Oncogene (2003) 22, 3655–3668

www.nature.com/onc

MLL-mediated transcriptional gene regulation investigated by gene expression profiling

Dorothe Schraets[1], Tanja Lehmann[1], Theo Dingermann[1] and Rolf Marschalek*[,1]

[1]*Institute of Pharmaceutical Biology, Biocenter N230, University of Frankfurt/Main, Marie-Curie-Str. 9, D-60439 Frankfurt/Main, Germany*

The human mixed lineage leukemia (*MLL*) gene is involved in about 50 different chromosomal translocations, associated with the disease phenotype of acute leukemia. However, the normal function of MLL is less understood. Homozygous knockouts of murine *Mll* were embryonal lethal, while heterozygous disruption led to aberrant *hox* gene expression associated with skeletal malformations, growth retardation, and impaired hematopoiesis. To understand MLL functions on the molecular level, gene expression profiling experiments were performed with a pair of murine cell lines ($MLL^{+/+}$ and $MLL^{-/-}$). Microarray hybridization experiments revealed 197 potential target genes that are differentially expressed, providing new and important clues about MLL functions.
Oncogene (2003) **22,** 3655–3668. doi:10.1038/sj.onc.1206438

Keywords: MLL; expression profiling; development; differential gene expression

1. Introduction

MLL translocations are identified in a subset of acute leukemia patients (Cimino *et al*., 1998; DiMartino and Cleary, 1999), and in patients with therapy-induced secondary leukemia (Felix, 1998). Unlike many other types of leukemia, the presence of an *MLL* translocation predicts early relapse and poor prognosis (for a review, see Cimino *et al*., 1998; DiMartino and Cleary, 1999). Frequently, leukemic blasts are characterized by the expression of lymphatic and myeloid surface markers, defining *MLL* or *mixed lineage leukemia* gene (Ziemin van der Poel *et al*., 1991). Recently, patients carrying *MLL* translocations were shown to exhibit a specific gene expression profile of activated and inactivated genes, suggesting a novel and specific disease phenotype (Armstrong *et al*., 2002).

MLL protein belongs to the Trithorax protein family and is part of a nuclear regulatory mechanism that establishes an epigenetic transcriptional memory system (Francis and Kingston, 2001). This system is based on high molecular weight protein complexes that exert histone/chromatin-modifying/remodeling and transactivating/repressing activities. In general, 'chromatin-modifier complexes' are composed of different combinations of Polycomb- and Trithorax-group proteins, as well as general transcription factors (e.g. GATA proteins) in combination with protein complexes that exert chromatin remodeling and histone-modifying activities. As a result, specific gene expression patterns will be established and maintained throughout subsequent mitotic cell cycles, which in turn will allow cells to cope with their cellular fate, or their specific developmental or differentiation pathways.

The importance of MLL protein for *hox* gene expression has been demonstrated in the embryos of heterozygous and homozygous *Mll* knockout mice. Expression of MLL protein was not necessary for turning on transcription of certain *hox* genes, but for the maintenance of their transcription (Yu *et al*., 1998). Insufficient Hox protein concentrations led to impaired development ($Mll^{+/-}$), while the homozygous knockout of the *Mll* gene was embryonal lethal at day 10.5 p.c. (Yu *et al*., 1995).

The human *MLL* gene became of interest because of its genetic alterations that are strictly correlated with human acute leukemias (acute myeloid leukemia or acute lymphoblastic leukemia). An illegitimate recombination event between the *MLL* gene and any of its many translocation partner genes results in reciprocal *MLL* fusion genes. Comparative twin analysis and quantitative analysis of leukemic cell clones on Guthrie test cards suggested that *MLL* translocations occur between the second and third month of pregnancy. By that time point, organogenesis has already taken place and fetal liver hematopoiesis has been activated (Wiemels *et al*., 1999). Obviously, *MLL* translocations are only permissive in cells predisposed to develop in hematopoietic cells, while they were never detected in any other tissue of the human body.

Based on our assumptions of MLL protein function, the disruption of an *MLL* wild-type allele and the additional expression of aberrant *MLL* fusion gene products are likely to result in the establishment of incorrect gene expression patterns during development of hematopoietic stem cells or progenitors.

Recently, experimental data became available that explained MLL protein function in more detail: Firstly,

*Correspondence: R Marschalek;
E-mail: Rolf.Marschalek@em.uni-frankfurt.de
Received 14 October 2002; revised 29 January 2003; accepted 30 January 2003

the MLL protein is rapidly proteolyzed into two large peptides, named MLL-N (amino acids 1–2666; p300) and MLL-C (amino acids 2719–3679; p180); the proteolytical sites were identified within two conserved amino-acid motifs (CS1: Q × D · GADD; CS2: Q × D · GVDD) and proteolytical cleavage occurs post-translationally between amino acids 2666/2667 and 2718/2719, respectively (Nakamura *et al*., 2002; Yokoyama *et al*., 2002; Hsieh *et al*., 2003). The resulting MLL peptides interact via dimerization domains located in MLL-N (amino acids 1979–2130, FYRN) and MLL-C (amino acids 3656–3969; FYRC and SET domain). The two complex MLL peptides build the basic structure required for an MLL multiprotein supercomplex (MLL MPSC) that consists of both MLL peptides and at least 27 additional proteins (Nakamura *et al*., 2002). The MLL MPSC exerts chromatin remodeling and histone-modifying activities, as concluded from the composition of the MLL MPSC and from experimental data (Nakamura *et al*., 2002). The post-translational cleavage and binding of MLL-N to MLL-C is protecting the MLL-N peptide from degradation (Hsieh *et al*., 2003); Secondly, MLL translocations occur predominantly between the exons 9 and 12, thus disrupting the wild-type MLL protein between amino-acid positions 1362 and 1444, respectively. As a result of reciprocal translocations, the N-terminal fusion portion of MLL exhibits two different DNA binding domains (AT-hooks and MT-domain; Zeleznik-Le *et al*., 1994; Birke *et al*., 2002) and the critical regions for correct subnuclear localization (Yano *et al*., 1997). Therefore, fusions of the N-terminal portion of the MLL protein (amino acids 1–1362/1444) with various partner protein sequences (derivative 11 proteins) should target the same subnuclear compartment, as evidenced from independent studies that have demonstrated a nuclear punctate distribution in immunohistochemical studies (Yano *et al*., 1997; Caslini *et al*., 2000). However, all derivative 11 proteins are missing the critical dimerization domain FYRN. This may indicate that derivative 11 proteins may not enter the MLL MPSC, but may compete for DNA binding (AT-hook, MT domain) at control regions or promoters of MLL MPSC target genes. Finally, fusion proteins containing the C-terminal portion of the MLL protein (derivative Z protein) contain both the FYRN and FYRC/SET interaction domain, as well as the proteolytic cleavage sites CS1 and CS2. Therefore, two cleaved peptides can be expected; one containing partner protein sequences fused to the PHD domain and the FYRN dimerization domain, while the other should be identical to the MLL-C peptide encoded by the wild-type MLL protein. The latter one exhibits the transcriptional transactivation domain, the FYRC dimerization domain, and the conserved SET domain. Thus, both post-translation cleavage products from a given derivative Z protein are functionally competent to compete for assembly into the MLL MPSC.

In conclusion, reciprocal fusion proteins of a given *MLL* translocation may either be assembled into the MLL MPSC, or are competing for DNA binding sites at MLL MPSC target genes, and thus, they are able to alter or disable functions of the MLL MPSC, which in turn may result in altered or ectopic gene expression patterns.

Apart from *MLL* translocations in acute leukemia and the recent findings, our knowledge of wild-type MLL protein function is limited, in particular, our insights into potential target genes affected by the MLL MPSC. For this reason, we used an established cell system to learn more about the pivotal role of the MLL protein in a 'normal' cellular environment. Expression profile experiments were performed using murine fibroblast lines that carry either wild-type or homozygously Disrupted *Mll* alleles to identify potential target genes that are affected by the absence or presence of MLL protein.

Results

Hybridization experiments and data analysis

First, the two fibroblast cell lines were analysed for their genotype and transcriptional properties of wild-type and disrupted *Mll* alleles (see Figure 1). In a second step, three independent mRNA preparations (one a week) were isolated from the two lines and pooled to generate a normalized probe for the chip hybridization experiments. The complete experiment was repeated after 6 months of continuing cell culture to assure the significance of the data and to prevent arguments against genetic instability of the maintained lines. Both data sets were compared and the results indicated that both cell lines exhibit extremely divergent gene expression profiles, with many genes up- and downregulated (Pearson correlation coefficient r of experiment $1 = 0.8859$; experiment $2 = 0.8853$; see Figure 2). Using different filter set criteria (increase/decrease (A), signal $\log_2$ ratio >1.32 or <-1.32 (B) and change P-value >0.999 or <0.001 (C)), a total of 197 differentially expressed candidate genes were consistently identified in both experiments. In *Mll* knockout cells, 136 genes were overexpressed, while in *Mll* wild-type cells 61 candidate genes were overexpressed. The top 40 gene lists of both cell lines are summarized in Tables 1 and 2 and the $\log_2$ expression rates of these genes were visualized in heat maps (Figure 3).

Differential gene expression in Mll *knockout cells*

The most highly activated genes identified in the *Mll* knockout cells code for the p53 apoptosis effector related to Pmp22 (PERP), vascular smooth muscle α-actin (VSMA), CD34 stem cell antigen, HNF-3/forkhead homolog (Brain factor 1), serine protease inhibitor 4 (PN-1), RBP-associated molecule RAM 14-1, PE31/TALLA-1, Nestin, adaptor molecule SRCASM, and the Myelodysplasia/Myeloid leukemia factor 1.

The *PERP* gene has been identified as a direct p53 target gene upon induction of apoptosis in murine embryonal fibroblasts (MEFs). *PERP* gene transcrip-

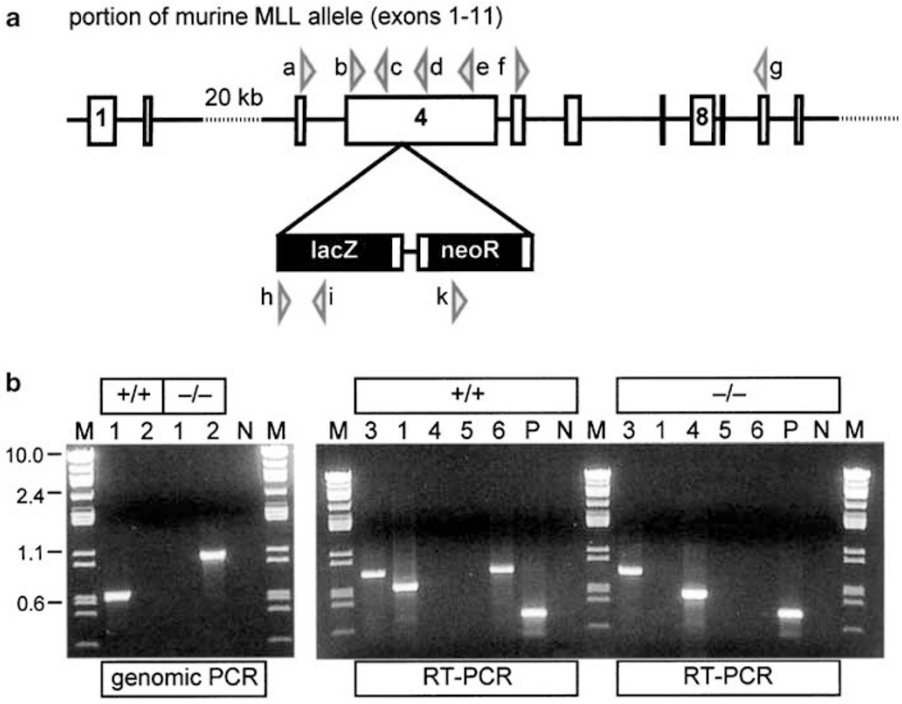

Figure 1 Integrity and expression of wildtype and knockout murine *Mll* alleles. (**a**) Scheme of the 5′-end of the murine *Mll* gene (exons 1–11). The *Mll* knockout cell line carries the in frame-insertion of the LacZ open reading frame into *Mll* exon 4. A series of oligonucleotides (a–k) is shown by triangles. (**b**) Genomic and RT–PCR analyses. Genomic DNA and isolated RNA from the $MLL^{+/+}$ and $MLL^{-/-}$ cell lines were used for PCR analyses. Left panel: genomic PCR using the oligonucleotide combinations b–d (1: 689 bp) and b–i (2: 1060 bp). Right panel: RT–PCR using the oligonucleotide combinations a–c (3: 837 bp), b–d (689 bp), h–i (4: 550 bp), k–e, f–g (6: 833 bp), positive (P: GAPDH: 432 bp), and negative controls, respectively. A PCR fragment in reaction 4 is indicative for the genomic LacZ cassette insertion, while reaction 6 verifies the expression of the *Mll* gene beyond the insertion site of the LacZ/NeoR cassette

tion correlates with p53-dependent apoptosis but not with p53-mediated cell cycle arrest. PERP predominantly localizes to the endoplasmic reticulum and Golgi apparatus, and displays sequence similarity to PMP-22/GAS-3 (peripheral myelin protein 22/growth arrest specific 3), a tetraspan membrane protein that is commonly mutated in human hereditary peripheral neuropathies (i.e. Charcot–Marie–Tooth). PERP appears to function in part by regulating cell proliferation and apoptosis (Attardi *et al*., 2000).

vSMA is an important protein for smooth muscle contractions and involved in the process of wound healing and differentiation into myofibroblasts (Cogan *et al*., 2002). The overexpression of vSMA was verified by immunohistological experiments and led to an alteration in the cellular morphology of $MLL^{-/-}$ and $MLL^{+/+}$ cells (unpublished data).

Relatively, little is known about the murine CD34 stem cell antigen. The murine counterpart of the human CD34 stem cell antigen has less importance for developmental processes of the hematopoietic system than in the human system. However, it is interesting to note that the CD34 antigen is highly expressed in $MLL^{-/-}$ fibroblasts, while transcription of CD34 is shut down in fibroblasts expressing the MLL protein (First experiment >100-fold; second experiment 74.5-fold). This may argue for a model in which stem cells that express high levels of the CD34 antigen have their *Mll* gene turned off. During hematopoietic differentiation, the *Mll* gene is transcriptionally turned on, and thus, CD34 production is lowered and finally shut down. If this is also true for human cells, long-term cultivation of hematopoietic stem cells, without losing differentiation potential, may be performed by blocking *MLL* gene expression using RNAi technologies.

HNF-3/forkhead homolog (*Brain factor 1*; *BF-1*) belongs to the *winged helix* gene family that encodes transcriptional repressor proteins. BF1 plays a pleiotropic role during development of the cerebral hemispheres of the brain. Mice lacking BF1 have defects in the morphogenesis of the dorsal telencephalon (e.g. neocortex) and the ventral telencephalon (Hanashima *et al*., 2002). BF1 interacts with several SMAD proteins via the FAST-2 protein (forkhead activin signal transducer-2) protein. This prevents SMAD2/4 proteins to bind to their cognate DNA response elements, located in several promotors of TGF-β inducible genes. Moreover, BF-1 antagonizes TGF-β signaling (Dou *et al*., 2000).

Serine protease inhibitor 4 (protease-nexin 1, PN-1) is a serine protease inhibitor (serpin) that inactivates several proteases, including thrombin, urokinase, plasminogen activators (PA), and plasmin. It also plays a role in regulating proteolytic activity generated by the

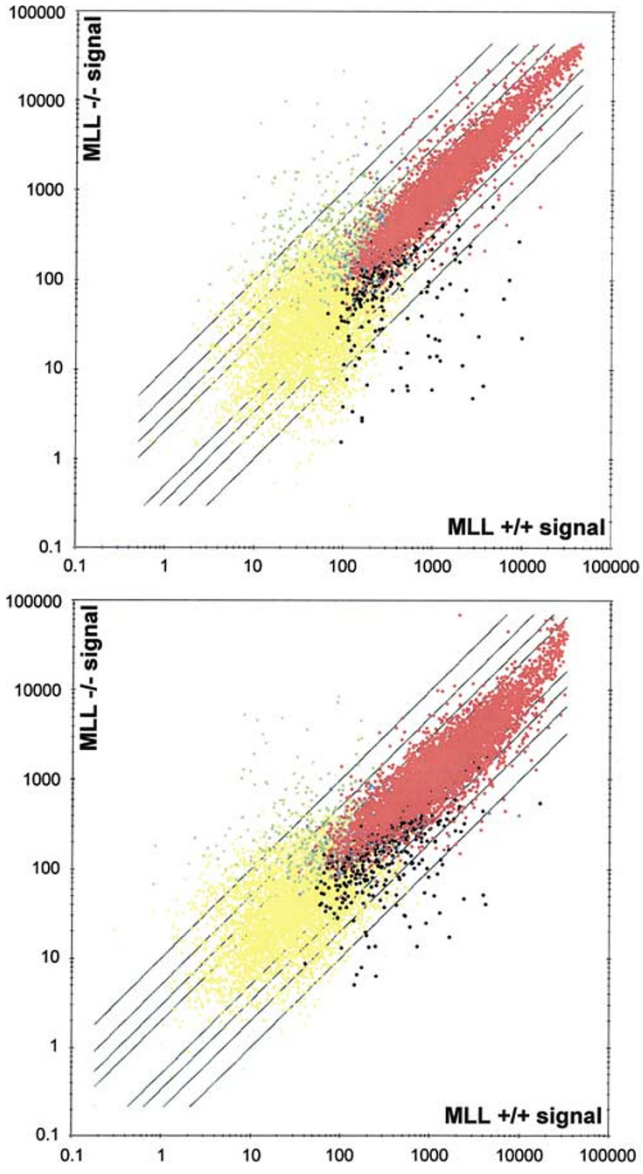

Figure 2 Scatter data of the chip hybridization experiment. Scatter blots showing the raw data of the first and second chip hybridization experiment. The second experiments was performed 6 month after the first experiment. Red dots: present in both population but unaffected; yellow dots: absent or marginal in both populations; green dots: genes only present in $MLL^{-/-}$ cells; black dots: only present in $MLL^{+/+}$ cells. The Pearson correlation coefficients were 0.8859 and 0.8853, respectively. The diagnonal lines indicate two-, three-, five-, and 10-fold overexpression for both subsets

PA system. PN-1 is known to be involved in tissue remodeling, cellular invasiveness, matrix degradation, and tumor growth. PN-1 is highly secreted by astrocytes in the brain but overexpression of PN-1 leads to progressive neuronal and motor dysfunction in post-mitotic neurons because of altered postsynaptic transmission (Meins *et al*., 2001).

RBP-Jk associated molecule RAM 14-1 (*Kyo-T2*) mRNA represents a splice variant of the *Kyo-T* gene. It encodes two LIM domains and a specific interaction domain for the RBP-Jk protein. RAM 14-1 protein was identified as a specific inhibitor of the Notch-1 pathway by preventing the Notch-1 IC/RBP-Jk heterodimeric protein from binding to its cognate DNA target site (Taniguchi *et al*., 1998). Thus, the Notch-1 and subsequent differentiation pathways are blocked. The currently known target genes affected by the Notch-1 pathway are the *hairy/enhancer of split* (*hes*) genes, the *HERP1* and *HERP2* genes, and the *NFkB2* gene. They are important for somatic recombination, MHC expression and are involved in the processes of neuronal and sensory organ development (de la Pompa *et al*., 1997). RBP-Jk protein heterodimerizes also with EBNA2 and EBNA3A-C, and thus establishes ectopic gene expression patterns leading to cellular transformation after EBV infection (summarized in Figure 4). More interestingly, both the *notch-1* and the *RBP-Jk* knockouts are embryonal lethal at day 10.5 p.c., the same day of development when homozygous *Mll* knockout mice die *in utero*. This argues for a link of the *Mll* knockout phenotype with impaired Notch-1 or RBP-Jk functions.

The four and a half LIM domain 1 gene transcript (*FLH1*; *Kyo-T1*) was also found to be activated. This transcript represents another splice form of the *Kyo-T* gene that is predominantly found in the cardiovascular system and in striated muscles (Chu *et al*., 2000). *FLH1* expression is linked to key events in myogenesis and functions as transcriptional coactivator for CREB or CREM transcription factors.

PE31/TALLA-1 (*A15*) is a member of the *transmembrane 4* gene superfamily and was isolated by differential screening of the cDNAs that are preferentially expressed on immature T-cells. TALLA-1 antigen has been found to be expressed exclusively on T-ALL cell lines, but was completely absent on peripheral blood mononuclear cells, including fresh and activated T cells (Takagi *et al*., 1995). *A15* encodes 244 amino acids with four hydrophobic domains. Northern blot analysis of the RNA samples from various mouse tissues disclosed that the *A15* transcripts are expressed most strongly in the brain, and are detectable in the colon, muscle, heart, kidney, and spleen. *In situ* hybridization of the mouse brain with an *A15* ribo-probe suggests that A15 protein may have a fundamental role within the higher nervous system (Hosokawa *et al*., 1999).

Nestin is an embryonic intermediate filament protein that represents a neural precursor/embryonal stem cell marker and is correlated with a differentiation fate to astroglial cells. Nestin may play an important role in remodeling and repairing the postnatal and adult central nervous system. Nestin has also been implicated in myogenesis (Wei *et al*., 2002).

The SRCASM protein is an adapter protein that interacts with and is phosphorylated by Fyn, a Src family tyrosine kinase. Phosphorylated SRCASM protein is able to interact with Grb-2 and the regulatory subunit of phosphoinositol 3-kinase, p85, and thus links a cell surface receptor to several signal cascades (Seykora *et al*., 2002).

Myelodysplasia/myeloid leukemia factor 1 (Mlf1) is a nuclear proto-oncoprotein involved in the

Table 1 Expression profile of *Mll* knockout cells. Top 40 list includes Genbank accession codes (GAC#), gene names (target gene), and log_2 factors of differential gene expression (signal log_2 ratios)

GAC#	*Target gene*	*Sig log_2 ratios*	
AI854029	p53 apoptosis effector related to Pmp22 (PERP)	8.57	9.11
X13297	mRNA for vascular smooth muscle α-actin	7.86	7.97
S69293	CD34 stem cell antigen	7.49	6.22
U36760	HNF-3/forkhead homolog, Brain factor 1	6.45	7.21
X70296	Serine protease inhibitor 4	6.32	7.31
U41739	RBP-associated molecule RAM 14-1 mRNA	6.68	6.18
D26483	mRNA for PE31/TALLA-1	4.92	7.87
AW061260	Nestin (neuronal stem cell protein)	6.33	6.01
AI840130	Adaptor molecule SRCASM (Src activating signaling molecule)	5.62	5.33
AF100171	Myelodysplasia/myeloid leukemia factor 1 (Mlf1)	3.83	6.94
AI153421	Erythroid diff. regulator	2.73	7.18
D76440	Necdin	5.32	4.54
X56304	Tenascin C	3.47	6.29
Y13185	Stromelysin-2	4.75	4.67
AF061260	Nectin-like protein 2	4.49	4.45
M31418	Interferon-activated gene 202	6.09	2.83
AF038939	Peg-3 zinc-finger	4.38	4.29
L47600	Cardiac Troponin T (subtype A3b)	4.95	3.48
AB021966	mRNA for Adhesion protein RA175C	4.23	4.15
AF297645	Stromelysin-3	5.20	3.06
AI839950	Four and a half LIM domain 1 (Fhl1)	4.17	4.08
M83649	Fas antigen	4.74	3.37
U74683	Cathepsin C	3.24	4.65
X56561	Homeo box D8	3.29	4.47
AF004428	Tumor protein D52-like 1	3.86	3.60
X59252	Homeo box msh-like 2	3.07	4.02
AJ242663	mRNA for Cathepsin Z precursor (ctsZ gene)	3.81	3.27
X53929	Decorin	4.64	2.28
AI839175	Serum deprivation response	3.90	3.00
U09816	GM2 activator protein (Gm2a) mRNA	3.79	2.73
D83745	ANA	2.79	3.58
V00727	FBJ osteosarcoma oncogene	3.71	2.34
U38967	Prothymosin beta 4	3.21	2.83
U43319	Frizzled homolog 6	3.47	2.09
AW120986	Suppressor of mif two 3 (SMT3); SUMO	2.58	2.96
M22115	Homeo box A 1	2.96	2.03
X83577	Glypican 4	2.21	2.23
M31419	Interferon activated gene 204	2.42	1.42
AA763466	Collagen type I, alpha 1	2.71	1.53
D78382	TOB	1.74	2.20

t(3;5)(q25.1;q34) translocation (NPM–MLF1 fusion protein) associated with acute myeloid leukemia (Yoneda-Kato *et al*., 1996). Mlf1 tissue distribution is restricted during development and postnatal life, with high levels present only in skeletal, cardiac, and selected smooth muscle, gonadal tissues, and rare epithelial tissues including the nasal mucosa and the ependyma/choroid plexus in the brain. Mlf1 transcripts were undetectable in the lympho-hematopoietic organs (embryonic and adult mouse), suggesting that the NPM-MLF1 fusion protein contributes to MDS/AML by ectopic overexpression of parts of the MLF1 protein in hematopoietic tissues. MLF1 protein is able to interact with the myeloid nuclear differentiation factor (MNDA) that is also a nuclear protein expressed specifically in the developing cells of the myelomonocytic lineage, including the end-stage monocytes/macrophages and granulocytes. Moreover, elevated MLF1 expression is correlated with malignant progression from myelodysplastic syndrome.

Some *hox* genes were identified to be transcriptionally activated in $MLL^{-/-}$ cells: the homeo box genes *hoxA1*, *hoxD8*, and *msh-like 2* (*hox8.1*) were found to be upregulated 5.6-fold, 14.8-fold, and 11.7-fold, respectively. HoxA1 protein is involved in patterning the hindbrain and neuronal crest cells. Targeted disruption of *hoxA1* led to severe abnormalities of the optic capsule and membranous labyrinth (Pasqualetti *et al*., 2001) and to malformations of the rhombomeric organization of the hindbrain. HoxD8 protein was found to be expressed in mesoderm and neuroectodermal structures, as well as in limbs and gonads (Izpisua-Belmonte *et al*., 1990). Hox8.1 protein is expressed in the surface ectoderm and in the optical vesicle before invagination (Monaghan *et al*., 1991).

SUMO proteins are small ubiquitin-related polypeptides that are reversibly conjugated to many nuclear proteins. SUMOylation of target proteins does not lead to degradation but instead appears to regulate protein–protein interactions, intracellular localization and protects some modified targets from ubiquitin-dependent

Table 2 Expression profile of *Mll* wild-type cells. Top 40 list includes Genbank accession codes (GAC#), gene names (target gene), and $\log_2$ factors of differential gene expression (signal $\log_2$ ratios)

GAC#	*Target gene*	*Sig $\log_2$ ratios*	
X16009	Mitogen regulated protein/proliferin (Mrp/plf3 gene) exon 1	9.23	6.61
NM_008597	Matrix gamma-carboxyglutamate protein (MGP)	6.34	6.56
L32838	IL-1R antagonist	9.10	3.26
NP_056559	Ectonucleotide pyrophosphatase/phosphodiesterase 2 (Enpp2)	8.85	2.53
AF064749	Collagen type VI α 3	5.92	4.99
K03235	Proliferin	4.89	6.01
NW_000027	Dimerization cofactor of HNF-1 α (DCoH)	7.22	3.43
L00653	Mast cell protease 7 (mcp7)	5.81	4.70
AF045887	Angiotensinogen	4.83	5.58
AF045801	Gremlin	5.45	4.43
AA673487	EST cDNA	6.87	2.75
U70132	Paired-like homeodomain transcription factor 2/RIEG/ARPl/Ptx2	5.25	4.23
AB008174	HNF1 beta	4.00	4.78
AW214136	EST cDNA	5.45	2.70
L06047	Glutathion-S-transferase α 1	3.03	5.04
L19932	TGF-β induced β ig-h3 gene, 68 kDa	4.62	3.31
Z18272	Procollagen, type VI, α 2	2.83	4.37
AF081789	Surface antigen AA4	4.40	2.62
NM_011844	Monoglyceride lipase (Mgll)	3.75	2.79
X14432	Thrombomodulin (Thbd)	4.62	1.79
L07918	GDP diss. inhibitor, preferentially expr. in hematopoietic cells	4.45	1.95
U09968	p27 kipl	3.63	2.76
X66405	Procollagen, type VI, α 1	2.59	3.35
AI843571	EST cDNA	3.10	2.37
AF057156	Small proline-rich protein 1A (Sprr1a) gene	3.13	2.32
U06119	Cathepsin H	2.53	2.81
X74040	Mesenchyme fork head-1	3.44	1.66
L07803	Thrombospondin 2	3.13	1.95
AB009377	CAD (DNA fragmentation factor)	2.95	2.09
L12367	Adenylyl cyclase-associated protein (CAP)	3.67	1.34
U36384	Twist-related bHLH protein Dermo-1	2.36	2.60
X07439	Homeo box Hox 3.1	3.40	1.47
U49513	Small inducible cytokine A9	2.21	2.60
X65128	Growth arrest specific 1; GAS1	1.62	3.05
X97227	Cell surface glycoprotein CD53	2.66	1.66
AW122653	Mevalonate kinase	1.37	2.78
U51335	GATA-6	1.93	2.17
NM_030704	Cryac heat-shock protein 20-like protein	2.23	1.61
U15159	LIM-domain containing, protein kinase (LIMK)	1.64	1.88
U07159	Acetyl coenzyme A dehydrogenase	1.55	1.39

degradation. SUMO alters the interaction properties of its targets and affects their subcellular localization. Examination of the PML nuclear bodies, whose components are SUMOylated, has revealed this modification to be essential for structural and functional integrity (Seeler and Dejean, 2001).

TOB and ANA are both members of the TOB/BTG antiproliferative protein family. The *BTG* gene family includes *Pc3*/*Btg2*, *Btg1*, *TOB*, *TOB2*, *ANA*/*Btg3*, *Pc3k*, and others. Unphosphorylated TOB negatively regulates proliferation of osteoblasts and T cells by inhibiting G0/G1 to S transitions. Activation of the ras/raf cascade leads to ERK1/2-dependent phosphorylation of TOB (Suzuki *et al*., 2002) that reduces its ability to block cell cycle progression (affects Rb, cyclin D1).

Differential gene expression in Mll *wild-type cells*

The most activated genes identified in the *Mll* wild-type cells code for Mrp/plf3 proliferin, matrix γ-carboxyglutamate protein (MGP), IL-1R antagonist, ectonucleotide pyrophosphatase, collagen type VI alpha 3(COL6A3), dimerization cofactor of hepatocyte nuclear factor 1 α (DCoH), mast cell protease 7, and angiotensinogen.

Proliferin has been described as a growth factor of the prolactin/growth hormone family and as an angiogenic placental hormone. It is normally expressed during cutaneous wound healing processes and is temporally expressed during the hair follicle cycle. Recently, it has been identified to be overexpressed in the end stage of fibrosarcomas, indicating a potential role as angiogenic hormone for malignantly transformed tumor cells (Toft *et al*., 2001).

MGP has been identified as the key regulatory molecule after PTH-mediated inhibition of the bone mineralization process by osteoblasts (Gopalakrishnan *et al*., 2001). MGP overexpression in developing limbs inhibits cartilage mineralization, delays chondrocyte maturation, and blocks endochondral ossification. MGP is a powerful and developmentally regulated inhibitor of cartilage mineralization, and appears to

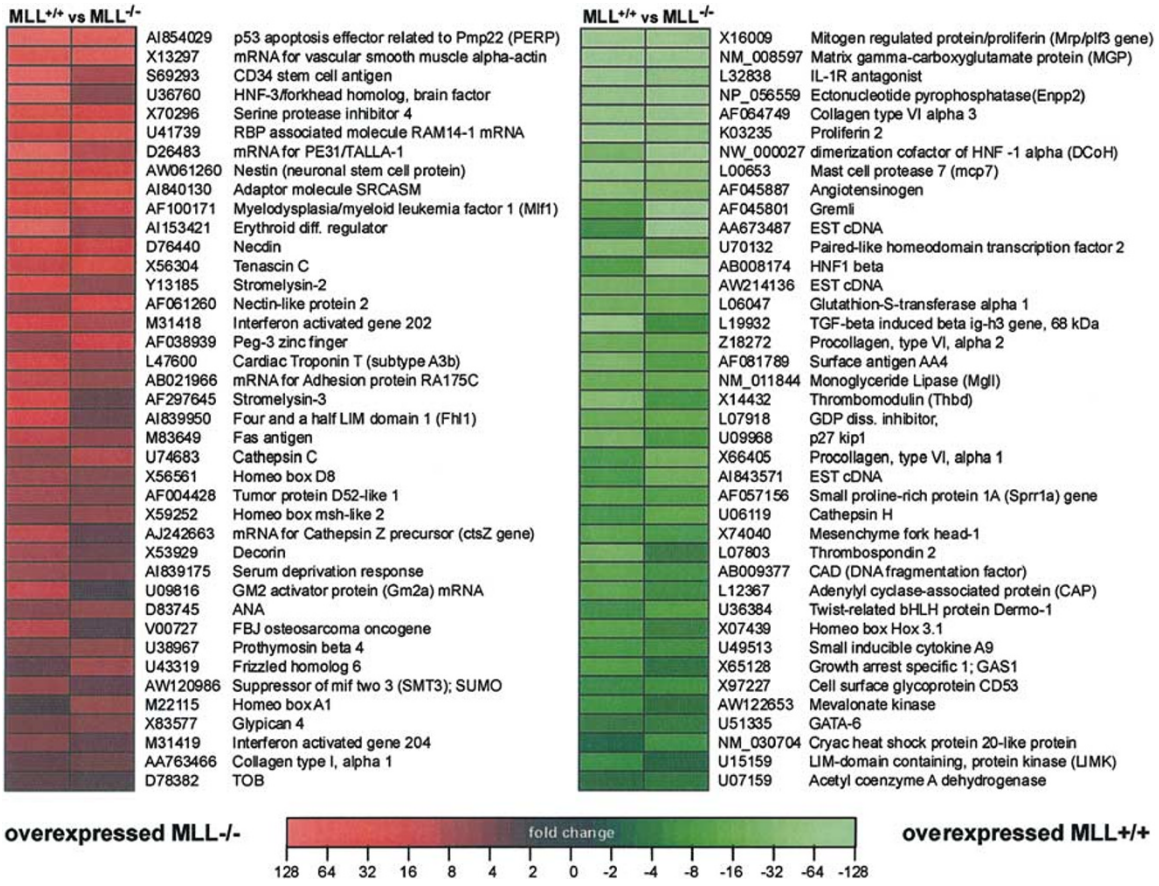

Figure 3 Heat maps of gene expression changes in MLL wildtype versus MLL knock-out cells. The top 40 genes expressed in $MLL^{-/-}$ and $MLL^{+/+}$ cells of both hybridization experiments are shown as heat maps. Color-codes: red-scale = overexpressed in $MLL^{-/-}$ cells; green-scale = overexpressed in $MLL^{+/+}$ cells. Genbank accession codes and gene names are given next to colored rectangles. Color scales represent eight $\log_2$ steps

have a role in regulating chondrocyte maturation and ossification processes (Yagami *et al*., 1999). In addition, MGP mRNA is constitutively expressed in normal vascular smooth muscle cells and plays a role in preventing vascular calcification processes.

IL-1R antagonist is a competitive inhibitor of the binding of IL-1 to the IL-1 receptor (IL-1R). Thus, IL-1R antagonist is a powerful inflammatory inhibitor that counteracts IL-1-mediated inflammation, septic shocks, and hematopoietic defects. Mice lacking the *IL-1R antagonist* (*IL-1Ra*) gene show abnormal development and homeostasis as well as altered responses to infectious and inflammatory stimuli (Irikura *et al*., 2002). Absence of IL-1 signaling appears to assist the early embryonic blastomer cleavage (Spandorfer, 2000). In addition, Il-1R antagonist is highly expressed in embryos during implantation into the endometrial surface.

Ectonucleotide pyrophosphatase belongs to a family of membrane bound enzymes that degrade extracellular ATP and ATP derivatives, which are known to influence cell signaling, cell motility, inflammation, and apoptosis. It is also known to be an activation marker of basophil granulocytes (CD203c) after allergic response and can be transcriptionally upregulated by TGF-β. Overexpression of ectonucleotide pyrophosphatase serves to prevent inflammation and influences cellular communication (for a review, see Bollen *et al*., 2000).

COL6A3 is a primary constituent of the extracellular matrix and disruption of the *COL6A3* gene has been implicated in muscular dystrophies. It also plays an important role at later stages of neural crest development and promotes neural crest cell migration and attachment (Perris *et al*., 1993). COL6A3 shows specific interaction with decorin as well as interactions with other matrix proteins, for example, hyaluronan, heparan sulfate, and NG-2 proteoglycans (Burg *et al*., 1996). Significantly, the NG-2 surface protein has been identified in MLL translocation patients (Armstrong *et al*., 2002). *COL6A3* can be transcriptionally upregulated by TGF-β.

The dimerization cofactor of hepatocyte nuclear factor 1 alpha, the bifunctional DCoH protein, stabilizes HNF-1 alpha dimers and enhances their transcriptional activity (Mendel *et al*., 1991). In addition, DCoH protein is a pterin-4-α-carbinolamine dehydratase that is involved in the regeneration of the cofactor tetrahydrobiopterin. HNF-1 α (TCF-1) has been identified to be biallelic inactivated in hepatic adenomas (Bluteau *et al*., 2002) and plays an important role as nuclear

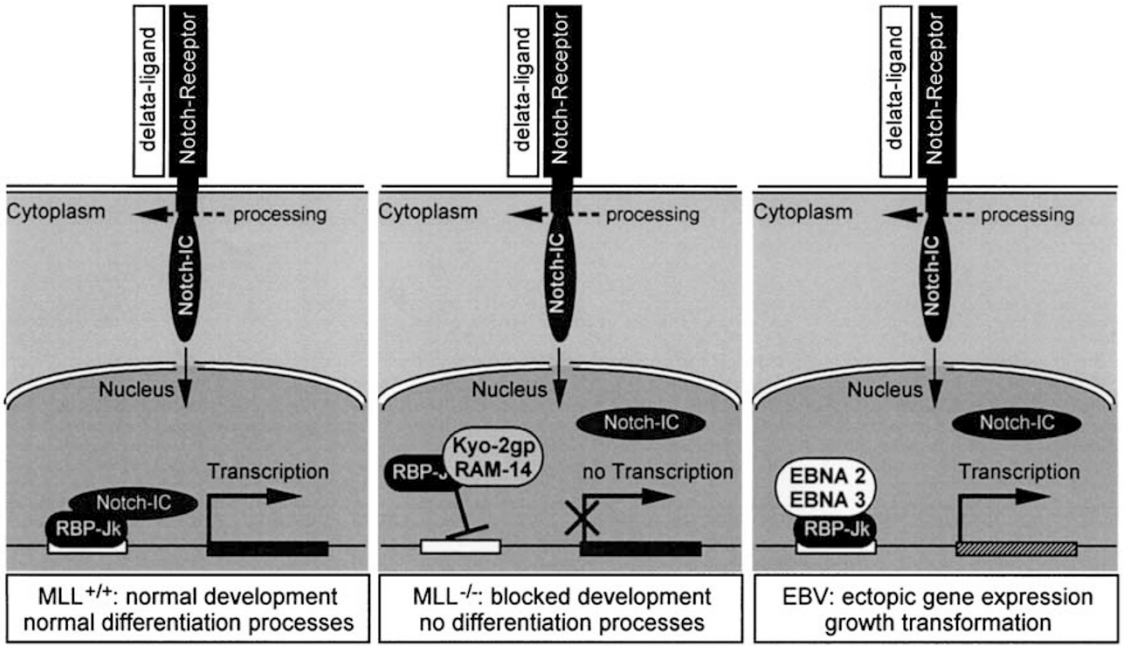

Figure 4 Notch-1/RBP-Jk pathways in mammalian cells. Left: intracellular Notch-1 (Notch-IC) is cleaved after ligand binding to the Notch-1 receptor. The cleaved intracellular domain enters the nucleus and heterodimerizes with RBP-Jk to activate specificly target gene transcription. Middle: in the presence of RAM 14-1 protein, Notch-IC is prevented from binding to RBP-Jk; the RBP-Jk/RAM14-l heterodimer is unable to bind to its cognate DNA response element, and thus, transcription of target genes is blocked. Right: after EBV infection, EBNA proteins compete with Notch-IC for RBP-Jk binding; the heterodimers of different EBNA proteins and RBP-Jk lead to ectopic transcription of genes that are important for the EBV life cycle and subsequent transformation

target protein of the Wnt/β-catenin signaling pathway. During mouse development, limb bud formation, lung formation, nephrogenesis and thymocyte development are critically dependent on Wnt signaling and TCF-1. Interestingly, *HNF-1 β* (*TCF-2*) has also been identified as a differentially regulated candidate gene in MLL$^{+/+}$ cells.

Besides these, several transcription factors have been identified in *Mll* wild-type cells, which are involved in ecto- and mesodermal development during organogenesis.

Mesenchyme fork head-1 (MFH-1) is a winged helix/forkhead transcription factor that binds to the HNF-3 binding site and acts as positive transactivator. MFH-1 is expressed temporally in developing embryos, first in the non-notochordal mesoderm and later in areas of mesenchymal condensation in the prospective trunk region, in cephalic neural crest and cephalic mesoderm-derived mesenchymal cells of the prechordal region in early embryos. Moreover, BMP-2 induces the expression of MFH-1 in developing cartilaginous tissues, kidney, and arch arteries, and is essential for the normal development of the axial skeleton and aortic arch formation of mice. MFH-1-deficient mice die embryonically and perinatally, and exhibit an interrupted aortic arch, skeletal defects, defects in the neurocranium and the vertebral column (Iida *et al*., 1997).

Twist-related bHLH protein Dermo-1 is a multifunctional basic helix–loop–helix transcription factor binding to the E-box consensus sequence (CANNTG) and has been shown to be a potent negative regulator for gene transcription and apoptosis. Dermo-1 is a homolog to *Drosophila* Twist and is expressed in a subset of mesodermally and ectodermally derived tissues. Dermo-1 represses in a dose-dependent fashion the transcriptional activation mediated by the heterodimer MyoD/MEF-2 (myocyte enhancer factor 2; Gong and Li, 2002).

A set of *hox* genes was identified to be transcriptionally activated in MLL$^{+/+}$ cells. The genes *hoxA7*, *hoxB9*, *hoxC9*, *hox 3.1* (*hoxC8*), *iroquois*, and the *paired-like homeo-domain transcription factor 2* (*RIEG*, *Ptx2*, *ARP1*) were upregulated 14.8-, 5.4-, 9.0-, 5.4-, 3.9-, and 26.7-fold, respectively. Significantly, the four genes *hoxa7*, *hox C8*, *hoxC9*, and *ARP1* have already been identified as MLL target genes (Arakawa *et al*., 1998; Yu *et al*., 1998; Milne *et al*., 2002).

HoxB9 expression is correlated with myeloid and erythroid differentiation processes, while HoxC8 is found to be mainly involved in neuronal development. Homozygous knockout of *hoxC9* is correlated with anterior transformation of several skeletal vertebrae and an additional pair of ribs (Le Mouellic *et al*., 1992). The Iroquois homeobox protein 3 (IRX-3) plays a pivotal function in the initial specification of the vertebrate neuroectoderm, and in concert with other transcription factors for later subdivision of the anterior–posterior and dorso-ventral axis of the neuroectoderm. Iroquois proteins are effective early in development to specify the identity of diverse territories of the body, such as the dorsal head and dorsal mesothorax of Drosophila and the neural plate (Cavodeassi *et al*., 2001). In addition,

iroquois genes are known targets of Pc-G/Trx-G complexes, an additional hint that an MLL target gene has been identified. ARP1 is expressed highly in bone marrow cells, but is completely absent in leukemic blasts carrying an MLL translocation. The latter protein is also involved in developing processes of the brain (diencephalon, mesencephalon, and rhombencephalon), in postmitotic neuron development, as well as in the development of eyes, teeth and the umbilicus. Mutations lead to malformations similar to those described in Rieger syndrome (for a review, see Alward, 2000).

A number of genes were found to be activated in MLL-expressing cells that encode secreted proteins. Gremlin is a member of the Dan family of BMP antagonists that functions early in development. Gremlin is able to bind to BMP2, BMP4, and BMP7 and prevent them from binding to their cognate receptors. Binding to these BMPs influences many developmental processes, for example, mesoderm development, control of limb outgrowth, and lung development (Capdevila *et al*., 1999). Gremlin is downregulated in transformed cells, while overexpression of Gremlin in tumor cell lines has a tumor suppressive effect.

β ig-h3 protein is secreted from many different cells after TGF-β treatment (Skonier *et al*., 1992). Human fibroblasts secrete β ig-h3 into the extracellular matrix where it binds to fibronectin (FN) and type I collagen (Col I). It is involved in the tube formation process of blood vessels during angiogenesis, an important mechanism for embryonic development, wound healing, and tumor growth, In addition, transformed cells downregulate β ig-h3 protein.

The GDP dissociation inhibitor (D4) is a homolog of RhoGDI; both exhibit GDI activity against the Ras-related Rho family GTPases (CDC42Hs, rho and rac I proteins). During murine embryogenesis, GDP dissociation inhibitor (D4) transcripts are detected in yolk-sac cells, where the earliest hematopoietic precursors are found. When these precursors undergo proliferation and differentiation, a dramatic increase in GDP dissociation inhibitor (D4) expression is observed. Another argument for the significance of the GDP dissociation inhibitor (D4) arises from *in vitro* differentiation experiments; when totipotent murine embryonic stem cells develop into hematopoietic cells, the *GDP dissociation inhibitor* (*D4*) gene is activated with the onset of hematopoiesis and modulated during further development (Lelias *et al*., 1993).

Sprr1a gene overexpression has been identified in injured neurons, where it is upregulated up to 60-fold. Overexpressed *Sprr1a* colocalizes with F-actin in membrane ruffles and augments axonal outgrowth and is essential for successful nerve regeneration (Bonilla *et al*., 2002).

Growth arrest specific 1 (GAS-1) was identified in *Mll* wild-type cells. The gene product is known as a tumor suppressor and is able to inhibit growth transformation of tumor cells by blocking G0/G1 transition and DNA synthesis. The mechanism is p53-dependent and cannot be observed in p53-mutated cells or MDM-2-overexpressing cells. Besides its effects on cell cycle and proliferation, GAS proteins are developmentally expressed in spatial–temporal expression patterns. GAS-1 expression was identified in 8.5- to 14.5-day-old embryos where it is expressed in different abundance in most organ systems including the brain, heart, kidney, limb, lung, and gonad (Lee *et al*., 2001).

The surface protein CD53 is an indicator for the presence of the ets family transcription factor Pu.1. CD53 has been identified as one of the target genes that have been identified in Pu.1 knockout cells when stably transduced with a PU.1-expressing retrovirus (Henkel *et al*., 2002). In our study, the myeloid-specific genes coding for CD11a, CD18, GM-CSFR-α, c-fes (1.6-fold) as well as the B-cell specific genes like Btk, CD27, vav-1 (1.8-fold) are downstream targets of Pu.1 complexes that contain several other transcription factors (e.g. Sp1, C/EBP, AML1, Oct). Pu.1 is expressed very early in precursors for myeloid and lymphatic lineages. An invariant consequence of *Pu.1* mutations is a multilineage defect in the generation of progenitors for B and T lymphocytes, monocytes, and granulocytes.

GATA-6 protein is expressed in primitive endoderm structures and various mesoderm- and endoderm-derived tissues such as heart, liver, lung, gonad and gut where it plays a critical role – together with GATA-4 and GATA-5 protein – in regulating tissue-specific gene expression (Morrisey *et al*., 1996). The homozygous knockout of *GATA-6* led to embryonal lethality (days 5.5–7.5 p.c.), with defects in visceral endoderm function and subsequent extraembryonic development (Morrisey *et al*., 1998; Koutsourakis *et al*., 1999). Interestingly, GATA-6 protein plays a role in vascular smooth muscle cells by regulating proliferation in response to mitogenic or mechanical stimulation.

LIM-kinase (LimK-1) plays a significant role in brain development, synapse formation and functions by acting on Rac-mediated actin reorganization and reverses cofilin-induced actin depolymerization (Yang *et al*., 1998). Abnormal expression of LimK-1 is correlated with Williams syndrome, a mental disorder (Meng *et al*., 2002). Between days 7.5 and 8.5 p.c, LimK-1 is expressed in three broad domains within the embryo, the neuroectodermal of the prospective forebrain and midbrain regions, the cardiac mesoderm, and the newly formed definitive endodermal derivatives, the foregut and hindgut. By day 10.0, LimK-1 remains prominently expressed in the ventromedial regions of the developing forebrain and midbrain, with continued expression in the hindgut. In adults, LimK-1 is expressed most prominently in the brain.

Verification of the obtained chip hybridization data by RT–PCR experiments

A series of oligonucleotides was designed and used with total RNA (DNA free) to perform RT–PCR experiments to verify the hybridization data, In a first step, a total of 48 genes were tested and 41 reactions resulted in a specific PCR product, In a second step, 25 of these target genes were amplified with 25, 30, and 35 cycles to semiquantitatively reproduce the observed gene

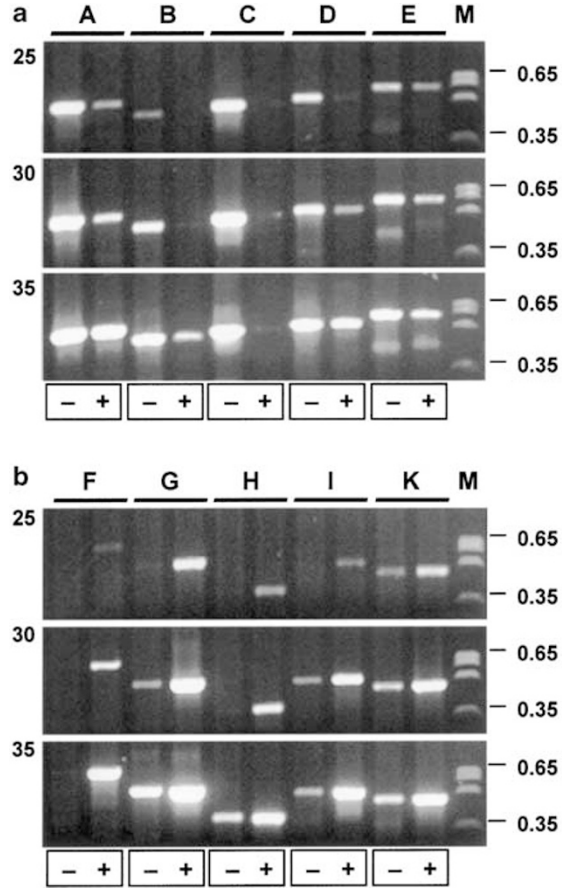

Figure 5 Verification of differential gene expression by RT–PCR. Several identified target genes were tested by semiquantitative RT–PCR experiment using 25–35 cycles. (**a**) RT–PCR of potential target genes overexpressed in MLL$^{-/-}$ cells: A: mRNA for vascular smooth muscle α-actin (#2 of 40); B: RBP associated molecule RAM14-1 mRNA /(#6 of 40); C: Serine protease inhibitor 4 (#4 of 40); D: Four and a half LIM domain 1 (#21 of 40); E: TOB (#40 of 40). M: DNA size marker. The number of cycles are indicated on the left. (**b**). RT–PCR of potential target genes overexpressed in MLL$^{+/+}$ cells: F: IL-1R antagonist (#3 of 40); G: Collagen type VI alpha 3 (#5 of 40); H: Matrix γ-carboxyglutamate protein (#2 of 40); I: Angiotensinogen (#9 of 40); K: TGF-β induced beta ig-h3 gene, 68 kDa (#16 of 40). M: DNA size marker. The number of cycles are indicated on the left. Sizes are given on the right

expression differences in MLL$^{-/-}$ and MLL$^{+/+}$ cells (all experiments were performed in triplicate). Representative results from 10 genes are shown in Figure 5. Although this type of analysis does not allow us to draw any accurate conclusions, the observed results of the RT–PCR experiment were in good agreement with the data obtained from the chip hybridization experiments. Specific oligonucleotides were designed for the human homolog of the interesting candidate gene *KyoT*, named *FHL-1*, enabling the analysis of cDNA derived from human cell lines. As shown in Figure 6, the Notch-1 1C inhibitor RAM14-1, named FHL-1C in the human system, was not found in the human T-cell line Jurkat, but was strongly expressed in the human t(4;11) cell line SEM (Greil *et al*., 1994; Marschalek *et al*., 1995). This demonstrates that the *KyoT*/*FHL-1* gene is not only highly transcribed in MLL$^{-/-}$ cells, but also alternatively

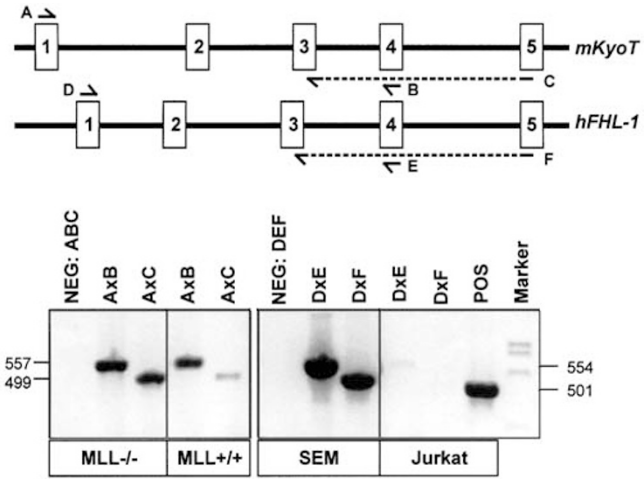

Figure 6 RAM14-1/FHL-1C is expressed in t(4;11) but not in Jurkat cells. RT–PCR experiments were performed on cDNA derived from MLL$^{-/-}$, MLL$^{+/+}$, SEM, and Jurkat cells. To distinguish between the splice variants of the *mouse KyoT* and the *human FHL-1* gene, oligonucleotides were designed to detect transcripts either containing exon 4 (primer combinations A–B, D–E) or missing exon 4 (primer combinations A–C, D–F). Top: scheme of the homolog genes *KyoT* and *FHL-1*. Bottom: RT–PCR analyses showing specific PCR amplimers after 35 cycles. Sizes of amplimers are given on the left (murine) and right (human). Negative controls were performed without template cDNA, while the positive control was endogenous GAPDH

spliced. Thus, alternative splicing and the steady-state amount of the *KyoT*/*FHL-1* RNA in both the mouse and human system seemed to be affected by the gene dose provided by the endogenous *Mll*/*MLL* alleles.

Discussion

Potential MLL target genes have been unveiled by gene profiling experiments in a biological model system. Comparison of two independent experimental data sets led to the discovery of 197 potential target genes of the MLL protein, constituting the MLL MPSC. Interestingly, in the absence of MLL protein, more target genes were identified than in the presence of MLL (136 versus 61 target genes). This suggests that the MLL multiprotein supercomplexes are not only involved in the process of chromatin activation, but also in the process of chromatin inactivation. This is in agreement with the published data about the composition of the MLL MPSC (Nakamura *et al*., 2002): the supercomplex is constituted in part by central components of the NuRD and Sin3A complex, indicating that histone deacetylase activity – associated with transcriptional downregulation – is presumably a central function of the MLL MPSC at specific target genes.

Several *hox* genes were transcriptionally influenced by the absence or presence of the *Mll* alleles. The three genes *hoxA7*, *hoxC8*, and *hoxC9* were described before as being transcriptionally maintained by MLL expression (Yu *et al*., 1998; Milne *et al*., 2002), thus verifying

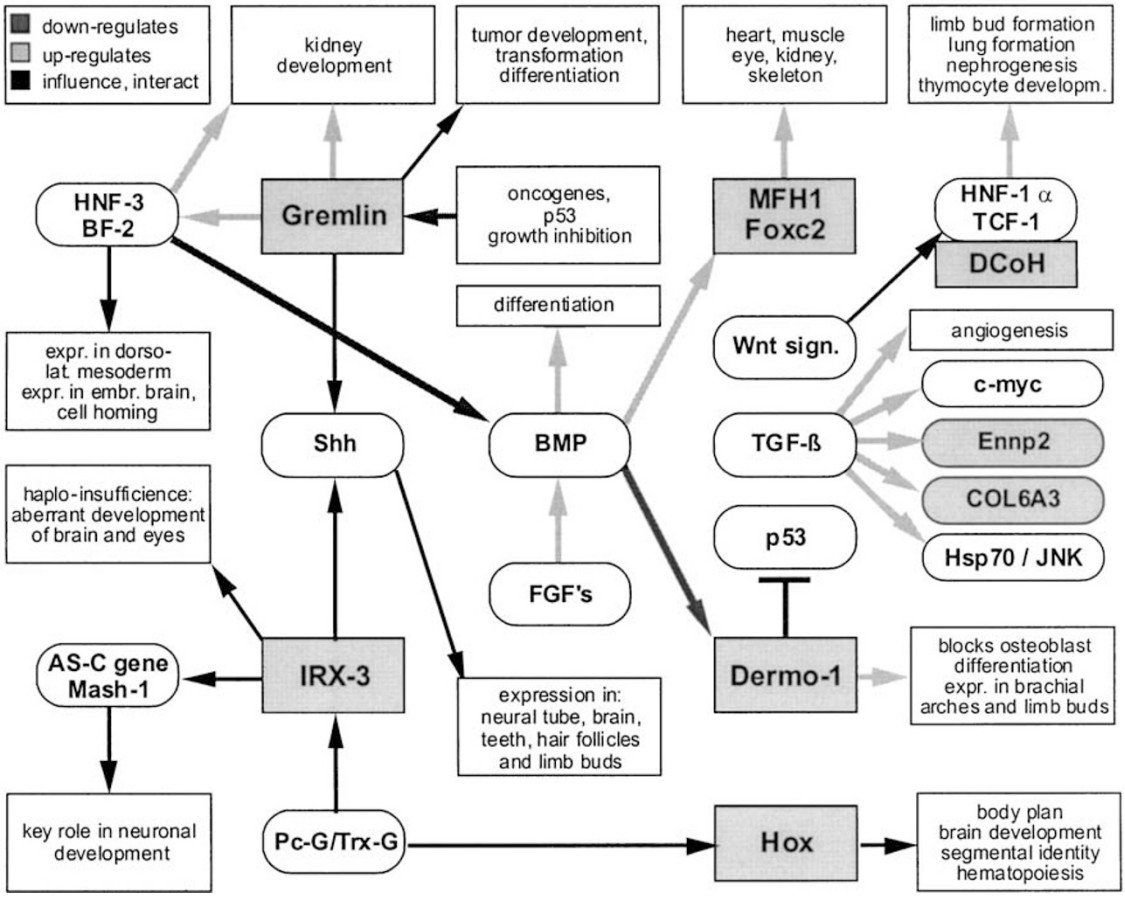

Figure 7 Interaction scheme of potential target gene products of $MLL^{+/+}$ cells. Scheme of known interactions between various proteins that have been identified in the $MLL^{+/+}$ cell line. Light gray rectangles: genes identified. White Ovals: genes affected or controlled by the light gray-marked genes. White rectangles: downstream effects of the light gray-marked or white oval-marked genes. Light gray arrows: 'upregulating'. Dark gray arrows: 'downregulating'. Black arrows: 'influencing' or 'directly interacting with'. *Abbreviations*: BMP, bone morphing proteins; TGF-ß, tumor growth factor-beta; Shh, Sonic hedgehog; Pc-G/Trx-G, Polycomb-group/Trithorax-group; Hox, *hox* genes influenced by Pc-G/Trx-G; AS-C, Antennapedia/bicoid complex; HNF-3/BF2, Hepatocyte nuclear factor-3/brain factor 2; C-myc, c-myc proto-oncoprotein; Hsp70/JNK, heatshock protein 70/Jun N-terminal kinase

the versatility of the biological system used in this study. The same is true for the *paired-like homeodomain transcription factor 2* gene (*RIEG*, *Ptx2*, *ARP1*) that has been identified to be transcriptionally activated by the MLL protein (Arakawa *et al*., 1998).

Genes found to be activated in the presence of MLL protein code for factors involved in differentiation pathways, specially in the development of mesodermal and ectodermal tissues or organogenesis (e.g. *COL6A3*, *DCoH*, *gremlin*, *β ig-h3*, *GDID4*, *Sprr1*, *Mfh-1*, *irx-3*, *dermo-1*, several *hox* genes, *GATA-6*, LIMK). These genes affect muscle, heart, kidney, liver, limbs, bone, and the development of nerve cells, sensory organs and hematopoietic cells (summarized in Figure 7). This is in agreement with data from heterozygous *MLL* knockout experiments, where the expression of a LacZ reporter protein demonstrated MLL expression predominantly in the central nervous system, spinal ganglia, epithelial cells, endothelial cells, heart and liver (Yu *et al*., 1995; Ayton *et al*., 2001). In summary, potential target genes identified in wild-type MLL cells are involved in differentiation processes, tissue repair, or tissue remodeling. In addition, p27kip1 and GAS-1 are known tumor suppressor proteins involved in cell cycle regulation.

The absence of endogenous MLL protein is linked to the expression of a variety of interesting candidate genes that might be important for leukemogenesis. The overexpression of the *KyoT* splice variant, coding for the RAM 14-1 protein, interferes with correct differentiation by inhibiting the Notch-1 pathway. The human homolog of the *KyoT* splice variant, FHL-1C, is overexpressed in the human t(4;11) cell line SEM, but was completely absent in the human T-cell line Jurkat. This argues for the possibility that the reduction of the *MLL* gene dose (haplo-insufficiency) is mimicking the *Mll* knockout situation, and moreover, in the absence of at least one *Mll*/*MLL* allele, splicing of certain mRNA species can be altered.

Other interesting candidate genes identified in the *Mll* knockout cell line, including *HNF-3*/*BF-1*, *Mlf1*, *FBJ osteosarcoma oncogene*, *Tenascin C*, *PE31*/*TALLA-1*, and *Tumor protein D52-like* gene, are all known to be overexpressed in malignant cells. Some other genes code

for proteins that modify the extracellular matrix and are important for tumor cell invasion (e.g. *Serin protease inhibitor 4*, *Stromelysin-2/3*). Thus, it seems that *Mll* knockout cells have upregulated genes that are linked to an oncogenic phenotype.

The fact that genes with oncogenic properties have been identified in *Mll* knockout cells has to be discussed also in the context of the available information of potential functions of MLL and MLL fusion proteins.

First, the common denominator of *MLL* translocations is the derivative 11 protein. As a result of a chromosomal translocation that occurs between *MLL* exons 9–12, only the AT hooks, the subnuclear localization sequences, and the MT domain are present in a given derivative 11 protein. All derivative 11 proteins are missing from the FYRN dimerization domain, and thus, should be unable to enter the MLL MPSC. However, derivative 11 proteins should target the same subnuclear compartment as the MLL MPSCs because of their subnuclear localization sequences (Yano *et al*., 1997). Therefore, derivative 11 proteins may compete with MLL MPSCs for binding sites at control regions or promoters of MLL target genes. As a direct result, potential MLL target genes may be transcriptionally deregulated which mimic the *MLL* knockout situation to an extent. If this is true, the identified target genes in the *Mll* knockout cells become highly interesting candidates for further analysis and may be important for leukemogenesis.

Conversely, reciprocal derivative Z fusion proteins should be processed into an N-terminal peptide (comprising partner protein sequences, the PHD domain, and the dimerization domain FYRN) and an MLL-C peptide that is identical to the MLL-C peptide derived from the wild-type MLL protein. The N-terminal peptide may enter the MLL MPSC (through FYRN) and compete with wild-type MLL-N peptides for binding to MLL-C peptides. If this is true, the assembly of an MLL MPSC will be altered or prevented. Moreover, if a given partner protein offers domains with DNA binding activities, for example AF4 (unpublished data), an altered MPSC will be targeted to partner protein — specific DNA binding sites elsewhere in the genome, leading to ectopic gene activations or inactivations. In conclusion, the presence of a specific chromosomal translocation of the *MLL* gene correlated with human acute leukemias results in the expression of a derivative 11 and/or Z protein. The presence of derivative proteins may not only affect certain *hox* genes in their transcriptional activation or maintenance, but also may lead to broader changes in the genome. Such changes completely alter the gene expression pattern, similar to that observed in *Mll* knockout cells.

The results presented here do not claim that all possible MLL target genes were identified in this biological system. Moreover, we are aware of the fact that most genes identified in this study are presumably secondary and not primary targets of the MLL protein. However, the results of this study have helped to identify potential target genes affected by the presence or absence of MLL protein. A complete list of all 197 identified target genes, their log expression data, and the heatmap can be retrieved from our web site (MLL+.pdf; MLL-.pdf; Table1.pdf, Table2.pdf, HeatMap.pdf, Scatter.pdf; Oligo.pdf): http://www.biozentrum.uni-frankfurt.de/Pharmazeutische_Biologie/Mitarbeiter/Marschalek/Main.html. Many of the identified target genes are currently being investigated in the human system for their transcriptional activity to unravel their functional importance.

Materials and methods

Establishment of cell lines and cell culture

Fibroblasts from isogenic $MLL^{+/+}$ and $MLL^{-/-}$ mice were isolated from dissected murine embryos at day 10.5 p.c. Cells were laced into cell culture and stable lines were obtained after transfection with an expression construct coding for the Polyoma large T antigen. Stable cell lines were cultured in Dulbecco's MEM (25 mM HEPES; Gibco-BRL), supplemented with 10% FCS Gold (PAA, Austria), 100 U/ml penicillin, 100 U/ml streptomycin, 2 mM glutamax and 50 μM ß-mercaptoethanol (all Gibco-BRL). Both cell lines were incubated in a humidified atmosphere (95%) with 5% CO_2 at 37°C.

Chip hybridization: RNA purification, labeling and hybridization

We used 1×10^7 cells to prepare total RNA using the Trizol (Gibco-BRL) purification method. This generally yielded between 80 and 100 μg total RNA. This RNA was further purified and DNase digested (RNase-Free DNase Set, Qiagen) on RNeasy columns (Qiagen) and the quality was examined by gel electrophoresis. If the ribosomal bands were intact, the RNA was determined to be of good quality and used for subsequent synthesis of biotinylated antisense RNA. In addition, RNA was tested for the absence of genomic DNA by specific PCR reactions using sets of GAPDH oligonucleotides.

For first-strand cDNA synthesis, 9 μl (13.5 μg) of total RNA were mixed with 1 μl of a mixture of three polyadenylated control RNAs, 1 μl 100 μM T7-oligo d(T)24 primer (5′-GGCCAGTGAATTGTAATACGACTCACTATAGGGAG GCGG-(dT24)-3′), incubated at 70°C for 10 min and put on ice. Next, 4 μl of 5× first-strand buffer, 2 μl 0.1 M DTT and 1 μl 10 mM dNTPs were added and the reaction was preincubated at 42°C for 2 min. Then, 2 μl (200 U) Superscript II (Life Technologies) were added and incubation continued at 42°C for 1 h.

For second-strand synthesis, 30 μl of 5× second-strand buffer, 91 μl of RNase-free water, 3 μl 10 mM dNTPs, 4 μl (40 U) *Escherichia coli* DNA polymerase I (Life Technologies), 1 μl (12 U) *E. coli* DNA ligase (TaKaRa), 1 μl (2 U) RNase H (TaKaRa) were added and the reaction was incubated at 16°C for 2 h. Then, 2.5 μl (10 U) T4 DNA polymerase I (TaKaRa) were added at 16°C for 5 min. The reaction was stopped by adding 10 μl 0.5 M EDTA, and the double-stranded cDNA was extracted with phenol/chloroform and the aqueous phase was recovered by phase lock gel separation (Eppendorf). After precipitation, the cDNA was restored in 12 μl of RNase-free water.

Double-stranded cDNA (5 μl) was used to synthesize biotinylated cRNA using the BioArray High Yield RNA Transcript Labeling Kit (Enzo Diagnostics, Santa Clara). Labeled cRNA was purified using the RNeasy mini kit (Qiagen). Fragmentation of cRNA, hybridization to

MG_U74Av2 microarrays (Affymetrix), washing and staining as well as scanning of the arrays in a GeneArray scanner (Agilent) were performed as recommended in the Affymetrix Gene Expression Analysis Technical Manual. Absolute and comparison analyses were done using the Affymetrix GeneChip Analysis Suite 5.0 software. A scaling across all probe sets of a given array was included to compensate for variations in the amount and quality of the cRNA samples and other experimental variables.

Data analysis

After scanning, absolute and comparison analyses were carried out using the Affymetrix GeneChip Analysis Suite 5.0 software. A scaling across all probe sets of a given array was included to compensate for variations in the amount and quality of the cRNA samples and other experimental variables. To isolate differentially expressed genes, comparison files were further filtered using the Affymetrix Data Mining Tool 3.0 software. Filter criteria for robustly up- and downregulated genes included Change = 'I' and Signal Log Ratio >1.32 and Change P-value >0.001 or Change = 'D' and Signal Log Ratio <-1.32 and Change P-value >0.999.

RT–PCR analyses

RT–PCR analyses were performed as described (Löchner *et al*., 1996). All oligonucleotides used for this study are summarized in a document (Oligo.pdf) that can be obtained from our web sites (see above).

Acknowledgements
We thank Dr Jay Hess for providing the MLL wild-type and knockout cell 1ines, Dr Peter Oliver for a critical reading of the manuscript, and Dr Ludger Klein-Hitpass for Affymetrix chip hybridzation and computer analysis. This work was funded in part by the DFG-Grant MA 1876/3-1 and the BMBF cancer network Grant 01GS0104/K-S04T07.

References

Alward WL. (2000). *Am. J. Ophthalmol.* **130,** 107–115.
Arakawa H, Nakamura T, Zhadanov AB, Fidanza V, Yano T, Bullrich F, Shimizu M, Blechman J, Mazo A, Canaani E and Croce CM. (1998). *Proc. Natl. Acad. Sci. USA*, **95,** 4573–4578.
Armstrong SA, Stounton JE, Silverman LB, Pieters R, den Boer ML, Minden MD, Sallan SE, Lander ES, Golub TR and Korsmeyer SJ. (2002). *Nat. Genet.*, **30,** 41–47.
Attardi LD, Reczek EE, Cosmas C, Demicco EG, McCurrach ME, Lowe SW and Jacks T. (2000). *Genes Dev.*, **14,** 704–718.
Ayton P, Sneddon SF, Palmer DB, Rosewell IR, Owen MJ, Young B, Presley R and Subramanian V. (2001). *Genesis*, **30,** 201–212.
Birke M, Schreiner S, Garcia-Cuellar MP, Mahr K, Titgemeyer F and Slany RK. (2002). *Nucleic Acids Res.*, **30,** 958–965.
Bluteau O, Jeannot E, Bioulac-Sage P, Marques JM, Blanc JF, Bui H, Beaudoin JC, Franco D, Balabaud C, Laurent-Puig P and Zucman-Rossi J. (2002). *Nat. Genet.*, **32,** 312–315.
Bollen M, Gijsbers R, Ceulemans H, Stalmans W and Stefan C. (2000). *Crit. Rev. Biochem. Mol. Biol.*, **35,** 393–432.
Bonilla IE, Tanabe K and Strittmatter SM. (2002). *J. Neurosci*, **22,** 1303–1315.
Burg MA, Tillet E, Timpl R and Stallcup WB. (1996). *J. Biol. Chem.*, **271,** 26110–26116.
Capdevila J, Tsukui T, Rodriquez Esteban C, Zappavigna V and Izpisua Belmonte JC. (1999). *Mol. Cell*, **4,** 839–849.
Caslini C, Shilatifard A, Yang L and Hess JL. (2000). *Proc. Natl. Acad. Sci. USA*, **91,** 2797–2802.
Cavodeassi F, Modolell J and Gomez-Skarmeta JL. (2001). *Development*, **128,** 2847–2855.
Chu PH, Ruiz-Lozano P, Zhou Q, Cai C and Chen J. (2000). *Mech. Dev.*, **95,** 259–265.
Cimino G, Rapanotti MC, Sprovieri T and Elia L. (1998). *Haematologica*, **83,** 350–357.
Cogan JG, Subramanian SV, Polikandriotis JA, Kelm Jr RJ and Strauch AR. (2002). *J. Biol. Chem.*, **277,** 36433–36442.
de la Pompa JL, Wakeham A, Correia KM, Samper E, Brown S, Aguilera RJ, Nakano T, Honjo T, Mak TW, Rossant J and Conlon RA. (1997). *Development*, **124,** 1139–1148.
DiMartino JF and Cleary ML. (1999). *Br. J. Haematol.*, **106,** 614–626.
Dou C, Lee J, Liu B, Liu F, Massague J, Xuan S and Lai E. (2000). *Mol. Cell Biol.*, **20,** 6201–6211.
Felix CA. (1998). *Biochim. Biophys. Acta*, **1400,** 233–255.
Francis NJ and Kingston RE. (2001). *Nat. Rev. Mol. Cell. Biol.*, **2,** 409–421.
Gong XQ and Li L. (2002). *J. Biol. Chem.*, **277,** 12310–12317.
Gopalakrishnan R, Ouyang H, Somerman MJ, McCauley LK and Franceschi RT. (2001). *Endocrinology*, **142,** 4379–4388.
Greil J, Gramatzki M, Burger R, Marschalek R, Peltner M, Trautmann U, Hansen-Hagge TE, Bartram CR, Fey GH, Stehr K and Beck J. (1994). *Br. J. Haematol.*, **86,** 275–283.
Hanashima C, Shen L, Li SC and Lai E. (2002). *J. Neurosci.*, **22,** 6526–6536.
Henkel GW, McKercher SR and Maki RA. (2002). *Int. Immunol.*, **14,** 723–732.
Hosokawa Y, Ueyama E, Morikawa Y, Maeda Y, Seto M and Senba E. (1999). *Neurosci. Res.*, **35,** 281–290.
Hsieh JJ, Ernst P, Erdjument-Bromage H, Tempst P and Korsmeyer SJ. (2003). *Mol. Cell. Biol.*, **23,** 186–194.
Iida K, Koseki H, Kakinuma H, Kato N, Mizutani-Koseki Y, Ohuchi H, Yoshioka H, Noji S, Kawamura K, Kataoka Y, Ueno F, Taniguchi M, Yoshida N, Sugiyama T and Miura N. (1997). *Development*, **124,** 4627–4638.
Irikura VM, Lagraoui M and Hirsh D. (2002). *J. Immunol.*, **169,** 393–398.
Izpisua-Belmonte JC, Dolle P, Renucci A, Zappavigna V, Falkenstein H and Duboule D. (1990). *Development*, **110,** 733–745.
Koutsourakis M, Langeveld A, Patient R, Beddington R and Grosveld F. (1999). *Development*, **126,** 723–732.
Lee KK, Leung AK, Tang MK, Cai DQ, Schneider C, Brancolini C and Chow PH. (2001). *Dev. Biol.*, **234,** 188–203.
Lelias JM, Adra CN, Wulf GM, Guillemot JC, Khagad M, Caput D and Lim B. (1993). *Proc. Natl. Acad. Sci. USA*, **90,** 1479–1483.
Le Mouellic H, Lallemand Y and Brulet P. (1992). *Cell*, **69,** 251–264.
Löchner K, Siegler G, Führer M, Greil J, Beck JD, Fey GH and Marschalek R. (1996). *Cancer Res.*, **56,** 2171–2177.
Marschalek R, Greil J, Löchner K, Nilson I, Siegler G, Zweckbronner I, Beck JD and Fey GH. (1995). *Br. J. Haematol.*, **90,** 308–320.

Meins M, Piosik P, Schaeren-Wiemers N, Franzoni S, Troncoso E, Kiss JZ, Brosamle C, Schwab ME, Molnar Z and Monard D. (2001). *J. Neurosci.*, **21,** 8830–8841.

Mendel DB, Khavari PA, Conley PB, Graves MK, Hansen LP, Admon A and Crabtree GR. (1991). *Science*, **254,** 1762–1767.

Meng Y, Zhang Y, Tregoubov V, Janus C, Cruz L, Jackson M, Lu WY, MacDonald JF, Wang JY, Falls DL and Jia Z. (2002). *Neuron*, **35,** 121–133.

Milne TA, Briggs SD, Brock HW, Martin ME, Gibbs D, Allis CD and Hess JL. (2002). *Mol. Cell*, **10,** 1107–1117.

Monaghan AP, Davidson DR, Sime C, Graham E, Baldock R, Bhattacharya SS and Hill RE. (1991). *Development*, **112,** 1053–1061.

Morrisey EE, Ip HS, Lu MM and Parmacek MS. (1996). *Dev. Biol.*, **177,** 309–322.

Morrisey EE, Tang Z, Sigrist K, Lu MM, Jiang F, Ip HS and Parmacek MS. (1998).*Genes Dev.*, **12,** 3579–3590.

Nakamura T, Mori T, Tada S, Krajewski W, Rozovskaia T, Wassell R, Dubois G, Mazo A, Croce CM and Canaani E. (2002). *Mol. Cell.*, **10,** 1119–1128.

Pasqualetti M, Neun R, Davenne M and Rijli FM. (2001). *Nat Genet.*, **29,** 34–39.

Perris R, Kuo HJ, Glanville RW and Bronner-Fraser M. (1993). *Dev. Dyn.*, **198,** 135–149.

Seeler JS and Dejean A. (2001). *Oncogene*, **20,** 7243–7249.

Seykora JT, Mei L, Dotto GP and Stein PL. (2002). *J. Biol. Chem.*, **277,** 2812–2822.

Skonier J, Neubauer M, Madisen L, Bennett K, Plowman GD and Purchio AF. (1992). *Cell. Biol.*, **11,** 511–522.

So CW and Cleary ML. (2002). *Mol. Cell. Biol.*, **22,** 6542–6552.

Spandorfer SD. (2000). *Am. J. Reprod. Immunol.*, **43,** 6–11.

Suzuki T, K-Tsuzuku J, Ajima R, Nakamura T, Yoshida Y and Yamamoto T. (2002). *Genes Dev.*, **16,** 1356–1370.

Takagi S, Fujikawa K, Imai T, Fukuhara N, Fukudome K, Minegishi M, Tsuchiya S, Konno T, Hinuma Y and Yoshie O. (1995). *Int. J. Cancer*, **61,** 706–715.

Taniguchi Y, Furukawa T, Tun T, Han H and Honjo T. (1998). *Mol. Cell. Biol.*, **18,** 644–654.

Toft DJ, Rosenberg SB, Bergers G, Volpert O and Linzer DI. (2001). *Proc. Natl. Acad. Sci. USA*, **98,** 13055–13059.

Wei LC, Shi M, Chen LW, Cao R, Zhang P and Chan YS. (2002). *Brain Res. Dev. Brain Res.*, **139,** 9–17.

Wiemels JL, Cazzaniga G, Daniotti M, Eden OB, Addison GM, Masera G, Saha V, Biondi A and Greaves MF. (1999). *Lancet*, **354,** 1499–1503.

Yagami K, Suh JY, Enomoto-Iwamoto M, Koyama E, Abrams WR, Shapiro IM, Pacifici M and Iwamoto M. (1999). *J. Cell. Biol*, **147,** 1097–1108.

Yang N, Higuchi O, Ohashi K, Nagata K, Wada A, Kangawa K, Nishida E and Mizuno K. (1998). *Nature*, **393,** 809–812.

Yano T, Nakamura T, Blechman J, Sorio C, Dang CV, Geiger B and Canaani E. (1997). *Proc. Natl. Acad. Sci. USA*, **94,** 7286–7291.

Yokoyama A, Kitabayashi I, Ayton PM, Cleary ML and Ohki M. (2002). *Blood*, **100,** 3710–3718.

Yoneda-Kato N, Look AT, Kirstein MN, Valentine MB, Raimondi SC, Cohen KJ, Carroll AJ and Morris SW. (1996). *Oncogene*, **12,** 265–275.

Yu BD, Hanson RD, Hess JL, Horning SE and Korsmeyer SJ. (1998). *Proc. Natl. Acad. Sci. USA*, **95,** 10632–10636.

Yu BD, Hess JL, Homing SE, Brown GA and Korsmeyer SJ. (1995). *Nature*, **378,** 505–508.

Zeleznik-Le NJ, Harden AM and Rowley JD. (1994). *Proc. Natl. Acad. Sci. USA*, **91,** 10610–10614.

Ziemin van der Poel S, McCabe NR, Gill HJ, Espinosa III R, Patel Y, Harden A, Rubinelli P, Smith SD, LeBeau MM, Rowley JD and Diaz MO. (1991). *Proc. Natl. Acad. Sci. USA*, **88,** 10735–10739.