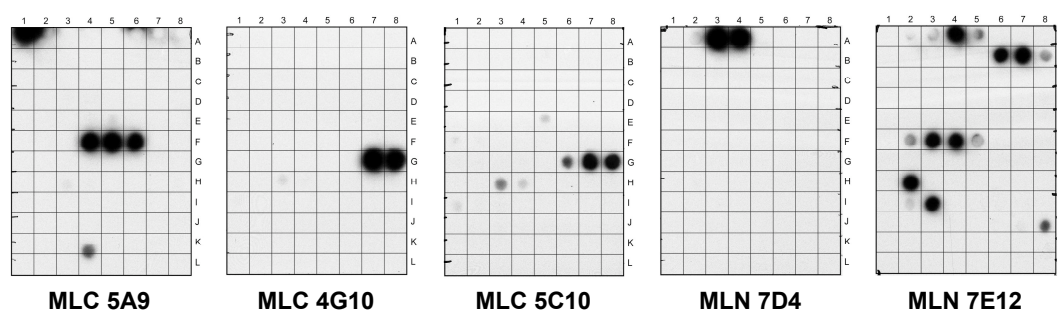


Figure 7. Epitope mapping for monoclonal antibodies against MLL. Two different types of peptide arrays were used comprising the peptides for either the N-terminal (MLC) or the C-terminal (MLN) antigen. Western Blots using monoclonal antibodies against MLL led to identification of the corresponding epitopes.

Table 7. Epitopes of the monoclonal antibodies against MLL.

clone	antigen	epitope	aa sequence
5A9	aa 612-935	aa784-806	VSPLATSALNPTFTFSPSHSLTQS
4G10	aa 612-935	aa828-846	SSSSPTPLFPWFTPGSQTE
5C10	aa 612-935	aa824-846	AEPFSSSSPTPLFPWFTPGSQTE
7D4	aa 3620-3969	aa3628-3646	ESAEPKTVEEEEENFSSPL
7E12	aa 3620-3969	aa3628-3650	ESAEPKTVEEEEENFSSPLMLWL
		aa3672-3694	EISSDDGFQICAESIEDAWKSLT
		aa3784-3810	KHRQPPEYNPNDEEEEEEVQLKSARRAT
		aa3848-3862	NIDAGEMVIEYAGNV
		aa3884-3898	MFRIDDSEVVDATMH

Stringency of IP and washing buffers is one way to increase purity of the isolated protein complex. Another method is the specific elution of precipitated complexes from the column. If the epitope of an antibody is known a peptide comprising this amino acid sequence can be used for elution. The peptide is offered in large excess and competes with the target protein for binding sites on the antibody molecule. This technique can be applied to standard Chromatin-IP protocols in order to increase specificity. Determination of the epitopes was performed using peptide arrays. For this purpose each of the two antigens was covered by 20mers that are overlapping by 15 amino acids each. These peptides were immobilized on a membrane, which was then subjected to standard Western Blot procedures for each of the monoclonal antibodies (Fig. 7). Monoclonals 4G10 and 5C10 turned out to have at least partially if not identical epitopes (Table 7). Also 5A9 and 7D4 showed a clearly defined target region. The epitope of the antibody 7E12, however, comprises other peptides in addition to the epitope of 7D4 and seems to be discontinuous (Fig. 7).

3.2 An inducible MLL fusion model system

The MLL gene is known to be involved in translocations with a host of fusion partner genes, many of which contain an activation domain (Daser and Rabbits 2005). This observation gave rise to speculations that fusion of the aminoterminal MLL domains to an activation domain only might be sufficient for tumorigenesis. It could be shown that a synthetic fusion construct consisting of the activation domain of Herpes simplex virus protein 16 (VP16) and the MLL N-terminus has transforming capacity in myeloid progenitor cells (So and Cleary 2003; Zeisig et al. 2003).

To investigate the general mechanism through which MLL-activator fusions influence transcriptional activation a model system was established using an inducible variant of MLL-VP16. The mutant ligand binding domain (ER^{T2}) of the human estrogen receptor (Feil et al. 1997) was fused to MLL-VP16 (Fig. 8). The resulting protein (MLL-VP16-ER-HA) is kept inactive through the ER-moiety. It is generally believed that cytoplasmic sequestration of the estrogen receptor fused proteins of interest by the Hsp90 complex inhibits the biological functions of nuclear proteins like MLL (Mattioni et al. 1994). There is some evidence, however, that there is an additional mechanism of inhibition, which presumably involves sterical hindrance of functional domains of the protein of interest (unpublished results).

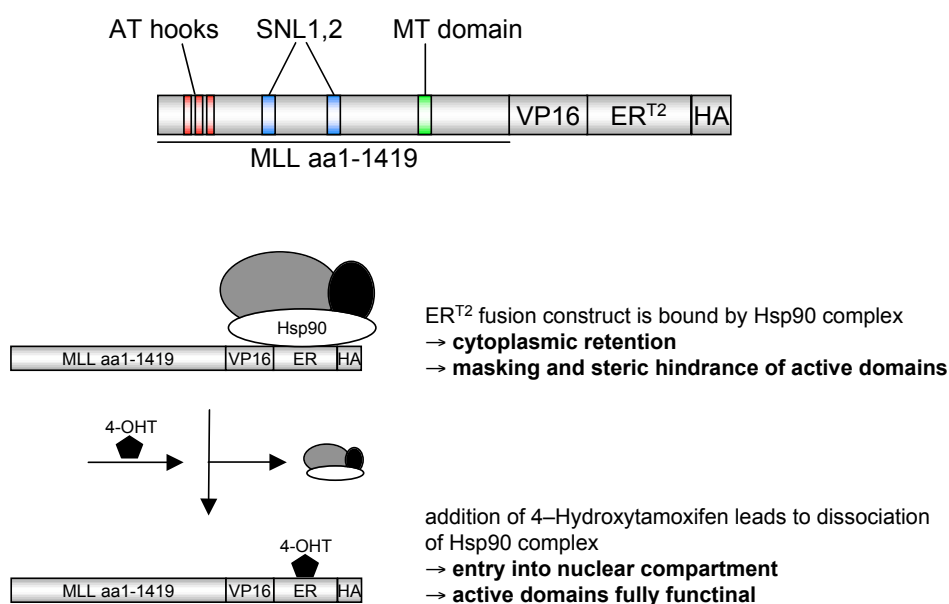


Figure 8. Presumptive mechanism of activation of MLL-VP16-ER-HA. The fusion of ER to the MLL N-terminus renders it a binding partner for the Hsp90 complex, which prevents it from migrating into the nucleus in absence of Tamoxifen.

3.2.1 Generations of cell lines stably expressing MLL–VP16–ER–HA or VP16–ER–HA

The human promonocytic cell line U937 (Sundstrom and Nilsson 1976) was chosen to create stable cell lines expressing MLL–VP16–ER–HA. U937 cells have been reported to retain a differentiation potential for final maturation into macrophages (Koren et al. 1979) suggesting that important myeloid developmental pathways are still functional in this cell line. This together with convenient culturing and transfection conditions made it an appropriate cell line for the inducible MLL–VP16 model system.

Cells were transfected with linearized plasmid DNA coding for MLL–VP16–ER–HA and stable clones were selected expressing the fusion protein (Fig. 9A). In parallel a control cell line was established that expresses VP16–ER–HA (Fig. 9B). This construct was designed to rule out that effects on gene transcription observed as a consequence of MLL–VP16–ER–HA are due to unspecific binding of the fusion protein mediated through the ER–moiety.

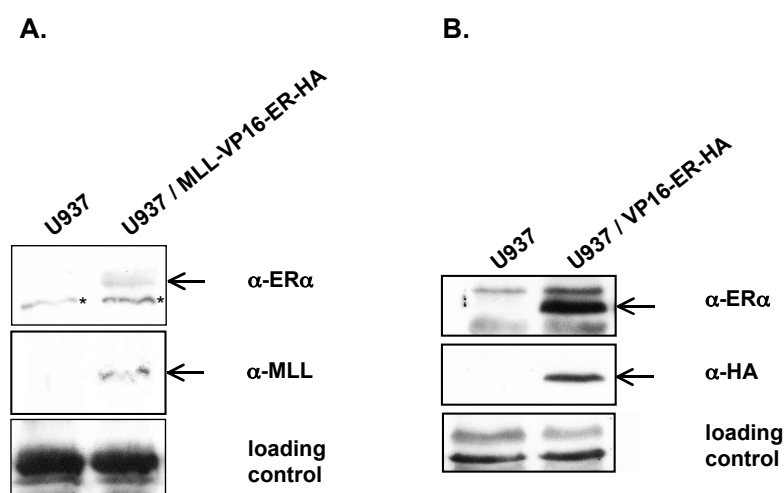


Figure 9. Expression of MLL–VP16–ER–HA and ER–VP–16–HA in U937. Whole cell lysates from stable cell lines expressing MLL–VP16–ER–HA (A) or ER–VP–16–HA (B), respectively, were analyzed in a Western Blot experiment using antibodies directed against the human estrogen receptor α , MLL and HA. Correct loading of SDS gels was controlled using PonceauS protein stain.

3.2.2 Transcription of human Hoxa9 is upregulated following induction of MLL–VP16–ER–HA

Having established transgenic cell lines it was important to check whether the induction of MLL–VP16–ER–HA mimicks the effect of MLL fusion proteins on target genes. Homeotic genes and in particular Hoxa9 have been reported to be targets for MLL and Hox gene overexpression is a hallmark of MLL leukemogenesis (Yu et al. 1995; Armstrong et al. 2002). Therefore Hoxa9 expression was chosen as a biological read–out for the biological activity of the transgene. RT–PCR analysis showed a 6–fold increase of Hoxa9 mRNA following addition of Tamoxifen to the MLL–VP16–ER–HA cell line (Fig. 10A, hoxa9 endogenous) indicating that the fusion protein is functional both in recruitment to target promoters as well as in gene activation.

One objective of the generation of this cell line was the analysis of the Hoxa9 promoter. For the following experiments an episomal reporter system was established that allows for comparison of the effect of MLL–VP16–ER–HA on different Hoxa9 promoter variants. For this purpose Hoxa9 promoter–luciferase constructs have been cloned into an Epstein–Barr–Virus (EBV)–based vector. It has been shown that these vectors can be propagated stably as episomal plasmids in cell lines (Mackey and Sugden 1999; Sugden and Leight 2001). EBNA-1, which is expressed from these vectors recruits the eukaryotic DNA replication machinery to the oriP region on the same vector thereby ensuring the propagation of this episome. For this work an EBV–based vector was used that could be selected for through the presence of a hygromycin resistance cassette (Fig. 10B). In contrast to transient reporter assays where several hundred copies of the reporter plasmid can be present in one cell EBV–based reporters are maintained at a relatively low copy number of 5–20. Since nucleosome formation is taking place on these plasmids they resemble cellular chromosomes much closer making them a suitable model system for analysis of chromatin–related events.

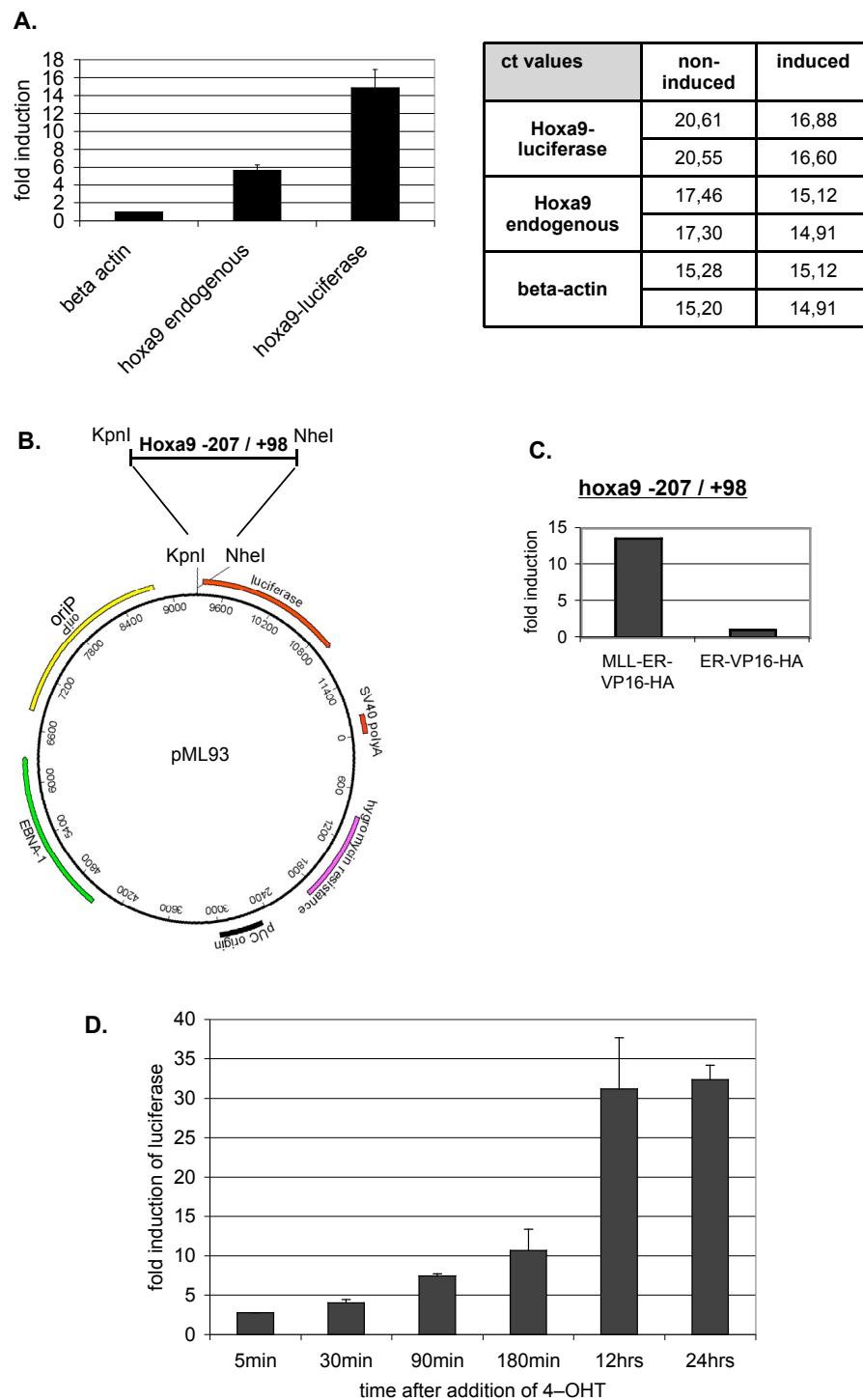


Figure 10. Hoxa9 upregulation by MLL–VP16–ER–HA in U937 cells. A. Upregulation of endogenous Hoxa9 mRNA as well as luciferase mRNA as determined by RT–PCR. Values were normalized for beta-actin mRNA levels. Absolute ct values are shown in the table. B. Schematic drawing of an episomal reporter construct. The regions oriP (yellow) and EBNA–1 (green) are required for host–cell dependent replication of the plasmid. The luciferase gene and the corresponding polyA sequence is depicted in red. KpnI and NheI sites were used for insertion of target promoter sequences. Not to scale. C. Induction rates of luciferase activity are shown for stable U937 cell lines. D. Luciferase induction levels at different time points after addition of Tamoxifen.

Addition of Tamoxifen led to robust induction of an episomal Hoxa9–luciferase reporter in MLL–VP16–ER–HA cells but not in a control cell line expressing VP16–ER–HA (Fig. 10C). This indicates that MLL–VP16–ER–HA is specifically targeted to the episomal Hoxa9 promoter activating reporter gene transcription while VP16–ER–HA has virtually no effect on this process. The elevated luciferase activity was reflected in an increase of luciferase mRNA–levels (Fig. 10A). The stronger increase of luciferase mRNA–levels as compared to Hoxa9 mRNA might be due to negative regulation of MLL–VP16–ER–HA activated transcription, which is taking place only at the endogenous locus, but not on the episome. One possible explanation could be heterochromatin formation taking place on the chromosome but not on the episome. This would render the reporter more accessible to the transcription machinery. Absolute Hoxa9 mRNA–levels before induction of MLL–VP16–ER–HA, however, seem to be higher than luciferase mRNA–levels (Fig. 10A, table ct values). These values rather argue for solid transcription of the Hoxa9 gene in U937 cells even without MLL–VP16–ER–HA activity. Therefore another explanation for the higher induction levels of the episomal reporter gene seems more likely: the short promoter fragment present in the reporter plasmid might simply lack inhibitory elements, which might negatively influence MLL–VP16–ER–HA activated transcription of the endogenous Hoxa9 gene.

One reason to choose the ER system for the inducible cell lines had been reports about the fast induction kinetics of ER–fused proteins (Mattioni et al. 1994). To estimate the time–spans needed for the induction of MLL–VP16–ER–HA a time–course experiment was performed where luciferase samples were measured after different induction periods. In fact already after 90 minutes a significant increase in luciferase activity could be measured (Fig. 10D) hinting at transcriptional processes that must start even earlier. Maximum levels of reporter gene activity, however, could be measured only after 12 hours and 24 hours, respectively.

3.2.3 MLL–VP16–ER–HA is targeted to the Hoxa9 promoter upon induction

Since an increase of transcriptional activity following MLL–VP16–ER–HA induction was observed the next question to be addressed was whether this is a direct effect mediated through binding of the fusion protein to the Hoxa9 promoter. Therefore ChIP experiments were performed analysing endogenous and episomal Hoxa9 promoter occupancy. U937 cells stably expressing MLL–VP16–ER–HA and stably propagating a Hoxa9 (–207 / +46)–luciferase reporter plasmid were induced with Tamoxifen and proteins were cross–linked to DNA after 24 hours by addition of formaldehyde. A mixture of three monoclonal antibodies against the MLL N–terminus (5A9, 5C10, 4G10; see 3.1) was used to precipitate MLL–VP16–ER–HA. PCR reactions were performed using primers specific for the chromosomal Hoxa9 promoter region and another primer pair specific for the episomal Hoxa9 sequence. In both cases an increased signal was obtained after induction of MLL–VP16–ER–HA indicating its binding not only to the chromosomal gene promoter but also to the promoter present in the episomal reporter construct (Fig. 11). This argues that the upregulation of transcription from these promoters is a direct effect of MLL–VP16–ER–HA. The observed promoter occupancy 24 hours after induction taken together with elevated reporter gene levels already 90 minutes after addition of Tamoxifen argues for fast induction kinetics leading to early recruitment of MLL–VP16–ER–HA to the promoter where the protein is still present after 24 hours.

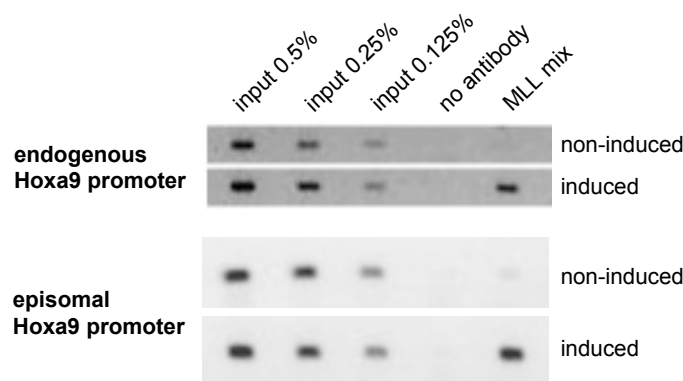


Figure 11. Induced binding of MLL–VP16–ER–HA to Hoxa9 promoter sequences. ChIPs were performed in a cell line expressing MLL–VP16–ER–HA and a Hoxa9 (–207 / +46)–luciferase reporter plasmid. Following IP with a mixture of antibodies specific for MLL (MLL mix) PCR reactions were performed with primers specific either for the endogenous (top panel) or the episomal Hoxa9 promoter (lower panel). Different amounts of IP input material were used to control the PCR reaction. IP reactions performed with sepharose beads only were used as a specificity control for the IP of MLL.

3.3 Identification of a promoter region critical for regulation of Hoxa9 by MLL-VP16-ER-HA

To define more precisely the region of the Hoxa9 promoter through which MLL-VP16-ER-HA is acting a deletion analysis of the human Hoxa9 promoter was carried out. Different promoter variants were generated via PCR and cloned upstream of a luciferase gene. U937 cells stable for MLL-VP16-ER-HA or for VP16-ER-HA, respectively, were transfected with five different variants of the human Hoxa9 promoter and hygromycin resistant cell lines were selected. No significant change in luciferase expression was observed after induction of VP16-ER-HA (data not shown). Strong increase of luciferase activity was seen, however, with all five Hoxa9 constructs after MLL-VP16-ER-HA induction (Fig. 12A). Even though the larger promoter segments had the strongest effect the major capacity for activation through MLL-VP16-ER-HA seems to reside in the region -207 / +46, which in this experiment led to almost 15-fold increase in luciferase activity. Extending this region further downstream (-207 / +98) did not augment the activation.

Therefore the -207 / +46 region was chosen for further investigation and shorter constructs were designed for fine-mapping of the MLL-responsive region. Shortening the region to -118 / +46 led to a reduction of luciferase activity of about 34 percent in comparison with the original construct (Fig. 12B). Hoxa9 -78 / +46, however, showed a dramatic loss not of basal activity (Fig. 12C) but of induction by MLL-VP16-ER-HA: the induction rate dropped to 11 percent in comparison to Hoxa9 -118 / +46, showing that the region from -118 to -79 (Fig. 12D) is critical for regulation of MLL-VP16-ER-HA induced transcription of Hoxa9.

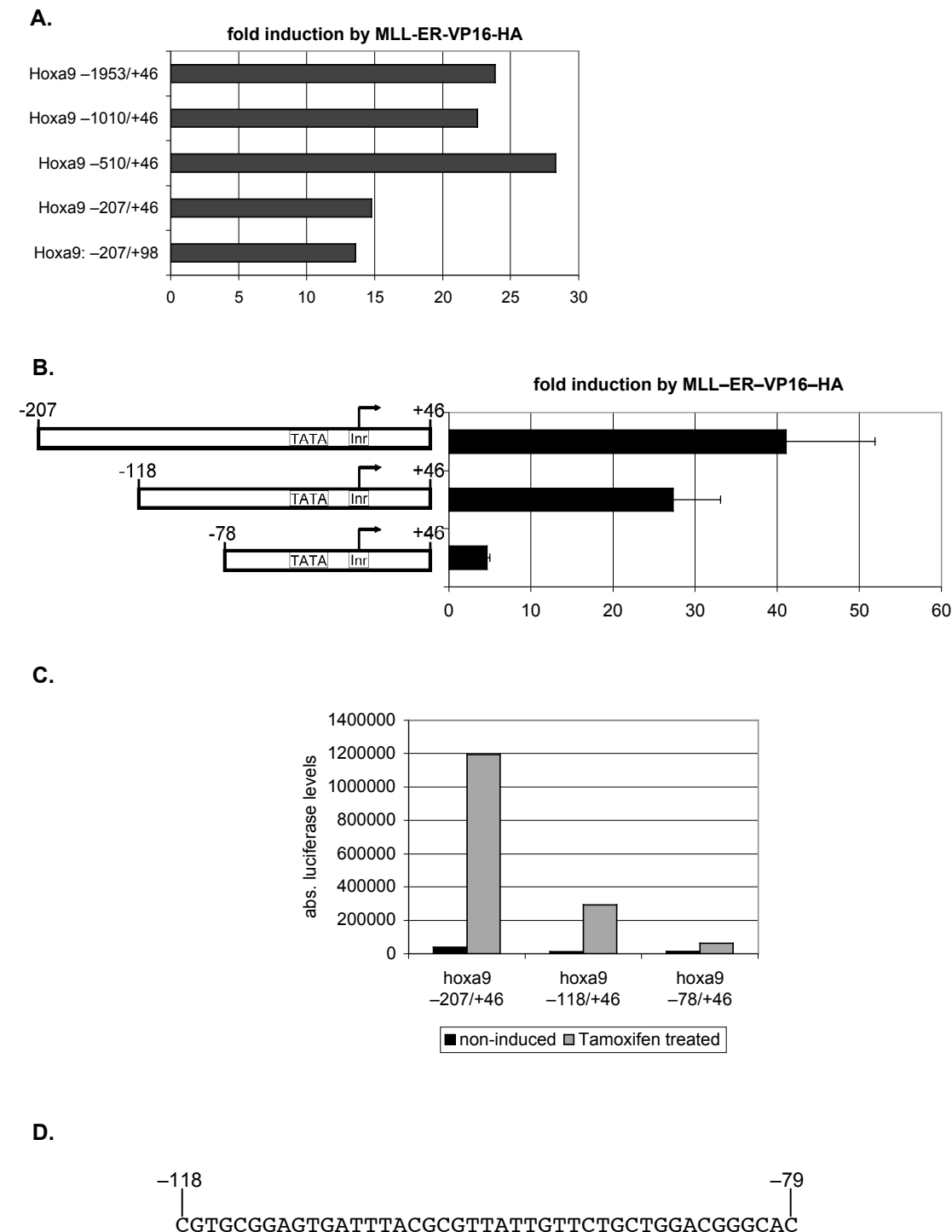


Figure 12. Hoxa9 promoter deletion analysis. A. and B. Reporter gene induction levels as determined by luciferase measurements in U937 cells. Numbers are referring to the start point of transcript NM_152739. C. Basal levels of the experiment shown in B. D. Sequence of the critical region -118 / -79.

3.4 Site-directed mutagenesis of the human Hoxa9 promoter region

In the next step individual transcription factor binding sites and other defined motifs in the Hoxa9 promoter sequence were to be analyzed. Towards this goal point mutations were introduced into the longer (i.e. -207 / +46) and shorter (i.e. -118 / +46) fragments (Fig. 13).

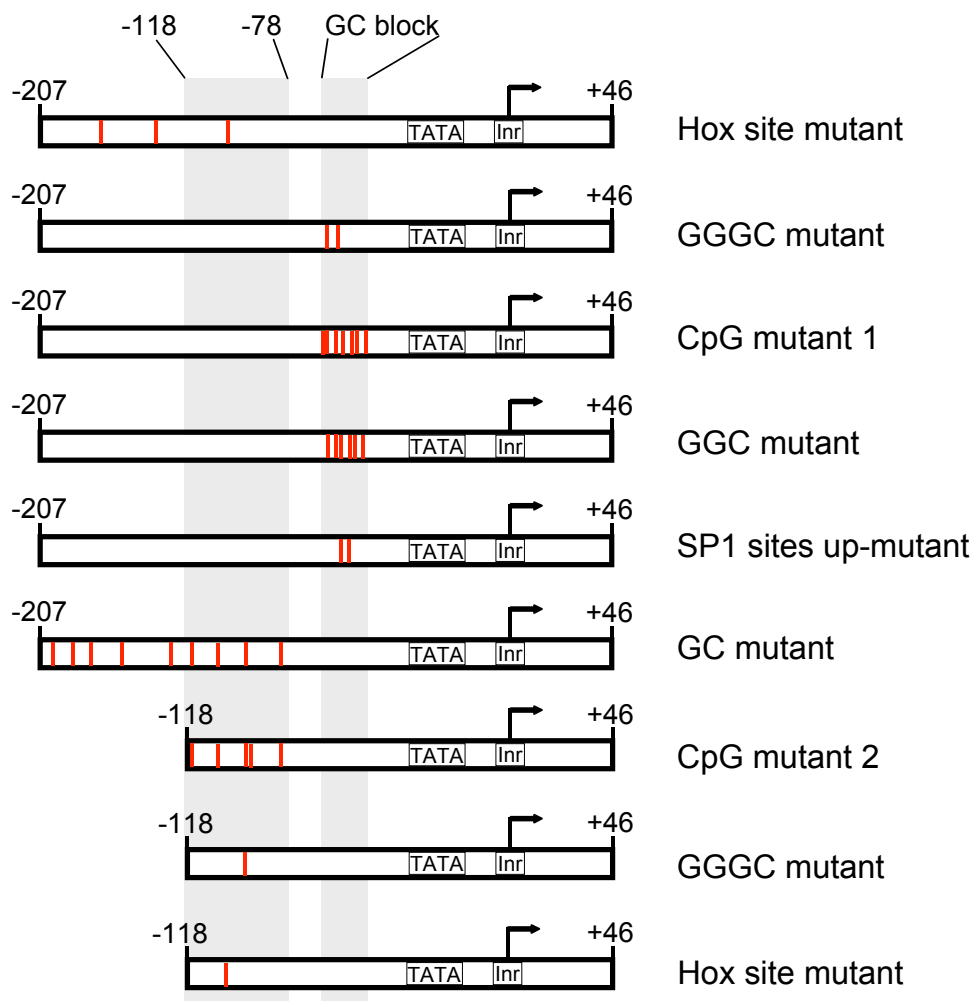


Figure 13. Overview of fragments used in the site-directed mutagenesis analysis. For sequence alignments see appendix 6.1. For the comprehensive plasmid list see 2.8.

3.4.1 Hox protein binding sites have a dual role for Hoxa9 regulation

One intriguing feature of the Hoxa9 promoter is the presence of binding sites for homeobox proteins. Their presence raises the possibility that Hoxa9 underlies autoregulatory control by Hox factors. Therefore, the significance of these sites for basal transcription as well as for MLL–VP16–ER–HA dependent transcription was investigated. Of three putative Hox factor binding sites H3 is the most well-conserved (consensus: ATGATTTAT). It is furthermore the only one present in the shorter construct (Fig. 13; H1, H2, H3).

Luciferase levels were measured before and after induction of stable cell lines. Interestingly basal transcriptional levels were raised for both of the mutant variants indicating that homeobox proteins might be involved in Hoxa9 transcriptional regulation (Fig. 14A). The increased reporter gene levels in the mutants, however, indicate that Hox factor binding to the Hoxa9 promoter exerts an inhibitory effect on gene activation. In contrast to this finding the mutations seemed to affect gene activation by MLL–VP16–ER–HA in a negative way: mutation of all three sites in Hoxa9 –207 / +46 reduced the induction rate by about 20 percent (Fig. 14B). Mutation of only H3 in Hoxa9 –118 / +46 led to an approximately 2-fold decrease of induction. This supports the previous conclusion of a critical role of the –118 / +46 region. Furthermore together with the increased basal transcription rate observed in these mutants these findings indicate that Hox factors might function as negative regulators of Hoxa9 transcription. MLL–VP16–ER–HA seems to act through the same target sites on the DNA to activate transcription, possibly by competition with these negative regulators.

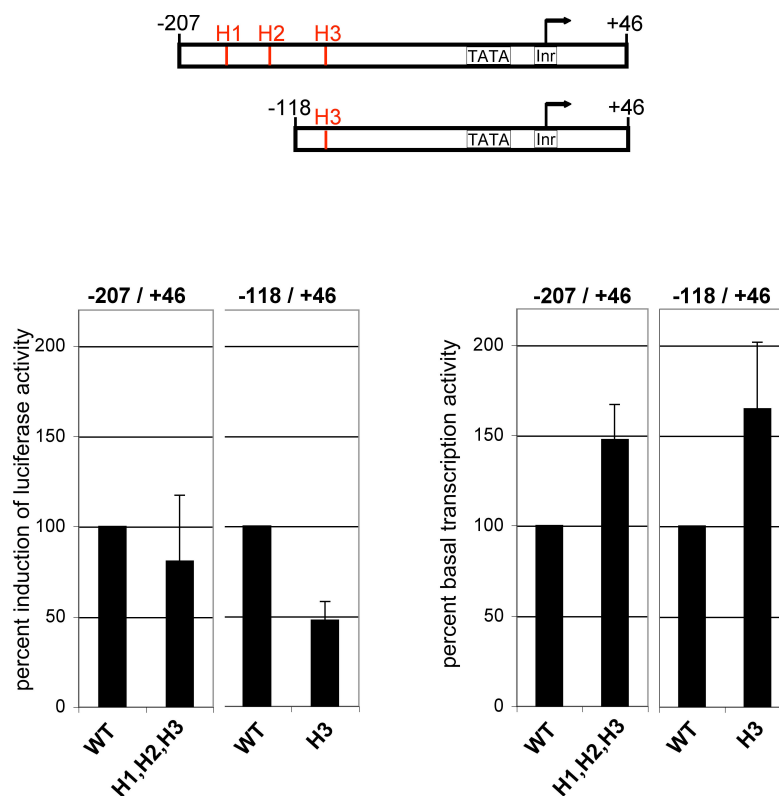


Figure 14. Effects of Hox protein binding site mutations on *Hoxa9* transcription. Two mutant constructs of different lengths were compared with the corresponding wild type sequences. A. Induction rates of the mutant constructs after activation of MLL–VP16–ER–HA are shown in comparison to wild type induction levels. B. Basal transcription rates of mutant reporter constructs are shown in comparison to wild type levels. WT: wild type. H1, H2, H3: Hox protein binding sites (for alignment see appendix 6.1).

3.4.2 A GC rich region in the *Hoxa9* promoter is mediating epigenetic regulation

It has been published that the MLL MT domain binds to GC rich DNA (Birke et al. 2002; Ayton et al. 2004). Along this line one striking feature of the *Hoxa9* promoter region is the presence of a sequence stretch with very high GC content (Fig. 13, GC block). This raises the question whether this is critical for the effect of MLL–VP16–ER–HA on this promoter. The highest GC density is found in the region between residues –78 and –36 that has been termed "GC block" (Fig. 15). Notably the *Hoxa9* –78 / +46 construct retains a residual activity that is sufficient to mediate five-fold induction of transcription by MLL–VP16–ER–HA (Fig. 10B). Furthermore this region might function cooperatively with other sequences that are positioned

more 5'. Hence four different mutant variants affecting several motifs were employed to investigate the specificity of this process (Fig. 15 and appendix 6.1).

GGGC mutant	CAC GTGACGCGCA CGGCCAATGG <u>GTGCGCGTGC</u> GCCGGCAACT
WT	CAC GTGACGCGCA CGGCCAATGG GGGC GC GGGC GCCGGCAACT
CpG mutant 1	CAC GTGAG <u>GGG</u> GCA <u>GGG</u> CGCCAAATGG GGG <u>GG</u> GGGGG <u>GGG</u> GGCAACT
WT	CAC GTGAC <u>CG</u> CGCA CGG CGCCAAATGG GGG <u>CG</u> CGGG C CG GGCAACT
SP1 sites upmutant	CAC GTGACGCGCA CGGCCAATGG GGGCG <u>G</u> GGGG <u>GGG</u> TCAACT
WT	CAC GTGACGCGCA CGGCCAATGG G GGGCGCGGGC GCCGG CAACT
GGC mutant	CAC GTG <u>T</u> CGCGCA CG <u>A</u> CCAATGG GG <u>A</u> CGGG <u>A</u> GTCG <u>A</u> CAACT
WT	CAC GTG <u>A</u> CGCGCA CGG CGCCAAATGG G GGC CGCG GGC GCC GGC CAACT

Figure 15. Mutants affecting the GC block region of the human Hoxa9 promoter. Nucleotide sequence alignments of the point mutants in the GC block region and the Hoxa9 wild type promoter region -78 / -36. Motifs are indicated in red, mutations were underlined in black. Asterisks are indicating the CpG dinucleotides that are affected by both the CpG mutant 1 and the SP1 sites upmutant.

3.4.2.1 CpG mutations in the Hoxa9 promoter lead to strong activation of transcription by MLL-VP16-ER-HA

All MLL fusion proteins contain a domain of high homology with DNA methyltransferases (Ma et al. 1993). *In vitro* this domain binds DNA with a preference for non-methylated CpG rich sequences (Birke et al. 2002; Ayton et al. 2004). It was hence particularly interesting to test whether CpG rich sequences that are present in the Hoxa9 promoter play a critical role in transcriptional control by MLL fusion proteins.

The GC block (Fig. 15) contained in Hoxa9 -78 / +46 is harbouring seven CpG dinucleotide sequences all of which have been mutated in the construct "CpG mutant 1" (Fig. 13 and 15). In order not to change the GC content of the DNA sequence CpGs were mutated to GpG. Reporter gene levels were measured before and after induction of MLL-VP16-ER-HA activity. In contrast to the reporter constructs described so far great variations were observed between different cell batches: while some batches behaved in a manner very similar to wild type, other samples that were transfected and selected independently, diverged dramatically

from wild type and showed very high induction rates. Furthermore inducibility seemed to change over the culturing period. Figure 16 is showing representative experiments. While the basal transcription level is hardly affected in the case of CpG mutant 1, induction rates after OHT–addition show an approximately four–fold increase over wild type (Fig. 16, CpG mutant 1, 23 days) indicating that elimination of CpG dinucleotides facilitates activation by MLL–VP16–ER–HA.

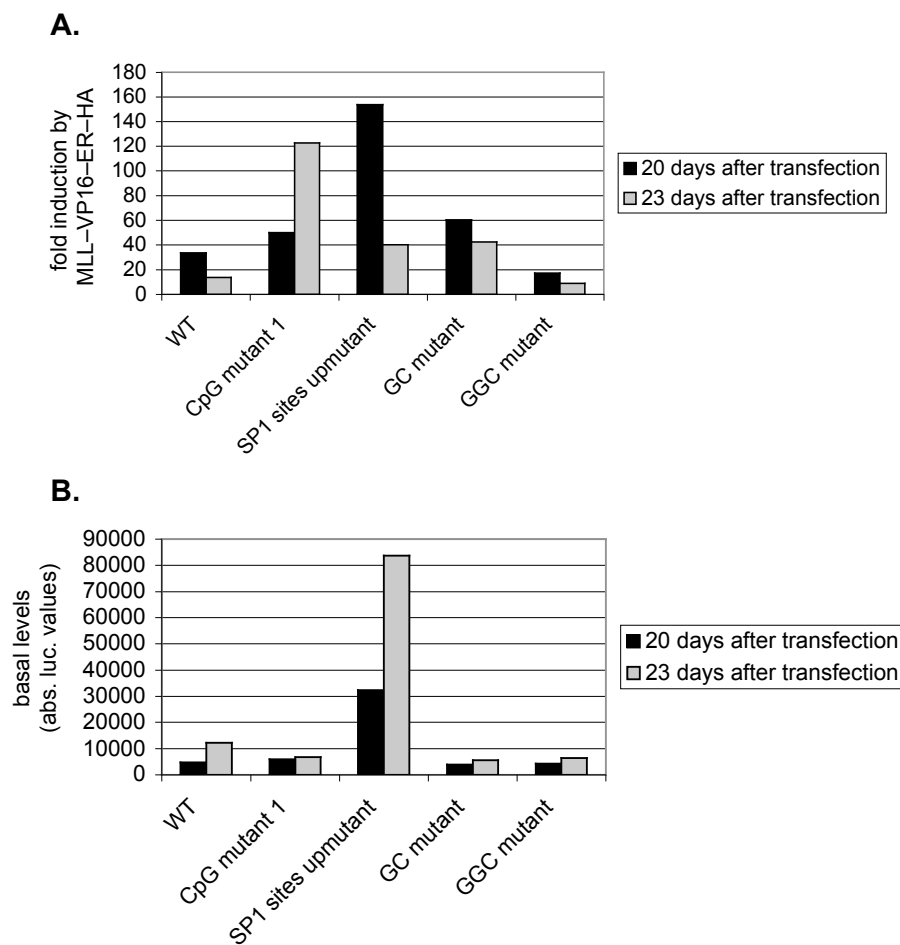


Figure 16. CpG mutant analysis in *Hoxa9* –207 / +46 background. Stable cell lines were generated containing *Hoxa9*–luciferase reporter carrying different point mutations affecting CpG dinucleotides. Luciferase levels were measured before and 24 hours after addition of Tamoxifen. The figure shows two measurements, that were performed 20 and 23 days after transfection, respectively. A. Induction levels following activation of MLL–VP16–ER–HA. B. Basal reporter gene levels.

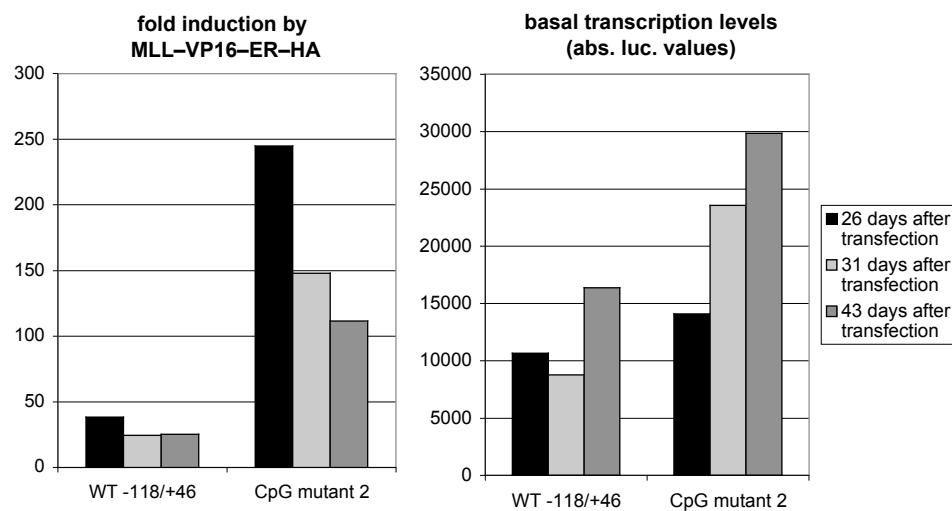


Figure 17. Effect of CpG mutations in Hoxa9 -118 / 46. Luciferase levels were measured before and 24 hours after induction of MLL-VP16-ER-HA. Reporter gene activity was analysed 26, 31 and 43 days after transfection.

Similar effects were observed with other mutant reporter constructs where CpG dinucleotides were affected: "SP1 sites upmutant" (Fig. 13 and 15) was initially designed to investigate whether SP1 sites might mediate activation through MLL-VP16-ER-HA. The same region was targeted as in CpG mutant 1. In SP1 sites upmutant, however, the GC block (Fig. 13 and 15) was changed at four nucleotide positions to create two consensus SP1 sites. These mutations affected two of the seven CpG dinucleotides that were also changed in CpG mutant 1 (Fig. 15) providing for an overlap between these two constructs and adding another aspect to SP1 site upmutant: this mutant does not only represent a better target for proteins binding to SP1 sites but independently from this effect it might also affect binding of factors that specifically recognize CpG dinucleotides and their methylation status, respectively. As already described for the CpG mutant 1 a dramatic increase of induction by MLL-VP16-ER-HA could be measured (Fig. 16A), which decreased after three more days of incubation, however. Basal levels were raised in this case as well and showed an increasing tendency over the culturing period (Fig. 16B). These results are basically mirrored in another mutant construct where CpG dinucleotides were targeted that are located between -118 and -78 (CpG mutant 2, Fig. 13). Similar to the SP1 sites upmutant this construct leads to increase of activated as well as basal transcription levels (Fig. 17). Three measurements over a period of 17 days showed a decrease of induction rates as well as an increase of basal transcription levels.

These three constructs (CpG mutant 1 and 2, SP1 sites upmutant) have in common that CpG dinucleotides were mutated to GpG dinucleotides that cannot be methylated by cellular DNMTs. The increased basal transcription levels observed in two of these mutants (SP1 sites upmutant, CpG mutant 2) argue for a loss of silencing due to a lack of DNA methylation. This effect might be augmented by the presence of two SP1 sites in one of the constructs. For all three constructs the induction levels were increased indicating that CpG dinucleotides negatively regulate activation through MLL–VP16–ER–HA possibly via reduced binding to methylated DNA. In all three constructs cytosine residues were replaced by guanosine residues. This argues for a model, where MLL is recruited to G-rich sequences, while methylated CpG dinucleotides exert an inhibitory function. Support for this hypothesis comes from another Hoxa9 promoter mutant: in "GC mutant" (Fig. 13) CpG dinucleotides were eliminated not by substitution through GpG, but by replacement with ApG reducing the G-content in comparison to the three mutants discussed above. Noteworthy activation rates by MLL–VP16–ER–HA were raised as well but did not reach the same levels that had been observed with the CpG to GpG mutants (Fig. 16A). This result might reflect the loss of repression of MLL–VP16–ER–HA activation mediated through the lack of methylated DNA without the positive effect of increased G-content.

Three additional mutants were designed to monitor the effect of changes in G-content in a background where integrity of CpG-dinucleotides was not affected (GGC mutant and GGGC mutants 1 and 2, Fig. 13 and 15). In both cases G-rich motifs were changed by replacing single G residues by A or T (see also alignments, appendix 6.1). In the case of GGC mutant slightly decreased induction rates in comparison to wild type were observed (Fig. 16A). Mutation of the G-rich motif "GGGC" (GGGC mutants 1 and 2) led to similar results. While mutation of GGGC motifs 2 and 3 negatively affected basal transcription rates, GGGC1 (located in the critical -118 / -79 region) was impeding activation of the reporter by MLL–VP16–ER–HA (Fig. 18A and B). This might indicate impaired recruitment of MLL–VP16–ER–HA in a background where the negative control through DNA methylation is still functional resulting in reduced transcription levels.

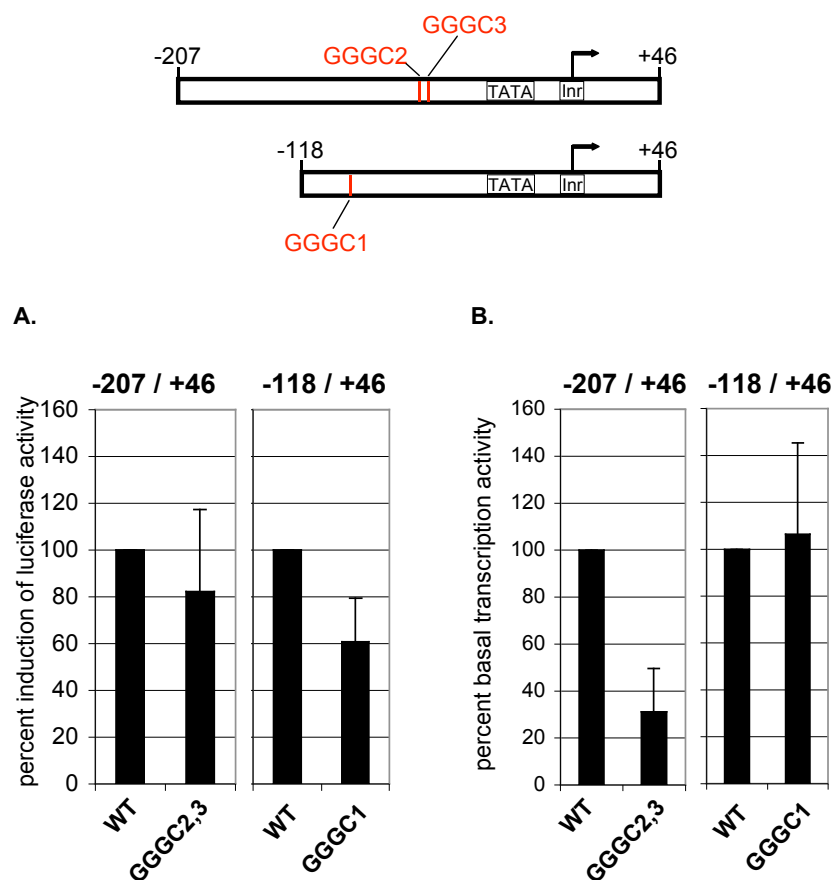


Figure 18. Mutation of GGGC motifs in the *Hoxa9* promoter sequence. Mutant promoter fragments of different lengths (–207 / +46 and –118 / +46) have been analysed in reporter gene assays. Luciferase levels of stable reporter cell lines were measured. A. Induction rates of the mutant constructs after activation of MLL–VP16–ER–HA are shown in comparison to wild type induction levels. B. Basal transcription rates of mutant reporter constructs are shown in comparison to wild type levels. For sequence alignment see appendix 6.1.

3.4.2.2 SP1 protein overexpression does not facilitate gene activation by MLL–VP16–ER–HA

The GC block present in *Hoxa9* –78 / +46 contains two non-consensus SP1 sites. As described above (see 3.4.2.1) the introduction of two consensus SP1 sites (which affects two CpG dinucleotides; see alignments, appendix 6.1) has dramatic effects on basal as well as activated transcription. Comparison with other mutant variants reveals similarities between SP1 sites upmutant and the constructs specifically mutated for CpG dinucleotides. This raises the question whether the activator SP1 might recruit MLL–VP16–ER–HA to the promoter or whether the effect is due to sequence-specific effects that are independent from SP1 protein.

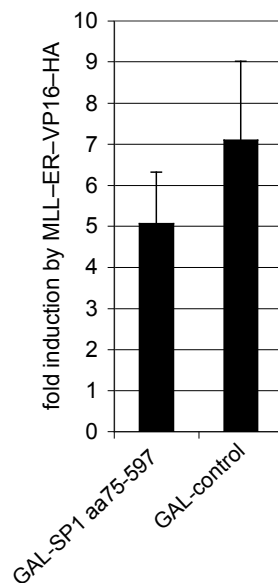


Figure 19. Effect of GAL-SP1 on MLL-VP16-ER-HA driven activation of a GAL-Hoxa9 promoter. Induction rates of luciferase activity after activation of MLL-VP16-ER-HA are shown for cells transfected with a GAL-SP1 expression vector or a GAL-control, respectively.

To clarify whether SP1 protein might be recruiting MLL-VP16-ER-HA to the Hoxa9 promoter a reporter construct was designed that contained the -78 / +46 fragment of the human Hoxa9 promoter, which was shown to be only weakly inducible in comparison to the -118 / +46 construct. Two GAL-binding sites were cloned 5' of this promoter sequence and stable cell lines were created by transfection of U937 / MLL-VP16-ER-HA cells with this DNA construct. Cells that were stably propagating the EBV reporter plasmid were then transfected transiently with an expression vector for GAL-SP1 or the empty vector, respectively. MLL-VP16-ER-HA activity was induced by addition of Tamoxifen 17 hours after transfection and luciferase levels were measured 24 hours later. In this experiment no significant influence of SP1 on MLL-VP16-ER-HA activated transcription could be detected (Fig. 19) indicating that SP1 protein is not involved in the activation of Hoxa9 through MLL-VP16-ER-HA. This result is rather suggesting that the positive effect of SP1 sites upmutant is more likely due to loss of DNA methylation or to sequence-specific binding events of either MLL-VP16-ER-HA itself to these sites or other factors that could in turn recruit MLL-VP16-ER-HA.

3.4.3 Introduction of synthetic binding sites into the Hoxa9 minimal promoter

Based on the finding that SP1 binding sites and GGGC motifs are important for Hoxa9 regulation it was to be tested whether these sequences can reconstitute activation by MLL–VP16–ER–HA in Hoxa9 –78 / +46. Deletion analyses showed that this promoter fragment retains a residual capacity for induction through MLL–VP16–ER–HA but the effect is minor in comparison to –118 / +46 (Fig 12B). Synthetic sequences were cloned 5' of the Hoxa9 –78 / +46–luciferase reporter in an episomal vector. Different oligonucleotides of 32 nucleotides length were designed that vary in GC content and binding sites that are present in the sequence. Three copies of the GGGC motif are present in MAZR1, MAZR2 and MAZR3. These constructs differ in spacing and orientation of the GGGC motifs as well as in the intervening sequences. Bioinformatic analysis revealed binding sites for different proteins (Table 8). Notably all three of these constructs contain at least one binding site for MAZR (myc–associated zinc finger protein related transcription factor). Stable reporter cell lines were generated containing these vectors and luciferase levels were measured before and after induction of MLL–VP16–ER–HA activity. Three different Hoxa9 constructs were measured as positive controls with Hoxa9 –78 / +46 representing the baseline (Fig. 20). As an unspecific control for the effect of introduction of synthetic sequences into the –78 / +46 construct GAL sites were used.

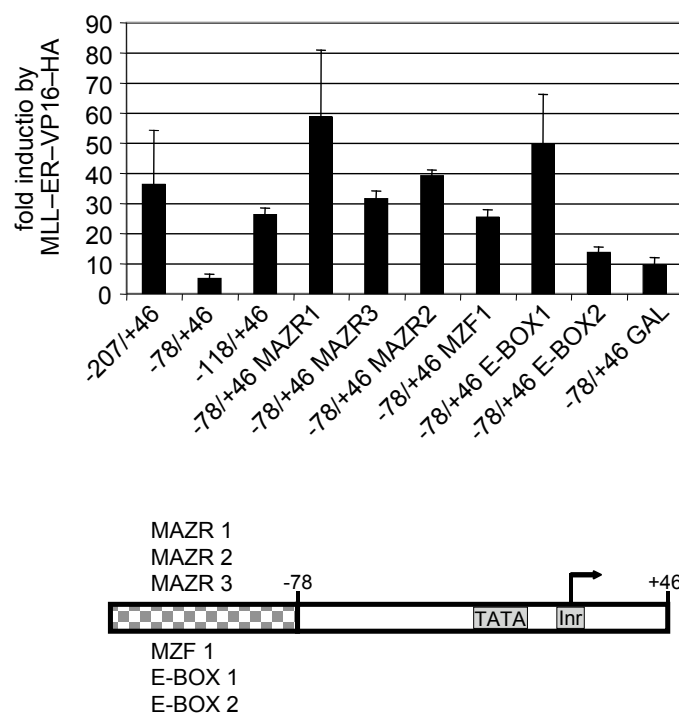
MAZR1, MAZR2 and MAZR3 all led to induction rates between 3 and 6–fold over the GAL–control levels. Induction of these constructs also exceeds the one of Hoxa9 –118 / +46 even though the synthetic constructs are 8 nucleotides shorter in comparison. Together the results from these three constructs are consistent with earlier observations that GGGC motifs positively modulate activation by MLL–VP16–ER–HA. Noteworthy, however, the highest induction levels have been measured with MAZR1 although considerable variations were noted in between experiments. The sequence of MAZR1 shows that GGGC motifs are located on the same DNA strand and are evenly spaced. Shifting the central GGGC motif to one side thereby generating uneven 5' and 3' intervals (MAZR2) leads to a decrease in activation by MLL–VP16–ER–HA. This effect is even stronger in MAZR3, where GGGC boxes are not only unevenly distributed but also located on different

DNA strands. These findings indicate that GGGC boxes act cooperatively to recruit MLL–VP16–ER–HA. Cooperativity furthermore seems to depend on regular spacing between these motifs and location on the same strand of the DNA.

Mutation analysis of the GC block in the Hoxa9 promoter led to the hypothesis that G-rich DNA might be important for MLL–VP16–ER–HA recruitment while cytosine residues might rather be required for modulation of this process via methylation (see 3.4.2.1). A series of GGGG motifs are present in the MAZR constructs as well, all of which showed increased induction rates by MLL–VP16–ER–HA. To address the question whether G nucleotide-repeats (GGGG) alone are sufficient to mediate activation by MLL–VP16–ER–HA a construct was designed containing three of these motifs in evenly distributed over the length of the oligonucleotide (MZF1). Like this two binding sites for the myeloid zinc finger protein 1 (MZF1) were generated as revealed by bioinformatic analysis. The induction rate of this construct was approximately 2.5-fold higher than the GAL control indicating that guanosine repeats positively affect activation via MLL–VP16–ER–HA. Direct comparison with constructs MAZR1, MAZR2 and MAZR3 is possible only to a limited extent because the intervening sequences between the motifs tested in each case are not identical.

A third class of constructs is represented by E–BOX1 and E–BOX2. As another representative of GC-rich motifs E-boxes should be tested for their capacity to facilitate induction by MLL–VP16–ER–HA. Two different constructs were designed, each of them containing three E-box motifs. While the spacing between these motifs is regular in E–BOX1, the motifs are distributed irregularly in E–BOX2. Interestingly, this seems to greatly affect activation through MLL–VP16–ER–HA: E–BOX2 induction levels were only slightly raised in comparison with the GAL-control, while E–BOX1 significantly augmented reporter gene transcription after MLL–VP16–ER–HA induction. These findings suggest that E-box motifs might be involved in MLL fusion protein recruitment to target gene promoters. Furthermore as described before for the GGGC motif also in this case cooperativity between several motifs seem to play a certain role. The fact that shifting one of the E-box motifs by only one nucleotide position leads to loss of activation indicates that spacing between single motifs is absolutely critical.

It is intriguing that introduction of either one of the three motifs (GGGC, GGGG, E-box) reconstitutes the induction levels that were observed with the Hoxa9 –118 / +46 reporter construct. A bioinformatic analysis of binding sites present in each of these constructs revealed no overlap between the three classes of synthetic constructs (Table 8, boxes in colors). Most strikingly, however, none of the binding sites present in the Hoxa9 –118 / –79 fragment was identified in any of the synthetic constructs.



GC	<u>GGGC</u> GGGGT	<u>GGGC</u> GGGGT	<u>GGGC</u> GGGGT	MAZR 1
GC	<u>GGGC</u> GGGGT	GCT <u>GGGC</u> AGC	<u>GGGC</u> GGGGT	MAZR 2
GC	<u>GGGC</u> GGGGT	ACCCC <u>GCCCC</u>	<u>GGGC</u> GGGGT	MAZR 3
GC	CAGT <u>GGGG</u> AA	CAGT <u>GGGG</u> AA	CAGT <u>GGGG</u> AA	MZF 1
GC	AC <u>CACGTG</u> AC	AC <u>CACGTG</u> AC	AC <u>CACGTG</u> AC	E-BOX 1
GC	TGC <u>CACGTG</u> C	GC <u>CACGTG</u> GC	CAG <u>CACGTG</u> G	E-BOX 2

Figure 20. Reconstitution of MLL-ER-VP16-HA activated transcription from the Hoxa9 –78 / +46 promoter. Synthetic oligonucleotides were cloned in front of Hoxa9 –78 / +46. Cells were selected that were stably replicating the EBV-based reporter constructs. Luciferase induction rates were determined following activation of MLL-ER-VP16-HA by addition of Tamoxifen. GGGC motifs were boxed in orange, GGGG motifs were boxed in light blue and E-box motifs were boxed in yellow. GGGG motifs present in the mutants MAZR1,2 and 3 are underlined in light blue.

Table 8. Predicted binding sites introduced into Hoxa9 -78 / +46.

	MAZR 1	MAZR 2	MAZR 3	MZF1	E-Box 1	E-Box 2	WT
CDE.01							●
CHREBP_MLX.01						●●	
E2F.01							●
XBP1.01					●●	●●	
MYCMAX.02					●●		
MYCMAX.03						●	
USF.01					●	●●●	
EGR1.02	●●						
BKLF.01	●						
KKLF.01	●●	●					
ZIC2.01		●					
HAND2_E12.01						●	
HELT.01					●●	●●●●	
ARNT.01					●●		
DEC2.01					●	●●●	
HIF1.01						●	
HNF1.03							●
PBX_HOXA9.01							●
HOXA9.01							●
MAZR.01	●●●●	●	●●				
MUSCLE_INI.02	●●						
VMYB.04							●
MZF1.01				●●			
PAX6.01							●
PBX1_MEIS1.02							●
PLAG1.01	●●						
RREB1.01	●●						
SOX5.01							●
GC.01	●●						
SP1.01	●		●				
ZF9.01	●●		●				

Oligonucleotide sequences present in MAZR1, MAZR2, MAZR3 (orange), MZF1 (light blue) and E-BOX1, E-BOX2 (yellow) have been analyzed using MatInspector. In parallel the region -118 / -79 of the Hoxa9 promoter has been analyzed (green). The number of black points symbolizes the number of binding sites identified.

3.4.4 A consensus TATA box stimulates activation of Hoxa9 by MLL-VP16-ER-HA

Besides the regulation of Hox gene expression a more global role has been proposed for MLL and it has been shown that MLL binds to regions around the transcriptional start point in many genes. To investigate whether core promoter features influence the effect of MLL fusion proteins on Hoxa9 expression a series of constructs was created containing point mutations in the region around the TATA box and the Initiator sequence (Fig. 21).

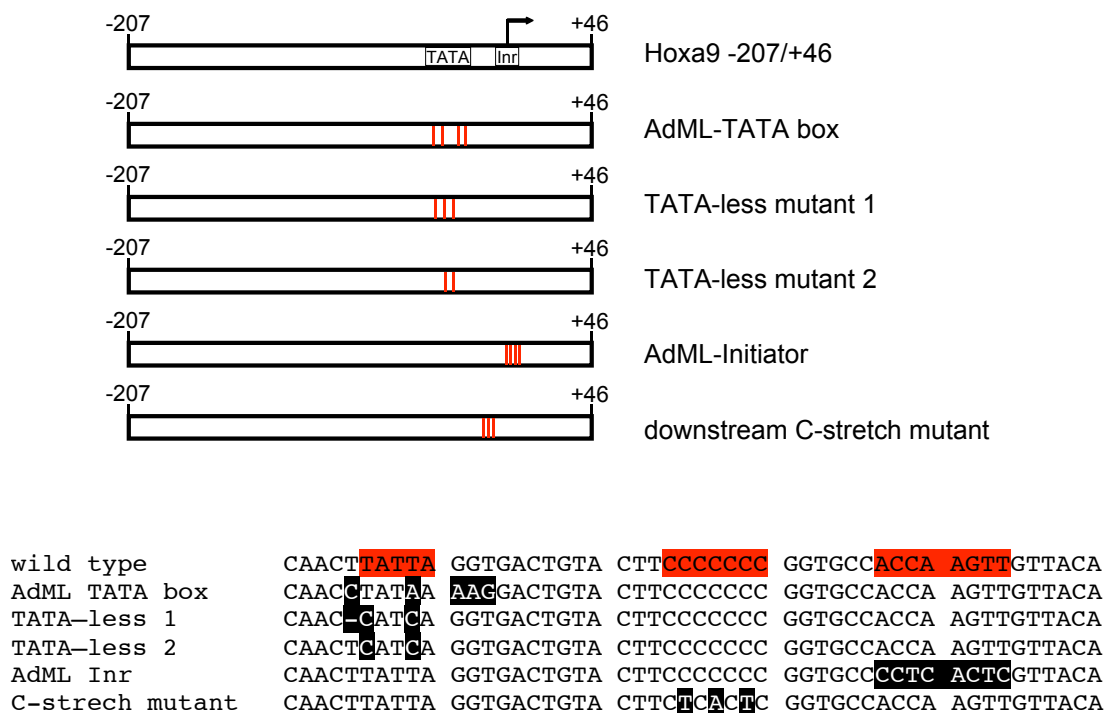


Figure 21. Hoxa9 promoter variants carrying mutations in core promoter sequences. TATA box, C-strech and Initiator are boxed in red in the wild type sequence. Point mutations are boxed in black.

Exchange of the non-consensus endogenous TATA box in the *Hoxa9* (Fig. 21) by the TATA box sequence from the Adenovirus Major Late Promoter (AdMLP) increased induction rates by MLL-VP16-ER-HA more than 2-fold in comparison to the wild type sequence (Fig. 22A). Also mutating the Initiator sequence (Inr) to the corresponding sequence of AdMLP led to an increase in transcriptional activity induced by MLL-VP16-ER-HA although large variations were observed in independent experiments. In order to eliminate a consensus TATA box sequence from the *Hoxa9* promoter sequence was changed from TATTA to CATCA (Fig. 21, TATA-less2). As a consequence the induction was diminished to 50% compared to the wild type promoter (Fig. 22, TATA-less-2). Interestingly the additional deletion of another cytosine residue immediately 5' of the *Hoxa9* TATA box (Fig. 21, TATA-less 1) could rescue this effect almost completely (Fig. 22A).

Based on the observation that core promoter sequences are capable of modulating *Hoxa9* regulation by MLL-VP16-ER-HA it was to be investigated whether the *Hoxa9* core promoter is containing a specific motif that is targeted by the fusion protein. In this respect a cytosine heptanucleotide repeat (Fig. 21, C-stretch) located between the TATA box and the Inr sequence raised the question whether this motif might be important for activation by MLL-VP16-ER-HA. Substitution of three of these C residues by A or T led to a decrease of the induction rate of about one third (Fig. 22A) indicating that this C-stretch might play a role in *Hoxa9* activation by MLL-VP16-ER-HA. Interestingly this mutation had the strongest impact on basal transcription levels (Fig. 22B). This is intriguing because it is not known that components of the basal machinery specifically recognize this motif. Together with decreased induction rates after MLL-VP16-ER-HA activation this might indicate direct targeting of both MLL and MLL fusion proteins to this site with full-length MLL maintaining low-level basal transcription and MLL-VP16-ER-HA leading to activated transcription on a high level.

Taken together these results suggest that both the presence of a consensus TATA box sequence and a cytosine heptanucleotide repeat between TATA box and Initiator facilitate recruitment of MLL-VP16-ER-HA to the *Hoxa9* promoter.

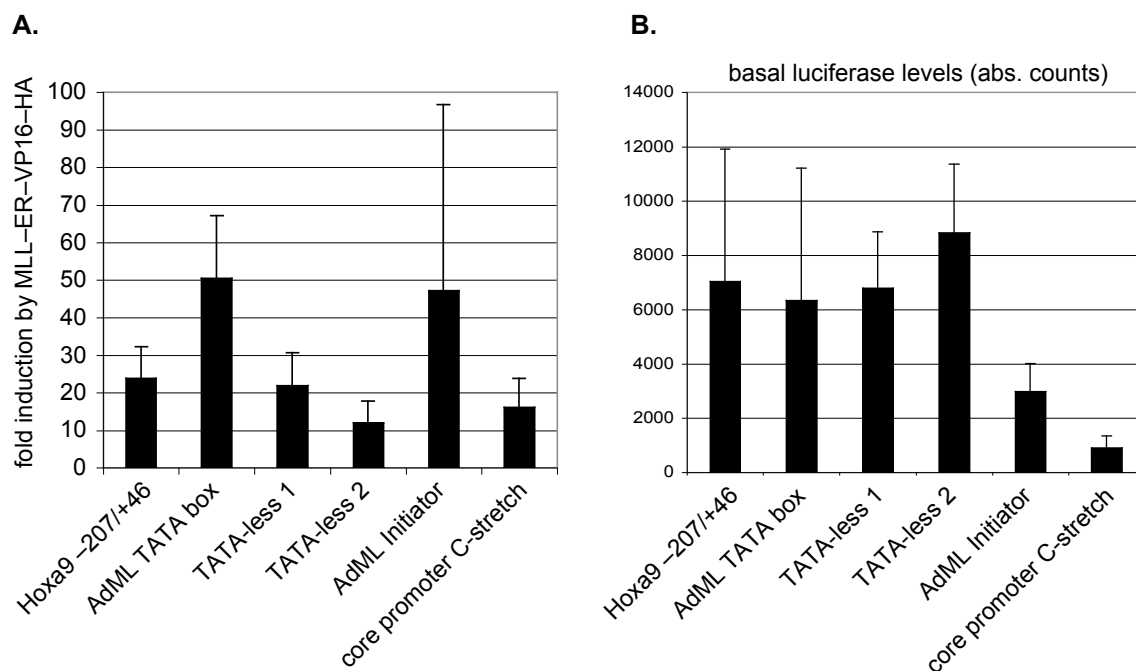


Figure 22. Mutation of core promoter sequences has an influence on activation through MLL-VP16-ER-HA. A. Reporter gene induction rates are shown derived from measurements performed in cells stably expressing MLL-VP16-ER-HA and episomally propagating one of the Hoxa9 reporter plasmids. B. Basal luciferase levels of the corresponding reporter constructs. Absolute luciferase counts are shown in the diagram.

3.5 Identification of new genomic target sites for MLL fusion proteins

As described so far the MLL-VP16-ER-HA model system was used to analyze the regulation of Hoxa9 by MLL fusion proteins. Critical features of this promoter were identified that mediate specifically the activation by MLL-VP16-ER-HA. Specific targeting of Hoxa9 by MLL fusion proteins is in line with many reports showing Hoxa9 and Meis1 overexpression in human leukemias. Recently, however, it has been suggested that the actual target gene population for MLL and MLL fusion proteins might be much larger (Guenther et al. 2005; Kohlmann et al. 2005).

The MLL-VP16-ER-HA system has been employed in an approach to identify new target genes for MLL fusion proteins. The ER system is particularly suitable for this application since induction of the gene of interest is not depending on de novo transcription and translation. As described earlier relocation (possibly in concert with a finishing step in protein folding) is enough to induce activity of the gene. The fast

kinetics of this process facilitates analysis of the early events following activation of an oncogene.

3.5.1 FDCPmix cells as a myeloid progenitor model system

MLL is described as a stem cell disease and it is clear that only cells of a certain population have the capacity to establish this type of leukemia. As an in vitro system of high physiological relevance hematopoietic progenitor cells were chosen for MLL fusion protein target gene analysis.

FDCP mix cells were transfected with the previously described expression constructs for MLL–VP16–ER–HA and VP16–ER–HA (see 3.2.1). After selection of cells that had stably integrated the construct single clones were selected as described in Material and Methods. Expression levels were very low in this case and no protein could be detected in Western Blot analysis (data not shown). Functional assays, however, showed that upon treatment of the cells with Tamoxifen some known MLL fusion protein target genes were upregulated.

A *Hoxa7*–luciferase reporter was induced more than 25–fold in cells transgenic for MLL–VP16–ER–HA, but not in stable cells harbouring the empty vector backbone (Fig. 23). Several different cell clones were behaving in the same way (data not shown). For practical reasons subsequent experiments were conducted with one clone each. The *Hoxa7*–luciferase experiment showed that MLL–VP16–ER–HA activity is under tight control. Luciferase levels measured in non-induced MLL–VP16–ER–HA cells were comparable to the levels detected in the control cell line. Titration of Tamoxifen showed that at a concentration of 1000 μM led to robust induction while 200 μM were not enough. In another set of experiments the concentration was raised further to 2500 μM Tamoxifen, which led to induction rates that were marginally higher (data not shown). In order to avoid toxic effects due to high drug levels subsequent experiments were performed at a Tamoxifen concentration of 1000 μM .

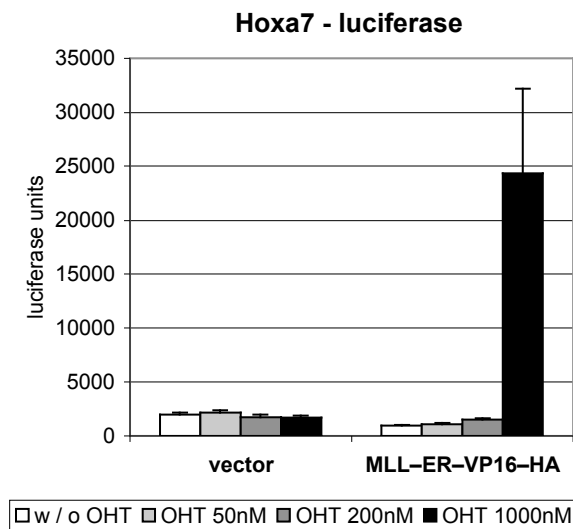


Figure 23. Hoxa7 reporter gene activation by MLL-VP16-ER-HA. FDCPmix cells stably expressing MLL-VP16-ER-HA were transfected with a Hoxa7-luciferase reporter plasmid and incubated with different concentrations of Tamoxifen (OHT). Depicted in the graph are absolute luciferase values measured 24 hours after induction.

Regulation of endogenous target genes by MLL-VP16-ER-HA was controlled by RT-PCR. For Meis1, a well-known MLL fusion protein target, it could be shown that mRNA amounts were significantly increased already 3 hours after addition of Tamoxifen (Fig. 24).

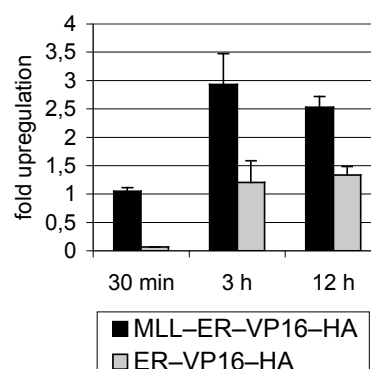


Figure 24. Upregulation of Meis1 mRNA concentrations by MLL-VP16-ER-HA but not by VP16-ER-HA. RT-PCR analysis of RNA samples derived from stable FDCPmix cells incubated with 1 μ M Tamoxifen for 30 min, 3 hours and 12 hours.

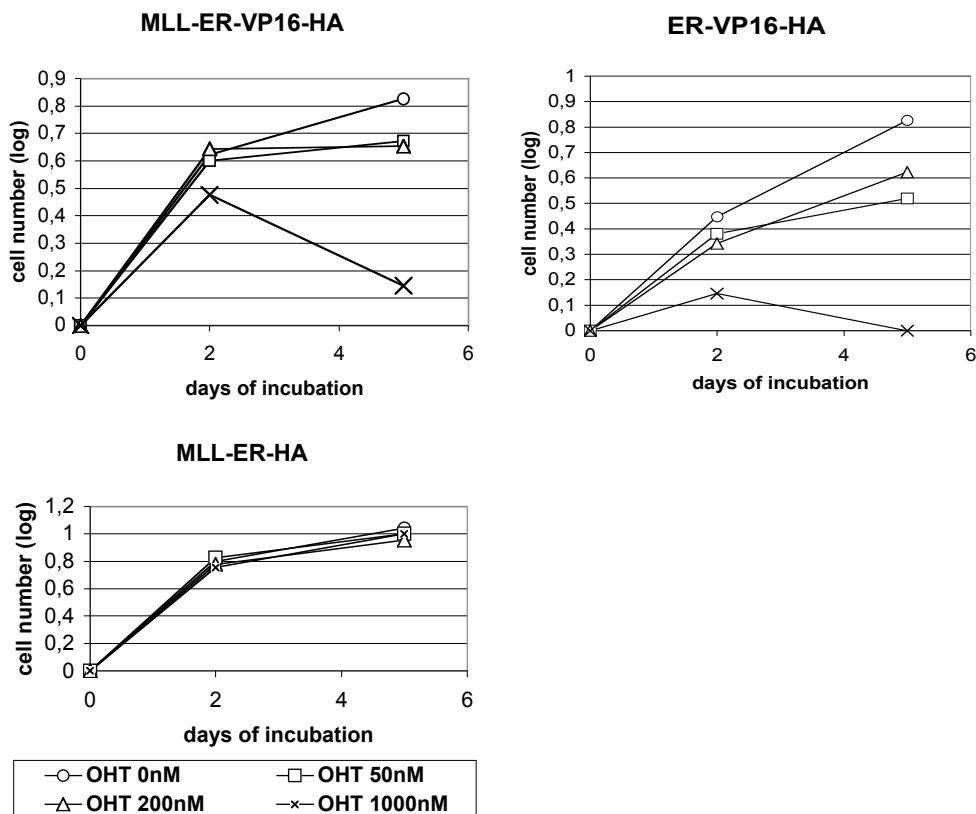


Figure 25. Induction of the ER-fused activation domain of VP16 has toxic effects in FDCPmix cells. Viable cells have been counted two and five days after addition of Tamoxifen. The graphs show a representative experiment.

It has been observed that treatment of the MLL-VP16-ER-HA cell line with Tamoxifen greatly slowed down the growth rate. Viable cell numbers decreased dramatically after 4 days if the growth medium contained 1 μ M Tamoxifen (Fig. 25). Comparison with two other cell lines provided evidence for the dependency of this effect on the induction of VP16-ER-HA. A control cell line expressing this variant showed the same toxic effects while a MLL-ER-HA cell line was not affected by the presence of Tamoxifen.

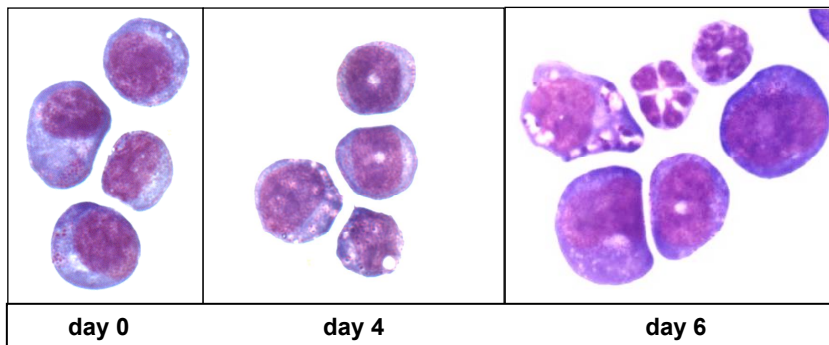
To ensure that the cell clone that was used for subsequent experiments was not compromised by manipulations required to establish the stable monoclonal cell line the cells were analyzed morphologically before and after differentiation. The myeloid differentiation potential of the transgenic MLL-VP16-ER-HA cell line was checked in an in vitro differentiation experiment. Non-transfected FDCP mix cells as well as the induced and non-induced MLL-VP16-ER-HA cell line were cultured in medium lacking IL 3 and containing GM-CSF and G-CSF. It has been shown that these

cytokines are required for differentiation of FDCP mix cells into granulocytes and macrophages and that this process is not impeded by the presence of Tamoxifen (Schroeder 2001). Cells were analyzed morphologically after Giemsa / May–Gruenwald staining (Fig. 26A). The relative numbers of undifferentiated blast cells, myelocytes / promyelocytes, granulocytes and macrophages / monocytes were counted at three different time points after induction of differentiation (Fig. 26B). After 11 days of culture under differentiation conditions the FDCP mix sample contained no undifferentiated cells (Table 9). More than half of the population was comprised of granulocytes and 20 percent of living cells were differentiated into macrophages or monocytes. The non Tamoxifen–treated MLL–VP16–ER–HA cell line showed a very similar distribution. Active MLL–VP16–ER–HA, however, led to a different outcome: almost half of the living cell population was still consisting of non–differentiated blast cells and only 15 percent of terminally differentiated cells could be observed in total.

Table 9. Cell numbers after GM differentiation of stable FDCPmix cells

		blasts	myelocytes/ promyelocytes	granulocytes	macrophages/ monocytes
FDCPmix WT	day0	88	10	0,5	1,5
	day11	0	28	52	20
MLL-ER-VP16-HA non-induced	day0	90	9	0	1
	day11	0	17	49	34
MLL-ER-VP16-HA induced	day0	90	9	0	1
	day11	41	47	9	6

A.



B.

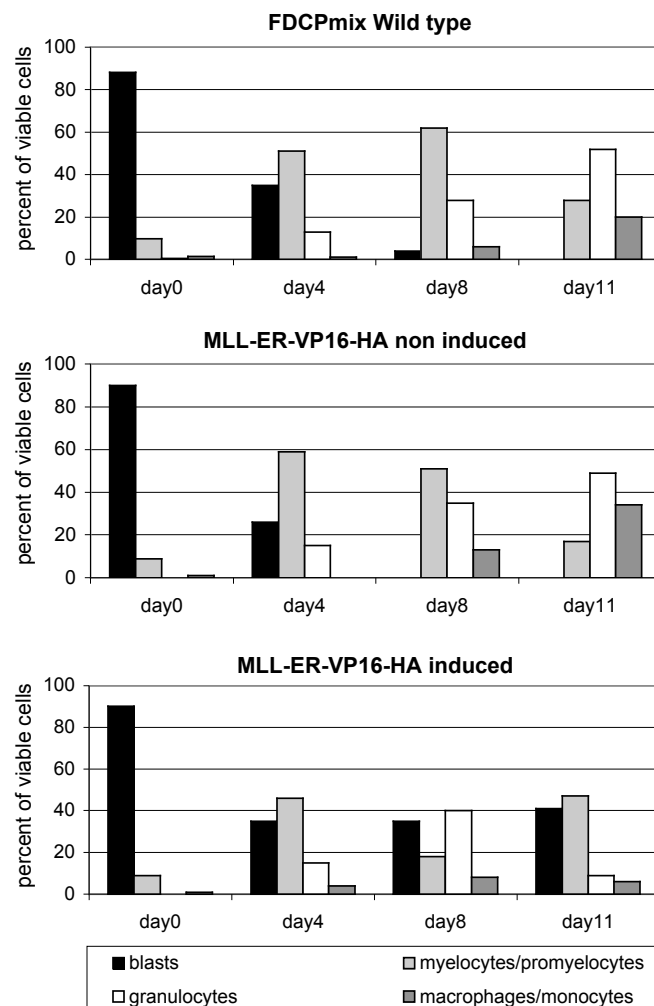


Figure 26. Induction of MLL-VP16-ER-HA blocks GM-differentiation. Stable FDCPmix cells were induced for GM differentiation and simultaneously treated with Tamoxifen. A. May-Gruenwald/Giemsa stain of cytopsin preparations of differentiated cells. B. Shown in the graph is the relative representation of cell types at different time points as determined by morphological analysis.

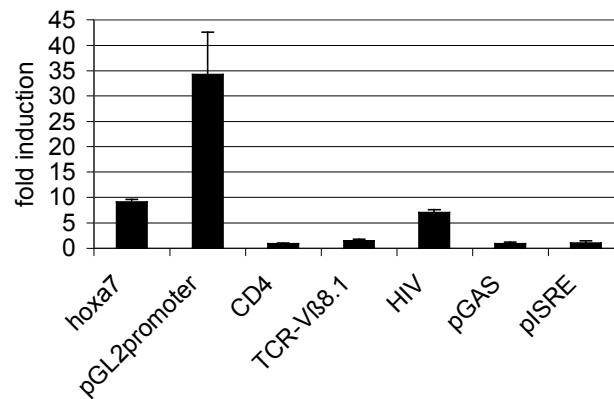


Figure 27. Effect of MLL–VP16–ER–HA activation on different reporter genes. Stable FDCPmix cells were transfected with reporter plasmids and induction of luciferase was measured 24 hours after induction with Tamoxifen.

This cell system was used for a series of different reporter experiments in order to investigate the influence of MLL fusion proteins on potential target gene promoters. One early experiment was performed in order to clarify whether the activation of the Hoxa7 reporter gene was specific for the Hoxa7 upstream sequences or whether any upstream region combined with a minimal promoter as found in the pGL2 basic vector would be sufficient to mediate the effect. A number of different luciferase reporter constructs were used for transfection of the stable MLL–VP16–ER–HA cell line (Fig. 27). While the Hoxa7 upstream region mediated a strong response upon induction of MLL–VP16–ER–HA by Tamoxifen some other promoters (CD4, TCR–Vb8.1, GAS, ISRE) showed only slightly upregulated transcription rates. Noteworthy, the pGL2 promoter plasmid, however, was induced by a factor of more than 30–fold indicating that the SV40 promoter contained in that vector is a suitable target for MLL–VP16–ER–HA and that the target gene spectrum for this protein might well exceed the range of so far known target genes.

3.5.2 c-myc promoter sequences as targets for MLL-VP16-ER-HA

Expression profiling of acute lymphoblastic and myeloblastic leukemias revealed upregulation of the proto-oncogene c-myc (Rozovskaia et al. 2003). Furthermore c-myc is one of the genes on whose promoters MLL was found to be present in a recent ChIP-chip study (Milne et al. 2005a). It was intriguing to follow these observations and investigate whether MLL plays a role in regulation of transcription from the c-myc promoter.

Using a series of c-myc luciferase reporter constructs it could be shown that the full-length construct (-2332 / +513) was induced by MLL-VP16-ER-HA more than 15-fold (Fig. 28A). Shorter constructs comprising only one of the two promoters of c-myc at a time were leading to similar rates, however, and there was no striking difference to be seen between P1 and P2. To find out whether the Igκ enhancer was playing a role in the induction of the c-myc reporter by MLL-VP16-ER-HA constructs were used in addition that contained the enhancer sequence which was cloned into the vector 3' of the luciferase gene. No dramatic change of the effect was observed with any of the constructs. Since the goal was to identify possible new MLL fusion target sites the shortest construct was also the most attractive object for the study since complexity could be reduced simply by the small size of the target sequence. Therefore the -101 / +66 construct was chosen for further investigation.

Mutant constructs of this promoter stretch, which for reasons of simplicity have been termed "distal" and "proximal" almost completely abolished the response to induction of MLL-VP16-ER-HA activity (Fig. 28B). In contrast basal transcription levels of the mutant reporter constructs were hardly affected.

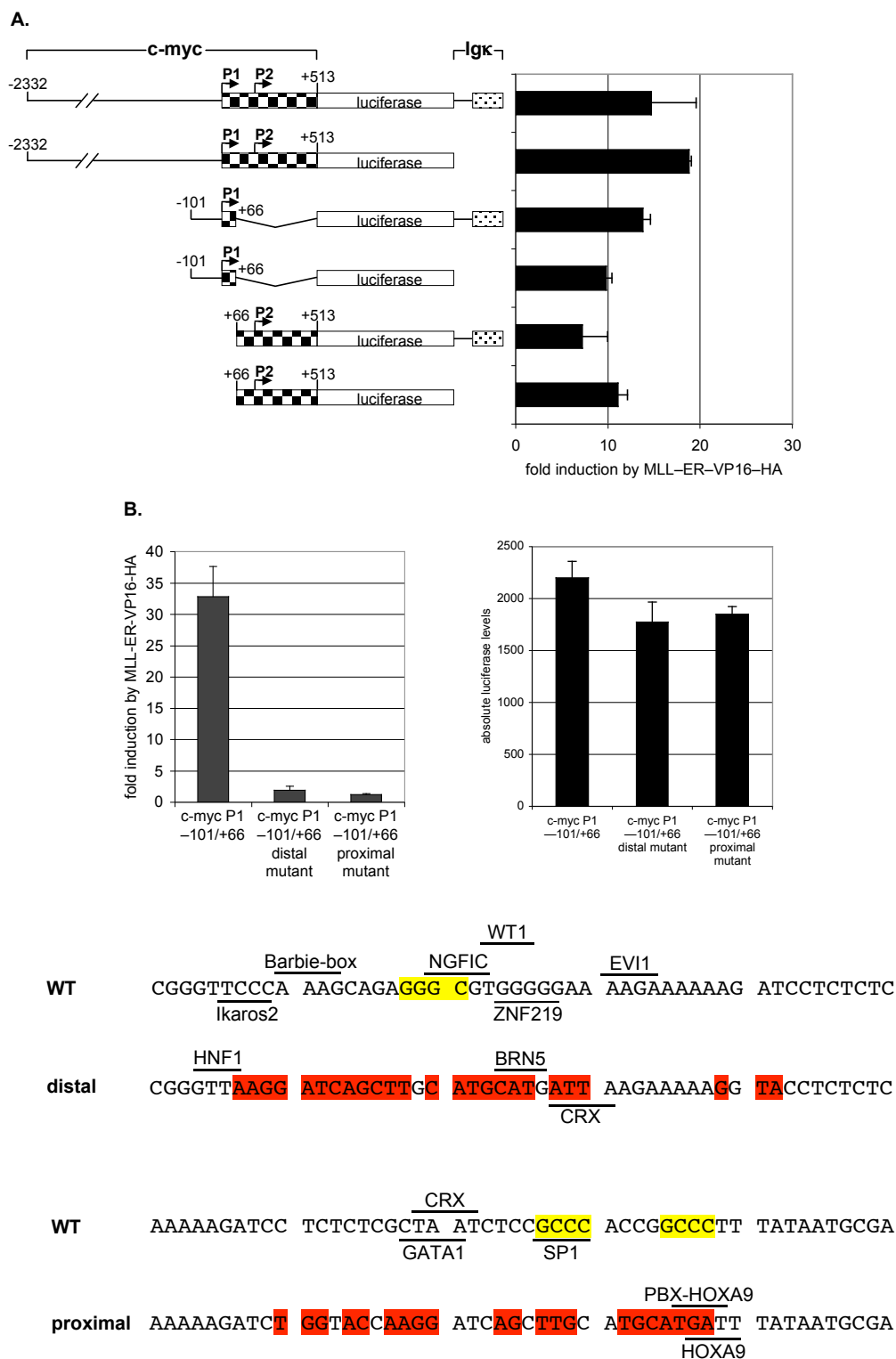


Figure 28. c-myc promoter analysis. A. and B. Deletion and point mutant c-myc luciferase reporter constructs were transfected into FDCPmix cells. The graphs show induction rates of luciferase activity after MLL-VP16-ER-HA activation and basal reporter gene levels, respectively. C. Mutated regions in the c-myc promoter were analyzed for protein binding sites using MatInspector. Nucleotide changes in mutant constructs are indicated in red. GGC motifs are underlined in yellow.

The complete loss of activation in both of these mutants could lead to the assumption that both of the mutated regions contain target sequences through which MLL fusion proteins exert their effects. The two sequences, however, do not display a high degree of homology and bioinformatic analysis could not reveal common binding motifs for transcription factors (Fig. 28C). Noteworthy the proximal mutant contains a binding site for Hoxa9 suggesting that presence of this motif alone is not sufficient for activation by MLL–VP16–ER–HA. Interestingly, a G–rich motif, mutation of which leads to impaired activation of the Hoxa9 promoter (see 3.4.2.1), is also present in the wild type c–myc promoter but not in the mutant constructs (Fig. 28C, yellow boxes). One of these GGGC motifs is contained in a SP1 binding site, which is lost upon mutation in the proximal mutant supporting previous findings where introduction of SP1 sites facilitated activation by MLL–VP16–ER–HA.

The fact that the two regions that were targeted in this mutagenesis are located next to each other (see alignment, appendix 6.1) and the presence of multiple copies of the GGGC motifs might also indicate cooperativity between two elements. This could also explain complete loss of activation as a consequence of changes in only one of the two regions.

3.5.3 p21 as MLL fusion protein target

3.5.3.1 MLL–VP16–ER–HA strongly activates a p21–luciferase reporter gene

It has been published that the cell cycle regulators p18 and p27 are MLL targets (Milne et al. 2005b; Xia et al. 2005). It is tempting to speculate that chromosomal translocation and expression of MLL fusion proteins interfere with normal transcriptional regulation and thereby destabilize cellular proliferation control. This raises the question whether this could be a general mechanism of transformation through MLL fusions and whether also other cell cycle regulators might be targeted. Based on the observation that GC rich DNA stretches mediate activation by MLL–VP16–ER–HA the human p21 promoter was to be tested. The upstream region of the p21 gene is not only rich in GC content but also contains SP1 sites, which had a positive effect on MLL–VP16–ER–HA driven transcription of Hoxa9 and c–myc.

Several p21 promoter luciferase reporter constructs were tested in transfection experiments using inducible FDCP mix cell lines. A 2.3 kb upstream fragment (–2325 / +8) of the human p21 gene led to robust induction of luciferase activity (Fig. 29A). A deletion mutant related to this construct was tested (deletion –122 / –66). The activation of transcription by MLL–VP16–ER–HA dropped dramatically with an observed induction rate that was about 17–fold lower than the one measured using the wild type sequence. A shorter construct (–215 / +8) was tested as well and proved to be sufficient for mediating the response to the MLL fusion protein. In fact the observed induction was almost 10–fold higher than that of the WT construct. This strong activation was also reflected in the absolute luciferase activity. The basal transcription rate of the –215 / +8 reporter was only about one fourth of the one of the long construct (Fig. 29B). Following induction of MLL–VP16–ER–HA, however, absolute luciferase counts were about two–fold higher using the short construct.

Further shortening of this construct by deletion of the region –215 / –144 led to a construct that showed induction rates which were close to the –2325 / +8 values. In comparison to –215 / +8, however, the activation rate by MLL–VP16–ER–HA is only 10 percent, showing that the –215 / –144 region is containing one or more elements that are responsive to the MLL fusion construct. Bioinformatic analysis of the two regions (–215 / –144 and –122 / –60) revealed binding sites for different subsets of transcription factors (Fig. 29C). It is also published that E2F activates p21 through noncanonical binding sites in both regions (Gartel et al. 1998). Noteworthy, a G–rich motif (GGGC), which has been shown to influence regulation of Hoxa9 by MLL–VP16–ER–HA (see 3.4.2.1) is present in both regions. Three copies of this motif were identified in the c–myc P1–promoter, which is strongly activated by MLL–VP16–ER–HA (see 3.5.2). Both mutant variants of this promoter that do not respond to induction of MLL–VP16–ER–HA lack these motifs. Taken together these data indicate a general role for the GGGC motif in MLL fusion protein recruitment.

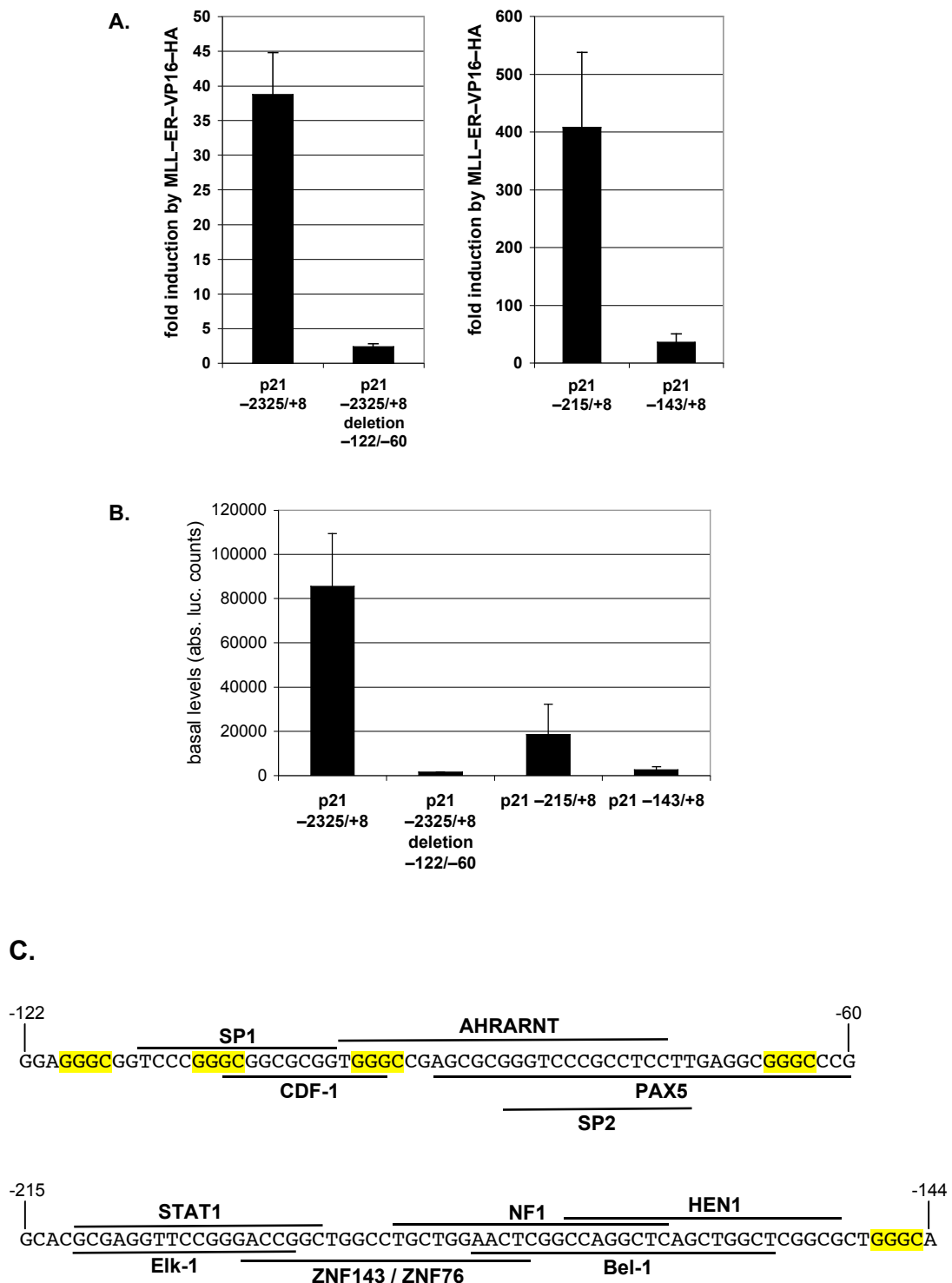


Figure 29. Induction of MLL-VP16-ER-HA in FDCPmix cells leads to strong transcriptional activation of p21 luciferase reporter constructs. FDCPmix cells stably expressing MLL-VP16-ER-HA were transiently transfected with reporter plasmids. 24 hours addition of Tamoxifen luciferase activity was measured. A. Fold induction rates after activation by MLL-VP16-ER-HA. B. Basal reporter gene activity (absolute values). C. Shown are the two regions -122 / -60 and -215 / -144. Transcription factor binding sites are indicated as identified by MatInspector (see appendix 6.3 and www.genomatix.de). GGGC motifs are boxed in yellow.

The p21 $-215 / +8$ reporter plasmid was also transfected into a control cell line expressing VP16–ER–HA where no significant change in luciferase activity could be detected after induction with Tamoxifen (data not shown). In a transient transfection experiment reporter response was compared in two batches transfected with an expression plasmid for either VP16–HA or MLL–VP16–HA. This control excludes an ER / Tamoxifen–dependent artifact. Robust induction was seen again with the MLL fusion construct in comparison to which the background activity due to the expression of VP16–HA was rather marginal (Fig. 30).

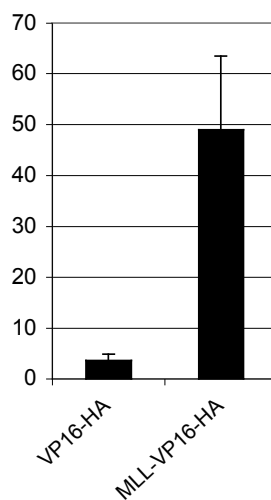


Figure 30. p21 reporter activation by transiently overexpressed MLL–VP16–HA in FDCPmix cells. Expression plasmids for MLL–VP16–HA as well as for the VP16–HA control construct were transiently overexpressed in FDCPmix and transcription from a $-215 / +8$ p21 reporter was measured by luciferase assay. The graph is showing the induction rates.

3.5.3.2 p21 promoter sequences are activated by MLL and MLL fusion proteins

One interesting question concerning MLL leukemogenesis is whether MLL and MLL fusion proteins are targeted to the same promoters. To address this question for the human p21 promoter expression plasmids for MLL, MLL–AF4 and MLL–AF9 were co-transfected with the p21 –215 / +8 reporter plasmid. Expression of each of the three constructs led to significant induction of p21 luciferase reporter activity (Fig. 31A). MLL–AF9 had the strongest effect. However, the induction rate (>12-fold) was lower than in the case of MLL–VP16–HA (>40-fold, Fig. 30). Expression of the full-length MLL protein was driven from a different type of expression vector (see 2.8), which may be responsible for the milder effect as compared with MLL fusion proteins. However, these data show that both MLL fusion and full-length protein affect the p21 promoter in a transient analysis.

HepG2, a hepatic cell line frequently used for analyses of p21 regulation was employed to verify these results within a different cellular background. Three different p21 reporter gene constructs were co-transfected together with expression plasmids coding for FLAG–MLL or for MLL–VP16–HA, respectively. While the two proteins performed almost identically in co-transfections with p21 –2325 / +8, MLL–VP16–HA had a much stronger effect on the –215 / +8 construct (Fig. 31B). FLAG–MLL in contrast led to higher induction rates than the fusion protein when used together with the full-length construct harboring the –122 / –60 deletion.

These findings suggest that both the fusion and the full-length protein to the promoter through the MLL aminoterminal. The differential effects as seen for the –215 / +8 promoter construct might indicate, however, the differences in promoter complexes that are formed upon recruitment of MLL or MLL–VP16–ER–HA, respectively. The activation domain present in the full-length MLL protein might contact other factors than the VP16 activation domain. Furthermore MLL–VP16–ER–HA lacks the C-terminal SET-domain, which might as well recruit factors that are important for modulation of the activating function of MLL.

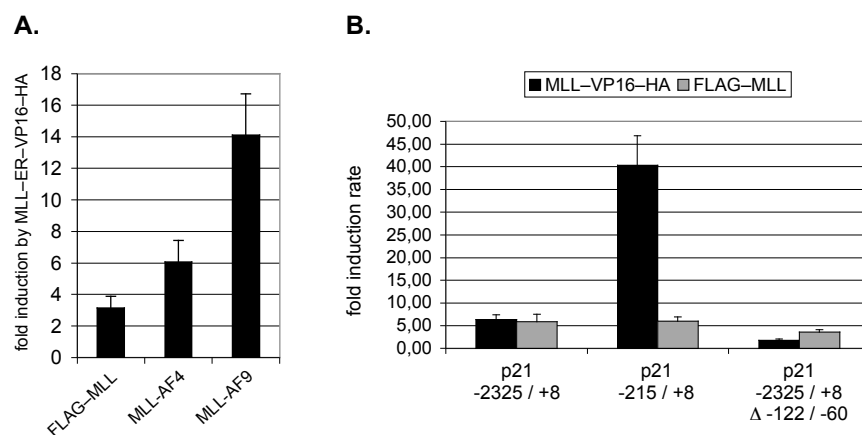


Figure 31. Full-length MLL as well as leukemic MLL fusion proteins activate p21 promoter sequences. A. Transcriptional activation of p21 $-215 / +8$ as shown in a transient transfection experiment in FDCPmix. B. MLL and MLL-VP16-HA were compared in their capacity to activate different regions of the p21 promoter. Reporter and expression constructs were transfected transiently into HepG2 cells.

3.5.3.3 SP1 overexpression does not affect p21 activation by MLL-VP16-ER-HA

Hoxa9, c-myc and p21 promoters all contain binding sites for SP1. However Gal-SP1 failed to recruit MLL-VP16-ER-HA to the Hoxa9 promoter (see 3.4.2.2). Using the p21 promoter this finding was to be controlled in a transient transfection experiment. U937 cells stably expressing MLL-VP16-ER-HA were transiently transfected with either one of two different p21 reporter plasmid variants and an expression plasmid for SP1 protein. The cells were incubated for 17 hours before induction of MLL-VP16-ER-HA activity via addition of Tamoxifen. Eventually luciferase levels were measured 24 hours post induction. No significant change in activation could be seen, however, depending on SP1 expression (Fig. 32). Taken together with results from the Hoxa9 and the c-myc promoter analyses this finding supports a model according to which MLL-VP16-ER-HA is recruited to SP1 sites in a SP1 protein independent manner.

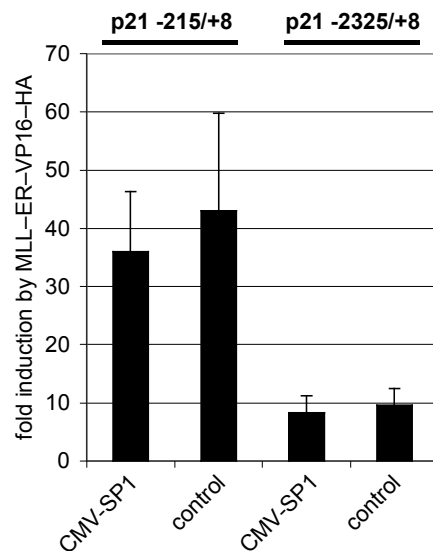


Figure 32. SP1-overexpression does not enhance MLL-VP16-ER-HA activated transcription of an episomal p21 reporter construct. U937 cells stably expressing MLL-VP16-ER-HA were co-transfected with an expression construct for SP1 and two different variants of p21 reporter plasmids. MLL-VP16-ER-HA activity was induced by Tamoxifen and luciferase expression was determined.

3.5.3.4 Expression of endogenous p21 mRNA after induction of MLL-VP16-ER-HA activity

Given that all p21 constructs were tested in transient reporter assays the obvious question was whether changes in transcriptional activity of the chromosomal p21 locus could be observed as a consequence of MLL-VP16-ER-HA induction. In order to determine p21 mRNA levels RT-PCR was performed. RNA was isolated from different cell lines and quantitative PCR technology was used for amplification of cDNA sequences. In the case of MLL-VP16-ER-HA FDCP mix cells only a minor increase of p21 mRNA synthesis could be detected following induction of the MLL fusion construct (Fig. 33) in comparison with a control cell line expressing VP16-ER-HA. A similar picture was obtained in U937 cells transgenic for MLL-VP16-ER-HA. Here the concentration of p21 mRNA increased by about 20 percent after induction (Fig. 33).

One explanation for the lack of activation of endogenous p21 by MLL-VP16-ER-HA could be heterochromatin formation on the chromosomal site, which would render the promoter inaccessible. RT-PCR results, however, showed that the endogenous p21 gene is not transcriptionally silenced as indicated by relatively low ct values (Fig. 33,

table). This led to the idea that constitutive transcription (possibly driven by endogenous MLL), which is taking place without induction of MLL fusion protein activity, is high that the MLL–VP16–ER–HA effect is not detectable in this situation. To circumvent this scenario mouse embryonic fibroblasts (MEFs) harbouring a homozygous MLL deletion were analyzed. Stable cell lines were created that expressed inducible MLL–VP16–ER–HA or VP16–ER–HA, respectively, in a MLL – / – background. These cells do not express any endogenous MLL protein, hence MLL dependent transcriptional activation is only expected after induction of MLL–VP16–ER–HA. However, no striking difference as compared to a wild type background could be observed. Induction of VP16–ER–HA led to a slight downregulation, while p21 mRNA amounts increased marginally after induction of MLL–VP16–ER–HA (Fig. 33). This argues against the hypothesis that endogenous MLL–driven transcription of the gene is responsible for the lack of induction by the fusion protein. These results underline the fact that the chromosomal gene obviously underlies an additional level of regulation, which does not apply to the reporter plasmids.

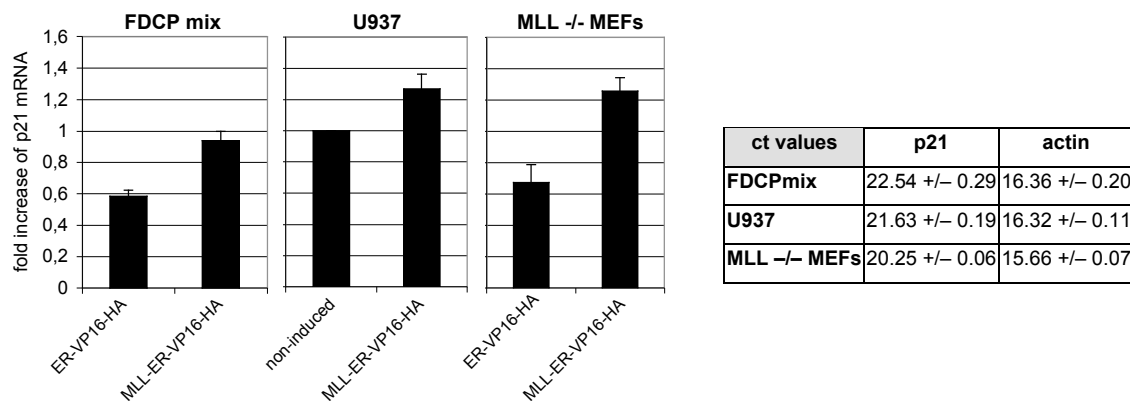


Figure 33. Expression of endogenous p21 measured by RT–PCR. Different cell lines all stably expressing MLL–VP16–ER–HA or ER–VP16–HA were induced with Tamoxifen, mRNA was isolated and p21 expression was analyzed using gene–specific primers for Real–Time–PCR. The table shows ct values of the Real–Time–PCR reactions of the non–induced samples. Values were normalized to β –actin levels and induction rates were calculated, which are shown in the diagrams.

3.5.3.5 MLL–VP16–ER–HA activated transcription of episomally stable p21 reporters

The striking difference between the effect of MLL–VP16–ER–HA in transient reporter assays and the transcriptional regulation of the endogenous p21 gene was the starting point for another experiment. It is known that EBV–based vector DNA that is maintained stably in mammalian cell lines is bound by histones and nucleosomes are formed (Zhou et al. 2005). These vectors can therefore be considered an intermediate between transient DNA and chromosomal loci and provide for a genetic tool in mammalian cells that can be used to study chromatin–dependent processes. Both the longer and the shorter versions of the human p21 promoter have been cloned into an EBV–based luciferase vector. U937 cells stably expressing MLL–VP16–ER–HA were transfected with these constructs and luciferase levels were measured after selection of stable cells. The shorter –215 / +8 construct was induced approximately 30–fold, while the longer –2325 / +8 construct was induced about 75–fold (Fig. 34). This is in striking contrast with the results of the transient analysis where the shorter construct showed a 10–fold higher induction rate than the –2325 / +8 construct (Fig. 29). Furthermore the overall induction levels observed in the transient analysis were much higher. It is tempting to speculate that the reduced overall induction rates of the episomal reporters as compared to the transient analysis are a consequence of chromatin formation. Obviously the p21 promoter constructs contained in the episomal constructs are subject to an additional level of transcriptional control that leads to lower expression rates. This might reflect the intermediate state of the EBV–based constructs between transient reporter plasmids, which are dramatically induced, on the one hand and the chromosomal gene, whose transcription is hardly upregulated on the other hand. Considering the role of the full–length MLL protein as transcription maintenance factor it is intriguing that gene activation through MLL–VP16–ER–HA is controlled on the epigenetic level.

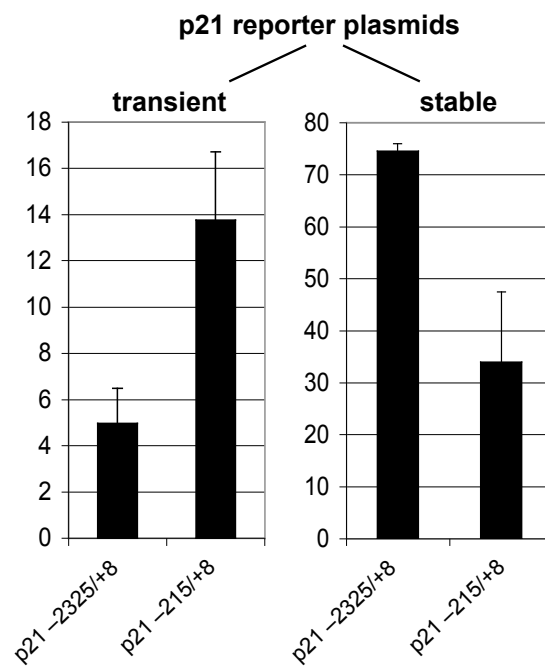


Figure 34. MLL–VP16–ER–HA is activating the p21 promoter on episomally stable reporter vectors. EBV–based reporter plasmids containing p21 promoter sequences were used for transient reporter assays as well as for generation of cell batches stably propagating these reporters. Luciferase levels were measured after induction of MLL–VP16–ER–HA activity.

3.5.3.6 Influence of DNA methylation on activation of p21 transcription

To further explore the epigenetic control of activation by MLL–VP16–ER–HA it should be tested whether DNA methylation might have an effect on this process. DNA methylation presents a major cellular pathway for gene silencing. The MT domain present in the N–terminus of MLL in particular suggests specificity for binding to DNA depending on its methylation status (Birke et al. 2002). Drugs are available that block DNA methylation. Deoxy–azacytidine, a nucleotide analog that is incorporated into DNA during replication and that, unlike cytosine, cannot be methylated, was applied to test whether DNA methylation affects p21 activation by MLL–VP16–ER–HA.

U937 cells stably expressing MLL–VP16–ER–HA and propagating the EBV reporter plasmid containing the p21 –215 / +8 fragment were cultured in presence of Azacytidine for 72 hours before induction of MLL–VP16–ER–HA activity by addition of Tamoxifen. Luciferase levels were measured and RNA was prepared for analysis of transcriptional activity via RT–PCR. As expected the basal transcription rate of both p21 and luciferase increased following Deoxy–azacytidine treatment (Fig. 35A).

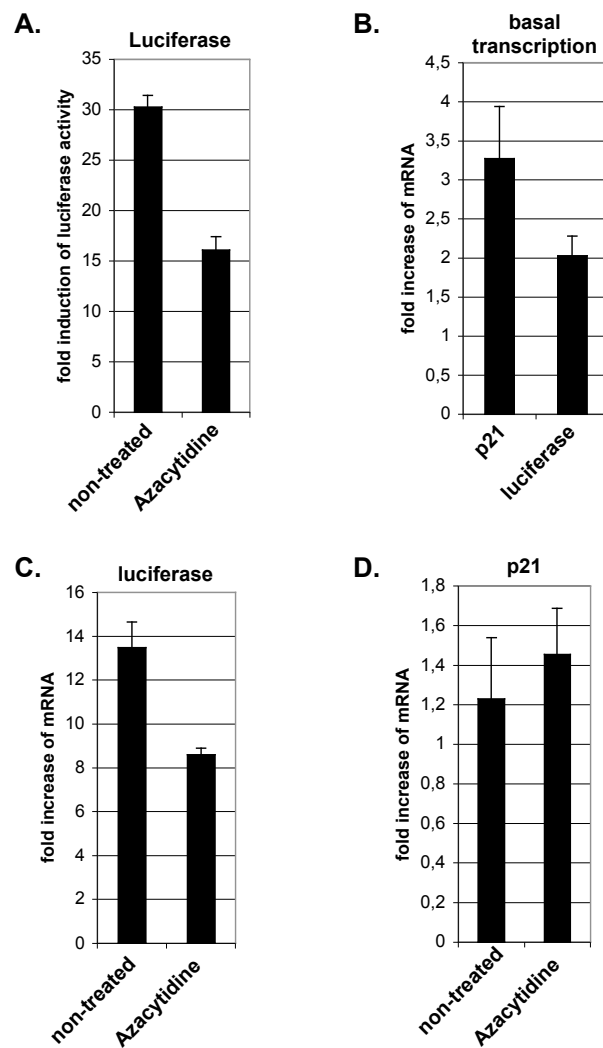


Figure 35. MLL-VP16-ER-HA dependent transcription of p21 after inhibition of DNA methylation. U937 cells stably expressing MLL-VP16-ER-HA and episomally replicating a reporter construct containing p21 -215 / +8 were treated with Deoxy-azacytidine for 72 hours before induction with Tamoxifen. A. RT-PCR analysis of the influence of Deoxy-azacytidine on basal expression levels of p21 and luciferase. B. Luciferase levels measured in Tamoxifen-induced cells that were cultured in presence or absence of Deoxy-azacytidine. C. and D. RT-PCR analysis of the effect of MLL-VP16-ER-HA on mRNA levels of luciferase and p21, respectively. The graphs are showing the fold increase of mRNA after induction of MLL-VP16-ER-HA activity.

The luciferase induction rate, however, was decreased (Fig. 35B), while absolute levels increased. Also luciferase mRNA induction levels were reduced (Fig. 35C). The impairment of MLL-VP16-ER-HA activation of the reporter gene is unexpected since the MLL MT-domain is thought to bind CpG dinucleotides with a preference for non-methylated DNA, which would supposedly lead to higher induction rates. One explanation might be impaired binding of MLL-VP16-ER-HA to Deoxy-azacytidine containing DNA as compared to cytidine-containing DNA.

p21 induction rates were very low as observed earlier and not significantly increased after treatment of the cells with Deoxy-azacytidine (Fig. 35D) suggesting that methylation alone of the endogenous p21 is not sufficient to block activation through MLL-VP16-ER-HA. Considering the close link of DNA methylation and other silencing mechanisms like histone deacetylation it might be possible that more than one regulatory mechanism render the p21 gene non-responsive to MLL-VP16-ER-HA.

3.5.4 Overview: results from reporter gene experiments

In an effort to summarize the results from the analyses of three different target gene promoters described in this work the different promoter variants were compared with the wild type sequences using a bioinformatic tool called MatInspector. Based on an archive of binding matrices (appendix 6.3) this program assigns all known transcription factors to a given input sequence. By comparison of luciferase induction levels the mutants were rated as "gain of function" or "loss of function" mutants, respectively. Any binding site that was introduced in a gain of function mutant was assumed to have a positive regulatory role for activation by MLL-VP16-ER-HA (orange boxes) and binding sites lost in the same mutant in comparison with the WT sequence were assumed to have a negative function (green boxes, Table 10). In the same way binding sites of the loss of function mutants were treated: sites only present in the WT sequence were rated as positive regulatory elements (orange boxes), sites only present in the mutant sequences as negative regulatory elements (green boxes). This classification per se is simplifying and therefore any picture emerging from this analysis has to be considered only in context of the corresponding experimental evidence. It can be used, however, as a preliminary approximation.

V\$RREB/RREB1.01	Orange																		
V\$RXRF/VDR_RXR.03									Orange										
V\$RXRF/VDR_RXR.05									Orange										
V\$SATB/SATB1.01										Orange									
V\$SORY/SOX5.01								Orange											
V\$SORY/SOX9.01																		Green	
V\$SP1F/GC.01	Orange								Orange										
V\$SP1F/SP1.01	Orange		Orange						Orange										
V\$SP1F/SP1.02								Orange											Orange
V\$SP1F/SP2.01								Orange											
V\$STAF/ZNF76_143.01								Orange											
V\$STAT/STAT1.01								Orange											
V\$ZBPF/ZBP89.01																			Orange
V\$ZBPF/ZF9.01	Orange		Orange																Orange
V\$ZBPF/ZNF219.01									Orange					Orange					Orange
V\$ZF5F/ZF5.01									Green					Green					Orange

Matrices that have been rated "activating" (orange) in at least three mutant constructs have been highlighted in blue. These matrices comprise candidate binding sites for MLL–VP16–ER–HA (Table 11). With the exception of E2F.01 there is a clear bias for GC rich matrices. Most of the sites are characterized by the presence of either a GGG or a GGGG motif confirming the earlier findings that G-rich motifs mediate activation by MLL–VP16–ER–HA.

Alternatively the proteins binding to these sequences can be considered candidates for interactors or recruitment factors, respectively, of MLL fusions. In this respect it seems to be intriguing that most of the proteins in this list are zinc finger proteins raising the possibility MLL proteins generally contact DNA through zinc finger proteins.

Table 11. Candidate matrices for MLL–ER–VP16–HA action.

Matrix ¹	Factor	IUPAC consensus sequence ²
V\$E2FF/E2F.01	E2F, p107 protein	gcgcGAAAa
V\$EGRF/WT1.01	Wilms Tumor Suppressor	gYGGGgg
V\$EKLF/BKLF.01	Basic krueppel-like factor (KLF3)	GGGTg
V\$EKLF/KKLF.01	Kidney-enriched kruppel-like factor, KLF15	GGGGmg
V\$MAZF/MAZR.01	MYC-associated zinc finger protein related transcription factor	gggGGGG
V\$PLAG/PLAG1.01	Pleomorphic adenoma gene (PLAG) 1	GRGGsncnnnnnrggg
V\$SP1F/SP1.01	Stimulating protein 1	GGGCggg
V\$ZBPF/ZF9.01	Core promoter-binding protein (CPBP) with 3 Krueppel-type zinc fingers	CCRCccc
V\$ZBPF/ZNF219.01	Kruppel-like zinc finger protein 219	CCCCc

1: for more detail see www.genomatix.de

2: according to MatInspector. Core matrices are indicated by capital letters. G–rich sequences present in this core are shaded in red.

3.6 Microarray analysis of MLL–VP16–ER–HA target genes

A genome-wide expression analysis was carried out using microarray technology. The Affymetrix GeneChip 430 2.0 was used, representing more than 39.000 transcripts from the mouse genome. For detection of target gene populations of physiological relevance the FDCP mix cell line stably expressing MLL–VP16–ER–HA was chosen for the experiment. Total RNA was isolated 30 minutes, 3 hours and 12 hours after induction of the fusion protein. RT–PCR analysis was performed using Meis 1–specific primer pairs in order to check for the response of a known physiological target of MLL fusion proteins. After 3 hours Meis 1 mRNA amounts had already increased 3–fold (Fig. 24). After 12 hours the mRNA concentration was still at a higher level than before induction. Interestingly, in the case of the control cell line (VP16–ER–HA) a even faster kinetic could be observed: after 3 hours Meis 1 was already downregulated. It has been reported many times that overexpression of transcriptional activators lead to a phenomenon called "squenching": limiting factors are titrated out by the large amount of activation domains and the general transcription rate is reduced. The strong downregulation after 30 minutes of

induction, however, is illustrating nicely the fast response of the ER-system. It also suggests that Meis 1 mRNA molecules have a rather short half-life. In contrast to the MLL-VP16-ER-HA cell line mRNA levels were back to pre-induction levels after 3 hours. Together these results show that both protein variants can be activated quickly with a certain variation possibly depending on size differences and higher expression levels. The fact that a known target gene is activated by induction of MLL-VP16-ER-HA but not by another ER fusion protein is providing evidence for the physiological relevance of the experimental set-up.

The numbers of deregulated genes suggest that MLL-VP16-ER-HA after induction exerts mainly an activating effect (Table 12). Even after 12 hours the expression of only 57 genes was reported to be downregulated as opposed to 223 upregulated genes. At the earlier time points this becomes even more obvious. For complete lists of genes see appendix 6.2.

Table 12. Total number of deregulated genes following induction of MLL-ER-VP16-HA.

	30 min	3 hours	12 hours
upregulated	23	112	223
downregulated	1	6	57

One way to analyze microarray data is to classify the target genes according to Gene Ontology terms. There were some biological processes whose representation changed over the course of induction. The numbers of upregulated genes involved in morphogenesis, organ development and cell communication increased from 3 to 12 hours after induction while the number of genes with negative regulatory functions decreased in comparison with the total number of upregulated genes (Fig. 36 A). However, it should be noted that the absolute number of negative regulators for example does not change and most of those, which are upregulated after 3 hours are also upregulated after 12 hours, e. g. the cell cycle regulators p27 and p18. The observed upregulation of these kinases is in one line with earlier reports that these genes are targeted by MLL.

Target genes can also be sorted according to the molecular function which is assigned by Gene Ontology. An upregulation of proteins with nucleic acid binding capacity was observed as well as an transient increase in expression of

transcriptional activators after 3 hours of induction (Fig. 36 B). This result might indicate the upregulation of activators and DNA binding proteins through MLL–VP16–ER–HA, which as a consequence could lead to subsequent reprogramming of transcriptional activity which has to be considered a secondary effect. Especially in the 12 hours sample a considerable number of genes might not be subject to direct regulation through MLL fusion proteins.

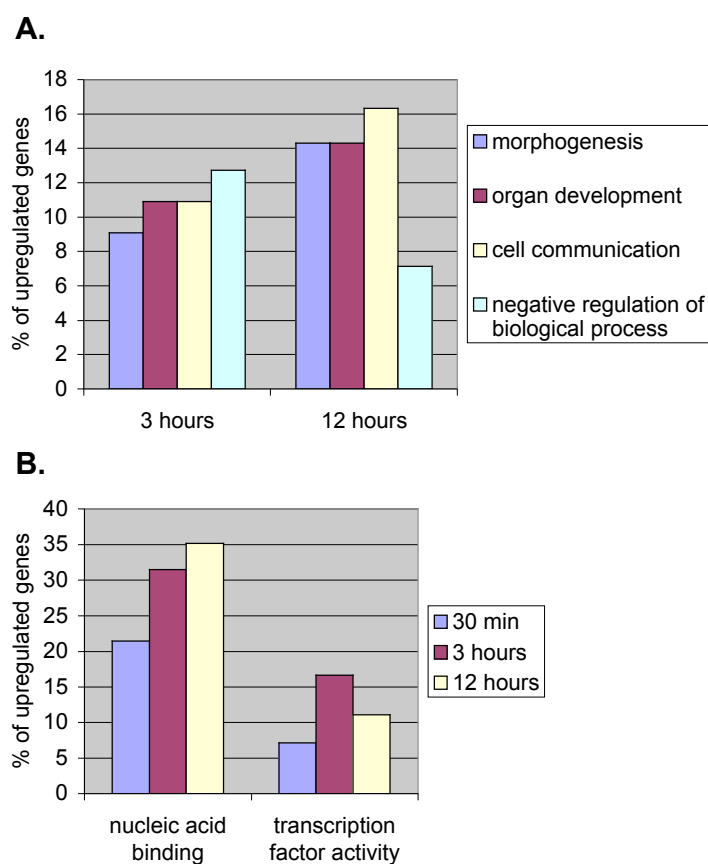


Figure 36. Classification of microarray targets according to Gene Ontology terms. The graphs are showing the relative numbers of certain classes at a given time point. A. Genes grouped due to assignment to certain biological processes. B. Classification according to molecular functions assigned to the genes by the Gene Ontology system.

In an independent microarray experiment analysis of gene expression patterns 23 hours after induction revealed downregulation of several genes, which are related to myeloid differentiation (Table 13). This is in line with the general property of oncogenes to support proliferative processes and counteract final differentiation of stem cells.

Table 13. Genes that are downregulated 23 hours after MLL–ER–VP16–HA induction

Mpo	myeloperoxidase
MMP8	neutrophil collagenase precursor matrix metalloproteinase 8
Ngp	neutrophilic granule protein
Lcn2	Neutrophil gelatinase-associated lipocalin precursor (NGAL) (P25) lipocalin 2
EL2	neutrophil elastase
Ly6c	lymphocyte antigen 6 complex, locus C
IL3R; IL3RB2	Interleukin-3 receptor class II beta chain precursor
Mbp2	eosinophil major basic protein 2

Upregulated gene populations were analyzed using Bibliosphere (www.genomatix.de), a bioinformatic tool developed recently to facilitate literature-based array data-mining. This web-based search engine is sorting genes according to co-citation in the NCBI database. Analysing the list for genes upregulated after 0.5 hours a correlation was detected with four transcription factors, one of which is c-jun. It is intriguing that out of 15 analysed target genes 6 were co-cited together with c-jun and analysis of their promoter regions also revealed binding sites for the protein (Table 14). Using these data a network was generated proposing a central role for c-jun (Fig. 37A). Analysis of later time points after induction of MLL-VP16-ER-HA led to a much higher complexity of the proposed regulatory networks (Fig. 37B,C).

Table 14. Microarray target genes that were co-cited together with c-Jun. Presence of a binding site for c-Jun or AP1, respectively, is indicated by shaded boxes and +.

Gene	Jun / AP1 site
Tox	
Ptpn12	+
D230007K08Rik	+
Pfn3	
Mbnl1	
Gpbp1	
Krit1	
Zfx1b	+
Ern1	+
Nedd4l	
Dgkd	
Myo1b	
Pacs1	+
Clpx	
Exoc4	+

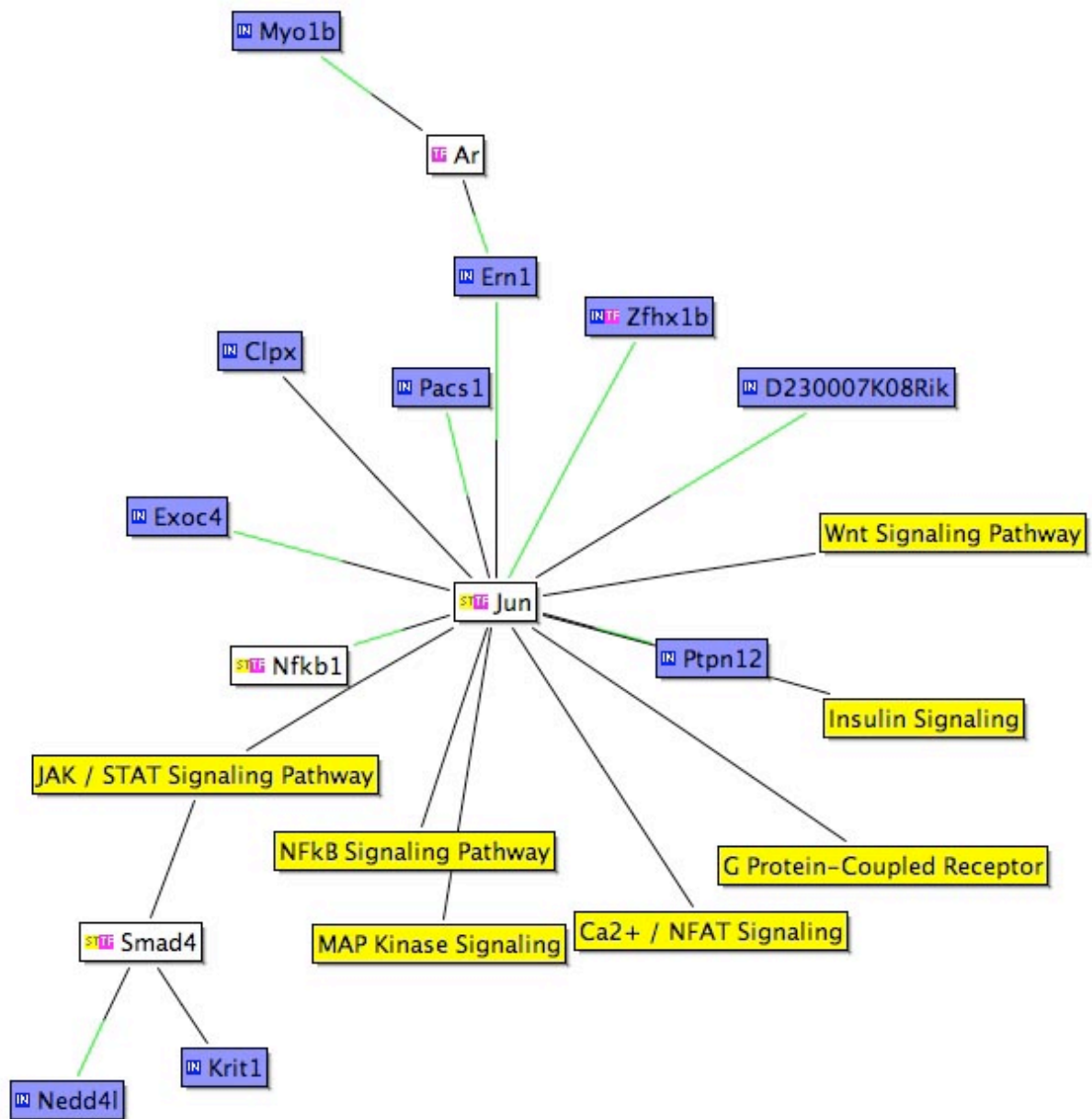


Figure 37. Predicted regulatory network based on microarray data leading to upregulation of genes whose expression increased following MLL-ER-VP16-HA induction. A. Genes upregulated after 30 min. B. Genes upregulated after 3 hours. C. Genes upregulated after 12 hours. Blue boxes: genes identified via microarray. White boxes: transcription factors that are co-cited together with target genes. Green lines: positive regulatory effects.

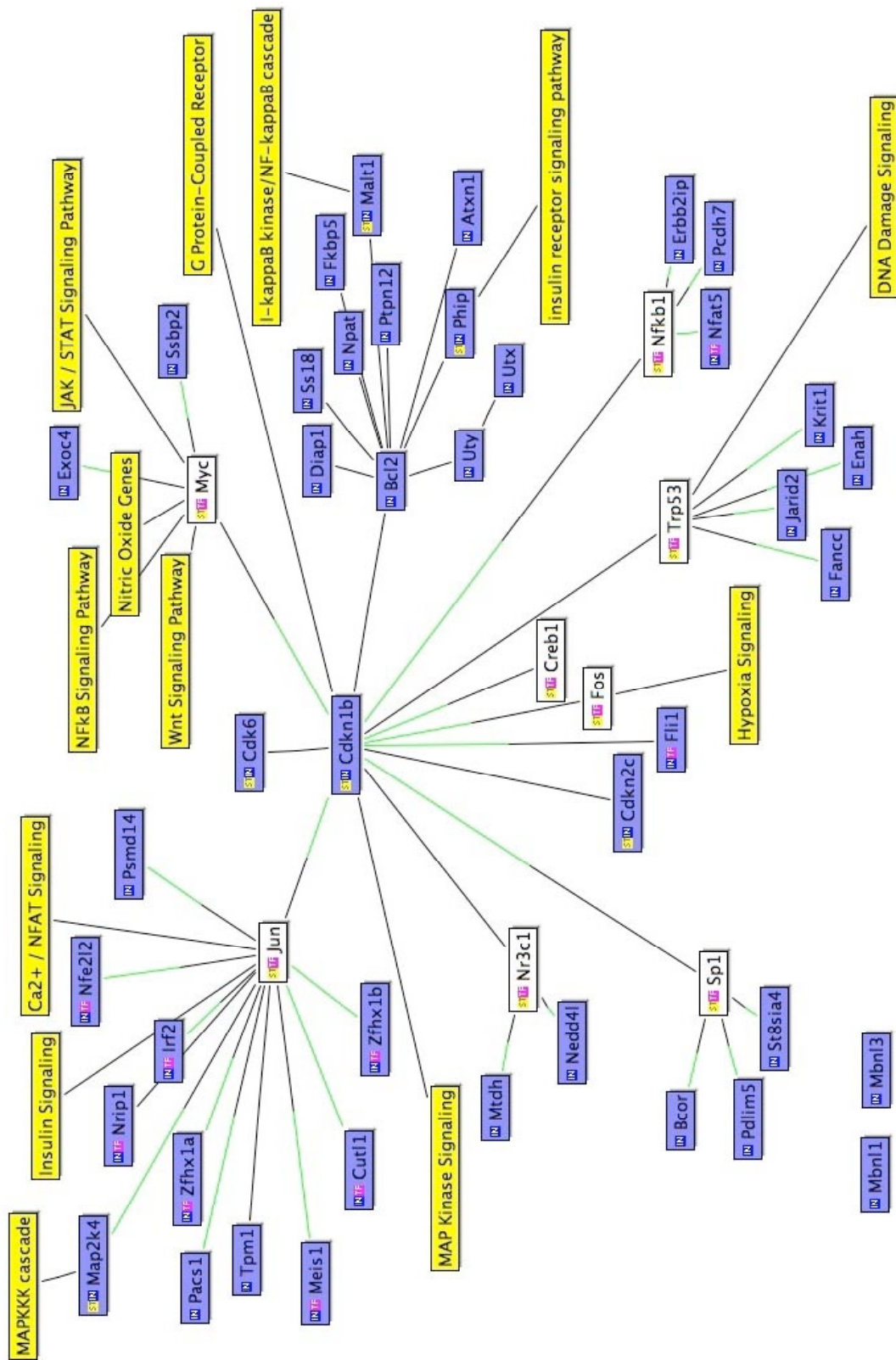


Figure 37B.

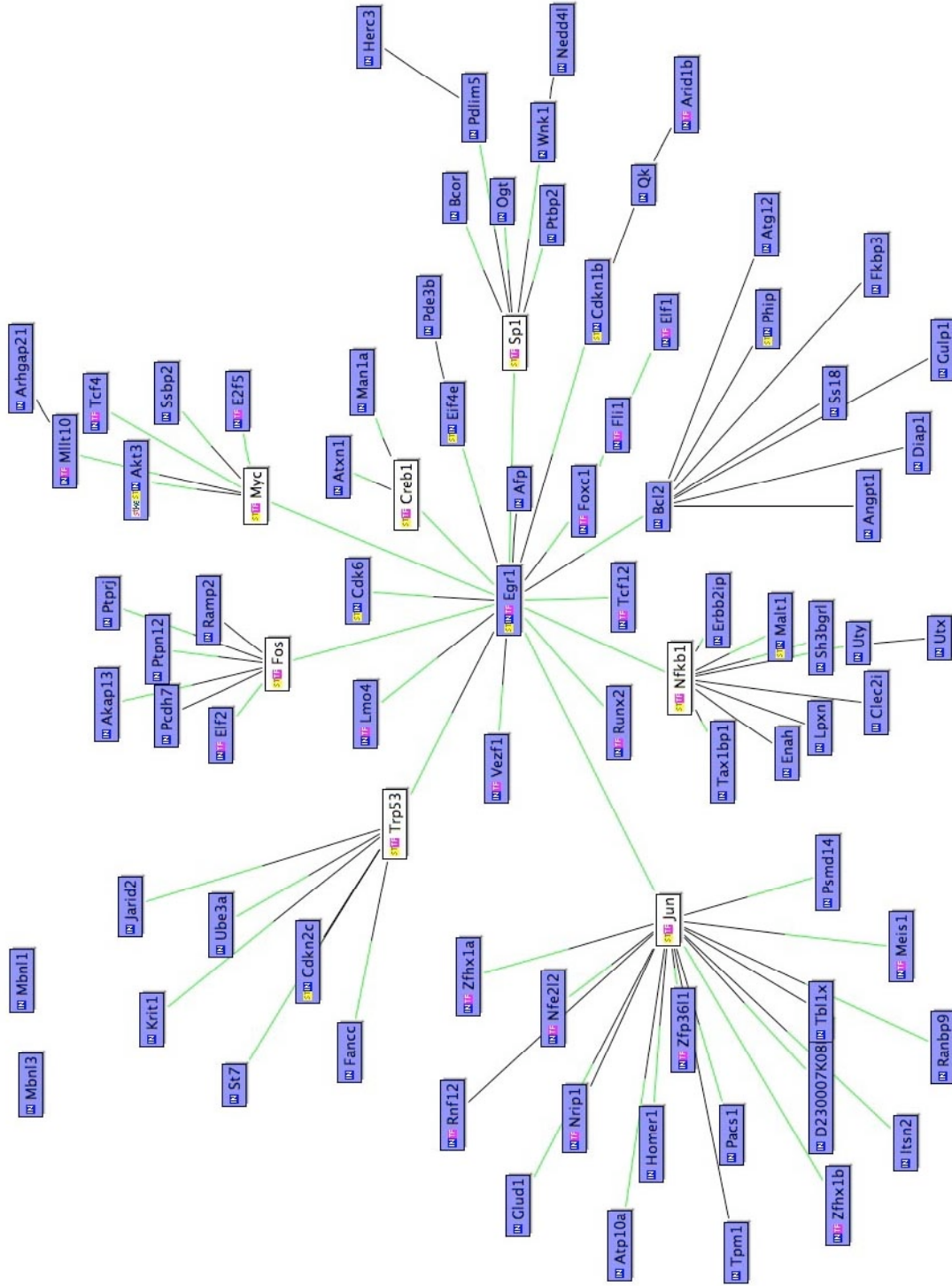


Figure 37C.

A more direct bioinformatic approach is to analyze promoter sequences with respect to binding matrices for transcription factors. Using the Gene2Promoter software the lists of upregulated genes were analyzed in this way. Interestingly matrices that were statistically over-represented in the 30 min sample as well as in the 3 hours sample were the matrix families for cut-like homeobox (CLOX) and for homeobox (HOXF) factors, the latter of which could also be found in the 12 hours sample (Table 15). The same type of analysis was performed using three known targets of MLL-VP16-ER-HA, Hoxa9, p21 and c-myc, as an input. It was intriguing to find amongst the matrices that were identified to be statistically over-represented also the CLOX and HOXF matrices. The CLOX matrix family comprises several binding sites that are recognized by the Cut-like homeodomain protein Cutl1. Cutl1 codes for a transcriptional repressor that binds to CCAAT motifs in eukaryotic promoters negatively regulating gene expression. The fact that Cutl1 binding sites are present in the majority of genes upregulated by MLL-VP16-ER-HA could indicate that these motifs are also targeted by MLL fusion proteins. In this case they would compete for these binding site with a transcriptional repressor leading to upregulation of the gene. The HOXF matrix family contains binding sites for several homeobox proteins amongst which notably is Hoxa9. The presence of these sites in genes upregulated after induction of MLL-VP16-ER-HA activity could indicate either direct targeting of these genes by the MLL fusions or by Hoxa9, which is itself targeted by MLL fusions.

Table 15. Predicted matrix families common to target gene promoters

matrix family ¹ :	microarray 30 min	microarray 3 hours	microarray 12 hours	Hoxa9 p21 c-myc
AP1R				
CLOX				
CREB				
E2FF				
EGRF				
EKLF				
FKHD				
GATA				
HOXF				
INSM				
MZF1				
NKXH				
NR2F				
PAX5				
SP1F				
ZBPF				

1: as identified by Gene2Promoter (www.genomatix.de)

4. Discussion

4.1 Definition of critical regions in the Hoxa9 promoter

Hox genes are the paradigm for MLL target genes. Early mouse knockout studies indicated a role for MLL in the transcriptional regulation of the developmental gene Hoxa9 (Yu et al. 1995). Over the last decade a body of data was collected showing that Hoxa9 upregulation by MLL fusion proteins is a key feature of 11q23 leukemias (Kawagoe et al. 1999; Rozovskaia et al. 2001; Armstrong et al. 2002; Imamura et al. 2002; Ferrando et al. 2003; Zeisig et al. 2004b). Using ChIP technology physical association of the MLL protein with the Hoxa9 upstream region could be demonstrated (Milne et al. 2002; Nakamura et al. 2002).

However, the mechanism of MLL recruitment to this promoter is largely unknown. No functional data were available so far as to which upstream sequences are required for activation by MLL and MLL fusions. MLL target sites had not been identified and no detailed information was available up to this point as to what are the elements in the Hoxa9 promoter sequence that mediate the response to MLL proteins. In this study the effect of an inducible MLL fusion protein on Hoxa9 reporter constructs was investigated. The fusion with the viral activation domain VP16 was reported to be functional as an oncogene (So and Cleary 2003; Zeisig et al. 2003). In concordance with this finding the inducible MLL–VP16–ER–HA protein used in this study not only strongly activated MLL target gene reporters (e.g. Hoxa9, Hoxa7) but also upregulated endogenous target genes as assessed by RT–PCR (Hoxa9, Meis1).

Episomally stable reporter cell lines allowed the analysis of reporter gene effects on a more physiological level than possible in transient transfection assays. The stable propagation has some consequences that render results more biologically relevant:

1. The transfected constructs are replicated by the host cell DNA replication machinery and therefore chromatin formation takes place like on mammalian chromosomes making these vectors more similar to a genomic template (Zhou et al. 2005).
2. Cells can be selected that propagate the EBV–based reporter plasmid. Since also the inducible MLL fusion protein is expressed stably in the same cells analysis of a homogenous cell population is possible. The copy numbers of these episomes range between 5 and 20 which also resembles the situation of a chromosomal target gene

much more than a transiently transfected reporter plasmid, which is usually present at much higher copy numbers (Mackey and Sugden 1999).

ChIP experiments employing monoclonal antibodies raised against the MLL aminoterminal showed induced binding of MLL–VP16–ER–HA not only to the endogenous *Hoxa9* promoter but also to the episomal reporter plasmid.

Using a series of different *Hoxa9* promoter fragments it was possible to identify a genomic region of only 164 bp (–118 / +46) as being critical for the induction by MLL–VP16–ER–HA. Fine-mapping revealed that the major activation capacity resides in a short sequence stretch of only 40 nucleotides (–118 / –79). Deletion of this region decreased dramatically the response to the MLL fusion protein indicating that MLL fusion protein binding is taking place in this region.

4.2 MLL–VP16–ER–HA acts on the core promoter region of *Hoxa9*

The shortened construct (–78 / +46), however, retains some capacity to mediate the response to MLL–VP16–ER–HA. This might hint at MLL fusion protein recruitment to this fragment as well, even if only to a lesser extent, raising the question whether MLL fusion proteins act at core promoter regions. Recently a role for MLL as a general transcription factor has been proposed since ChIP–chip studies showed co-occupancy with RNA polymerase II on the majority of gene promoters (Guenther et al. 2005). In fact both the aminoterminal and the carboxyterminal moiety of MLL interact with the CTD of RNA polymerase II in vitro (Milne et al. 2005a). On some genes MLL was seen to be required for RNA polymerase II recruitment, as examined by single gene ChIP analysis, but this does not hold true on a global scale (Milne et al. 2005a). There are genes where binding of MLL seems to be a prerequisite for RNA polymerase II binding and transcription but other genes are transcribed independently of MLL binding.

In this study site-directed mutagenesis of core promoter features revealed the involvement of *Hoxa9* core promoter sequences in MLL fusion protein recruitment. It was demonstrated that upmutation of the endogenous *Hoxa9* TATA box to the consensus sequence as present in the AdML promoter augments transcription suggesting that MLL fusion proteins are recruited to the core promoter region through

proteins specifically binding to TATA-box motifs. TFIIA, which associates with TBP-TATA box complexes is a good candidate since it has been reported that TFIIA is processed by Taspase (Zhou et al. 2006), the same protease that is also involved in MLL protein processing (Hsieh et al. 2003a). It could be speculated that Taspase establishes a contact between TFIIA and MLL. This contact could stabilize binding of MLL proteins to upstream activating sequences and facilitate transcriptional activation (Fig. 38A). Alternatively, the interaction of MLL with the CTD of RNA polymerase may be the link between motifs in the core promoter and MLL fusion protein driven activation of transcription.

It is not the case, however, that core promoter mutations lead to induction rate changes that mirror the effect of MLL-VP16-ER-HA on basal transcription. This would be expected for a model where impaired MLL function is a direct consequence of impaired binding of general transcription factors as described above. The differential effects of core promoter mutations on basal and MLL-VP16-ER-HA driven transcription favor a model in which MLL fusion protein binding to Hoxa9 promoter sequences is taking place independent of components of the basal machinery. Particularly intriguing are the effects observed with mutants where a cytosine heptarepeat was targeted that is located between TATA box and Inr. No general transcription factors are known that specifically bind to this sequence element, nevertheless basal transcription levels are significantly decreased. The same mutation is also affecting activated transcription induced by MLL-VP16-ER-HA. Based on these observations a model can be envisaged where both the wild type and MLL fusion proteins bind to this motif in a competitive manner. In normal hematopoiesis full-length MLL might bind this sequence and modulate Hoxa9 expression leading to maintenance of transcription on a low level. In the heterozygous state MLL fusion proteins might compete with wild type protein for this motif. Given that these two MLL protein variants very likely have a different spectrum of interaction partners the protein complexes assembled on the promoter might be completely different. It has been reported that full-length MLL can be found in a complex together with a variety of complexes and proteins involved in transcriptional regulation. Also negative regulators like the NuRD complex and Sin3A are found in this complex (Nakamura et al. 2002). These interactions have not been mapped precisely, hence one could speculate that the C-terminal MLL moiety is binding

these factors. To date no protein complex has been purified clarifying the interactions that MLL fusion proteins are involved in. It is likely, however, that those binding partners, which contact the MLL C-terminus do not interact with MLL fusion proteins. Factors might be lost that in the wild type situation ensure a delicate equilibrium of transcription of *Hoxa9* (Fig. 38B). Furthermore the C-terminal MLL moiety might itself exert negative effects on promoter complexes. Loss of these repressing functions in MLL fusion protein complexes and the presence of strong activation domains could lead to overexpression of *Hoxa9* as observed in leukemias (Fig. 38C).

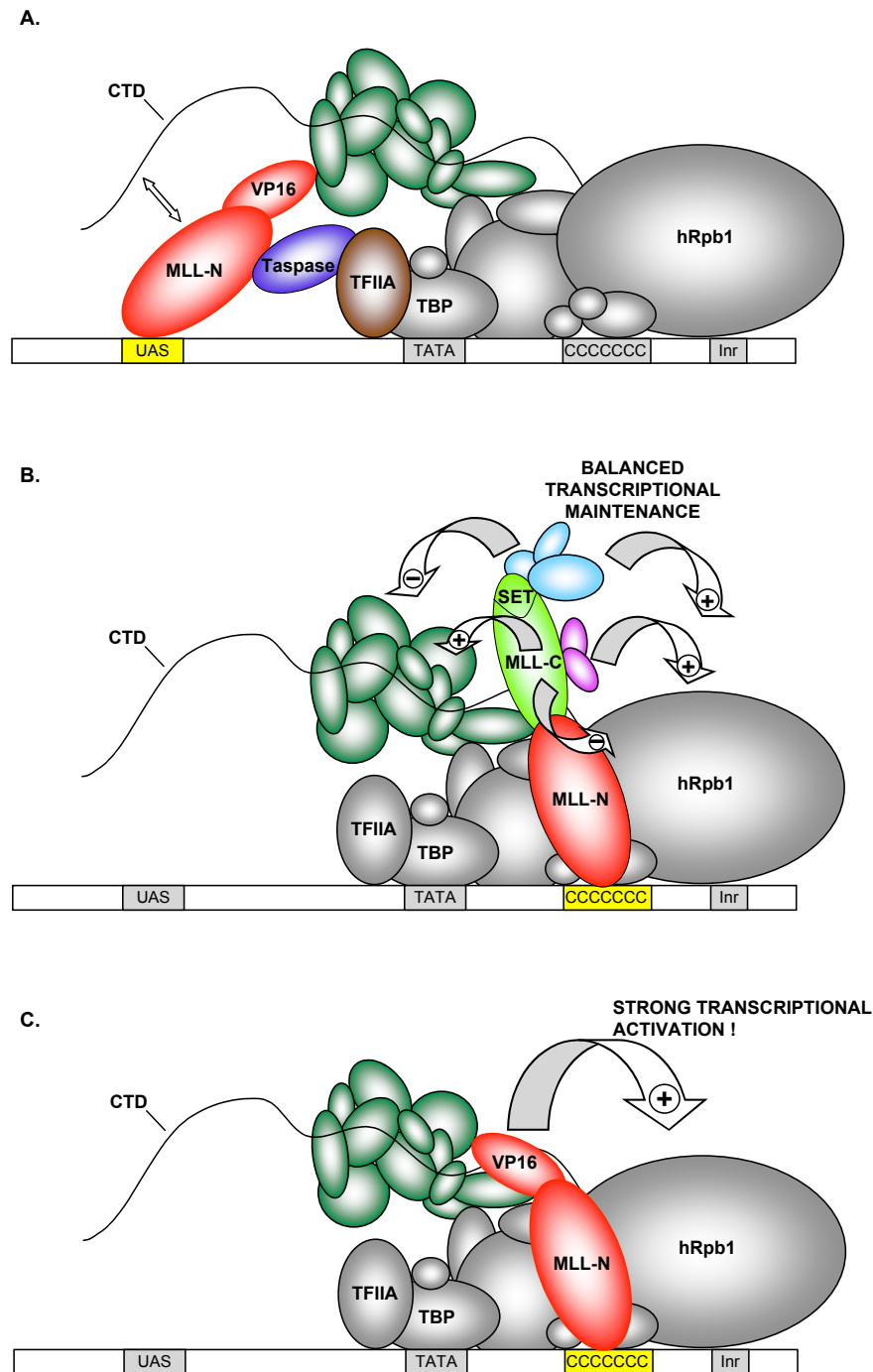


Figure 38. Model for the interplay between MLL and core promoter sequences. A. Model for an indirect contact through Taspase. Binding of the basal transcriptional machinery (grey) to the TATA box via TBP. TFIIA (brown) interacts with Taspase (violet), which might simultaneously contact the MLL N-terminus (red), which is recruited to the promoter by upstream activating sequences (UAS, yellow). The contact to Taspase might stabilize MLL fusion protein binding so that transcription can be activated through interaction with coactivator complexes (dark green). The binding of MLL fusion proteins to UAS might also be stabilized by interaction with the CTD of hRpb1. B. Direct binding of full-length MLL protein to core promoter features like the C-stretch (yellow) identified in *Hoxa9* might maintain transcription on a moderate level through multiple interactions with transcriptional co-repressors and co-activators (light blue and pink). The MLL C-terminus (light green) might itself exert regulatory effects on transcription contributing to a balanced transcription level. C. In the case of MLL-VP16-ER-HA binding to these core promoter sequences overexpression might result from the presence of the activation domain without the compensating effects of factors found in the promoter complexes containing wild type MLL.

4.3 A model for MLL binding to upstream activating sequences

4.3.1 Heterogeneity in the sequence motifs recruiting MLL–VP16–ER–HA

The models proposed above for MLL protein action on core promoter sequences illustrate a general uncertainty in this field of research as to what might be the binding specificities of MLL proteins. Structural features of the MLL aminoterminal have led to a model according to which this part of the protein is responsible for DNA binding. The presence of three AT hooks is intriguing and it was reported that the MLL AT hooks bind DNA depending less on the sequence than on cruciform structure of the target (Zelevnik-Le et al. 1994). The MLL MT domain on the other hand was shown to bind DNA with a preference for non-methylated CpGs (Birke et al. 2002). No sequence elements, however, have been identified in mammals that recruit MLL or other TrxG or PcG proteins. In *Drosophila* genomic fragments of typically some kilobases in length have been identified that are responsible for the transcriptional maintenance of a given region (Orlando 2003). Polycomb/Trithorax Response Elements (PRE/TREs) have been identified that can be transferred to another genomic context where they exert similar effects. It has been difficult so far to discover a consensus sequence in these response elements. Using bioinformatics it has been possible, however, to identify short sequence motifs that are common to all these elements (Ringrose et al. 2003).

In the study presented here the promoter sequences of *Hoxa9* and two other MLL targets, *p21* and *c-myc*, have been analyzed in detail to understand the mechanism of MLL and MLL fusion protein recruitment to promoters. Several motifs could be identified that mediate the response to MLL–VP16–ER–HA:

- Predicted Homeobox protein binding sites (in particular H3; TGATTTA) in the *Hoxa9* promoter facilitate activation by MLL–VP16–ER–HA. Interestingly mutation of these sites led to increased basal levels suggesting that *Hoxa9* protein binding to its own promoter might negatively regulate transcription. Instead of recruitment of MLL–VP16–ER–HA through *Hoxa9* protein direct binding of MLL–VP16–ER–HA to these motifs and competition with Homeobox proteins seems more likely.
- E-box sites (CTCGAG) were identified that could reconstitute MLL–VP16–ER–HA driven activation of a *Hoxa9* promoter mutant. Known

interactors with these sequence motifs include for example the transcription factor USF or the myc/Max-heterodimer none of which have been reported to interact with MLL or MLL fusion proteins.

- Noteworthy it could be demonstrated that a G-rich motif, which is present in the promoters of Hoxa9, p21 and c-myc plays a pivotal role in activation through MLL-VP16-ER-HA. Three copies of the motif GGGC are present in the Hoxa9 -207 / +46 promoter and it could be shown that point mutation of these motifs leads to a reduction in activation by MLL-VP16-ER-HA. Two regions were identified in the human p21 promoter (-122 / -60 and -215 / -144) that are essential for transcriptional activation of a luciferase reporter gene through MLL-VP16-ER-HA. Strikingly multiple copies of the same GGGC motif can be found in these regions. Also the promoter of the cellular proto-oncogene c-myc contains three GGGC motifs, which are absent in the mutant promoter versions that could no longer be activated by MLL-VP16-ER-HA indicating that MLL fusion proteins are targeted to promoters through GGGC motifs. Mutation of these motifs from GGGC to GGGG in the context of Hoxa9 -207 / +46 did not impair activation by MLL-VP16-ER-HA, however (see CpG mutant 1, alignments). Interestingly, the promoter sequences of Cutl1, a new candidate target gene for MLL-VP16-ER-HA identified in this work and also the promoter sequences of the MLL targets p18 and p27 are characterized by multiple copies of GGGC / GGGG motifs (suppl. Fig. 1, appendix 6.4). Furthermore, the GGGG sequence motif is present in the artificial Hoxa9 variant constructs MZF1 and MAZR1,2 and 3, which showed increased induction rates. Taken together these data suggest that MLL fusions bind to both GGGC and GGGG motifs.

4.3.2 MLL–VP16–ER–HA might compete with other transcription factors.

The presence of these motifs in the consensus sequences of other transcription factors is intriguing and leads to the question whether MLL fusion proteins are recruited through other DNA binding proteins. The GGGC motif is found in the binding matrix for SP1 (GGGCg_{gg}) and one could speculate about targeting of MLL fusions to promoters via interaction with SP1 protein. The data presented here, however, do not support this hypothesis. Targeting of GAL–SP1 to a GAL–Hoxa9 promoter did not affect activation by MLL–VP16–ER–HA. Furthermore no increase of induction of a p21 reporter gene by MLL–VP16–ER–HA was observed after overexpression of SP1 protein. The results of this study rather argue for a direct binding mechanism, where MLL contacts DNA directly, e.g. through its MT domain. In such a scenario SP1 would probably compete with MLL and MLL fusion proteins for binding to DNA target sequences. In particular for promoters, in which multiple SP1 binding sites are present this might mean that normal transcription of these genes is driven by both SP1 and MLL binding to the promoter (Fig. 39A). Competition of the two factors will, however, in some cases lead to occupancy of the promoter by MLL only. Due to the balancing functions that full-length MLL might exert in contrast to MLL fusions (see also 4.2) this might not significantly change transcription rates of the gene in wild type cells (Fig. 39C). In leukemic cells expressing a MLL fusion protein this statistic event in which both sites are occupied by MLL fusion proteins might lead to a different outcome: lacking the balancing effects of the wild type protein that could counteract the strong transcriptional activation these proteins might induce overexpression of the gene (Fig. 39D).

The GGGG motif is contained in both the binding matrices of myc-associated zinc finger protein related transcription factor (MAZR, matrix: gggGGGG) and the myeloid zinc finger protein 1 (MZF1, matrix: GGGGa). MAZR contains a BTB/POZ domain (Bardwell and Treisman 1994; Zollman et al. 1994), which has been shown to be involved in formation of homo-oligomers and hetero-oligomers together with Bach2 (Kobayashi et al. 2000). The BTB/POZ domain has been reported to contact the co-repressor complexes SMRT and N-CoR (Deweindt et al. 1995; Huynh and Bardwell 1998; Wong and Privalsky 1998). Recently it was shown that MAZR binds

to the Cd8 enhancer and negatively regulates transcription through recruitment of N-CoR (Bilic et al. 2006).

MZF1 is expressed in myeloid progenitor cells and binds to DNA through its zinc finger domains (Morris et al. 1994). Deletion of MZF1 in mice leads to lethal neoplasia and MZF1 $-/-$ hematopoietic progenitor cells show an increase in autonomous proliferation indicating a role for MZF1 as a tumor suppressor (Gaboli et al. 2001).

In the light of the biological functions of MAZR and MZF1 a model might be proposed where MLL fusion proteins bind to GGGC and GGGG in competition with other transcription factors. Leukemogenesis could be triggered through MLL fusion proteins binding to GGGG motifs instead of MAZR: repression might be substituted by strong transcriptional activation leading to overexpression of MLL target genes like *Hoxa9*. This might happen in a similar fashion to the one described for SP1 above. Figure 39B illustrates a situation where normal transcription is regulated by MLL and MAZR, which is acting as a repressor. The consequence of MLL and MLL fusion binding would be the same as described above: negative regulation of the gene through MAZR might be lost, while the MLL fusion protein is providing for strong activation (Fig. 39D). If MLL fusions compete efficiently with MZF1 protein the functions of the tumor suppressor might be impaired leading to a similar phenotype as observed in MZF1 $-/-$ mice: autonomous proliferation of myeloid cells that could ultimately lead to leukemia. As discussed before for MLL-VP16-ER-HA targeting core promoter sequences the differences between the effector domains associated with wild type MLL and MLL fusions might be critical. Disturbances in a delicate equilibrium maintained by wild type MLL might be enough for transformation of heterozygous cells. Additional experiments are required to clarify the relationship between MLL fusion proteins and these proteins. Co-transfection with MLL-VP16-ER-HA and overexpression of MAZR and MZF1 might provide insights into the mechanism by which MLL fusions are targeted to the G-rich motifs.

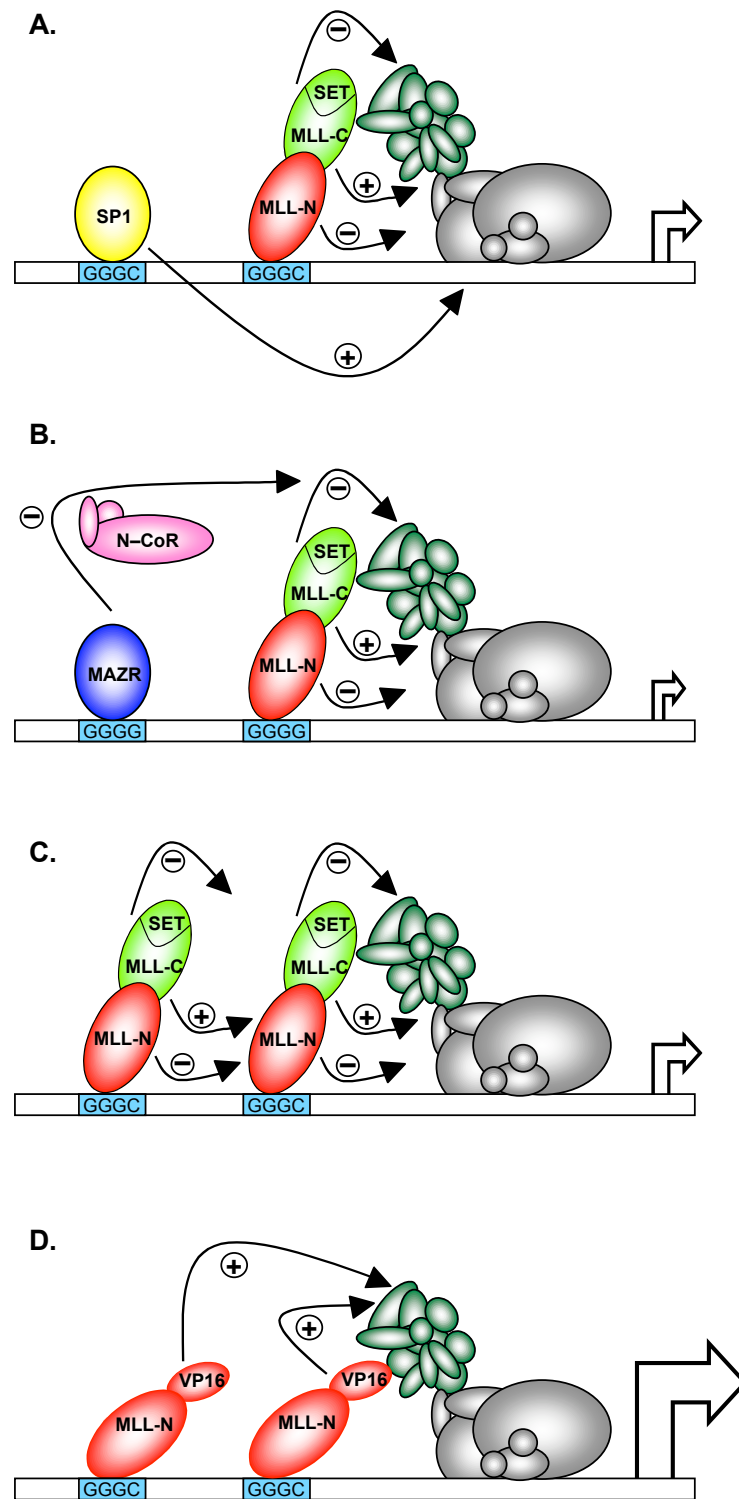


Figure 39. Competition of MLL with other transcription factors. A. SP1 (yellow) binding to its binding site in the promoter (light blue) positively regulates transcription of the gene. MLL wild type protein (red and light green) exerts both positive and negative effects and stabilizes transcription on a moderate level. The basal machinery and the mediator complex are shown in grey and dark green, respectively. B. Binding of MAZR (dark blue) to its binding site negatively affects the transcription rate through interaction with the corepressor complex N-CoR (pink), while MLL sustains transcription at a low level. C. MLL wild type protein binding without SP1 or MAZR does not lead to overexpression due to the balancing effects that are mediated through the C-terminal part of MLL and the complexes interacting with it (not shown). D. Promoter occupancy by MLL-VP16-ER-HA only leads to loss of transcriptional control and overexpression of the target gene.

4.3.3 A two-module hypothesis for MLL binding

An interesting feature of the G-rich motifs is their presence in multiple copies in each promoter. The analysis of the promoters of c-myc and p21 was particularly instructive in this respect. Deletion of GGC motifs in the p21 promoter through deletion of either the region -215 / -144 or -122 / +60 led to dramatic loss of induction by MLL-VP16-ER-HA. In the case of c-myc there were two different point mutants analyzed, both of which lacked at least one GGC motif present in the wild type sequence.

The mutations are located in two adjacent regions of about 60 nucleotides each, hence the spatial distribution of the mutated motifs resembles the one in the p21 promoter (see above). Just like in the case of the p21 mutation of only one of these regions leads to loss of induction by MLL-VP16-ER-HA. If each of the regions would be sufficient to bind MLL-VP16-ER-HA independent from the binding event taking place on other such motifs in the promoter such a dramatic effect would not be expected. The results demonstrated here rather argue for co-operative binding of MLL-VP16-ER-HA to two adjacent regions of the p21 or the c-myc promoter, respectively. Based on this hypothesis a model is proposed in which recruitment of MLL-VP16-ER-HA requires two adjacent modules in a promoter sequence that contain binding motifs. As illustrated by the schematic drawing each of these modules might contain one or more binding sites for MLL-VP16-ER-HA (Fig. 40A). In a simplified approximation two different mechanisms are possible:

A. Considering the large size of MLL and MLL fusion proteins and the flexibility of DNA it is theoretically possible that one molecule contacts both modules (Fig. 40B). The affinity of MLL-VP16-ER-HA to GGC motifs is not known. The MLL MT domain alone binds to DNA with a relatively low K_D in the range of 10^{-8} M (Birke et al. 2002). One might speculate, however, that two contacts are necessary to tether the protein tightly to the promoter. Another reason for the requirement of two contacts could be correct positioning of MLL-VP16-ER-HA. This could be the case if MLL-VP16-ER-HA has to be oriented in a certain way for promoter complex assembly.

B. Two or more molecules of MLL-VP16-ER-HA might be involved (Fig. 40C). Synergism in this case might be due to interaction of MLL-VP16-ER-HA with itself.

Binding of one molecule to one module might greatly facilitate binding of other molecules to the adjacent module.

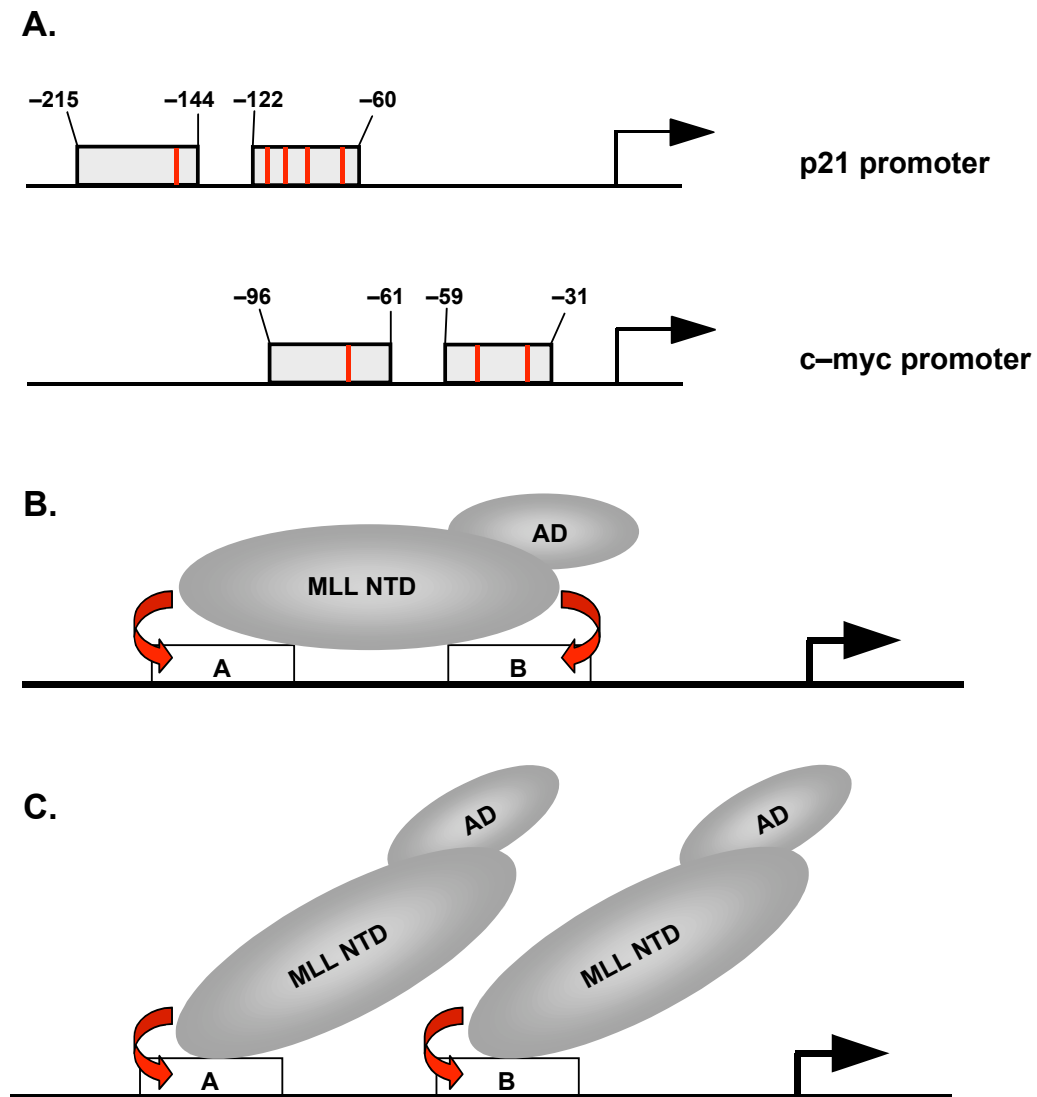


Figure 40. Two-module binding model for MLL fusion proteins. A. Modular structure of promoter elements mediating activation by MLL fusions. Grey boxes indicate regions that have been targeted by mutagenesis and loss of which impairs activation. Red bars indicate the presence of putative MLL fusion binding motifs. B. Model for the interaction of one MLL fusion protein molecule with both modules at the same time. C. Model for promoter occupancy by more than one MLL fusion protein molecule at a time. Both molecules could bind individual motifs, while binding of one molecule might facilitate binding of another molecule in the close vicinity.

4.3.4 Spatial distribution of binding sites is critical for MLL fusion protein function

The analysis presented here indicates that not only GGGC can function as a binding site for MLL–VP16–ER–HA. Also the related motif GGGG seems to recruit MLL fusion proteins to target gene promoters. Furthermore other motifs like the Hoxa9 binding site or E–box motifs have been shown in this study to stimulate activation by MLL–VP16–ER–HA. These observations argue for moderate sequence specificity of the MLL–DNA interaction. There seem to be multiple ways, however, to compensate for the lack of stringency in sequence specific binding. The model described above is postulating a requirement for two modules in a MLL target gene promoter both of which contain one or more binding sites for MLL. This adds another level of specificity to the process of promoter recognition by MLL, which is based on combinatorial constraints. Using oligonucleotides, which resemble one of these modules, however, an additional level of control was observed: three copies of an E–box motif evenly distributed over 26 nucleotides led to significant increase of transcriptional activation by MLL–VP16–ER–HA in comparison with a control sequence. Shifting the 3' E–box only one nucleotide further downstream led to induction rates only slightly over background. Irregular spacing of GGGC motifs in the MAZR constructs decreased induction rates moderately. These findings indicate that multiple copies of binding sites in MLL target modules can co-operate in MLL recruitment. This co–operativity, however seems to depend on regular spacing between the single binding motifs.

4.3.5 Epigenetic control of Hoxa9 transcriptional activation by MLL–VP16–ER–HA

Published binding studies using the MLL MT domain as a bait have shown that this part of the protein binds preferentially to non–methylated CpG rich DNA (Birke et al. 2002). Using site–directed mutagenesis of CpG dinucleotides in the Hoxa9 promoter sequence it was demonstrated in this study that cytosine residues are not essential for MLL recruitment. G–rich motifs (GGGC, GGGG), E–box motifs and a binding site for Hoxa9 present in the Hoxa9 promoter could be shown to mediate activation by MLL–VP16–ER–HA. Interestingly, mutation of cytosine residues in CpG rich regions

of the *Hoxa9* promoter dramatically increased induction by MLL–VP16–ER–HA. Since methylation of DNA by cellular enzymes is blocked by substitution of cytosine with guanosine nucleotides this phenomenon appears to provide a link between DNA methylation and MLL fusion protein binding. This is in line with reports about methylated DNA at the *Hoxa* locus (Hershko et al. 2003). The observed increase in transcriptional activation by MLL–VP16–ER–HA might indicate a negative regulatory role for cytosine residues in MLL target promoters. Specific binding to non-methylated sites on the DNA has been reported for a number of transcription factors like USF, NFκB and SP1 (Cowled et al. 2005; Fujii et al. 2006; Mulero-Navarro et al. 2006). Based on the observation that MLL binding to the *Hoxa9* promoter is mediated by GGGG as well as by GGGC a model is proposed in which guanosine, not cytosine residues recruit MLL to CpG rich DNA. Cytosine residues in these genomic regions might rather function as modulators. Some of the stable cell lines containing CpG mutant *Hoxa9* constructs showed induction rates similar to wild type while other (genetically identical) cell lines showed dramatically increased rates. This could indicate a role for methylated cytosine residues in regulation of MLL binding to the *Hoxa9* promoter. The human *Hoxa9* promoter is obviously targeted very efficiently by MLL proteins. In order to maintain moderate expression levels negative regulation of MLL binding might be required. A certain number of methylated CpGs could be necessary to prevent excessive MLL recruitment. If this critical threshold would be crossed by mutation of CpG to GpG uncontrolled recruitment of MLL to the promoter could lead to overexpression.

A recent publication demonstrated that disruption of the MLL SET domain leads to abnormal DNA methylation patterns at the *Hoxd4* locus and homeotic transformation in transgenic mice (Terranova et al. 2006). These results suggest that this truncated protein variant somehow interferes with DNA methylation. There is no evidence so far that MLL fusion constructs influence methylation patterns. The MT domain, however, which is likely to be involved in this process, is retained in all fusion proteins. Therefore one could think of transcriptional deregulation of certain MLL target genes as a consequence of promoter methylation changes following 11q23 translocation.

4.4 Identification and characterization of new target promoters

4.4.1 The human p21 promoter is targeted by MLL and MLL fusion proteins

Using inducible hematopoietic cell lines of human and murine origin in combination with luciferase reporter constructs two new target genes for MLL fusion proteins were identified.

It has been reported earlier that cyclin dependent kinase inhibitors are targets for transcriptional regulation by MLL (Milne et al. 2005b). In the present study a new target for MLL and MLL fusion proteins was identified that belongs to the same class: experiments using different variants of the human p21 promoter could show that MLL fusions as well as MLL full length protein activate transcription from these sequences. The specificity of this process was shown not only by parallel experiments using a stable control cell line expressing ER-VP-16-HA but also by results from transient transfection experiments carried out in several cell lines that all support the observations made initially in stable FDCPmix cells. RT-PCR analysis of several cell lines could show only marginal increases in endogenous p21 mRNA amounts, however. One possible reason for this could be the cell cycle dependent transcription of the gene, which makes it difficult to detect transcriptional changes in non-synchronized cell populations (Gartel and Radhakrishnan 2005). Alternatively, epigenetic modifications of the endogenous locus might render it non-responsive to induction of MLL-VP16-ER-HA activity.

A short sequence element (-215 / +8) conferred the highest induction rate, whereas robust induction was also seen with a longer promoter variant (-2325 / +8). Via mutation analysis two critical regions could be identified: -122 / -60 and -215 / +8. Deletion of either one of these two sites was sufficient to reduce induction levels to less than 10 percent. GGGC motifs are present in both of these regions and might directly target MLL-VP16-ER-HA to these modules (see 4.3.3).

It could be shown that not only the synthetic protein MLL-VP16 fusion but also MLL-AF4, MLL-AF9 and MLL full length activate transcription from the p21 promoter. Interestingly, the Δ -122 / -60 mutation had a much stronger effect on activation by MLL-VP16-HA than on activation by FLAG-MLL. This could be hinting at different binding specificities of the full-length protein in comparison with MLL fusion constructs. These differences in specificity are closely linked to the

question whether there are certain "leukemogenic" subsets of genes that are only targeted by the oncogenic fusion proteins but not by MLL.

4.4.2 Activation of p21 by MLL–VP16–ER–HA is dependent from the chromatin environment

Analysis of p21 expression levels after induction of MLL–VP16–ER–HA in different cell lines revealed a different response of the chromosomal p21 promoter in comparison to the episomal reporters. No significant increase of transcription could be observed analyzing mRNA derived from the chromosomal gene. In the same cells, however, the increase of transcription from transient and episomal reporter plasmids was dramatic. It has been reported that transcription from the p21 locus can be inhibited by promoter methylation (Allan et al. 2000). In the same report it has been demonstrated that introduction of the genomic p21 locus in a P–1 derived artificial chromosome (PAC) can restore p21 expression showing that transcription is taking place on the non-methylated vector but not on the methylated genomic locus in the same cell. As demonstrated in the present analysis treatment of a stable p21 reporter cell line with Azacytidine increased basal transcription levels of the chromosomal gene as well as of the episomal reporter illustrating the relief from transcriptional repression following loss of DNA methylation. No significant induction of the chromosomal p21 gene by MLL–VP16–ER–HA could be seen under both conditions, however. Following Azacytidine treatment the induction rate of the reporter decreased as assessed by measurement of mRNA amounts and luciferase activity. This could be due to a secondary effect of the drug that interferes with the activation process through MLL–VP16–ER–HA. Another explanation would be reduced recruitment of MLL–VP16–ER–HA to the promoter as a consequence of decreased affinity to the cytidine analog.

4.4.3 MLL–VP16–ER–HA activates transcription from the c–myc promoter

Another new target for MLL identified in this work is the c–myc promoter sequence. c–myc was identified as the human homolog for an avian oncogene that was found to be involved in chromosomal translocations in Burkitt lymphoma (Marcu et al. 1992). Immunoglobulin loci like the Igk enhancer have been identified as translocation partners that were leading to c–myc overexpression (Erikson et al. 1983). Transcription of the c–myc gene is subject to complex regulation. Four promoters (P0, P1, P2, P3) drive transcription of the gene. Under normal conditions, however most of c–myc transcription is initiated from the P2 promoter. P1 driven transcription usually accounts for only 10-25% of the transcripts (Stewart et al. 1984; Taub et al. 1984; Bentley and Groudine 1986).

Reporter analysis revealed that both promoters of the gene (P1 and P2) are strongly induced by MLL–VP16–ER–HA. A short fragment of the P1 promoter (–101 / +66), however, was enough to mediate this effect. The presence of the Igk enhancer did not affect transcriptional activation by MLL–VP16–ER–HA underlining the fact that cis–elements in the c–myc P1 promoter mediate the effect. Both mutant promoter variants that were used for the transfection experiments almost completely abolished the activation through MLL–VP16–ER–HA (see 4.3.3). Interestingly the G–rich sequences affected by the proximal mutation were reported early on to be critical for proper transcription initiation and are conserved between human and mouse (Nishikura 1986).

Given the function of c–myc as a gene that drives cell proliferation it is intriguing that the oncogene MLL–VP16–ER–HA is targeted to its promoter. Support for a model in which MLL and MLL fusion proteins are involved in regulation of c–myc comes not only from expression analysis of leukemias but also from a recent CHIP-chip study where MLL was found on the c–myc promoter (Rozovskaia et al. 2003; Guenther et al. 2005). Further studies will be required, however, to clarify whether MLL and MLL fusion proteins regulate chromosomal c–myc. CHIP analyses can be used to analyze promoter occupancy by MLL–VP16–ER–HA on wild type versus mutant promoter constructs.

4.4.4 A hematopoietic stem cell system for MLL fusion target gene analysis

One of the major open questions in the MLL research area still is the genome-wide identification of direct target genes. Human leukemias that arise as a consequence of chromosomal rearrangement in the MLL locus in hematopoietic stem cells illustrate the fact that the resulting MLL fusions can only exert their full oncogenic potential in a certain subset of pluripotent cells (Cozzio et al. 2003; So et al. 2003).

In this work FDCPmix cells were chosen as a model system for MLL fusion target gene activation. It has been shown earlier that this pluripotent cell line has the potential to differentiate into the majority of myeloid cell lineages (Schroeder 2001). In the present study it could be shown that also after introduction of an inducible form of MLL-VP16 cells could be differentiated to the terminal stages of granulocytic and monocytic pathways. The estrogen receptor mutant fused variant used for most of the experiments proved to be a valid research tool: target gene induction as shown via RT-PCR and reporter gene assays was both fast and specific. This was demonstrated by the upregulation of Meis1 only 3 hours after induction in the MLL-VP16-ER-HA cell line but not in the control cell line. In a genome-wide microarray also other known target genes like the cell cycle regulators p27 and p18 were shown to be induced. Downregulation of genes important for myeloid differentiation was observed after 23 hours of MLL-VP16-ER-HA induction, which is reflected in impaired terminal differentiation of these cells. These findings are in line with the general model of the biological activity of an oncogene, which counteracts differentiation and stimulates proliferation.

The induction of ER-VP16 fusions by Tamoxifen induced cellular toxicity. For analysis of the immediate early events after induction of the fusion construct, however, this hardly matters. Even though the clinical phenotypes of 11q23 translocations are rather diverse gene expression profiling of leukemias derived from different fusions display a common signature (Armstrong et al. 2002; Kohlmann et al. 2005). The rather abstract approach taken here using the acidic activation domain of VP16 as a paradigm for acidic activators represents an effort to generate a tool for detection of the largest possible population of target genes. In fact in transient transfections the VP16 fusion did not behave in a less specific manner than for example MLL-AF9 but induction levels of reporter genes were considerably higher.

Furthermore there are reports showing that fusion of the VP16 activation domain to the MLL aminotermminus leads to transformation of myeloid progenitor cells and to leukemia in mice (So and Cleary 2003; Zeisig et al. 2003). Taken together these properties render the system a specific and sensitive biological tool to investigate the early events following activation of a leukemogenic fusion protein. The advantage of having a cell line greatly facilitates the analysis of target gene regulation. Biological material can be provided for biochemical experiments or ChIP, where larger cell numbers are usually required. For this type of experiments the monoclonal antibodies generated in this work that specifically recognize and precipitate both moieties of MLL will prove useful.

4.4.5 A genome-wide screen for MLL fusion target genes

The results of microarray analysis provide new candidate target genes. A series of studies have been published reporting aberrant gene expression patterns in leukemia patients carrying 11q23 translocations (Armstrong et al. 2002; Schoch et al. 2002; Ferrando et al. 2003; Rozovskaia et al. 2003; Kohlmann et al. 2005). An inherent problem of these studies is the high heterogeneity of biological material and the secondary events taking place in cancer cells that prevent discrimination of direct targets for MLL fusion proteins from genes that are upregulated as a consequence of leukemogenesis. The approach presented in this work combines a hematopoietic stem cell background with an oncogenic MLL fusion that can be induced very fast. Working with a monoclonal cell line ensures a very high degree of homogeneity and allows for analysis of the early events taking place after activation of MLL fusion proteins. This increases the chance of finding target genes whose expression is directly regulated by these oncogenes.

Amongst the targets identified in this study are some genes that had been reported to be regulated by MLL. Meis1, which is considered a hallmark of MLL leukemogenesis was stably upregulated already 30 min after induction of MLL–VP16–ER–HA (Kawagoe et al. 1999; Rozovskaia et al. 2001; Imamura et al. 2002; Zeisig et al. 2004b). At the same time cyclin–dependent kinase inhibitors p18 and p27 were upregulated, which recently have been described as MLL target

genes (Schraets et al. 2003; Milne et al. 2005b; Xia et al. 2005). Comparison with published array data yielded moderate overlap. The chromatin remodeling factor Runx2 and the transcriptional corepressor Zfx1B have been identified as common MLL target genes in acute lymphoblastic and acute myeloid leukemias (ALL, AML) by microarray analysis of leukemia cells (Kohlmann et al. 2005). Runx2 normally plays a role in osteogenesis but there reports about the oncogenic capacity of the protein (Stewart et al. 1997; Ito 2004). Zfx1B encoding Smad-interacting protein 1 (SIP1) directly represses E-cadherin gene transcription and activates cancer invasion via the upregulation of the matrix metalloproteinase gene family (Miyoshi et al. 2004). The upregulation of these genes in leukemias is underlining the physiological relevance of these genes. The fact, however, that they were independently identified in the model system presented in this study is indicating that they are probably direct targets of MLL rather than secondary effects of leukemogenesis. Expression of three other genes, however, that have been identified recently in an MLL-ENL model (Flt3, Lmo2 and N-myc) did not change significantly in the system presented here (Zeisig et al. 2004a).

One way of proceeding from this point is to investigate the promoters of some of these genes using a classical mutagenesis and reporter gene approach like it was done e.g. for Hoxa9 in this study. The alternative is a bioinformatic approach employing databases for the identification of common pathways and common elements in target gene promoters. For the two early time points (30 min. and 3 hours) the predicted common features in target gene promoters were binding sites for Homeobox proteins and the Cut-like homeodomain protein Cutl1. The potential role of Hoxa9 binding sites for the activation by MLL-VP16-ER-HA has been discussed earlier. Taking into account that Hoxa9 is upregulated by MLL fusions one obvious conclusion would be that these are not direct MLL-VP16-ER-HA targets but genes that are upregulated as a secondary effect by Hoxa9. The fact, however, that also in the 30 min. sample genes where upregulated whose promoters contain Hox binding sites indicates that MLL-VP16-ER-HA is directly targeting these sequences. A similar situation is found for Cut-like homeodomain proteins and their binding sites in target gene promoters, respectively. Cutl1 binding sites are significantly overrepresented already in the promoters of target genes upregulated after 30 min, arguing either for direct targeting by MLL-VP16-ER-HA or a very fast secondary

response. However, since Cutl1 itself is one of the genes that are upregulated after 3 hours this could also be interpreted as a secondary effect: MLL–VP16–ER–HA would target Cutl1 and the upregulation of this gene would subsequently lead to deregulation of secondary target genes.

In any case the upregulation of Cutl1 by itself is an interesting observation. It has been reported that Cutl1 binds to the promoter of p21 and inhibits transcription through recruitment of the histone methyltransferase G9a (Coqueret et al. 1998; Nishio and Walsh 2004). This finding is shedding a different light again on the lack of activation of the chromosomal p21 gene by MLL–VP16–ER–HA. Possibly induction of MLL–VP16–ER–HA leads to increased expression of Cutl1, a repressor of p21 transcription that exerts repression through recruitment of a chromatin modifier. It has to be tested in CHIP whether H3K9 methylation is present on the chromosomal but not on the episomal locus. If this would be the case this could indicate that the p21 promoter sequence is a target for MLL fusion. The chromosomal gene would, however, not be activated because MLL–VP16–ER–HA is upregulating a potent repressor of p21 transcription at the same time.

Besides p21 several other genes have been reported to be silenced by Cutl1 or its mouse homolog Cux (Nepveu 2001). There are some reports, however, indicating a role as a transcriptional activator (Nepveu 2001; Truscott et al. 2003). Interestingly Cutl1 seems to be involved in tumorigenesis (Zhu et al. 2004; Michl et al. 2005; Michl et al. 2006), which further reasons for the biological relevance of the gene as a possible target for MLL fusions.

5. References

- Allan, L.A., T. Duhig, M. Read, and M. Fried. 2000. The p21(WAF1/CIP1) promoter is methylated in Rat-1 cells: stable restoration of p53-dependent p21(WAF1/CIP1) expression after transfection of a genomic clone containing the p21(WAF1/CIP1) gene. *Mol Cell Biol* **20**: 1291-8.
- Alvarez-Venegas, R. and Z. Avramova. 2002. SET-domain proteins of the Su(var)3-9, E(z) and trithorax families. *Gene* **285**: 25-37.
- Armstrong, S.A., J.E. Staunton, L.B. Silverman, R. Pieters, M.L. den Boer, M.D. Minden, S.E. Sallan, E.S. Lander, T.R. Golub, and S.J. Korsmeyer. 2002. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet* **30**: 41-7.
- Asturias, F.J. 2004. RNA polymerase II structure, and organization of the preinitiation complex. *Curr Opin Struct Biol* **14**: 121-9.
- Ayton, P.M., E.H. Chen, and M.L. Cleary. 2004. Binding to nonmethylated CpG DNA is essential for target recognition, transactivation, and myeloid transformation by an MLL oncoprotein. *Mol Cell Biol* **24**: 10470-8.
- Bardwell, V.J. and R. Treisman. 1994. The POZ domain: a conserved protein-protein interaction motif. *Genes Dev* **8**: 1664-77.
- Beisel, C., A. Imhof, J. Greene, E. Kremmer, and F. Sauer. 2002. Histone methylation by the Drosophila epigenetic transcriptional regulator Ash1. *Nature* **419**: 857-62.
- Bell, A.C., A.G. West, and G. Felsenfeld. 2001. Insulators and boundaries: versatile regulatory elements in the eukaryotic. *Science* **291**: 447-50.
- Bentley, D.L. and M. Groudine. 1986. Novel promoter upstream of the human c-myc gene and regulation of c-myc expression in B-cell lymphomas. *Mol Cell Biol* **6**: 3481-9.
- Bilic, I., C. Koesters, B. Unger, M. Sekimata, A. Hertweck, R. Maschek, C.B. Wilson, and W. Ellmeier. 2006. Negative regulation of CD8 expression via Cd8 enhancer-mediated recruitment of the zinc finger protein MAZR. *Nat Immunol* **7**: 392-400.
- Bird, A.P. and A.P. Wolffe. 1999. Methylation-induced repression--belts, braces, and chromatin. *Cell* **99**: 451-4.
- Birke, M., S. Schreiner, M.P. Garcia-Cuellar, K. Mahr, F. Titgemeyer, and R.K. Slany. 2002. The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. *Nucleic Acids Res* **30**: 958-65.
- Blazek, E., G. Mittler, and M. Meisterernst. 2005. The mediator of RNA polymerase II. *Chromosoma* **113**: 399-408.
- Bourbon, H.M., A. Aguilera, A.Z. Ansari, F.J. Asturias, A.J. Berk, S. Bjorklund, T.K. Blackwell, T. Borggreffe, M. Carey, M. Carlson, J.W. Conaway, R.C. Conaway, S.W. Emmons, J.D. Fondell, L.P. Freedman, T. Fukasawa, C.M. Gustafsson, M. Han, X. He, P.K. Herman, A.G. Hinnebusch, S. Holmberg, F.C. Holstege,

- J.A. Jaehning, Y.J. Kim, L. Kuras, A. Leutz, J.T. Lis, M. Meisterernest, A.M. Naar, K. Nasmyth, J.D. Parvin, M. Ptashne, D. Reinberg, H. Ronne, I. Sadowski, H. Sakurai, M. Sipiczki, P.W. Sternberg, D.J. Stillman, R. Strich, K. Struhl, J.Q. Svejstrup, S. Tuck, F. Winston, R.G. Roeder, and R.D. Kornberg. 2004. A unified nomenclature for protein subunits of mediator complexes linking transcriptional regulators to RNA polymerase II. *Mol Cell* **14**: 553-7.
- Boyes, J. and A. Bird. 1991. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell* **64**: 1123-34.
- Brock, H.W. and C.L. Fisher. 2005. Maintenance of gene expression patterns. *Dev Dyn* **232**: 633-55.
- Buratowski, S., S. Hahn, L. Guarente, and P.A. Sharp. 1989. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56**: 549-61.
- Caligiuri, M.A., M.P. Strout, S.A. Schichman, K. Mrozek, D.C. Arthur, G.P. Herzig, M.R. Baer, C.A. Schiffer, K. Heinonen, S. Knuutila, T. Nousiainen, T. Ruutu, A.W. Block, P. Schulman, J. Pedersen-Bjergaard, C.M. Croce, and C.D. Bloomfield. 1996. Partial tandem duplication of ALL1 as a recurrent molecular defect in acute myeloid leukemia with trisomy 11. *Cancer Res* **56**: 1418-25.
- Calvo, K.R., D.B. Sykes, M. Pasillas, and M.P. Kamps. 2000. Hoxa9 immortalizes a granulocyte-macrophage colony-stimulating factor-dependent promyelocyte capable of biphenotypic differentiation to neutrophils or macrophages, independent of enforced meis expression. *Mol Cell Biol* **20**: 3274-85.
- Cao, R., L. Wang, H. Wang, L. Xia, H. Erdjument-Bromage, P. Tempst, R.S. Jones, and Y. Zhang. 2002. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**: 1039-43.
- Cartharius, K., K. Frech, K. Grote, B. Klocke, M. Haltmeier, A. Klingenhoff, M. Frisch, M. Bayerlein, and T. Werner. 2005. MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* **21**: 2933-42.
- Chen, Z., J. Zang, J. Whetstone, X. Hong, F. Davrazou, T.G. Kutateladze, M. Simpson, Q. Mao, C.H. Pan, S. Dai, J. Hagman, K. Hansen, Y. Shi, and G. Zhang. 2006. Structural insights into histone demethylation by JMJD2 family members. *Cell* **125**: 691-702.
- Coqueret, O., G. Berube, and A. Nepveu. 1998. The mammalian Cut homeodomain protein functions as a cell-cycle-dependent transcriptional repressor which downmodulates p21WAF1/CIP1/SDI1 in S phase. *Embo J* **17**: 4680-94.
- Courey, A.J. and S. Jia. 2001. Transcriptional repression: the long and the short of it. *Genes Dev* **15**: 2786-96.
- Cowled, P., I. Kanter, L. Leonardos, and P. Jackson. 2005. Uroplakin Ib gene transcription in urothelial tumor cells is regulated by CpG methylation. *Neoplasia* **7**: 1091-103.
- Cozzio, A., E. Passegue, P.M. Ayton, H. Karsunky, M.L. Cleary, and I.L. Weissman. 2003. Similar MLL-associated leukemias arising from self-renewing stem cells and short-lived myeloid progenitors. *Genes Dev* **17**: 3029-35.

- Cramer, P. 2004. RNA polymerase II structure: from core to functional complexes. *Curr Opin Genet Dev* **14**: 218-26.
- Cramer, P., D.A. Bushnell, J. Fu, A.L. Gnatt, B. Maier-Davis, N.E. Thompson, R.R. Burgess, A.M. Edwards, P.R. David, and R.D. Kornberg. 2000. Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* **288**: 640-9.
- Cui, X., I. De Vivo, R. Slany, A. Miyamoto, R. Firestein, and M.L. Cleary. 1998. Association of SET domain and myotubularin-related proteins modulates growth control. *Nat Genet* **18**: 331-7.
- Czermin, B., R. Melfi, D. McCabe, V. Seitz, A. Imhof, and V. Pirrotta. 2002. Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**: 185-96.
- Daser, A. and T.H. Rabbitts. 2005. The versatile mixed lineage leukaemia gene MLL and its many associations in leukaemogenesis. *Semin Cancer Biol* **15**: 175-88.
- Deweindt, C., O. Albagli, F. Bernardin, P. Dhordain, S. Quief, D. Lantoine, J.P. Kerckaert, and D. Leprince. 1995. The LAZ3/BCL6 oncogene encodes a sequence-specific transcriptional inhibitor: a novel function for the BTB/POZ domain as an autonomous repressing domain. *Cell Growth Differ* **6**: 1495-503.
- Dikstein, R., S. Ruppert, and R. Tjian. 1996. TAFII250 is a bipartite protein kinase that phosphorylates the basal transcription factor RAP74. *Cell* **84**: 781-90.
- Djabali, M., L. Selleri, P. Parry, M. Bower, B.D. Young, and G.A. Evans. 1992. A trithorax-like gene is interrupted by chromosome 11q23 translocations in acute leukaemias. *Nat Genet* **2**: 113-8.
- Dobson, C.L., A.J. Warren, R. Pannell, A. Forster, and T.H. Rabbitts. 2000. Tumorigenesis in mice with a fusion of the leukaemia oncogene Mll and the bacterial lacZ gene. *Embo J* **19**: 843-51.
- Dou, Y., T.A. Milne, A.J. Tackett, E.R. Smith, A. Fukuda, J. Wysocka, C.D. Allis, B.T. Chait, J.L. Hess, and R.G. Roeder. 2005. Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell* **121**: 873-85.
- Dynlacht, B.D., T. Hoey, and R. Tjian. 1991. Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. *Cell* **66**: 563-76.
- Eastman, Q. and R. Grosschedl. 1999. Regulation of LEF-1/TCF transcription factors by Wnt and other signals. *Curr Opin Cell Biol* **11**: 233-40.
- Erikson, J., K. Nishikura, A. ar-Rushdi, J. Finan, B. Emanuel, G. Lenoir, P.C. Nowell, and C.M. Croce. 1983. Translocation of an immunoglobulin kappa locus to a region 3' of an unrearranged c-myc oncogene enhances c-myc transcription. *Proc Natl Acad Sci U S A* **80**: 7581-5.
- Ernst, P., J. Wang, M. Huang, R.H. Goodman, and S.J. Korsmeyer. 2001. MLL and CREB bind cooperatively to the nuclear coactivator CREB-binding protein. *Mol Cell Biol* **21**: 2249-58.

- Fair, K., M. Anderson, E. Bulanova, H. Mi, M. Tropschug, and M.O. Diaz. 2001. Protein interactions of the MLL PHD fingers modulate MLL target gene regulation in human cells. *Mol Cell Biol* **21**: 3589-97.
- Feil, R., J. Wagner, D. Metzger, and P. Chambon. 1997. Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochem Biophys Res Commun* **237**: 752-7.
- Fernandez, L.A., M. Winkler, and R. Grosschedl. 2001. Matrix attachment region-dependent function of the immunoglobulin mu enhancer involves histone acetylation at a distance without changes in enhancer occupancy. *Mol Cell Biol* **21**: 196-208.
- Ferrando, A.A., S.A. Armstrong, D.S. Neuberg, S.E. Sallan, L.B. Silverman, S.J. Korsmeyer, and A.T. Look. 2003. Gene expression signatures in MLL-rearranged T-lineage and B-precursor acute leukemias: dominance of HOX dysregulation. *Blood* **102**: 262-8.
- Flanagan, P.M., R.J. Kelleher, 3rd, M.H. Sayre, H. Tschochner, and R.D. Kornberg. 1991. A mediator required for activation of RNA polymerase II transcription in vitro. *Nature* **350**: 436-8.
- Fondell, J.D., H. Ge, and R.G. Roeder. 1996. Ligand induction of a transcriptionally active thyroid hormone receptor coactivator complex. *Proc Natl Acad Sci U S A* **93**: 8329-33.
- Francastel, C., D. Schubeler, D.I. Martin, and M. Groudine. 2000. Nuclear compartmentalization and gene activity. *Nat Rev Mol Cell Biol* **1**: 137-43.
- Fujii, G., Y. Nakamura, D. Tsukamoto, M. Ito, T. Shiba, and N. Takamatsu. 2006. CpG methylation at the USF-binding site is important for the liver-specific transcription of the chipmunk HP-27 gene. *Biochem J* **395**: 203-9.
- Fuks, F., P.J. Hurd, R. Deplus, and T. Kouzarides. 2003a. The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Res* **31**: 2305-12.
- Fuks, F., P.J. Hurd, D. Wolf, X. Nan, A.P. Bird, and T. Kouzarides. 2003b. The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J Biol Chem* **278**: 4035-40.
- Gaboli, M., P.A. Kotsi, C. Gurrieri, G. Cattoretti, S. Ronchetti, C. Cordon-Cardo, H.E. Broxmeyer, R. Hromas, and P.P. Pandolfi. 2001. Mzf1 controls cell proliferation and tumorigenesis. *Genes Dev* **15**: 1625-30.
- Gartel, A.L., E. Goufman, S.G. Tevosian, H. Shih, A.S. Yee, and A.L. Tyner. 1998. Activation and repression of p21(WAF1/CIP1) transcription by RB binding proteins. *Oncogene* **17**: 3463-9.
- Gartel, A.L. and S.K. Radhakrishnan. 2005. Lost in transcription: p21 repression, mechanisms, and consequences. *Cancer Res* **65**: 3980-5.
- Ge, Y.Z., M.T. Pu, H. Gowher, H.P. Wu, J.P. Ding, A. Jeltsch, and G.L. Xu. 2004. Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. *J Biol Chem* **279**: 25447-54.

- Glaser, S., J. Schaft, S. Lubitz, K. Vintersten, F. van der Hoeven, K.R. Tufteland, R. Aasland, K. Anastassiadis, S.L. Ang, and A.F. Stewart. 2006. Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Development* **133**: 1423-32.
- Goodrich, J.A., T. Hoey, C.J. Thut, A. Admon, and R. Tjian. 1993. Drosophila TAFII40 interacts with both a VP16 activation domain and the basal transcription factor TFIIB. *Cell* **75**: 519-30.
- Goppelt, A., G. Stelzer, F. Lottspeich, and M. Meisterernst. 1996. A mechanism for repression of class II gene transcription through specific binding of NC2 to TBP-promoter complexes via heterodimeric histone fold domains. *Embo J* **15**: 3105-16.
- Gu, Y., G. Cimino, H. Alder, T. Nakamura, R. Prasad, O. Canaani, D.T. Moir, C. Jones, P.C. Nowell, C.M. Croce, and et al. 1992. The (4;11)(q21;q23) chromosome translocations in acute leukemias involve the VDJ recombinase. *Proc Natl Acad Sci U S A* **89**: 10464-8.
- Guenther, M.G., R.G. Jenner, B. Chevalier, T. Nakamura, C.M. Croce, E. Canaani, and R.A. Young. 2005. Global and Hox-specific roles for the MLL1 methyltransferase. *Proc Natl Acad Sci U S A* **102**: 8603-8.
- Hampsey, M. 1998. Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol Mol Biol Rev* **62**: 465-503.
- Hanson, R.D., J.L. Hess, B.D. Yu, P. Ernst, M. van Lohuizen, A. Berns, N.M. van der Lugt, C.S. Shashikant, F.H. Ruddle, M. Seto, and S.J. Korsmeyer. 1999. Mammalian Trithorax and polycomb-group homologues are antagonistic regulators of homeotic development. *Proc Natl Acad Sci U S A* **96**: 14372-7.
- Harrison, S.C. 1991. A structural taxonomy of DNA-binding domains. *Nature* **353**: 715-9.
- Hendrich, B. and A. Bird. 1998. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol* **18**: 6538-47.
- Hershko, A.Y., T. Kafri, A. Fainsod, and A. Razin. 2003. Methylation of HoxA5 and HoxB5 and its relevance to expression during mouse development. *Gene* **302**: 65-72.
- Hsieh, J.J., E.H. Cheng, and S.J. Korsmeyer. 2003a. Taspase1: a threonine aspartase required for cleavage of MLL and proper HOX gene expression. *Cell* **115**: 293-303.
- Hsieh, J.J., P. Ernst, H. Erdjument-Bromage, P. Tempst, and S.J. Korsmeyer. 2003b. Proteolytic cleavage of MLL generates a complex of N- and C-terminal fragments that confers protein stability and subnuclear localization. *Mol Cell Biol* **23**: 186-94.
- Huyen, Y., O. Zgheib, R.A. Ditullio, Jr., V.G. Gorgoulis, P. Zacharatos, T.J. Petty, E.A. Sheston, H.S. Mellert, E.S. Stavridi, and T.D. Halazonetis. 2004. Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. *Nature* **432**: 406-11.

- Huynh, K.D. and V.J. Bardwell. 1998. The BCL-6 POZ domain and other POZ domains interact with the co-repressors N-CoR and SMRT. *Oncogene* **17**: 2473-84.
- Ikeda, K., T. Stuehler, and M. Meisterernst. 2002. The H1 and H2 regions of the activation domain of herpes simplex virion protein 16 stimulate transcription through distinct molecular mechanisms. *Genes Cells* **7**: 49-58.
- Imamura, T., A. Morimoto, M. Takanashi, S. Hibi, T. Sugimoto, E. Ishii, and S. Imashuku. 2002. Frequent co-expression of HoxA9 and Meis1 genes in infant acute lymphoblastic leukaemia with MLL rearrangement. *Br J Haematol* **119**: 119-21.
- Ito, M., C.X. Yuan, S. Malik, W. Gu, J.D. Fondell, S. Yamamura, Z.Y. Fu, X. Zhang, J. Qin, and R.G. Roeder. 1999. Identity between TRAP and SMCC complexes indicates novel pathways for the function of nuclear receptors and diverse mammalian activators. *Mol Cell* **3**: 361-70.
- Ito, Y. 2004. Oncogenic potential of the RUNX gene family: 'overview'. *Oncogene* **23**: 4198-208.
- Jaenisch, R. and A. Bird. 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **33 Suppl**: 245-54.
- Jaskelioff, M. and C.L. Peterson. 2003. Chromatin and transcription: histones continue to make their marks. *Nat Cell Biol* **5**: 395-9.
- Jenuwein, T. 2001. Re-SET-ting heterochromatin by histone methyltransferases. *Trends Cell Biol* **11**: 266-73.
- Jenuwein, T. and C.D. Allis. 2001. Translating the histone code. *Science* **293**: 1074-80.
- Johnson, C.N., N.L. Adkins, and P. Georgel. 2005. Chromatin remodeling complexes: ATP-dependent machines in action. *Biochem Cell Biol* **83**: 405-17.
- Jones, P.L., G.J. Veenstra, P.A. Wade, D. Vermaak, S.U. Kass, N. Landsberger, J. Strouboulis, and A.P. Wolffe. 1998. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* **19**: 187-91.
- Kaiser, K. and M. Meisterernst. 1996. The human general co-factors. *Trends Biochem Sci* **21**: 342-5.
- Kawagoe, H., R.K. Humphries, A. Blair, H.J. Sutherland, and D.E. Hogge. 1999. Expression of HOX genes, HOX cofactors, and MLL in phenotypically and functionally defined subpopulations of leukemic and normal human hematopoietic cells. *Leukemia* **13**: 687-98.
- Kennison, J.A. 2004. Introduction to Trx-G and Pc-G genes. *Methods Enzymol* **377**: 61-70.
- Khochbin, S., A. Verdell, C. Lemercier, and D. Seigneurin-Berny. 2001. Functional significance of histone deacetylase diversity. *Curr Opin Genet Dev* **11**: 162-6.
- Kim, Y.J., S. Bjorklund, Y. Li, M.H. Sayre, and R.D. Kornberg. 1994. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**: 599-608.

- Kobayashi, A., H. Yamagiwa, H. Hoshino, A. Muto, K. Sato, M. Morita, N. Hayashi, M. Yamamoto, and K. Igarashi. 2000. A combinatorial code for gene expression generated by transcription factor Bach2 and MAZR (MAZ-related factor) through the BTB/POZ domain. *Mol Cell Biol* **20**: 1733-46.
- Kobayashi, N., T.G. Boyer, and A.J. Berk. 1995. A class of activation domains interacts directly with TFIIA and stimulates TFIIA-TFIID-promoter complex assembly. *Mol Cell Biol* **15**: 6465-73.
- Kohlmann, A., C. Schoch, M. Dugas, S. Schnittger, W. Hiddemann, W. Kern, and T. Haferlach. 2005. New insights into MLL gene rearranged acute leukemias using gene expression profiling: shared pathways, lineage commitment, and partner genes. *Leukemia* **19**: 953-64.
- Koren, H.S., S.J. Anderson, and J.W. Larrick. 1979. In vitro activation of a human macrophage-like cell line. *Nature* **279**: 328-31.
- Kretzschmar, M., K. Kaiser, F. Lottspeich, and M. Meisterernst. 1994. A novel mediator of class II gene transcription with homology to viral immediate-early transcriptional regulators. *Cell* **78**: 525-34.
- Krogan, N.J., J. Dover, S. Khorrani, J.F. Greenblatt, J. Schneider, M. Johnston, and A. Shilatifard. 2002. COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. *J Biol Chem* **277**: 10753-5.
- Kroon, E., J. Kros, U. Thorsteinsdottir, S. Baban, A.M. Buchberg, and G. Sauvageau. 1998. Hoxa9 transforms primary bone marrow cells through specific collaboration with Meis1a but not Pbx1b. *Embo J* **17**: 3714-25.
- Kuzmichev, A., K. Nishioka, H. Erdjument-Bromage, P. Tempst, and D. Reinberg. 2002. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev* **16**: 2893-905.
- Lee, T.I., R.G. Jenner, L.A. Boyer, M.G. Guenther, S.S. Levine, R.M. Kumar, B. Chevalier, S.E. Johnstone, M.F. Cole, K. Isono, H. Koseki, T. Fuchikami, K. Abe, H.L. Murray, J.P. Zucker, B. Yuan, G.W. Bell, E. Herbolsheimer, N.M. Hannett, K. Sun, D.T. Odom, A.P. Otte, T.L. Volkert, D.P. Bartel, D.A. Melton, D.K. Gifford, R. Jaenisch, and R.A. Young. 2006. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**: 301-13.
- Lee, T.I. and R.A. Young. 2000. Transcription of eukaryotic protein-coding genes. *Annu Rev Genet* **34**: 77-137.
- Lehnertz, B., Y. Ueda, A.A. Derijck, U. Braunschweig, L. Perez-Burgos, S. Kubicek, T. Chen, E. Li, T. Jenuwein, and A.H. Peters. 2003. Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr Biol* **13**: 1192-200.
- Lemon, B. and R. Tjian. 2000. Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev* **14**: 2551-69.
- Li, Q., K.R. Peterson, X. Fang, and G. Stamatoyannopoulos. 2002. Locus control regions. *Blood* **100**: 3077-86.
- Luger, K., A.W. Mader, R.K. Richmond, D.F. Sargent, and T.J. Richmond. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**: 251-60.

- Ma, Q., H. Alder, K.K. Nelson, D. Chatterjee, Y. Gu, T. Nakamura, E. Canaani, C.M. Croce, L.D. Siracusa, and A.M. Buchberg. 1993. Analysis of the murine All-1 gene reveals conserved domains with human ALL-1 and identifies a motif shared with DNA methyltransferases. *Proc Natl Acad Sci U S A* **90**: 6350-4.
- Mackey, D. and B. Sugden. 1999. Applications of oriP plasmids and their mode of replication. *Methods Enzymol* **306**: 308-28.
- Marcu, K.B., S.A. Bossone, and A.J. Patel. 1992. myc function and regulation. *Annu Rev Biochem* **61**: 809-60.
- Marmorstein, R. and S.Y. Roth. 2001. Histone acetyltransferases: function, structure, and catalysis. *Curr Opin Genet Dev* **11**: 155-61.
- Martin, M.E., T.A. Milne, S. Bloyer, K. Galoian, W. Shen, D. Gibbs, H.W. Brock, R. Slany, and J.L. Hess. 2003. Dimerization of MLL fusion proteins immortalizes hematopoietic cells. *Cancer Cell* **4**: 197-207.
- Mattioni, T., J.F. Louvion, and D. Picard. 1994. Regulation of protein activities by fusion to steroid binding domains. *Methods Cell Biol* **43 Pt A**: 335-52.
- Meisterernst, M., A.L. Roy, H.M. Lieu, and R.G. Roeder. 1991. Activation of class II gene transcription by regulatory factors is potentiated by a novel activity. *Cell* **66**: 981-93.
- Michl, P., B. Knobel, and J. Downward. 2006. CUTL1 is phosphorylated by protein kinase A, modulating its effects on cell proliferation and motility. *J Biol Chem*.
- Michl, P., A.R. Ramjaun, O.E. Pardo, P.H. Warne, M. Wagner, R. Poulsom, C. D'Arrigo, K. Ryder, A. Menke, T. Gress, and J. Downward. 2005. CUTL1 is a target of TGF(beta) signaling that enhances cancer cell motility and invasiveness. *Cancer Cell* **7**: 521-32.
- Milne, T.A., S.D. Briggs, H.W. Brock, M.E. Martin, D. Gibbs, C.D. Allis, and J.L. Hess. 2002. MLL targets SET domain methyltransferase activity to Hox gene promoters. *Mol Cell* **10**: 1107-17.
- Milne, T.A., Y. Dou, M.E. Martin, H.W. Brock, R.G. Roeder, and J.L. Hess. 2005a. MLL associates specifically with a subset of transcriptionally active target genes. *Proc Natl Acad Sci U S A* **102**: 14765-70.
- Milne, T.A., C.M. Hughes, R. Lloyd, Z. Yang, O. Rozenblatt-Rosen, Y. Dou, R.W. Schnepf, C. Krankel, V.A. Livolsi, D. Gibbs, X. Hua, R.G. Roeder, M. Meyerson, and J.L. Hess. 2005b. Menin and MLL cooperatively regulate expression of cyclin-dependent kinase inhibitors. *Proc Natl Acad Sci U S A* **102**: 749-54.
- Milne, T.A., M.E. Martin, H.W. Brock, R.K. Slany, and J.L. Hess. 2005c. Leukemogenic MLL fusion proteins bind across a broad region of the Hox a9 locus, promoting transcription and multiple histone modifications. *Cancer Res* **65**: 11367-74.
- Mittler, G., T. Stuhler, L. Santolin, T. Uhlmann, E. Kremmer, F. Lottspeich, L. Berti, and M. Meisterernst. 2003. A novel docking site on Mediator is critical for activation by VP16 in mammalian cells. *Embo J* **22**: 6494-504.

- Miyoshi, A., Y. Kitajima, K. Sumi, K. Sato, A. Hagiwara, Y. Koga, and K. Miyazaki. 2004. Snail and SIP1 increase cancer invasion by upregulating MMP family in hepatocellular carcinoma cells. *Br J Cancer* **90**: 1265-73.
- Mizzen, C.A., X.J. Yang, T. Kokubo, J.E. Brownell, A.J. Bannister, T. Owen-Hughes, J. Workman, L. Wang, S.L. Berger, T. Kouzarides, Y. Nakatani, and C.D. Allis. 1996. The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. *Cell* **87**: 1261-70.
- Morris, J.F., R. Hromas, and F.J. Rauscher, 3rd. 1994. Characterization of the DNA-binding properties of the myeloid zinc finger protein MZF1: two independent DNA-binding domains recognize two DNA consensus sequences with a common G-rich core. *Mol Cell Biol* **14**: 1786-95.
- Mulero-Navarro, S., J.M. Carvajal-Gonzalez, M. Herranz, E. Ballestar, M.F. Fraga, S. Ropero, M. Esteller, and P.M. Fernandez-Salguero. 2006. The dioxin receptor is silenced by promoter hypermethylation in human acute lymphoblastic leukemia through inhibition of Sp1 binding. *Carcinogenesis* **27**: 1099-1104.
- Muller, J., C.M. Hart, N.J. Francis, M.L. Vargas, A. Sengupta, B. Wild, E.L. Miller, M.B. O'Connor, R.E. Kingston, and J.A. Simon. 2002. Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell* **111**: 197-208.
- Naar, A.M., P.A. Beurang, S. Zhou, S. Abraham, W. Solomon, and R. Tjian. 1999. Composite co-activator ARC mediates chromatin-directed transcriptional activation. *Nature* **398**: 828-32.
- Naar, A.M., B.D. Lemon, and R. Tjian. 2001. Transcriptional coactivator complexes. *Annu Rev Biochem* **70**: 475-501.
- Nagy, P.L., J. Griesenbeck, R.D. Kornberg, and M.L. Cleary. 2002. A trithorax-group complex purified from *Saccharomyces cerevisiae* is required for methylation of histone H3. *Proc Natl Acad Sci U S A* **99**: 90-4.
- Nakamura, T., T. Mori, S. Tada, W. Krajewski, T. Rozovskaia, R. Wassell, G. Dubois, A. Mazo, C.M. Croce, and E. Canaani. 2002. ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. *Mol Cell* **10**: 1119-28.
- Nan, X., H.H. Ng, C.A. Johnson, C.D. Laherty, B.M. Turner, R.N. Eisenman, and A. Bird. 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**: 386-9.
- Neely, K.E., A.H. Hassan, A.E. Wallberg, D.J. Steger, B.R. Cairns, A.P. Wright, and J.L. Workman. 1999. Activation domain-mediated targeting of the SWI/SNF complex to promoters stimulates transcription from nucleosome arrays. *Mol Cell* **4**: 649-55.
- Nepveu, A. 2001. Role of the multifunctional CDP/Cut/Cux homeodomain transcription factor in regulating differentiation, cell growth and development. *Gene* **270**: 1-15.
- Neufeld, E.J., D.G. Skalnik, P.M. Lievens, and S.H. Orkin. 1992. Human CCAAT displacement protein is homologous to the Drosophila homeoprotein, cut. *Nat Genet* **1**: 50-5.

- Ng, H.H., F. Robert, R.A. Young, and K. Struhl. 2003. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell* **11**: 709-19.
- Ng, H.H., Y. Zhang, B. Hendrich, C.A. Johnson, B.M. Turner, H. Erdjument-Bromage, P. Tempst, D. Reinberg, and A. Bird. 1999. MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat Genet* **23**: 58-61.
- Nishikura, K. 1986. Sequences involved in accurate and efficient transcription of human c-myc genes microinjected into frog oocytes. *Mol Cell Biol* **6**: 4093-8.
- Nishio, H. and M.J. Walsh. 2004. CCAAT displacement protein/cut homolog recruits G9a histone lysine methyltransferase to repress transcription. *Proc Natl Acad Sci U S A* **101**: 11257-62.
- Orlando, V. 2003. Polycomb, epigenomes, and control of cell identity. *Cell* **112**: 599-606.
- Orphanides, G. and D. Reinberg. 2002. A unified theory of gene expression. *Cell* **108**: 439-51.
- Ottolenghi, S., R. Mantovani, S. Nicolis, A. Ronchi, and B. Giglioni. 1989. DNA sequences regulating human globin gene transcription in nondeletional hereditary persistence of fetal hemoglobin. *Hemoglobin* **13**: 523-41.
- Pabo, C.O. and R.T. Sauer. 1992. Transcription factors: structural families and principles of DNA recognition. *Annu Rev Biochem* **61**: 1053-95.
- Palancade, B. and O. Bensaude. 2003. Investigating RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation. *Eur J Biochem* **270**: 3859-70.
- Peterson, C.L. and M.A. Laniel. 2004. Histones and histone modifications. *Curr Biol* **14**: R546-51.
- Pinhero, R., P. Liaw, K. Bertens, and K. Yankulov. 2004. Three cyclin-dependent kinases preferentially phosphorylate different parts of the C-terminal domain of the large subunit of RNA polymerase II. *Eur J Biochem* **271**: 1004-14.
- Qiu, C., K. Sawada, X. Zhang, and X. Cheng. 2002. The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. *Nat Struct Biol* **9**: 217-24.
- Quandt, K., K. Frech, H. Karas, E. Wingender, and T. Werner. 1995. MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res* **23**: 4878-84.
- Ringrose, L., M. Rehmsmeier, J.M. Dura, and R. Paro. 2003. Genome-wide prediction of Polycomb/Trithorax response elements in *Drosophila melanogaster*. *Dev Cell* **5**: 759-71.
- Roeder, R.G. and W.J. Rutter. 1969. Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature* **224**: 234-7.
- Rozovskaia, T., E. Feinstein, O. Mor, R. Foa, J. Blechman, T. Nakamura, C.M. Croce, G. Cimino, and E. Canaani. 2001. Upregulation of Meis1 and HoxA9 in acute lymphocytic leukemias with the t(4 : 11) abnormality. *Oncogene* **20**: 874-8.

- Rozovskaia, T., O. Ravid-Amir, S. Tillib, G. Getz, E. Feinstein, H. Agrawal, A. Nagler, E.F. Rappaport, I. Issaeva, Y. Matsuo, U.R. Kees, T. Lapidot, F. Lo Coco, R. Foa, A. Mazo, T. Nakamura, C.M. Croce, G. Cimino, E. Domany, and E. Canaani. 2003. Expression profiles of acute lymphoblastic and myeloblastic leukemias with ALL-1 rearrangements. *Proc Natl Acad Sci U S A* **100**: 7853-8.
- Santos-Rosa, H., R. Schneider, B.E. Bernstein, N. Karabetsou, A. Morillon, C. Weise, S.L. Schreiber, J. Mellor, and T. Kouzarides. 2003. Methylation of histone H3 K4 mediates association of the Isw1p ATPase with chromatin. *Mol Cell* **12**: 1325-32.
- Sarraf, S.A. and I. Stancheva. 2004. Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Mol Cell* **15**: 595-605.
- Sato, S., C. Tomomori-Sato, T.J. Parmely, L. Florens, B. Zybilov, S.K. Swanson, C.A. Banks, J. Jin, Y. Cai, M.P. Washburn, J.W. Conaway, and R.C. Conaway. 2004. A set of consensus mammalian mediator subunits identified by multidimensional protein identification technology. *Mol Cell* **14**: 685-91.
- Scherf, M., A. Epple, and T. Werner. 2005. The next generation of literature analysis: integration of genomic analysis into text mining. *Brief Bioinform* **6**: 287-97.
- Schneider, R., A.J. Bannister, F.A. Myers, A.W. Thorne, C. Crane-Robinson, and T. Kouzarides. 2004. Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat Cell Biol* **6**: 73-7.
- Schnittger, S., U. Kinkelin, C. Schoch, A. Heinecke, D. Haase, T. Haferlach, T. Buchner, B. Wormann, W. Hiddemann, and F. Griesinger. 2000. Screening for MLL tandem duplication in 387 unselected patients with AML identify a prognostically unfavorable subset of AML. *Leukemia* **14**: 796-804.
- Schoch, C., A. Kohlmann, S. Schnittger, B. Brors, M. Dugas, S. Mergenthaler, W. Kern, W. Hiddemann, R. Eils, and T. Haferlach. 2002. Acute myeloid leukemias with reciprocal rearrangements can be distinguished by specific gene expression profiles. *Proc Natl Acad Sci U S A* **99**: 10008-13.
- Schraets, D., T. Lehmann, T. Dingermann, and R. Marschalek. 2003. MLL-mediated transcriptional gene regulation investigated by gene expression profiling. *Oncogene* **22**: 3655-68.
- Schroeder. 2001. Untersuchung der Funktion von aktiviertem Notch bei der Regulation von Proliferation und Differenzierung haematopoietischer Zellen. In *Faculty of Science*. Friedrich-Alexander-Universitaet, Erlangen-Nuernberg.
- Shi, Y., F. Lan, C. Matson, P. Mulligan, J.R. Whetstine, P.A. Cole, and R.A. Casero. 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**: 941-53.
- Silverman, N. and T. Maniatis. 2001. NF-kappaB signaling pathways in mammalian and insect innate immunity. *Genes Dev* **15**: 2321-42.
- Slany, R.K., C. Lavau, and M.L. Cleary. 1998. The oncogenic capacity of HRX-ENL requires the transcriptional transactivation activity of ENL and the DNA binding motifs of HRX. *Mol Cell Biol* **18**: 122-9.

- Smale, S.T. and J.T. Kadonaga. 2003. The RNA polymerase II core promoter. *Annu Rev Biochem* **72**: 449-79.
- So, C.W. and M.L. Cleary. 2003. Common mechanism for oncogenic activation of MLL by forkhead family proteins. *Blood* **101**: 633-9.
- So, C.W., H. Karsunky, E. Passegue, A. Cozzio, I.L. Weissman, and M.L. Cleary. 2003. MLL-GAS7 transforms multipotent hematopoietic progenitors and induces mixed lineage leukemias in mice. *Cancer Cell* **3**: 161-71.
- So, C.W., H. Karsunky, P. Wong, I.L. Weissman, and M.L. Cleary. 2004. Leukemic transformation of hematopoietic progenitors by MLL-GAS7 in the absence of Hoxa7 or Hoxa9. *Blood* **103**: 3192-9.
- Stewart, M., A. Terry, M. Hu, M. O'Hara, K. Blyth, E. Baxter, E. Cameron, D.E. Onions, and J.C. Neil. 1997. Proviral insertions induce the expression of bone-specific isoforms of PEBP2alphaA (CBFA1): evidence for a new myc collaborating oncogene. *Proc Natl Acad Sci U S A* **94**: 8646-51.
- Stewart, T.A., A.R. Bellve, and P. Leder. 1984. Transcription and promoter usage of the myc gene in normal somatic and spermatogenic cells. *Science* **226**: 707-10.
- Strehl, S., A. Borkhardt, R. Slany, U.E. Fuchs, M. Konig, and O.A. Haas. 2003. The human LASP1 gene is fused to MLL in an acute myeloid leukemia with t(11;17)(q23;q21). *Oncogene* **22**: 157-60.
- Sugden, B. and E.R. Leight. 2001. EBV's plasmid replicon: an enigma in cis and trans. *Curr Top Microbiol Immunol* **258**: 3-11.
- Sundstrom, C. and K. Nilsson. 1976. Establishment and characterization of a human histiocytic lymphoma cell line (U-937). *Int J Cancer* **17**: 565-77.
- Taub, R., C. Moulding, J. Battey, W. Murphy, T. Vasicek, G.M. Lenoir, and P. Leder. 1984. Activation and somatic mutation of the translocated c-myc gene in burkitt lymphoma cells. *Cell* **36**: 339-48.
- Terranova, R., H. Agherbi, A. Boned, S. Meresse, and M. Djabali. 2006. Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll. *Proc Natl Acad Sci U S A* **103**: 6629-34.
- Thomas, J.O. and R.D. Kornberg. 1975. An octamer of histones in chromatin and free in solution. *Proc Natl Acad Sci U S A* **72**: 2626-30.
- Tkachuk, D.C., S. Kohler, and M.L. Cleary. 1992. Involvement of a homolog of *Drosophila trithorax* by 11q23 chromosomal translocations in acute leukemias. *Cell* **71**: 691-700.
- Triezenberg, S.J., R.C. Kingsbury, and S.L. McKnight. 1988. Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression. *Genes Dev* **2**: 718-29.
- Truscott, M., L. Raynal, P. Premdas, B. Goulet, L. Leduy, G. Berube, and A. Nepveu. 2003. CDP/Cux stimulates transcription from the DNA polymerase alpha gene promoter. *Mol Cell Biol* **23**: 3013-28.

- Tsukada, Y., J. Fang, H. Erdjument-Bromage, M.E. Warren, C.H. Borchers, P. Tempst, and Y. Zhang. 2006. Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**: 811-6.
- Uesugi, M., O. Nyanguile, H. Lu, A.J. Levine, and G.L. Verdine. 1997. Induced alpha helix in the VP16 activation domain upon binding to a human TAF. *Science* **277**: 1310-3.
- Utle, R.T., K. Ikeda, P.A. Grant, J. Cote, D.J. Steger, A. Eberharter, S. John, and J.L. Workman. 1998. Transcriptional activators direct histone acetyltransferase complexes to nucleosomes. *Nature* **394**: 498-502.
- Van Dyke, M.W., R.G. Roeder, and M. Sawadogo. 1988. Physical analysis of transcription preinitiation complex assembly on a class II gene promoter. *Science* **241**: 1335-8.
- Verrijzer, C.P. and R. Tjian. 1996. TAFs mediate transcriptional activation and promoter selectivity. *Trends Biochem Sci* **21**: 338-42.
- Wade, P.A., A. Geronne, P.L. Jones, E. Ballestar, F. Aubry, and A.P. Wolffe. 1999. Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. *Nat Genet* **23**: 62-6.
- Warnmark, A., E. Treuter, A.P. Wright, and J.A. Gustafsson. 2003. Activation functions 1 and 2 of nuclear receptors: molecular strategies for transcriptional activation. *Mol Endocrinol* **17**: 1901-9.
- Wassarman, D.A. and F. Sauer. 2001. TAF(II)250: a transcription toolbox. *J Cell Sci* **114**: 2895-902.
- Watt, F. and P.L. Molloy. 1988. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev* **2**: 1136-43.
- Wong, C.W. and M.L. Privalsky. 1998. Components of the SMRT corepressor complex exhibit distinctive interactions with the POZ domain oncoproteins PLZF, PLZF-RARalpha, and BCL-6. *J Biol Chem* **273**: 27695-702.
- Xia, Z.B., M. Anderson, M.O. Diaz, and N.J. Zeleznik-Le. 2003. MLL repression domain interacts with histone deacetylases, the polycomb group proteins HPC2 and BMI-1, and the corepressor C-terminal-binding protein. *Proc Natl Acad Sci U S A* **100**: 8342-7.
- Xia, Z.B., R. Popovic, J. Chen, C. Theisler, T. Stuart, D.A. Santillan, F. Erfurth, M.O. Diaz, and N.J. Zeleznik-Le. 2005. The MLL fusion gene, MLL-AF4, regulates cyclin-dependent kinase inhibitor CDKN1B (p27kip1) expression. *Proc Natl Acad Sci U S A* **102**: 14028-33.
- Xiao, H., A. Pearson, B. Coulombe, R. Truant, S. Zhang, J.L. Regier, S.J. Triezenberg, D. Reinberg, O. Flores, C.J. Ingles, and et al. 1994. Binding of basal transcription factor TFIIH to the acidic activation domains of VP16 and p53. *Mol Cell Biol* **14**: 7013-24.
- Yano, T., T. Nakamura, J. Blechman, C. Sorio, C.V. Dang, B. Geiger, and E. Canaani. 1997. Nuclear punctate distribution of ALL-1 is conferred by distinct elements at the N terminus of the protein. *Proc Natl Acad Sci U S A* **94**: 7286-91.

- Yokoyama, A., I. Kitabayashi, P.M. Ayton, M.L. Cleary, and M. Ohki. 2002. Leukemia proto-oncoprotein MLL is proteolytically processed into 2 fragments with opposite transcriptional properties. *Blood* **100**: 3710-8.
- Yokoyama, A., T.C. Somervaille, K.S. Smith, O. Rozenblatt-Rosen, M. Meyerson, and M.L. Cleary. 2005. The menin tumor suppressor protein is an essential oncogenic cofactor for MLL-associated leukemogenesis. *Cell* **123**: 207-18.
- Yokoyama, A., Z. Wang, J. Wysocka, M. Sanyal, D.J. Aufiero, I. Kitabayashi, W. Herr, and M.L. Cleary. 2004. Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. *Mol Cell Biol* **24**: 5639-49.
- Yu, B.D., R.D. Hanson, J.L. Hess, S.E. Horning, and S.J. Korsmeyer. 1998. MLL, a mammalian trithorax-group gene, functions as a transcriptional maintenance factor in morphogenesis. *Proc Natl Acad Sci U S A* **95**: 10632-6.
- Yu, B.D., J.L. Hess, S.E. Horning, G.A. Brown, and S.J. Korsmeyer. 1995. Altered Hox expression and segmental identity in Mll-mutant mice. *Nature* **378**: 505-8.
- Zeisig, B.B., T. Milne, M.P. Garcia-Cuellar, S. Schreiner, M.E. Martin, U. Fuchs, A. Borkhardt, S.K. Chanda, J. Walker, R. Soden, J.L. Hess, and R.K. Slany. 2004a. Hoxa9 and Meis1 are key targets for MLL-ENL-mediated cellular immortalization. *Mol Cell Biol* **24**: 617-28.
- Zeisig, B.B., T. Milne, M.P. Garcia-Cuellar, S. Schreiner, M.E. Martin, U. Fuchs, A. Borkhardt, S.K. Chanda, J. Walker, R. Soden, J.L. Hess, R.K. Slany, K.R. Calvo, D.B. Sykes, M.P. Pasillas, and M.P. Kamps. 2004b. Hoxa9 and Meis1 are key targets for MLL-ENL-mediated cellular immortalization
- Nup98-HoxA9 immortalizes myeloid progenitors, enforces expression of Hoxa9, Hoxa7 and Meis1, and alters cytokine-specific responses in a manner similar to that induced by retroviral co-expression of Hoxa9 and Meis1. *Mol Cell Biol* **24**: 617-28.
- Zeisig, B.B., S. Schreiner, M.P. Garcia-Cuellar, and R.K. Slany. 2003. Transcriptional activation is a key function encoded by MLL fusion partners. *Leukemia* **17**: 359-65.
- Zeleznik-Le, N.J., A.M. Harden, and J.D. Rowley. 1994. 11q23 translocations split the "AT-hook" cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the mixed-lineage leukemia (MLL) gene. *Proc Natl Acad Sci U S A* **91**: 10610-4.
- Zhang, X., Z. Yang, S.I. Khan, J.R. Horton, H. Tamaru, E.U. Selker, and X. Cheng. 2003. Structural basis for the product specificity of histone lysine methyltransferases. *Mol Cell* **12**: 177-85.
- Zhang, Y., H.H. Ng, H. Erdjument-Bromage, P. Tempst, A. Bird, and D. Reinberg. 1999. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev* **13**: 1924-35.
- Zhou, H., S. Spicuglia, J.J. Hsieh, D.J. Mitsiou, T. Hoiby, G.J. Veenstra, S.J. Korsmeyer, and H.G. Stunnenberg. 2006. Uncleaved TFIIA is a substrate for taspase 1 and active in transcription. *Mol Cell Biol* **26**: 2728-35.

- Zhou, J., C. Chau, Z. Deng, W. Stedman, and P.M. Lieberman. 2005. Epigenetic control of replication origins. *Cell Cycle* **4**: 889-92.
- Zhu, H., V. Joliot, and R. Prywes. 1994. Role of transcription factor TFIIIF in serum response factor-activated transcription. *J Biol Chem* **269**: 3489-97.
- Zhu, Q., U. Maitra, D. Johnston, M. Lozano, and J.P. Dudley. 2004. The homeodomain protein CDP regulates mammary-specific gene transcription and tumorigenesis. *Mol Cell Biol* **24**: 4810-23.
- Ziemin-van der Poel, S., N.R. McCabe, H.J. Gill, R. Espinosa, III, Y. Patel, A. Harden, P. Rubinelli, S.D. Smith, M.M. LeBeau, J.D. Rowley, and et al. 1991. Identification of a gene, MLL, that spans the breakpoint in 11q23 translocations associated with human leukemias. *Proc Natl Acad Sci U S A* **88**: 10735-9.
- Zollman, S., D. Godt, G.G. Prive, J.L. Couderc, and F.A. Laski. 1994. The BTB domain, found primarily in zinc finger proteins, defines an evolutionarily conserved family that includes several developmentally regulated genes in *Drosophila*. *Proc Natl Acad Sci U S A* **91**: 10717-21.

6. Appendix

6.1 Alignment of Hoxa9 promoter variants

WT -78/+46	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
WT -118/+46	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
WT -207/+46	C	C	C	A	T	C	G	T	A	G	A	G	C	G	G	C	A	C	G	A
Hox site mut1 -207/+46	C	C	C	A	T	C	G	T	A	G	A	G	C	G	G	C	A	C	G	A
GGGC mut1 -207/+46	C	C	C	A	T	C	G	T	A	G	A	G	C	G	G	C	A	C	G	A
CpG mut1 -207/+46	C	C	C	A	T	C	G	T	A	G	A	G	C	G	G	C	A	C	G	A
GGC mut -207/+46	C	C	C	A	T	C	G	T	A	G	A	G	C	G	G	C	A	C	G	A
SP1 sites mut -207/+46	C	C	C	A	T	C	G	T	A	G	A	G	C	G	G	C	A	C	G	A
GC mut -207/+46	C	C	T	A	T	C	G	T	A	G	A	G	C	T	G	C	A	C	G	A
CpG mut2 -118/+46	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GGGC mut2 -118/+46	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Hox site mut2 -118/+46	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AdML-TATA -207/+46	C	C	C	A	T	C	G	T	A	G	A	G	C	G	G	C	A	C	G	A
TATA-less1 -207/+46	C	C	C	A	T	C	G	T	A	G	A	G	C	G	G	C	A	C	G	A
TATA-less2 -207/+46	C	C	C	A	T	C	G	T	A	G	A	G	C	G	G	C	A	C	G	A
AdML-Inr -207/+46	C	C	C	A	T	C	G	T	A	G	A	G	C	G	G	C	A	C	G	A
C-stretch mut -207/+46	C	C	C	A	T	C	G	T	A	G	A	G	C	G	G	C	A	C	G	A

WT -78/+46	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
WT -118/+46	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
WT -207/+46	T	C	C	C	T	T	T	A	C	A	T	A	A	A	A	A	C	A	T	A
Hox site mut1 -207/+46	T	C	C	C	T	T	T	A	C	A	T	A	A	A	A	A	C	A	T	A
GGGC mut1 -207/+46	T	C	C	C	T	T	T	A	C	A	T	A	A	A	A	A	C	A	T	A
CpG mut1 -207/+46	T	C	C	C	T	T	T	A	C	A	T	A	A	A	A	A	C	A	T	A
GGC mut -207/+46	T	C	C	C	T	T	T	A	C	A	T	A	A	A	A	A	C	A	T	A
SP1 sites mut -207/+46	T	C	C	C	T	T	T	A	C	A	T	A	A	A	A	A	C	A	T	A
GC mut -207/+46	T	C	A	C	T	T	T	A	C	A	T	A	A	A	A	A	C	A	T	A
CpG mut2 -118/+46	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GGGC mut2 -118/+46	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Hox site mut2 -118/+46	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AdML-TATA -207/+46	T	C	C	C	T	T	T	A	C	A	T	A	A	A	A	A	C	A	T	A
TATA-less1 -207/+46	T	C	C	C	T	T	T	A	C	A	T	A	A	A	A	A	C	A	T	A
TATA-less2 -207/+46	T	C	C	C	T	T	T	A	C	A	T	A	A	A	A	A	C	A	T	A
AdML-Inr -207/+46	T	C	C	C	T	T	T	A	C	A	T	A	A	A	A	A	C	A	T	A
C-stretch mut -207/+46	T	C	C	C	T	T	T	A	C	A	T	A	A	A	A	A	C	A	T	A

WT -78/+46	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
WT -118/+46	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
WT -207/+46	T	G	G	C	T	T	T	T	G	C	T	A	T	A	A	A	A	A	T	T
Hox site mut1 -207/+46	T	G	G	C	T	T	T	T	G	C	T	A	T	A	A	A	A	A	G	C
GGGC mut1 -207/+46	T	G	G	C	T	T	T	T	G	C	T	A	T	A	A	A	A	A	T	T
CpG mut1 -207/+46	T	G	G	C	T	T	T	T	G	C	T	A	T	A	A	A	A	A	T	T
GGC mut -207/+46	T	G	G	C	T	T	T	T	G	C	T	A	T	A	A	A	A	A	T	T
SP1 sites mut -207/+46	T	G	G	C	T	T	T	T	G	C	T	A	T	A	A	A	A	A	T	T
GC mut -207/+46	T	G	A	C	T	T	T	T	G	C	T	A	T	A	A	A	A	A	T	T
CpG mut2 -118/+46	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GGGC mut2 -118/+46	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Hox site mut2 -118/+46	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AdML-TATA -207/+46	T	G	G	C	T	T	T	T	G	C	T	A	T	A	A	A	A	A	T	T
TATA-less1 -207/+46	T	G	G	C	T	T	T	T	G	C	T	A	T	A	A	A	A	A	T	T
TATA-less2 -207/+46	T	G	G	C	T	T	T	T	G	C	T	A	T	A	A	A	A	A	T	T
AdML-Inr -207/+46	T	G	G	C	T	T	T	T	G	C	T	A	T	A	A	A	A	A	T	T
C-stretch mut -207/+46	T	G	G	C	T	T	T	T	G	C	T	A	T	A	A	A	A	A	T	T

WT -78/+46	C	C	G	T	G	G	G	T	C	G	G	G	C
WT -118/+46	C	C	G	T	G	G	G	T	C	G	G	G	C
WT -207/+46	C	C	G	T	G	G	G	T	C	G	G	G	C
Hox site mut1 -207/+46	C	C	G	T	G	G	G	T	C	G	G	G	C
GGGC mut1 -207/+46	C	C	G	T	G	G	G	T	C	G	G	G	C
CpG mut1 -207/+46	C	C	G	T	G	G	G	T	C	G	G	G	C
GGC mut -207/+46	C	C	G	T	G	G	G	T	C	G	G	G	C
SP1 sites mut -207/+46	C	C	G	T	G	G	G	T	C	G	G	G	C
GC mut -207/+46	C	C	G	T	G	G	G	T	C	G	G	G	C
CpG mut2 -118/+46	C	C	G	T	G	G	G	T	C	G	G	G	C
GGGC mut2 -118/+46	C	C	G	T	G	G	G	T	C	G	G	G	C
Hox site mut2 -118/+46	C	C	G	T	G	G	G	T	C	G	G	G	C
AdML-TATA -207/+46	C	C	G	T	G	G	G	T	C	G	G	G	C
TATA-less1 -207/+46	C	C	G	T	G	G	G	T	C	G	G	G	C
TATA-less2 -207/+46	C	C	G	T	G	G	G	T	C	G	G	G	C
AdML-Inr -207/+46	C	C	G	T	G	G	G	T	C	G	G	G	C
C-stretch mut -207/+46	C	C	G	T	G	G	G	T	C	G	G	G	C

6.2 Microarray results

Table 16. Genes upregulated 30 min after induction of MLL–VP16–ER–HA.

		fold change:
Ccm1	cerebral cavernous malformations 1	3,2
Phf14	PHD finger protein 14	3,0
Zfx1b	RIKEN cDNA 5830411K21 gene	2,6
Pacs1	RIKEN cDNA D430030G11 gene	2,5
Nedd4l	Neural precursor cell expressed, developmentally down-regulated gene 4-like	2,5
D230007K08Rik	RIKEN cDNA D230007K08 gene	2,3
Pfn3	profilin 3	2,3
Myo1b	myosin IB	2,3
Sec81	SEC8-like 1 (<i>S. cerevisiae</i>)	2,1
3110048L19Rik	RIKEN cDNA 3110048L19 gene	2,1
Mbnl1	Muscleblind-like 1 (<i>Drosophila</i>)	2,1
4930428J16Rik	RIKEN cDNA 4930428J16 gene	2,1
Ern1	Endoplasmic reticulum (ER) to nucleus signalling 1	2,0
A130071D04Rik	RIKEN cDNA A130071D04 gene	2,0
1700034P14Rik	RIKEN cDNA 1700034P14 gene	2,0
Tox	Thymocyte selection-associated HMG box gene	2,0
Trim59	Tripartite motif-containing 59	2,0
5930436O19Rik	RIKEN cDNA 5930436O19 gene	2,0
D230040A04Rik	RIKEN cDNA D230040A04 gene	2,0
Ptpn12	Protein tyrosine phosphatase, non-receptor type 12	2,0
Ureb1	Upstream regulatory element binding protein 1e	2,0
Clpx	Caseinolytic protease X (<i>E.coli</i>)	2,0
Dgkd	Diacylglycerol kinase, delta	2,0

Table 17. Genes downregulated 30 min after induction of MLL–VP16–ER–HA

Osm	oncostatin M	2
-----	--------------	---

Table 18. Genes upregulated 3 hours after induction of MLL–VP16–ER–HA

Mbnl1	Muscleblind-like 1 (<i>Drosophila</i>)	4.6
Ccm1	cerebral cavernous malformations 1	4.6
Psme4	proteasome (prosome, macropain) activator subunit 4	3.7
2810013C04Rik	RIKEN cDNA 2810013C04 gene	3.2
Pcdh7	protocadherin 7	3.0
Erb2ip	Erb2 interacting protein	3.0

Ss18	synovial sarcoma translocation, Chromosome 18	2.8
Enah	Enabled homolog (Drosophila)	2.8
Dleu2	Deleted in lymphocytic leukemia, 2	2.8
Sdccag33	Serologically defined colon cancer antigen 33	2.6
Pacs1	RIKEN cDNA D430030G11 gene	2.6
D130037M23Rik	RIKEN cDNA D130037M23 gene	2.6
Ankib1	Ankyrin repeat and IBR domain containing 1	2.6
4921505C17Rik	RIKEN cDNA 4921505C17 gene	2.6
3110048L19Rik	RIKEN cDNA 3110048L19 gene	2.6
2900075B16Rik	RIKEN cDNA 2900075B16 gene	2.6
Zfhx1a	Zinc finger homeobox 1a	2.5
Ssbp2	single-stranded DNA binding protein 2	2.5
Smc4l1	SMC4 structural maintenance of chromosomes 4-like 1 (yeast)	2.5
Sca1	Spinocerebellar ataxia 1 homolog (human)	2.5
Phf14	PHD finger protein 14	2.5
Nfe2l2	nuclear factor, erythroid derived 2, like 2	2.5
Kcnq5	potassium voltage-gated channel, subfamily Q, member 5	2.5
Depdc5	DEP domain containing 5	2.5
Bcl2	B-cell leukemia/lymphoma 2	2.5
A130071D04Rik	RIKEN cDNA A130071D04 gene	2.5
A130012E19Rik	RIKEN cDNA A130012E19 gene	2.5
5830474E16Rik	RIKEN cDNA 5830474E16 gene	2.5
5330401F18Rik	RIKEN cDNA 5330401F18 gene	2.5
Zfhx1b	RIKEN cDNA 5830411K21 gene	2.3
Uty	Ubiquitously transcribed tetratricopeptide repeat gene, Y chromosome	2.3
Utx	Ubiquitously transcribed tetratricopeptide repeat gene, X chromosome	2.3
Usp47	ubiquitin specific protease 47	2.3
Tox	Thymocyte selection-associated HMG box gene	2.3
Tbl1x	Transducin (beta)-like 1 X-linked	2.3
Nrip1	nuclear receptor interacting protein 1	2.3
Nedd4l	Neural precursor cell expressed, developmentally down-regulated gene 4-like	2.3
Meis1	Myeloid ecotropic viral integration site 1	2.3
Lmo4	LIM domain only 4	2.3
Jmjd1c	Jumonji domain containing 1C	2.3
Herc4	hect domain and RLD 4	2.3
Galnt7	UDP-N-acetyl-alpha-D-galactosamine: polypeptide N-acetylgalactosaminyl-transferase 7	2.3

E130108L08Rik	RIKEN cDNA E130108L08 gene	2.3
Cdk6	Cyclin-dependent kinase 6	2.3
BC025872	cDNA sequence BC025872	2.3
BC013481	CDNA sequence BC013481	2.3
9630026M06Rik	RIKEN cDNA 9630026M06 gene	2.3
5930436O19Rik	RIKEN cDNA 5930436O19 gene	2.3
5830435C13Rik	RIKEN cDNA 5830435C13 gene	2.3
5730555F13Rik	RIKEN cDNA 5730555F13 gene	2.3
1810043J12Rik	RIKEN cDNA 1810043J12 gene	2.3
St8sia4	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 4	2.1
Ptpn12	Protein tyrosine phosphatase, non-receptor type 12	2.1
Phf3	PHD finger protein 3	2.1
Nfat5	Nuclear factor of activated T-cells 5	2.1
Mospd2	motile sperm domain containing 2	2.1
Map2k4	Mitogen activated protein kinase kinase 4	2.1
Malt1	mucosa associated lymphoid tissue lymphoma translocation gene 1	2.1
LOC432971	hypothetical gene supported by AK038224	2.1
Herc1	Hect (homologous to the E6-AP (UBE3A) carboxyl terminus) domain and RCC1 (CHC1)-like domain (RLD) 1	2.1
Ggnbp2	Gametogenetin binding protein 2	2.1
Fli1	Friend leukemia integration 1	2.1
Fancc	Fanconi anemia, complementation group C	2.1
D030065N23Rik	RIKEN cDNA D030065N23 gene	2.1
Cdkn2c	cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4)	2.1
Cdkn1b	Cyclin-dependent kinase inhibitor 1B (P27)	2.1
Cacnb2	Calcium channel, voltage-dependent, beta 2 subunit	2.1
Bhc80	BRAF35/HDAC2 complex	2.1
Arid1b	AT rich interactive domain 1B (Swi1 like)	2.1
Ankra2	ankyrin repeat, family A (RFXANK-like), 2	2.1
AA407452	EST AA407452	2.1
A330103N21Rik	RIKEN cDNA A330103N21 gene	2.1
9030227G01Rik	RIKEN cDNA 9030227G01 gene	2.1
5830407E08Rik	RIKEN cDNA 5830407E08 gene	2.1
4833416J08Rik	RIKEN cDNA 4833416J08 gene	2.1
2900001A12Rik	Ankyrin repeat domain 12	2.1
2810436B12Rik	RIKEN cDNA 2810436B12 gene	2.1
2600011C06Rik	RIKEN cDNA 2600011C06 gene	2.1

2010109K09Rik	RIKEN cDNA 2010109K09 gene	2.1
1700081L11Rik	RIKEN cDNA 1700081L11 gene	2.1
1110001A05Rik	RIKEN cDNA 1110001A05 gene	2.1
Wac	WW domain containing adaptor with coiled-coil	2.0
Tpm1	tropomyosin 1, alpha	2.0
Tes3	Testis derived transcript 3	2.0
Stk17b	serine/threonine kinase 17b (apoptosis-inducing)	2.0
Senp6	SUMO/sentrin specific protease 6	2.0
Sec81	SEC8-like 1 (<i>S. cerevisiae</i>)	2.0
Psmd14	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 14	2.0
Ppp3ca	Protein phosphatase 3, catalytic subunit, alpha isoform	2.0
Phip	pleckstrin homology domain interacting protein	2.0
Pctk2	PCTAIRE-motif protein kinase 2	2.0
Npat	nuclear protein in the AT region	2.0
Mtdh	Metadherin	2.0
Mpp6	Membrane protein, palmitoylated 6 (MAGUK p55 subfamily member 6)	2.0
Mbnl3	Muscleblind-like 3 (<i>Drosophila</i>)	2.0
Jarid2	Jumonji, AT rich interactive domain 2	2.0
Irf2	Interferon regulatory factor 2	2.0
Fndc3a	fibronectin type III domain containing 3a	2.0
Fkbp5	RIKEN cDNA 6030422H21 gene	2.0
Diap1	Diaphanous homolog 1 (<i>Drosophila</i>)	2.0
Cutl1	RIKEN cDNA 2600010L24 gene	2.0
Chm	Choroideremia	2.0
Bcor	Bcl6 interacting corepressor	2.0
BC017647	CDNA sequence BC017647	2.0
Axot	Axotrophin	2.0
A630072M18Rik	RIKEN cDNA A630072M18 gene	2.0
9430034F23Rik	RIKEN cDNA 9430034F23 gene	2.0
9030406N13Rik	RIKEN cDNA 9030406N13 gene	2.0
6720463M24Rik	RIKEN cDNA 6720463M24 gene	2.0
5832424M12Rik	RIKEN cDNA 5832424M12 gene	2.0
4631422O05Rik	RIKEN cDNA 4631422O05 gene	2.0
2610005L07Rik	RIKEN cDNA 2610005L07 gene	2.0

Table 19. Genes downregulated 3 hours after induction of MLL–VP16–ER–HA

Egr3	early growth response 3	3.2
Nmyc1	neuroblastoma myc-related oncogene 1	2.5
Cxcl2	chemokine (C-X-C motif) ligand 2	2.1
Il6	interleukin 6	2.0
Osm	oncostatin M	2.0
2310006J04Rik	RIKEN cDNA 2310006J04 gene	2.0

Table 20. Genes upregulated 12 hours after induction of MLL–VP16–ER–HA

D130037M23Rik	RIKEN cDNA D130037M23 gene	9.2
Ccm1	cerebral cavernous malformations 1	6.1
Pank1	pantothenate kinase 1	5.3
Mbnl1	Muscleblind-like 1 (Drosophila)	4.6
Lpxn	leupaxin	4.6
BC022960	cDNA sequence BC022960	4.6
1110034A24Rik	RIKEN cDNA 1110034A24 gene	4.6
Atp8a1	ATPase, aminophospholipid transporter (APLT), class I, type 8A, member 1	4.0
Erb2ip	Erb2 interacting protein	3.7
Dleu2	Deleted in lymphocytic leukemia, 2	3.7
Cdkn2c	cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4)	3.7
Pcdh7	protocadherin 7	3.5
LOC319225	hypothetical LOC319225	3.5
Jmjd1c	Jumonji domain containing 1C	3.5
Clec2i	C-type lectin domain family 2, member i	3.5
B930096F20Rik	RIKEN cDNA B930096F20 gene	3.5
Ankib1	Ankyrin repeat and IBR domain containing 1	3.5
Galnt7	UDP-N-acetyl-alpha-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase 7	3.2
Elf1	E74-like factor 1	3.2
D5Wsu152e	DNA segment, Chr 5, Wayne State University 152, expressed	3.2
5830435C13Rik	RIKEN cDNA 5830435C13 gene	3.2
3110048L19Rik	RIKEN cDNA 3110048L19 gene	3.2
Ssbp2	single-stranded DNA binding protein 2	3.0
Qk	quaking	3.0
Phf14	PHD finger protein 14	3.0
Nedd4l	Neural precursor cell expressed, developmentally down-regulated gene 4-like	3.0
Ndufs1	NADH dehydrogenase (ubiquinone) Fe-S protein 1	3.0

1700108L22Rik	RIKEN cDNA 1700108L22 gene	3.0
1110003F05Rik	RIKEN cDNA 1110003F05 gene	3.0
Vezf1	Vascular endothelial zinc finger 1	2.8
Utx	Ubiquitously transcribed tetratricopeptide repeat gene, X chromosome	2.8
Tes3	Testis derived transcript 3	2.8
Tbc1d5	TBC1 domain family, member 5	2.8
Smc4l1	SMC4 structural maintenance of chromosomes 4-like 1 (yeast)	2.8
Senp6	SUMO/sentrin specific protease 6	2.8
Sca1	Spinocerebellar ataxia 1 homolog (human)	2.8
Psme4	proteasome (prosome, macropain) activator subunit 4	2.8
Cdkn1b	Cyclin-dependent kinase inhibitor 1B (P27)	2.8
Afp	alpha fetoprotein	2.8
2610024H22Rik	RIKEN cDNA 2610024H22 gene	2.8
1810043J12Rik	RIKEN cDNA 1810043J12 gene	2.8
Sdccag33	serologically defined colon cancer antigen 33	2.6
Ptpn12	Protein tyrosine phosphatase, non-receptor type 12	2.6
Ptbp2	Polypyrimidine tract binding protein 2	2.6
Prkwnk1	Protein kinase, lysine deficient 1	2.6
L3mbtl3	l(3)mbt-like 3 (Drosophila)	2.6
Kcnq5	potassium voltage-gated channel, subfamily Q, member 5	2.6
Foxc1	forkhead box C1	2.6
Enah	Enabled homolog (Drosophila)	2.6
Depdc5	DEP domain containing 5	2.6
C130047D21Rik	RIKEN cDNA C130047D21 gene	2.6
AI314180	expressed sequence AI314180	2.6
AA407452	EST AA407452	2.6
A630072M18Rik	RIKEN cDNA A630072M18 gene	2.6
A430106J12Rik	RIKEN cDNA A430106J12 gene	2.6
A130078K24Rik	RIKEN cDNA A130078K24 gene	2.6
A130012E19Rik	RIKEN cDNA A130012E19 gene	2.6
4921524P20Rik	RIKEN cDNA 4921524P20 gene	2.6
4631422O05Rik	RIKEN cDNA 4631422O05 gene	2.6
2810436B12Rik	RIKEN cDNA 2810436B12 gene	2.6
2810013C04Rik	RIKEN cDNA 2810013C04 gene	2.6
2600011C06Rik	RIKEN cDNA 2600011C06 gene	2.6
2010109K09Rik	RIKEN cDNA 2010109K09 gene	2.6
Zfhx1b	RIKEN cDNA 5830411K21 gene	2.5

Tdrd3	tudor domain containing 3	2.5
Tbl1x	Transducin (beta)-like 1 X-linked	2.5
Rnf103	ring finger protein 103	2.5
Rdh10	Retinol dehydrogenase 10 (all-trans)	2.5
Meis1	Myeloid ecotropic viral integration site 1	2.5
Man1a	Mannosidase 1, alpha	2.5
Lmo4	LIM domain only 4	2.5
Gulp1	GULP, engulfment adaptor PTB domain containing 1	2.5
Fkbp3	FK506 binding protein 3	2.5
Elf2	E74-like factor 2	2.5
Egr1	early growth response 1	2.5
D030065N23Rik	RIKEN cDNA D030065N23 gene	2.5
9630026M06Rik	RIKEN cDNA 9630026M06 gene	2.5
5830411K21Rik	RIKEN cDNA 5830411K21 gene	2.5
5730555F13Rik	RIKEN cDNA 5730555F13 gene	2.5
4933421G18Rik	RIKEN cDNA 4933421G18 gene	2.5
Zswim6	zinc finger, SWIM domain containing 6	2.3
Zfp53	Zinc finger protein 53	2.3
Uty	Ubiquitously transcribed tetratricopeptide repeat gene, Y chromosome	2.3
Usp47	ubiquitin specific protease 47	2.3
Sh3d1B	SH3 domain protein 1B	2.3
SEP6	septin 6	2.3
Runx2	runt related transcription factor 2	2.3
Prkar2b	protein kinase, cAMP dependent regulatory, type II beta	2.3
Pde3b	phosphodiesterase 3B, cGMP-inhibited	2.3
Pctk2	PCTAIRE-motif protein kinase 2	2.3
Matr3	matrin 3	2.3
LOC432971	hypothetical gene supported by AK038224	2.3
Jarid2	Jumonji, AT rich interactive domain 2	2.3
Gm872	gene model 872, (NCBI)	2.3
Glud1	glutamate dehydrogenase 1	2.3
Fbxo30	F-box protein 30	2.3
D230040A04Rik	RIKEN cDNA D230040A04 gene	2.3
D230007K08Rik	RIKEN cDNA D230007K08 gene	2.3
Cstf3	cleavage stimulation factor, 3' pre-RNA, subunit 3	2.3
Bcor	Bcl6 interacting corepressor	2.3
Bcl2	B-cell leukemia/lymphoma 2	2.3

Arid2	AT rich interactive domain 2 (Arid-rfx like)	2.3
9930116O05Rik	RIKEN cDNA 9930116O05 gene	2.3
5930436O19Rik	RIKEN cDNA 5930436O19 gene	2.3
5830474E16Rik	RIKEN cDNA 5830474E16 gene	2.3
4833416J08Rik	RIKEN cDNA 4833416J08 gene	2.3
2700017A04Rik	RIKEN cDNA 2700017A04 gene	2.3
1110059E24Rik	RIKEN cDNA 1110059E24 gene	2.3
Zmynd11	Zinc finger, MYND domain containing 11	2.1
Zfhx1a	Zinc finger homeobox 1a	2.1
Ube3a	ubiquitin protein ligase E3A	2.1
Trim59	Tripartite motif-containing 59	2.1
Tox	Thymocyte selection-associated HMG box gene	2.1
Tcf4	transcription factor 4	2.1
Tcf12	RIKEN cDNA E430034C17 gene	2.1
Sypl	synaptophysin-like protein	2.1
SImap	sarcolemma associated protein	2.1
Sh3kbp1	SH3-domain kinase binding protein 1	2.1
Ranbp9	RAN binding protein 9	2.1
Rab27b	RAB27b, member RAS oncogene family	2.1
Phip	pleckstrin homology domain interacting protein	2.1
Peli1	pellino 1	2.1
Pacs1	RIKEN cDNA D430030G11 gene	2.1
Obfc1	oligonucleotide/oligosaccharide-binding fold containing 1	2.1
Med12l	mediator of RNA polymerase II transcription, subunit 12 homolog (yeast)-like	2.1
Lpin2	Lipin 2	2.1
Herc3	hect domain and RLD 3	2.1
Eif4e	Eukaryotic translation initiation factor 4E	2.1
D18Ert232e	DNA segment, Chr 18, ERATO Doi 232, expressed	2.1
Cdk6	Cyclin-dependent kinase 6	2.1
Cacnb2	Calcium channel, voltage-dependent, beta 2 subunit	2.1
C430003N24Rik	RIKEN cDNA C430003N24 gene	2.1
Bhc80	BRAF35/HDAC2 complex	2.1
BC017647	CDNA sequence BC017647	2.1
Bbx	bobby sox homolog (Drosophila)	2.1
Axot	Axotrophin	2.1
Apg12l	autophagy 12-like (S. cerevisiae)	2.1
Angpt1	Angiopoietin 1	2.1

9430091E24Rik /// LOC434350	RIKEN cDNA 9430091E24 gene /// hypothetical gene supported by AK080922	2.1
5832424M12Rik	RIKEN cDNA 5832424M12 gene	2.1
5830411O09Rik	RIKEN cDNA 5830411O09 gene	2.1
5830407E08Rik	RIKEN cDNA 5830407E08 gene	2.1
5730526G10Rik	RIKEN cDNA 5730526G10 gene	2.1
5730405M06Rik	RIKEN cDNA 5730405M06 gene	2.1
5530401J07Rik	RIKEN cDNA 5530401J07 gene	2.1
5033430I15Rik	RIKEN cDNA 5033430I15 gene	2.1
4833412E19Rik	RIKEN cDNA 4833412E19 gene	2.1
4632427E13Rik /// LOC435992	RIKEN cDNA 4632427E13 gene /// similar to 40S ribosomal protein S3a (V-fos transformation effector protein)	2.1
2810036L13Rik	Heterogeneous nuclear ribonucleoprotein L-like	2.1
2310061J03Rik	RIKEN cDNA 2310061J03 gene	2.1
1700034P14Rik	RIKEN cDNA 1700034P14 gene	2.1
1190002N15Rik	RIKEN cDNA 1190002N15 gene	2.1
1110001A05Rik	RIKEN cDNA 1110001A05 gene	2.1
Zfp608	zinc finger protein 608	2.0
Zfp361l1	zinc finger protein 36, C3H type-like 1	2.0
Tpm1	tropomyosin 1, alpha	2.0
Top2b	topoisomerase (DNA) II beta	2.0
Tax1bp1	Tax1 (human T-cell leukemia virus type I) binding protein 1	2.0
Stxbp6	syntaxin binding protein 6 (amisyn)	2.0
St7	Suppression of tumorigenicity 7	2.0
Ssa2	Sjogren syndrome antigen A2	2.0
Ss18	synovial sarcoma translocation, Chromosome 18	2.0
Sntb2	syntrophin, basic 2	2.0
Snopc3	small nuclear RNA activating complex, polypeptide 3	2.0
Sh3bgrl	SH3-binding domain glutamic acid-rich protein like	2.0
Sfrs11	splicing factor, arginine/serine-rich 11	2.0
Sesn1	Sestrin 1	2.0
Robo3	roundabout homolog 3 (Drosophila)	2.0
Rnf12	Ring finger protein 12	2.0
Rapgef2	RIKEN cDNA B930012P20 gene	2.0
Ramp2	receptor (calcitonin) activity modifying protein 2	2.0
Ptprj	protein tyrosine phosphatase, receptor type, J	2.0
Psmd14	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 14	2.0

Pitpnc1	phosphatidylinositol transfer protein, cytoplasmic 1	2.0
Phtf2	putative homeodomain transcription factor 2	2.0
Parp14	poly (ADP-ribose) polymerase family, member 14	2.0
Ogt	O-linked N-acetylglucosamine (GlcNAc) transferase (UDP-N-acetylglucosamine:polypeptide-N-acetylglucosaminyl transferase)	2.0
Nrip1	nuclear receptor interacting protein 1	2.0
Nfe2l2	nuclear factor, erythroid derived 2, like 2	2.0
Mospd2	motile sperm domain containing 2	2.0
Mllt10	Myeloid/lymphoid or mixed lineage-leukemia translocation to 10 homolog (Drosophila)	2.0
Mbnl3	Muscleblind-like 3 (Drosophila)	2.0
Malt1	mucosa associated lymphoid tissue lymphoma translocation gene 1	2.0
Homer1	homer homolog 1 (Drosophila)	2.0
Fndc3	Fibronectin type III domain containing 3	2.0
Fli1	Friend leukemia integration 1	2.0
Fancc	Fanconi anemia, complementation group C	2.0
E2f5	E2F transcription factor 5	2.0
E130308A19Rik	RIKEN cDNA E130308A19 gene	2.0
E130108L08Rik	RIKEN cDNA E130108L08 gene	2.0
E030024I16Rik	RIKEN cDNA E030024I16 gene	2.0
Diap1	Diaphanous homolog 1 (Drosophila)	2.0
D5ErtD798e	DNA segment, Chr 5, ERATO Doi 798, expressed	2.0
D10Ucla1	DNA segment, Chr 10, University of California at Los Angeles 1	2.0
Crlf3	cytokine receptor-like factor 3	2.0
C430010P07Rik	RIKEN cDNA C430010P07 gene	2.0
B930013M22Rik	RIKEN cDNA B930013M22 gene	2.0
B830007D08Rik	RIKEN cDNA B830007D08 gene	2.0
Atp10a	ATPase, class V, type 10A	2.0
Arid1b	AT rich interactive domain 1B (Swi1 like)	2.0
Arhgap21	Rho GTPase activating protein 21	2.0
Akt3	thymoma viral proto-oncogene 3	2.0
Akap13	A kinase (PRKA) anchor protein 13	2.0
Aim1	absent in melanoma 1	2.0
A730095J18Rik	Zinc finger protein, subfamily 1A, 2 (Helios)	2.0
A430001F24Rik	RIKEN cDNA A430001F24 gene	2.0
A330103N21Rik	RIKEN cDNA A330103N21 gene	2.0
A130071D04Rik	RIKEN cDNA A130071D04 gene	2.0

9530028C05	hypothetical protein 9530028C05	2.0
9430034F23Rik	RIKEN cDNA 9430034F23 gene	2.0
9130203F04Rik	RIKEN cDNA 9130203F04 gene	2.0
9030406N13Rik	RIKEN cDNA 9030406N13 gene	2.0
6430585N13Rik	RIKEN cDNA 6430585N13 gene	2.0
6430510B20Rik	RIKEN cDNA 6430510B20 gene	2.0
6330549H03Rik	General transcription factor II A, 1	2.0
6330505F04Rik	RIKEN cDNA 6330505F04 gene	2.0
5830472H07Rik	RIKEN cDNA 5830472H07 gene	2.0
5630401D06Rik	RIKEN cDNA 5630401D06 gene	2.0
2900084O13Rik	RIKEN cDNA 2900084O13 gene	2.0
2900001A12Rik	Ankyrin repeat domain 12	2.0
2700008G24Rik	RIKEN cDNA 2700008G24 gene	2.0
2210420N10Rik	RIKEN cDNA 2210420N10 gene	2.0
1700081L11Rik	RIKEN cDNA D030002E05 gene	2.0

Table 21. Genes downregulated 12 hours after induction of MLL–VP16–ER–HA.

Lipg	lipase, endothelial	4.3
Il6	interleukin 6	3.7
Ccr2	chemokine (C-C motif) receptor 2	3.7
Nmyc1	neuroblastoma myc-related oncogene 1	3.5
Thbs1	thrombospondin 1	3.2
Il1rn	interleukin 1 receptor antagonist	3.2
BC035044	cDNA sequence BC035044	3.2
5830411E10Rik	RIKEN cDNA 5830411E10 gene	3.2
Tcrg-V4	T-cell receptor gamma, variable 6	3.0
Tcf7	transcription factor 7, T-cell specific	2.8
Gimap3	GTPase, IMAP family member 3	2.8
Myliip	myosin regulatory light chain interacting protein	2.6
Ear11	eosinophil-associated, ribonuclease A family, member 11	2.6
Dctd	dCMP deaminase	2.6
Maf	avian musculoaponeurotic fibrosarcoma (v-maf) AS42 oncogene homolog	2.5
Ifitm6	interferon induced transmembrane protein 6	2.5
Idb2	inhibitor of DNA binding 2	2.5
Csf1	colony stimulating factor 1 (macrophage)	2.5

Cpsf2	cleavage and polyadenylation specific factor 2	2.5
St6galnac4	ST6 (alpha-N-acetyl-neuraminy-2,3-beta-galactosyl-1,3)-N-acetylgalactosaminide alpha-2,6-sialyltransferase 4	2.3
Slc19a1	solute carrier family 19 (sodium/hydrogen exchanger), member 1	2.3
Nolc1	nucleolar and coiled-body phosphoprotein 1	2.3
Mmp9	matrix metalloproteinase 9	2.3
Ly6e	lymphocyte antigen 6 complex, locus E	2.3
2610009I02Rik	RIKEN cDNA 2610009I02 gene	2.3
Slco4a1	solute carrier organic anion transporter family, member 4a1	2.1
MAR1	membrane-associated ring finger (C3HC4) 1	2.1
Ifrd2	interferon-related developmental regulator 2	2.1
Grwd1	glutamate-rich WD repeat containing 1	2.1
Ccr1	chemokine (C-C motif) receptor 1	2.1
Ccl3	chemokine (C-C motif) ligand 3	2.1
Alox5	arachidonate 5-lipoxygenase	2.1
Adam8	a disintegrin and metalloprotease domain 8	2.1
9430063L05Rik	RIKEN cDNA 9430063L05 gene	2.1
2310008H09Rik	RIKEN cDNA 2310008H09 gene	2.1
1110007M04Rik	RIKEN cDNA 1110007M04 gene	2.1
Thop1	thimet oligopeptidase 1	2.0
Tbl2	transducin (beta)-like 2	2.0
Srm	spermidine synthase	2.0
Spp1	secreted phosphoprotein 1	2.0
Rpo1-2	RNA polymerase 1-2	2.0
Pim3	proviral integration site 3	2.0
Myc	myelocytomatosis oncogene	2.0
MGI:2136405	glucuronyl C5-epimerase	2.0
MGC58382	similar to Normal mucosa of esophagus specific gene 1 protein	2.0
Krtap8-2	keratin associated protein 8-2	2.0
Kctd12	potassium channel tetramerisation domain containing 12	2.0
Hist1h1c	histone 1, H1c	2.0
Gm1960	gene model 1960, (NCBI)	2.0
Fzr1	Fizzy/cell division cycle 20 related 1 (Drosophila)	2.0
Cxcr4	chemokine (C-X-C motif) receptor 4	2.0
C230052I12Rik	RIKEN cDNA C230052I12 gene	2.0
Bspry	B-box and SPRY domain containing	2.0
BC056485	cDNA sequence BC056485	2.0

Atad3a	ATPase family, AAA domain containing 3A	2.0
4633402N23Rik	RIKEN cDNA 4633402N23 gene	2.0
1810015A11Rik	RIKEN cDNA 1810015A11 gene	2.0

6.3 List of available binding matrices (Genomatix)

Table 22. Matrices used from Genomatix (www.genomatix.de) for the detection of transcription factors

family	matrix	description
V\$AARF	V\$AARE.01	Amino acid response element, ATF4 binding site
V\$AHRR	V\$AHR.01	Aryl hydrocarbon / dioxin receptor
V\$AHRR	V\$AHRARNT.01	Aryl hydrocarbon receptor / Arnt heterodimers
V\$AHRR	V\$AHRARNT.02	Aryl hydrocarbon / Arnt heterodimers, fixed core
V\$AHRR	V\$NXF_ARNT.01	bHLH-PAS type transcription factors NXF/ARNT heterodimer
V\$AIRE	V\$AIRE.01	Autoimmune regulator
V\$AP1F	V\$AP1.01	Activator protein 1
V\$AP1F	V\$AP1.02	Activator protein 1
V\$AP1F	V\$AP1.03	Activator protein 1
V\$AP1R	V\$BACH1.01	BTB/POZ-bZIP transcription factor BACH1 forms heterodimers with the small Maf protein family
V\$AP1R	V\$BACH2.01	Bach2 bound TRE
V\$AP1R	V\$NFE2.01	NF-E2 p45
V\$AP1R	V\$NFE2L2.01	Nuclear factor (erythroid-derived 2)-like 2, NRF2
V\$AP1R	V\$TCF11MAFG.01	TCF11/MafG heterodimers, binding to subclass of AP1 sites
V\$AP1R	V\$VMAF.01	v-Maf
V\$AP2F	V\$AP2.01	Activator protein 2
V\$AP2F	V\$AP2.02	Activator protein 2 alpha
V\$AP4R	V\$AP4.01	Activator protein 4
V\$AP4R	V\$AP4.02	Activator protein 4
V\$AP4R	V\$LXL1_E12.01	LYL1-E12 heterodimeric complex
V\$AP4R	V\$PARAXIS.01	Paraxis (TCF15), member of the Twist subfamily of Class B bHLH factors, forms heterodimers with E12
V\$AP4R	V\$TAL1ALPHA E47.01	Tal-1alpha/E47 heterodimer
V\$AP4R	V\$TAL1BETA E47.01	Tal-1beta/E47 heterodimer
V\$AP4R	V\$TAL1BETAHEB.01	Tal-1beta/HEB heterodimer
V\$AP4R	V\$TAL1BETAITF2.01	Tal-1beta/ITF-2 heterodimer

V\$AP4R	V\$TH1E47.01	Thing1/E47 heterodimer, TH1 bHLH member specific expression in a variety of embryonic tissues
V\$AREB	V\$AREB6.01	AREB6 (Atp1a1 regulatory element binding factor 6)
V\$AREB	V\$AREB6.02	AREB6 (Atp1a1 regulatory element binding factor 6)
V\$AREB	V\$AREB6.03	AREB6 (Atp1a1 regulatory element binding factor 6)
V\$AREB	V\$AREB6.04	AREB6 (Atp1a1 regulatory element binding factor 6)
V\$ATBF	V\$ATBF1.01	AT-binding transcription factor 1
V\$BARB	V\$BARBIE.01	Barbiturate-inducible element
V\$BEL1	V\$BEL1.01	Bel-1 similar region (defined in Lentivirus LTRs)
V\$BCL6	V\$BCL6.01	POZ/zinc finger protein, transcriptional repressor, translocations observed in diffuse large cell lymphoma
V\$BCL6	V\$BCL6.02	POZ/zinc finger protein, transcriptional repressor, translocations observed in diffuse large cell lymphoma
V\$BNCF	V\$BNC.01	Basonuclin, cooperates with USF1 in rDNA Poll transcription)
V\$BRAC	V\$BRACH.01	Brachyury
V\$BRAC	V\$TBX5.01	T-Box factor 5 site (TBX5), mutations related to Holt-Oram syndrome
V\$BRNF	V\$BRN2.01	Brn-2, POU-III protein class
V\$BRNF	V\$BRN2.02	Brn-2, POU-III protein class
V\$BRNF	V\$BRN2.03	Brn-2, POU-III protein class
V\$BRNF	V\$BRN3.01	Brn-3, POU-IV protein class
V\$BRNF	V\$BRN3.02	Brn-3, POU-IV protein class
V\$BRNF	V\$BRN4.01	POU domain transcription factor brain 4
V\$BRNF	V\$BRN5.01	Brn-5, POU-VI protein class (also known as emb and CNS-1)
V\$BTBF	V\$KAISO.01	Transcription factor Kaiso, ZBTB33
V\$CAAT	V\$ACAAT.01	Avian C-type LTR CCAAT box
V\$CAAT	V\$CAAT.01	Cellular and viral CCAAT box
V\$CAAT	V\$NFY.01	Nuclear factor Y (Y-box binding factor)
V\$CAAT	V\$NFY.02	Nuclear factor Y (Y-box binding factor)
V\$CAAT	V\$NFY.03	Nuclear factor Y (Y-box binding factor)
V\$CABL	V\$CABL.01	Multifunctional c-Abl src type tyrosine kinase
V\$CART	V\$CART1.01	Cart-1 (cartilage homeoprotein 1)
V\$CART	V\$XVENT2.01	Xenopus homeodomain factor Xvent-2; early BMP signaling response
V\$CDEF	V\$CDE.01	Cell cycle-dependent element, CDF-1 binding site (CDE/CHR tandem elements regulate cell cycle dependent repression)
V\$CDXF	V\$CDX1.01	Intestine specific homeodomain factor CDX-1

V\$CDXF	V\$CDX2.01	Cdx-2 mammalian caudal related intestinal transcr. factor
V\$CEBP	V\$CEBP.02	CCAAT/enhancer binding protein
V\$CEBP	V\$CEBPB.01	CCAAT/enhancer binding protein beta
V\$CHOP	V\$CHOP.01	Heterodimers of CHOP and C/EBPalpha
V\$CHRE	V\$CHREBP_MLX.01	Carbohydrate response element binding protein (CHREBP) and Max-like protein X (Mlx) bind as heterodimers to glucose-responsive promoters
V\$CHRF	V\$CHR.01	Cell cycle gene homology region (CDE/CHR tandem elements regulate cell cycle dependent repression)
V\$CIZF	V\$NMP4.01	NMP4 (nuclear matrix protein 4) / CIZ (Cas-interacting zinc finger protein)
V\$CLOX	V\$CDP.01	Cut-like homeodomain protein
V\$CLOX	V\$CDP.02	Transcriptional repressor CDP
V\$CLOX	V\$CDPCR3.01	Cut-like homeodomain protein
V\$CLOX	V\$CDPCR3HD.01	Cut-like homeodomain protein
V\$CLOX	V\$CLOX.01	Cut-like homeo box
V\$CLOX	V\$CUT2.01	Cut repeat II
V\$COMP	V\$COMP1.01	COMP1, cooperates with myogenic proteins in multicomponent complex
V\$CP2F	V\$CP2.01	CP2
V\$CP2F	V\$CP2.02	LBP-1c (leader-binding protein-1c), LSF (late SV40 factor), CP2, SEF (SAA3 enhancer factor)
V\$CREB	V\$ATF.01	Activating transcription factor
V\$CREB	V\$ATF.02	Activating transcription factor
V\$CREB	V\$ATF2.01	Activating transcription factor 2
V\$CREB	V\$ATF6.02	Activating transcription factor 6, member of b-zip family, induced by ER stress
V\$CREB	V\$CJUN_ATF2.01	c-Jun/ATF2 heterodimers
V\$CREB	V\$CREB.01	cAMP-responsive element binding protein
V\$CREB	V\$CREB.02	cAMP-responsive element binding protein
V\$CREB	V\$CREB.03	cAMP-response element-binding protein
V\$CREB	V\$CREBP1.01	cAMP-responsive element binding protein 1
V\$CREB	V\$CREBP1CJUN.01	CRE-binding protein 1/c-Jun heterodimer
V\$CREB	V\$E4BP4.01	E4BP4, bZIP domain, transcriptional repressor
V\$CREB	V\$TAXCREB.01	Tax/CREB complex
V\$CREB	V\$TAXCREB.02	Tax/CREB complex

V\$CSEN	V\$DREAM.01	Downstream regulatory element-antagonist modulator, Ca ²⁺ -binding protein of the neuronal calcium sensors family that binds DRE (downstream regulatory element) sites as a tetramer
V\$DEAF	V\$NUDR.01	NUDR (nuclear DEAF-1 related transcriptional regulator protein)
V\$DICE	V\$DICE.01	Downstream Immunoglobulin Control Element, interacting factor: BEN (also termed Mus-TRD1 and WBSR11)
V\$DMTF	V\$DMP1.01	Cyclin D-interacting myb-like protein, DMTF1 - cyclin D binding myb-like transcription factor 1
V\$E2FF	V\$E2F.01	E2F, involved in cell cycle regulation, interacts with Rb p107 protein
V\$E2FF	V\$E2F.02	E2F, involved in cell cycle regulation, interacts with Rb p107 protein
V\$E2FF	V\$E2F.03	E2F, involved in cell cycle regulation, interacts with Rb p107 protein
V\$E2TF	V\$E2.01	BPV bovine papilloma virus regulator E2
V\$E4FF	V\$E4F.01	GLI-Krueppel-related transcription factor, regulator of adenovirus E4 promoter
V\$EBOR	V\$DELTAEF1.01	deltaEF1
V\$EBOR	V\$SIP1.01	Smad-interacting protein
V\$EBOR	V\$XBP1.01	X-box-binding protein 1
V\$EBOX	V\$ATF6.01	Member of b-zip family, induced by ER damage/stress, binds to the ERSE in association with NF-Y
V\$EBOX	V\$MAX.01	Max/Max dimer
V\$EBOX	V\$MYCMAX.01	c-Myc/Max heterodimer
V\$EBOX	V\$MYCMAX.02	c-Myc/Max heterodimer
V\$EBOX	V\$MYCMAX.03	MYC-MAX binding sites
V\$EBOX	V\$NMYC.01	N-Myc
V\$EBOX	V\$SREBP.01	Sterol regulatory element binding protein 1 and 2
V\$EBOX	V\$SREBP.02	Sterol regulatory element binding protein
V\$EBOX	V\$SREBP.03	Sterol regulatory element binding protein
V\$EBOX	V\$USF.01	Upstream stimulating factor
V\$EBOX	V\$USF.02	Upstream stimulating factor
V\$EBOX	V\$USF.03	Upstream stimulating factor
V\$EGRF	V\$CKROX.01	Collagen krox protein (zinc finger protein 67 - zfp67)
V\$EGRF	V\$EGR1.01	Egr-1/Krox-24/NGFI-A immediate-early gene product
V\$EGRF	V\$EGR1.02	EGR1, early growth response 1
V\$EGRF	V\$EGR2.01	Egr-2/Krox-20 early growth response gene product
V\$EGRF	V\$EGR3.01	Early growth response gene 3 product
V\$EGRF	V\$NGFIC.01	Nerve growth factor-induced protein C

V\$EGRF	V\$WT1.01	Wilms Tumor Suppressor
V\$EKLF	V\$BKLF.01	Basic krueppel-like factor (KLF3)
V\$EKLF	V\$EKLF.01	Erythroid krueppel like factor (EKLF)
V\$EKLF	V\$KKLF.01	Kidney-enriched kruppel-like factor, KLF15
V\$EREF	V\$ER.01	Estrogen receptor
V\$EREF	V\$ER.02	Canonical palindromic estrogen response element (ERE)
V\$EREF	V\$ERR.01	Estrogen related receptor
V\$ETSF	V\$CETS1P54.01	c-Ets-1(p54)
V\$ETSF	V\$ELF2.01	Ets - family member ELF-2 (NERF1a)
V\$ETSF	V\$ELK1.01	Elk-1
V\$ETSF	V\$ELK1.02	Elk-1
V\$ETSF	V\$ETS1.01	c-Ets-1 binding site
V\$ETSF	V\$ETS2.01	c-Ets-2 binding site
V\$ETSF	V\$FLI.01	ETS family member FLI
V\$ETSF	V\$GABP.01	GABP: GA binding protein
V\$ETSF	V\$NRF2.01	Nuclear respiratory factor 2
V\$ETSF	V\$PDEF.01	Prostate-derived Ets factor
V\$ETSF	V\$PU1.01	Pu.1 (Pu120) Ets-like transcription factor identified in lymphoid B-cells
V\$EVI1	V\$EVI1.01	Ecotropic viral integration site 1 encoded factor, amino-terminal zinc finger domain
V\$EVI1	V\$EVI1.02	Ecotropic viral integration site 1 encoded factor, amino-terminal zinc finger domain
V\$EVI1	V\$EVI1.03	Ecotropic viral integration site 1 encoded factor, amino-terminal zinc finger domain
V\$EVI1	V\$EVI1.04	Ecotropic viral integration site 1 encoded factor, amino-terminal zinc finger domain
V\$EVI1	V\$EVI1.05	Ecotropic viral integration site 1 encoded factor, amino-terminal zinc finger domain
V\$EVI1	V\$EVI1.06	Ecotropic viral integration site 1 encoded factor, amino-terminal zinc finger domain
V\$EVI1	V\$EVI1.07	Evi-1 zinc finger protein, carboxy-terminal zinc finger domain
V\$EVI1	V\$MEL1.01	MEL1 (MDS1/EVI1-like gene 1) DNA-binding domain 1
V\$EVI1	V\$MEL1.02	MEL1 (MDS1/EVI1-like gene 1) DNA-binding domain 2
V\$EVI1	V\$MEL1.03	MEL1 (MDS1/EVI1-like gene 1) DNA-binding domain 2
V\$FAST	V\$FAST1.01	FAST-1 SMAD interacting protein

V\$FKHD	V\$FKHRL1.01	Fkh-domain factor FKHRL1 (FOXO)
V\$FKHD	V\$FREAC2.01	Fork head related activator-2 (FOXF2)
V\$FKHD	V\$FREAC3.01	Fork head related activator-3 (FOXC1)
V\$FKHD	V\$FREAC4.01	Fork head related activator-4 (FOXD1)
V\$FKHD	V\$FREAC7.01	Fork head related activator-7 (FOXL1)
V\$FKHD	V\$HFH1.01	HNF-3/Fkh Homolog 1 (FOXQ1)
V\$FKHD	V\$HFH2.01	HNF-3/Fkh Homolog 2 (FOXD3)
V\$FKHD	V\$HFH3.01	HNF-3/Fkh Homolog 3 (FOX11, Freac-6)
V\$FKHD	V\$HFH8.01	HNF-3/Fkh Homolog-8 (FOXF1)
V\$FKHD	V\$HNF3B.01	Hepatocyte nuclear factor 3beta (FOXA2)
V\$FKHD	V\$ILF1.01	Winged-helix transcription factor IL-2 enhancer binding factor (ILF), forkhead box K2 (FO XK2)
V\$FKHD	V\$XFD1.01	Xenopus fork head domain factor 1 (FoxA4a)
V\$FKHD	V\$XFD2.01	Xenopus fork head domain factor 2 (Foxl1a)
V\$FKHD	V\$XFD3.01	Xenopus fork head domain factor 3 (FoxA2a)
V\$GABF	V\$GAGA.01	GAGA-Box
V\$GATA	V\$GATA.01	GATA binding factor
V\$GATA	V\$GATA1.01	GATA-binding factor 1
V\$GATA	V\$GATA1.02	GATA-binding factor 1
V\$GATA	V\$GATA1.03	GATA-binding factor 1
V\$GATA	V\$GATA1.04	GATA-binding factor 1
V\$GATA	V\$GATA1.05	GATA-binding factor 1
V\$GATA	V\$GATA1.06	Complex of Lmo2 bound to Tal-1, E2A proteins, and GATA-1, half-site 2
V\$GATA	V\$GATA2.01	GATA-binding factor 2
V\$GATA	V\$GATA2.02	GATA-binding factor 2
V\$GATA	V\$GATA3.01	GATA-binding factor 3
V\$GATA	V\$GATA3.02	GATA-binding factor 3
V\$GFI1	V\$GFI1.01	Growth factor independence 1 zinc finger protein acts as transcriptional repressor
V\$GFI1	V\$GFI1B.01	Growth factor independence 1 zinc finger protein Gfi-1B
V\$GKLF	V\$GKLF.01	Gut-enriched Krueppel-like factor
V\$GKLF	V\$GKLF.02	Gut-enriched Krueppel-like factor
V\$GLIF	V\$GLI1.01	Zinc finger transcription factor GLI1
V\$GLIF	V\$ZIC2.01	Zinc finger transcription factor, Zic family member 2 (odd-paired homolog, Drosophila)

V\$GREF	V\$ARE.01	Androgene receptor binding site
V\$GREF	V\$GRE.01	Glucocorticoid receptor, C2C2 zinc finger protein binds glucocorticoid dependent to GREs
V\$GREF	V\$PRE.01	Progesterone receptor binding site
V\$GZF1	V\$GZF1.01	GDNF-inducible zinc finger protein 1 (ZNF336)
V\$HAML	V\$AML1.01	AML1/CBFA2 Runt domain binding site
V\$HAML	V\$AML3.01	Runt-related transcription factor 2 / CBFA1 (core-binding factor, runt domain, alpha subunit 1)
V\$HAND	V\$HAND2_E12.01	Heterodimers of the bHLH transcription factors HAND2 (Thing2) and E12
V\$HEAT	V\$HSF1.01	Heat shock factor 1
V\$HEAT	V\$HSF1.02	Heat shock factor 1
V\$HEAT	V\$HSF1.03	Heat shock factor 1
V\$HEAT	V\$HSF2.01	Heat shock factor 2
V\$HEAT	V\$HSF2.02	Heat shock factor 2
V\$HEN1	V\$HEN1.01	HEN1
V\$HEN1	V\$HEN1.02	HEN1
V\$HESF	V\$HELT.01	Hey-like bHLH-transcriptional repressor
V\$HESF	V\$HES1.01	Drosophila hairy and enhancer of split homologue 1 (HES-1)
V\$HESF	V\$HES1.02	Drosophila hairy and enhancer of split homologue 1 (HES-1)
V\$HICF	V\$HIC1.01	Hypermethylated in cancer 1, transcriptional repressor containing five Kr ^v oppel-like C2H2 zinc fingers, for optimal binding multiple binding sites are required.
V\$HIFF	V\$ARNT.01	AhR nuclear translocator homodimers
V\$HIFF	V\$CLOCK_BMAL1.01	Binding site of Clock/BMAL1 heterodimer, NPAS2/BMAL1 heterodimer
V\$HIFF	V\$DEC1.01	Basic helix-loop-helix protein known as Dec1, Stra13 or Sharp2
V\$HIFF	V\$DEC2.01	Basic helix-loop-helix protein known as Dec2 or Sharp2
V\$HIFF	V\$HIF1.01	Hypoxia induced factor-1 (HIF-1)
V\$HIFF	V\$HIF1.02	Hypoxia inducible factor, bHLH / PAS protein family
V\$HMTB	V\$MTBF.01	Muscle-specific Mt binding site
V\$HNF1	V\$HNF1.01	Hepatic nuclear factor 1
V\$HNF1	V\$HNF1.02	Hepatic nuclear factor 1
V\$HNF1	V\$HNF1.03	Hepatic nuclear factor 1
V\$HNF6	V\$HNF6.01	Liver enriched Cut - Homeodomain transcription factor HNF6 (ONECUT)

V\$HNF6	V\$OC2.01	CUT-homeodomain transcription factor Onecut-2
V\$HOMF	V\$DLX1.01	DLX-1, -2, and -5 binding sites
V\$HOMF	V\$DLX3.01	Distal-less 3 homeodomain transcription factor
V\$HOMF	V\$EN1.01	Homeobox protein engrailed (en-1)
V\$HOMF	V\$MSX.01	Homeodomain proteins MSX-1 and MSX-2
V\$HOMF	V\$MSX2.01	Muscle segment homeo box 2, homologue of Drosophila (HOX 8)
V\$HOMF	V\$S8.01	Binding site for S8 type homeodomains
V\$HOXF	V\$BARX2.01	Barx2, homeobox transcription factor that preferentially binds to paired TAAT motifs
V\$HOXF	V\$CRX.01	Cone-rod homeobox-containing transcription factor / otx-like homeobox gene
V\$HOXF	V\$GSC.01	Vertebrate bicoid-type homeodomain protein Goosecoid
V\$HOXF	V\$GSH1.01	Homeobox transcription factor Gsh-1
V\$HOXF	V\$GSH2.01	Homeodomain transcription factor Gsh-2
V\$HOXF	V\$HOX1-3.01	Hox-1.3, vertebrate homeobox protein
V\$HOXF	V\$HOXA9.01	Member of the vertebrate HOX - cluster of homeobox factors
V\$HOXF	V\$HOXB9.01	Abd-B-like homeodomain protein Hoxb-9
V\$HOXF	V\$HOXC13.01	Homeodomain transcription factor HOXC13
V\$HOXF	V\$OTX2.01	Homeodomain transcription factor Otx2 (homolog of Drosophila orthodenticle)
V\$HOXF	V\$PHOX2.01	Phox2a (ARIX) and Phox2b
V\$HOXF	V\$PTX1.01	Pituitary Homeobox 1 (Ptx1, Pitx-1)
V\$HOXH	V\$MEIS1A_HOXA9.01	Meis1a and Hoxa9 form heterodimeric binding complexes on target DNA
V\$HOXH	V\$MEIS1B_HOXA9.01	Meis1b and Hoxa9 form heterodimeric binding complexes on target DNA
V\$HOXC	V\$HOX_PBX.01	HOX/PBX binding sites
V\$HOXC	V\$PBX1.01	Homeo domain factor Pbx-1
V\$HOXC	V\$PBX_HOXA9.01	PBX - HOXA9 binding site
V\$IKRS	V\$IK1.01	Ikaros 1, potential regulator of lymphocyte differentiation
V\$IKRS	V\$IK2.01	Ikaros 2, potential regulator of lymphocyte differentiation
V\$IKRS	V\$IK3.01	Ikaros 3, potential regulator of lymphocyte differentiation
V\$IKRS	V\$LYF1.01	LyF-1 (Ikaros 1), enriched in B and T lymphocytes
V\$INSM	V\$INSM1.01	Zinc finger protein insulinoma-associated 1 (IA-1) functions as a transcriptional repressor
V\$IRFF	V\$IRF1.01	Interferon regulatory factor 1

V\$IRFF	V\$IRF2.01	Interferon regulatory factor 2
V\$IRFF	V\$IRF3.01	Interferon regulatory factor 3 (IRF-3)
V\$IRFF	V\$IRF4.01	Interferon regulatory factor (IRF)-related protein (NF-EM5, PIP, LSIRF, ICSAT)
V\$IRFF	V\$IRF7.01	Interferon regulatory factor 7 (IRF-7)
V\$IRFF	V\$ISRE.01	Interferon-stimulated response element
V\$LEFF	V\$LEF1.01	TCF/LEF-1, involved in the Wnt signal transduction pathway
V\$LEFF	V\$LEF1.02	TCF/LEF-1, involved in the Wnt signal transduction pathway
V\$LHXF	V\$LHX3.01	Homeodomain binding site in LIM/Homeodomain factor LHX3
V\$LHXF	V\$LMX1B.01	LIM-homeodomain transcription factor
V\$LTUP	V\$TAACC.01	Lentiviral TATA upstream element
V\$MAZF	V\$MAZ.01	Myc associated zinc finger protein (MAZ)
V\$MAZF	V\$MAZR.01	MYC-associated zinc finger protein related transcription factor
V\$MEF2	V\$MEF2.01	Myocyte-specific enhancer factor 2
V\$MEF2	V\$MEF2.02	Myocyte-specific enhancer factor 2
V\$MEF2	V\$MEF2.03	Myocyte-specific enhancer factor 2
V\$MEF2	V\$MEF2.04	Myocyte-specific enhancer factor 2
V\$MEF2	V\$MEF2.05	Myocyte-specific enhancer factor 2
V\$MEF2	V\$MEF2.06	Myocyte-specific enhancer factor 2
V\$MEF2	V\$MEF2.07	Myocyte-specific enhancer factor 2
V\$MEF2	V\$RSRFC4.01	Related to serum response factor, C4
V\$MEF2	V\$SL1.01	Member of the RSRF (related to serum response factor) protein family from <i>Xenopus laevis</i>
V\$MEF3	V\$MEF3.01	MEF3 binding site, present in skeletal muscle-specific transcriptional enhancers
V\$MINI	V\$MUSCLE_INI.01	Muscle Initiator Sequence
V\$MINI	V\$MUSCLE_INI.02	Muscle Initiator Sequence
V\$MITF	V\$MIT.01	MIT (microphthalmia transcription factor) and TFE3
V\$MOKF	V\$MOK2.01	Ribonucleoprotein associated zinc finger protein MOK-2 (mouse)
V\$MOKF	V\$MOK2.02	Ribonucleoprotein associated zinc finger protein MOK-2 (human)
V\$MTF1	V\$MTF-1.01	Metal transcription factor 1, MRE
V\$MYBL	V\$CMYB.01	c-Myb, important in hematopoiesis, cellular equivalent to avian myoblastosis virus oncogene v-myb
V\$MYBL	V\$CMYB.02	c-Myb, important in hematopoiesis, cellular equivalent to avian myoblastosis virus oncogene v-myb
V\$MYBL	V\$VMYB.01	v-Myb

V\$MYBL	V\$VMYB.02	v-Myb
V\$MYBL	V\$VMYB.03	v-Myb, viral myb variant from transformed BM2 cells
V\$MYBL	V\$VMYB.04	v-Myb, AMV v-myb
V\$MYBL	V\$VMYB.05	v-Myb, variant of AMV v-myb
V\$MYOD	V\$E47.01	MyoD/E47 and MyoD/E12 dimers
V\$MYOD	V\$E47.02	E47 homodimer
V\$MYOD	V\$MYF5.01	Myf5 myogenic bHLH protein
V\$MYOD	V\$MYOD.01	Myogenic regulatory factor MyoD (myf3)
V\$MYOD	V\$MYOGENIN.01	Myogenic bHLH protein myogenin (myf4)
V\$MYOD	V\$TAL1_E2A.01	Complex of Lmo2 bound to Tal-1, E2A proteins, and GATA-1, half-site 1
V\$MYT1	V\$MYT1.01	MyT1 zinc finger transcription factor involved in primary neurogenesis
V\$MYT1	V\$MYT1.02	MyT1 zinc finger transcription factor involved in primary neurogenesis
V\$MYT1	V\$MYT1L.01	Myelin transcription factor 1-like, neuronal C2HC zinc finger factor 1
V\$MZF1	V\$MZF1.01	Myeloid zinc finger protein MZF1
V\$MZF1	V\$MZF1.02	Myeloid zinc finger protein MZF1
V\$NEUR	V\$NEUROD1.01	DNA binding site for NEUROD1 (BETA-2 / E47 dimer)
V\$NEUR	V\$NEUROG.01	Neurogenin 1 and 3 (ngn1/3) binding sites
V\$NF1F	V\$NF1.01	Nuclear factor 1
V\$NF1F	V\$NF1.02	Nuclear factor 1 (CTF1)
V\$NFAT	V\$NFAT.01	Nuclear factor of activated T-cells
V\$NFKB	V\$CREL.01	c-Rel
V\$NFKB	V\$HIVEP1.01	ZAS domain transcription factor: human immunodeficiency virus type 1 enhancer-binding protein-1 (HIVEP1), major histocompatibility complex-binding protein-1 (MBP-1), positive regulatory domain II-binding factor 1 (PRDII-BF1)
V\$NFKB	V\$NFKAPPAB.01	NF-kappaB
V\$NFKB	V\$NFKAPPAB.02	NF-kappaB
V\$NFKB	V\$NFKAPPAB50.01	NF-kappaB (p50)
V\$NFKB	V\$NFKAPPAB65.01	NF-kappaB (p65)
V\$NKXH	V\$HMX2.01	Hmx2/Nkx5-2 homeodomain transcription factor
V\$NKXH	V\$HMX2.02	Hmx2/Nkx5-2 homeodomain transcription factor
V\$NKXH	V\$HMX3.01	H6 homeodomain HMX3/Nkx5.1 transcription factor
V\$NKXH	V\$HMX3.02	Hmx3/Nkx5-1 homeodomain transcription factor

V\$NKXH	V\$NKX25.01	Homeo domain factor Nkx-2.5/Csx, tinman homolog, high affinity sites
V\$NKXH	V\$NKX25.02	Homeo domain factor Nkx-2.5/Csx, tinman homolog low affinity sites
V\$NKXH	V\$NKX31.01	Prostate-specific homeodomain protein NKX3.1
V\$NKXH	V\$NKX32.01	Homeodomain protein NKX3.2 (BAPX1, NKX3B, Bagpipe homolog)
V\$NOLF	V\$OLF1.01	Olfactory neuron-specific factor
V\$NR2F	V\$ARP1.01	Apolipoprotein AI regulatory protein 1, NR2F2
V\$NR2F	V\$COUP.01	COUP antagonizes HNF-4 by binding site competition or synergizes by direct protein - protein interaction with HNF-4
V\$NR2F	V\$HNF4.01	Hepatic nuclear factor 4
V\$NR2F	V\$HNF4.02	Hepatic nuclear factor 4
V\$NR2F	V\$HPF1.01	HepG2-specific P450 2C factor-1
V\$NR2F	V\$PNR.01	Photoreceptor-specific nuclear receptor subfamily 2, group E, member 3 (Nr2e3)
V\$NRF1	V\$NRF1.01	Nuclear respiratory factor 1 (NRF1), bZIP transcription factor that acts on nuclear genes encoding mitochondrial proteins
V\$NRLF	V\$NRL.01	Neural retinal basic leucine zipper factor (bZIP)
V\$NRSF	V\$NRSE.01	Neural-restrictive-silencer-element
V\$NRSF	V\$NRSF.01	Neuron-restrictive silencer factor
V\$OAZF	V\$ROAZ.01	Rat C2H2 Zn finger protein involved in olfactory neuronal differentiation
V\$OCT1	V\$OCT.01	Octamer binding site (OCT1/OCT2)
V\$OCT1	V\$OCT1.01	Octamer-binding factor 1
V\$OCT1	V\$OCT1.02	Octamer-binding factor 1
V\$OCT1	V\$OCT1.04	Octamer-binding factor 1
V\$OCT1	V\$OCT1.05	Octamer-binding factor 1
V\$OCT1	V\$OCT1.06	Octamer-binding factor 1
V\$OCTB	V\$TST1.01	POU-factor Tst-1/Oct-6
V\$OCTP	V\$OCT1P.01	Octamer-binding factor 1, POU-specific domain
V\$P53F	V\$P53.01	Tumor suppressor p53
V\$P53F	V\$P53.02	Tumor suppressor p53 (5' half site)
V\$P53F	V\$P53.03	Tumor suppressor p53 (3' half site)
V\$PARF	V\$DBP.01	Albumin D-box binding protein
V\$PAX1	V\$PAX1.01	Pax1 paired domain protein, expressed in the developing vertebral column of mouse embryos
V\$PAX2	V\$PAX2.01	Zebrafish PAX2 paired domain protein

V\$PAX3	V\$PAX3.01	Pax-3 paired domain protein, expressed in embryogenesis, mutations correlate to Waardenburg Syndrome
V\$PAX3	V\$PAX3.02	Pax-3 paired domain protein
V\$PAX4	V\$PAX4.01	Pax-4 homeodomain binding site, together with PAX-6 involved in pancreatic development
V\$PAX5	V\$PAX5.01	B-cell-specific activating protein
V\$PAX5	V\$PAX5.02	B-cell-specific activating protein
V\$PAX5	V\$PAX5.03	PAX5 paired domain protein
V\$PAX5	V\$PAX9.01	Zebrafish PAX9 binding sites
V\$PAX6	V\$PAX4_PD.01	PAX4 paired domain binding site
V\$PAX6	V\$PAX6.01	Pax-6 paired domain binding site
V\$PAX6	V\$PAX6.02	PAX6 paired domain and homeodomain are required for binding to this site
V\$PAX8	V\$PAX8.01	PAX 2/5/8 binding site
V\$PBXC	V\$PBX1_MEIS1.01	Binding site for a Pbx1/Meis1 heterodimer
V\$PBXC	V\$PBX1_MEIS1.02	Binding site for a Pbx1/Meis1 heterodimer
V\$PBXC	V\$PBX1_MEIS1.03	Binding site for a Pbx1/Meis1 heterodimer
V\$PDX1	V\$ISL1.01	Pancreatic and intestinal lim-homeodomain factor
V\$PDX1	V\$PDX1.01	Pdx1 (IDX1/IPF1) pancreatic and intestinal homeodomain TF
V\$PERO	V\$PPARA.01	PPAR/RXR heterodimers
V\$PIT1	V\$PIT1.01	Pit1, GHF-1 pituitary specific pou domain transcription factor
V\$PLAG	V\$PLAG1.01	Pleomorphic adenoma gene (PLAG) 1, a developmentally regulated C2H2 zinc finger protein
V\$PLZF	V\$PLZF.01	Promyelocytic leukemia zinc finger (TF with nine Krueppel-like zinc fingers)
V\$PRDF	V\$PRDM1.01	PRDI binding factor 1
V\$PXRF	V\$PXRCAR.01	Halfsite of PXR (pregnane X receptor)/RXR resp. CAR (constitutive androstane receptor)/RXR heterodimer binding site
V\$RARF	V\$RAR.01	Retinoic acid receptor, member of nuclear receptors, half site
V\$RARF	V\$RAR.02	Retinoic acid receptor, member of nuclear receptors, DR5 site
V\$RARF	V\$RTR.01	Retinoid receptor-related testis-associated receptor (GCNF/RTR)
V\$RBIT	V\$BRIGHT.01	Bright, B cell regulator of IgH transcription
V\$RBPF	V\$RBPJK.01	Mammalian transcriptional repressor RBP-Jkappa/CBF1
V\$RBPF	V\$RBPJK.02	Mammalian transcriptional repressor RBP-Jkappa/CBF1
V\$RCAT	V\$CLTR_CAAT.01	Mammalian C-type LTR CCAAT box
V\$REBV	V\$EBVR.01	Epstein-Barr virus transcription factor R

V\$RORA	V\$NBRE.01	Monomers of the nur subfamily of nuclear receptors (nur77, nurr1, nor-1)
V\$RORA	V\$RORA1.01	RAR-related orphan receptor alpha1
V\$RORA	V\$RORA2.01	RAR-related orphan receptor alpha2
V\$RORA	V\$T3R.01	vErbA, viral homolog of thyroid hormone receptor alpha1
V\$RORA	V\$TR2.01	Nuclear hormone receptor TR2, half site
V\$RORA	V\$TR4.01	Nuclear hormone receptor TR4 homodimer binding site
V\$RP58	V\$RP58.01	Zinc finger protein RP58 (ZNF238), associated preferentially with heterochromatin
V\$RREB	V\$RREB1.01	Ras-responsive element binding protein 1
V\$RXRF	V\$FXRE.01	Farnesoid X - activated receptor (RXR/FXR dimer)
V\$RXRF	V\$LXRE.01	Nuclear receptor involved in the regulation lipid homeostasis
V\$RXRF	V\$VDR_RXR.01	VDR/RXR Vitamin D receptor RXR heterodimer site
V\$RXRF	V\$VDR_RXR.02	VDR/RXR Vitamin D receptor RXR heterodimer site
V\$RXRF	V\$VDR_RXR.03	Bipartite binding site of VDR/RXR heterodimers without a spacer between directly repeated motifs
V\$RXRF	V\$VDR_RXR.04	Bipartite binding site of VDR/RXR heterodimers: 2 spacer nucleotides between the two directly repeated motifs
V\$RXRF	V\$VDR_RXR.05	Bipartite binding site of VDR/RXR heterodimers: 3 spacer nucleotides between the two directly repeated motifs
V\$RXRF	V\$VDR_RXR.06	Bipartite binding site of VDR/RXR heterodimers: 4 spacer nucleotides between the two directly repeated motifs
V\$SATB	V\$SATB1.01	Special AT-rich sequence-binding protein 1, predominantly expressed in thymocytes, binds to matrix attachment regions (MARs)
V\$SF1F	V\$FTF.01	Alpha (1)-fetoprotein transcription factor (FTF), liver receptor homologue-1 (LRH-1)
V\$SF1F	V\$SF1.01	SF1 steroidogenic factor 1
V\$SIXF	V\$SIX3.01	SIX3 / SIXdomain (SD) and Homeodomain (HD) transcription factor
V\$SMAD	V\$SMAD3.01	Smad3 transcription factor involved in TGF-beta signaling
V\$SMAD	V\$SMAD4.01	Smad4 transcription factor involved in TGF-beta signaling
V\$SNAP	V\$PSE.01	Proximal sequence element (PSE) of RNA polymerase II-transcribed snRNA genes
V\$SNAP	V\$PSE.02	Proximal sequence element (PSE) of RNA polymerase III-transcribed genes
V\$SORY	V\$HBP1.01	HMG box-containing protein 1
V\$SORY	V\$HMG1Y.01	HMG1(Y) high-mobility-group protein I (Y), architectural transcription factor organizing the framework of a nuclear protein-DNA

		transcriptional complex
V\$SORY	V\$SOX5.01	Sox-5
V\$SORY	V\$SOX9.01	SOX (SRY-related HMG box)
V\$SORY	V\$SRY.01	Sex-determining region Y gene product
V\$SP1F	V\$BTEB3.01	Basic transcription element (BTE) binding protein, BTEB3, FKLF-2
V\$SP1F	V\$GC.01	GC box elements
V\$SP1F	V\$SP1.01	Stimulating protein 1, ubiquitous zinc finger transcription factor
V\$SP1F	V\$SP1.02	Stimulating protein 1, ubiquitous zinc finger transcription factor
V\$SP1F	V\$SP2.01	Sp2, member of the Sp/XKLF transcription factors with three C2H2 zinc fingers in a conserved carboxyl-terminal domain
V\$SP1F	V\$TIEG.01	TGFbeta-inducible early gene (TIEG) / Early growth response gene alpha (EGRalpha)
V\$SRFF	V\$SRF.01	Serum response factor
V\$SRFF	V\$SRF.02	Serum response factor
V\$SRFF	V\$SRF.03	Serum response factor
V\$STAF	V\$STAF.01	Se-Cys tRNA gene transcription activating factor
V\$STAF	V\$STAF.02	Se-Cys tRNA gene transcription activating factor
V\$STAF	V\$ZNF76_143.01	ZNF143 is the human ortholog of Xenopus Staf, ZNF76 is a DNA binding protein related to ZNF143 and Staf
V\$STAT	V\$STAT.01	Signal transducers and activators of transcription
V\$STAT	V\$STAT1.01	Signal transducer and activator of transcription 1
V\$STAT	V\$STAT3.01	Signal transducer and activator of transcription 3
V\$STAT	V\$STAT5.01	STAT5: signal transducer and activator of transcription 5
V\$STAT	V\$STAT6.01	STAT6: signal transducer and activator of transcription 6
V\$TALE	V\$MEIS1.01	Binding site for monomeric Meis1 homeodomain protein
V\$TALE	V\$TGIF.01	TG-interacting factor belonging to TALE class of homeodomain factors
V\$TBPF	V\$ATATA.01	Avian C-type LTR TATA box
V\$TBPF	V\$LTATA.01	Lentivirus LTR TATA box
V\$TBPF	V\$MTATA.01	Muscle TATA box
V\$TBPF	V\$TATA.01	Cellular and viral TATA box elements
V\$TBPF	V\$TATA.02	Mammalian C-type LTR TATA box
V\$TCFF	V\$TCF11.01	TCF11/KCR-F1/Nrf1 homodimers
V\$TEAF	V\$HLF.01	Hepatic leukemia factor
V\$TEAF	V\$TEF.01	Thyrotrophic embryonic factor
V\$TEAF	V\$TEF1.01	TEF-1 related muscle factor

V\$TEAF	V\$TEF_HLF.01	Thyrotrophic embryonic factor / hepatic leukemia factor
V\$TTFF	V\$TTF1.01	Thyroid transcription factor-1 (TTF1) binding site
V\$VBPF	V\$VBP.01	PAR-type chicken vitellogenin promoter-binding protein
V\$WHZF	V\$WHN.01	Winged helix protein, involved in hair keratinization and thymus epithelium differentiation
V\$XBBF	V\$MIF1.01	MIBP-1 / RFX1 complex
V\$XBBF	V\$RFX1.01	X-box binding protein RFX1
V\$XBBF	V\$RFX1.02	X-box binding protein RFX1
V\$YY1F	V\$YY1.01	Yin and Yang 1 activator sites
V\$YY1F	V\$YY1.02	Yin and Yang 1 repressor sites
V\$ZBPF	V\$ZBP89.01	Zinc finger transcription factor ZBP-89
V\$ZBPF	V\$ZF9.01	Core promoter-binding protein (CPBP) with 3 Krueppel-type zinc fingers
V\$ZBPF	V\$ZNF202.01	Transcriptional repressor, binds to elements found predominantly in genes that participate in lipid metabolism
V\$ZBPF	V\$ZNF219.01	Kruppel-like zinc finger protein 219
V\$ZF5F	V\$ZF5.01	Zinc finger / POZ domain transcription factor
V\$ZFIA	V\$ZID.01	Zinc finger with interaction domain
V\$ZF35	V\$ZNF35.01	Human zinc finger protein ZNF35
V\$ZNFP	V\$SZF1.01	SZF1, hematopoietic progenitor-restricted KRAB-zinc finger protein
V\$ZNFP	V\$ZBRK1.01	Transcription factor with 8 central zinc fingers and an N-terminal KRAB domain

6.4 Promoter sequences of Cutl1, p18 and p27

Cutl1

```

-221   CTGCTGCATT TGGAGCTATT GTAGGACATC ACAGATCTTA GATGGGGAC CCCACCGCA
-161   GCTCACCAGA CATGATGCAA GGTCCCATTT AGCTCAAGTC ATGGGAAGGA AATGTTGGGA
-101   CCTGGCTCTC TGTCAGTACA GATGGGGCA GGGTACCACC ATGACACCCA AGCCGTTTCC
-41    TGCATCTCC TAGGACCCTC AATGTCATCC AGTGGGACCC

```

p18 (CDKN2C)

```

-221   TTTTTTTGAA AAGAAAAAAA TTGAGCGCTT TTTGAGTTGA AAAACCCGCC CCCATTTTAA
-161   CGGCAGAGTT TTAAGGAGGC TCCGCCGAGT TAGGGCGCCG CGGGAGCGGG AAGGCCGGGC
-101   CCCGCCACG CCCCGCCCAA CGCCCTGGCG GCGCAGTGGC CGGCAGCCCG CTCGCTCCGC
-41    CCTCGCGGCC CTGCGCCCTT CTCTGCCGC GCTGGGCTG

```

p27 (CDKN1B)

```

-221   GGTGTTCCT TATTTGCTTT GTTGTACTAC CTGTGTATAT AGTTTTTATC TTTTACTCTG
-161   TAGCACATAA ACGTTAGGGA TGGGAGGGCA GGGTGGGGT GAGGAGTCAG CATGGGGTG
-101   AAGAACTTGC TTCAATTTGT AGCAAGGAGA AAAATATTTG ACTTGCATAA AGAGAAGCAA
-41    CCTGGGTGGG GGTGGGGGG GCAAGGTTT GAATTCCTTT

```

Supplementary Figure 1. Promoter sequences of MLL–VP16–ER–HA target genes. Depicted are 220 bp of the upstream sequences of the genes Cutl1, p18 and p27. GGGC and GGGG motifs are highlighted in red.

Acknowledgements

I would like to thank Prof. Dr. Meisterernst for the opportunity to work on this thesis in his laboratory basically without any material constraint. I am grateful to Prof. Dr. Jansen for taking the part of referee and Head of the PhD committee.

I owe many thanks to Timm Schroeder and Ursula Just not only for introducing me to hematopoietic techniques and stem cell culture but also for always giving a helping hand and good advice whenever I needed it.

I am grateful to Elisabeth Kremmer and the members of her group for the generation of monoclonal antibodies used in this work. Marion Horsch was contributing to this study by doing part of the microarray analysis.

Thanks to Prof. Dr. Bohlander and Prof. Dr. Slany for interesting discussions, which were both helpful and inspiring for this project.

I would like to thank all members and former members of the Meisterernst lab for the pleasant working atmosphere I was experiencing in this group. Special thanks for help with the experimental work go to Florian Rieß, Christine Göbel and Gabrielle Gembe.

Last but not least I want to thank Angela Hofmann and Roswitha Neumann for always being of great help with all the bureaucratic aspects of research work.

Curriculum Vitae Michael Lehmbacher

Personal data:

date of birth: 29.08.1975
place of birth: Fürstenfeldbruck
nationality: Austrian

Dissertation:

since March 2002 Institute for Molecular Immunology, GSF München
Title of the thesis: Identification and characterization of genomic target sites for MLL fusion proteins

University education:

Oct. 1995 – Feb. 2002 Ludwig-Maximilians-Universität München
Focus: Molecular Biology
Feb. 2001 – Nov. 2001 Diploma thesis at the Biochemical Institute of the LMU (Gene Center)
Title: Ein immunchemisches Verfahren zur Analyse von Transkriptionsnetzwerken in eukaryotischen Subproteomen
Degree: Diploma in Biology, Grade: 1.1

School education:

Sept. 1982 – July 1986 Elementary school: Grundschule Emmering
Sept. 1986 – July 1995 High school: Graf-Rasso-Gymnasium
Fürstenfeldbruck
Degree: Abitur, Grade: 1.6

Internships:

March 1999 – May 1999 Lab intern with Prof. Famulok, Gene Center
Dec. 2000 – Jan. 2001 Internship at Medigenomix, Martinsried.
Department: High Throughput Genome Sequencing

Munich, 14.11.06