Oncogene (2002) 21, 148–152
© 2002 Nature Publishing Group All rights reserved 0950–9232/02 $25.00
www.nature.com/onc

SHORT REPORTS

Molecular cloning of ESET, a novel histone H3-specific methyltransferase that interacts with ERG transcription factor

Liu Yang[1,2], Li Xia[4], Daniel Y Wu[1,3], Hengbin Wang[4], Howard A Chansky[1,2], William H Schubach[1,3], Dennis D Hickstein[5] and Yi Zhang*[,4]

[1]Medical Research Service, VA Puget Sound Health Care System, University of Washington, Seattle, Washington, WA 98108, USA; [2]Department of Orthopedics, University of Washington, Seattle, Washington, WA 98108, USA; [3]Department of Medicine/Oncology, University of Washington, Seattle, Washington, WA 98108, USA; [4]Department of Biochemistry and Biophysics, Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina, NC 27599, USA; [5]Department of Experimental Transplantation and Immunology, National Cancer Institute, Bethesda, Maryland, MD 20892, USA

The *ets*-related gene *erg* encodes a transcription factor that is implicated in the control of cell growth and differentiation. To identify interacting partners of ERG, we screened a yeast two-hybrid cDNA library constructed from mouse hematopoietic cells using the N-terminal region of ERG as a bait. We isolated a 4.6 kb full-length mouse cDNA encoding a 1307-amino acid protein migrating as a 180 kD band, which was termed ESET (ERG-associated protein with SET domain). ESET is 92% identical to the human protein SETDB1 (SET domain, bifurcated 1). The interaction between ESET and ERG was supported by *in vitro* pull-down using glutathione S-transferase (GST) fusion protein, by transfection and co-immunoprecipitation experiments, and by association of endogenous SETDB1 with ERG. Since ESET possesses evolutionarily conserved SET, preSET, and postSET domains implicated in histone methylation, we tested the ability of ESET to methylate core histones. The results of these studies demonstrated that ESET is a histone H3-specific methyltransferase, and that mutations within ESET abolished its methyltransferase activity. Together, these findings raise the possibility that transcription factor ERG may participate in transcriptional regulation through ESET-mediated histone methylation.
Oncogene (2002) **21,** 148–152. DOI: 10.1038/sj/onc/1204998

Keywords: histone methyltransferase; ESET; ERG

The evolutionarily conserved ETS-family of transcription factors have been a subject of intensive study because of their ability to regulate important physiological processes including osteogenesis and regulation of extracellular matrix (Raouf and Seth, 2000; Trojanowska, 2000). The criterion for ETS family membership is the presence of an approximately 85-amino acid ETS domain named after the founding member ETS-1, originally identified from the avian leukemia retrovirus **E t**wenty **s**ix (E26) (Nunn *et al*., 1983; Leprince *et al*., 1983). ETS proteins usually bind to DNA as monomers via the ETS domain, however it has been shown that the DNA binding activity is enhanced or modulated through interactions with other transcription factors (Li *et al*., 2000).

The *ets*-related gene *erg* is located on chromosome 21 and was originally cloned through cross-hybridization to the *ets*-2 probe (Rao *et al*., 1987). In human myeloid leukemia, the *erg* gene is fused to the *tls* (translocation liposarcoma) gene located on chromosome 16 (Ichikawa *et al*., 1994). This reciprocal t(16;21) chromosomal translocation generates a fusion protein TLS/ERG. The TLS/ERG leukemia fusion protein retains the N-terminal domain of TLS, however the C-terminal domain of TLS is replaced by the DNA-binding domain of ERG. In subsets of **Ew**ing's **s**arcomas with a t(21;22) chromosomal translocation, the N-terminal domain of EWS protein is replaced by the same DNA-binding domain of ERG, resulting in the generation of chimeric EWS/ERG sarcoma fusion protein (Dunn *et al*., 1994; Giovannini *et al*., 1994; Sorensen *et al*., 1994; Zucman *et al*., 1993).

The TLS/ERG leukemia fusion protein and the EWS/ERG sarcoma fusion protein have been reported to interfere with important cellular processes such as transcription and RNA splicing (Im *et al*., 2000; Prasad *et al*., 1994; Yang *et al*., 2000). Although both fusion proteins are capable of cellular transformation, the underlying transformation mechanism is not understood. Understanding the mechanism of cellular transformation by these fusion proteins will likely require knowledge regarding the functions of the intact wild-type proteins. In this regard, it is notable that the N-terminal domain of ERG is replaced in both the TLS/ERG and EWS/ERG fusion proteins, suggesting that normal protein functions mediated through the N-terminal domain of ERG might be lost via oncogenic fusion.

To identify ERG-interacting proteins, we used the N-terminal domain of ERG (amino acids 1–114) as a bait to screen a yeast two-hybrid library derived from mouse multipotential hematopoietic cells. In the initial

*Correspondence: Y Zhang; E-mail: yi_zhang@med.unc.edu
Received 16 May 2001; revised 6 September 2001; accepted 18 September 2001

yeast two-hybrid screen of 1×10^7 transformants, we identified two independent clones that were positive for interaction with the ERG bait but negative for interactions with control proteins. The first clone encodes UBC9, a ubiquitin-conjugating enzyme that had previously been reported to interact with ERG protein (Hahn *et al*., 1997). The second clone represents a novel protein with the highest homology to human clone KIAA0067 (GenBank accession no. D31891). Full-length mouse cDNA for this ERG-interacting clone was isolated from an EML cDNA phage library. Nucleotide sequence revealed that the cDNA is 4.6 kilobases in length (GenBank accession no. AF091628), with the sequence surrounding the first ATG codon matching well to the Kozak translation initiation consensus (Kozak, 1991). The deduced amino acid sequence of the open reading frame corresponded to a protein of 1307 amino acids with a calculated molecular weight of 145 kD (Figure 1a). Since this protein contained an evolutionarily conserved SET (**s**uppressor of variegation, **e**nhancer of zeste and **t**rithorax) domain (Tschiersch *et al*., 1994) between amino acids 814 to 1307, it was therefore termed ERG-associated protein with SET domain (ESET).

A database search revealed that ESET is 92% identical to the 1291-amino acid protein deduced from the KIAA0067 clone isolated from KG-1cells, a human acute myelogenous leukemia cell line (Nomura *et al*., 1994). The human protein was named SETDB1 (SET domain, bifurcated 1) based on the fact that the SET domain was separated by a 347-amino acid insertion (Harte *et al*., 1999). In addition to the divided SET domain, both mouse ESET and human SETDB1 proteins contain the tudor domain (Ponting, 1997) and the methyl-CpG binding domain (Bird and Wolffe, 1999) (Figure 1b). A tudor domain in the **s**urvival of **m**otor **n**euron (SMN) protein is known to mediate protein–protein interactions that are crucial to the assembly of spliceosomal complexes (Buhler *et al*., 1999), whereas the methyl-CpG binding domain is found in proteins involved in methylated DNA silencing (Bird and Wolffe, 1999).

To verify that the deduced reading frame of ESET corresponds to the expressed ESET protein, pBS phagemid containing the full-length ESET cDNA was

A.

```
   1 MSSLPGCMSL AAAPAAADSA EIAELQQAVV EELGISMEEL RQYIDEELEK MDCIQQRKKQ
  61 LAELETWVLQ KESEVAYVDR LFDDASREVT NCESLVKDFY SKLGLQYHDS SSEDEASRPT
 121 EIIEIPDEDD DVLSIDSGDA GSRTPKDQKL REAMAALRKS AQDVQKFMDA VNKKSSSQDL
 181 HKGTLGQVSG ELSKDGDLIV SMRILGKKRT KTWHKGTLIA IQTVGLGKKY KVKFDNKGKS
 241 LLSGNHIAYD YHPPADKLFV GSRVVAKYKD GNQVWLYAGI VAETPNVKNK LRFLIFFDDG
 301 YASYVTQSEL YPICRPLKKT WEDIEDSSCR DFIEEYITAY PNRPMVLLKS GQLIKTEWEG
 361 TWWKSRVEEV DGSLVRILFL DDKRCEWIYR GSTRLEPMFS MKTSSASAME KKQGGQLRTR
 421 PNMGAVRSKG PVVQYTQDLT GTGIQFKPME PLQPIAPPAP LPIPPLSPQA ADTDLESQLA
 481 QSRKQVAKKS TSFRPGSVGS GHSSPTSSTL SENVSAGKLG INQTYRSPLA SVTSTPASAA
 541 PPVPPVPPGP PTPPGPPAPP GPLAPPAFHG MLERAPAEPS YRAPMEKLFY LPHVCSYTCL
 601 SRIRPMRNEQ YRGKNPLLVP LLYDFRRMTA RRRVNRKMGF HVIYKTPCGL CLRTMQEIER
 661 YLFETGCDFL FLEMFCLDPY VLVDRKFQPF KPFYYILDIT YGKEDVPLSC VNEIDTTPPP
 721 QVAYSKERIP GKGVFINTGP EFLVGCDCKD GCRDKSKCAC HQLTIQATAC TPGGQVNPNS
 781 GYQYKRLEEC LPTGVYECNK RCNCDPNMCT NRLVQHGLQV RLQLFKTQNK GWGIRCLDDI
 841 AKGSFVCIYA GKILTDDFAD KEGLEMGDEY FANLDHIESV ENFKEGYESD VPTSSDSSGV
 901 DMKDQEDGNS GSEDPEESND DSSDDNFCKD EDFSTSSVWR SYATRRQTRG QKENELSEMT
 961 SKDSRPPDLG PPHVPIPSSV SVGGCNPPSS EETPKNKVAS WLSCNSVSEG GFADSDSRSS
1021 FKTSEGGDGR AGGGRGEAER ASTSGLSFKD EGDNKQPKKE DPENRNKMPV VTEGSQNHGH
1081 NPPMKSEGLR RPASKMSVLQ SQRVVTSTQS NPDDILTLSS STESEGESGT SRKPTAGHTS
1141 ATAVDSDDIQ TISSGSDGDD FEDKKNLSGP TKRQVAVKST RGFALKSTHG IAIKSTNMAS
1201 VDKGESAPVR KNTRQFYDGE ESCYIIDAKL EGNLGRYLNH SCSPNLFVQN VFVDTHDLRF
1261 PWVAFFASKR IRAGTELTWD YNYEVGSVEG KELLCCCGAI ECRGRLL
```

B.

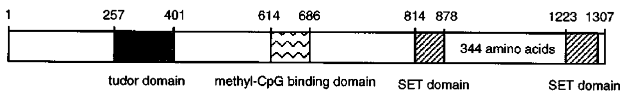
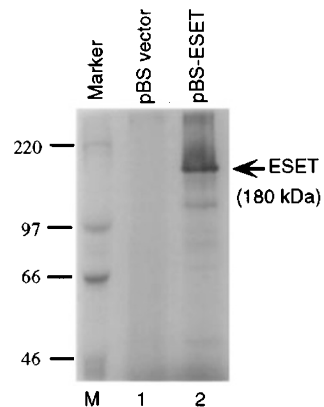


Figure 1 Amino acid sequence of ESET and its functional domains. (**a**) The deduced amino acid sequence of ESET. The protein consists of 1307 amino acids with a calculated molecular weight of 145 kDa. The 4.6 kb cDNA sequence for ESET has been deposited in the GenBank (accession no. AF091628). (**b**) Schematic representation of ESET functional domains. The tudor domain is shown in gray box, the methyl-CpG binding domain in box with wavy lines, and the SET domain in hatched box



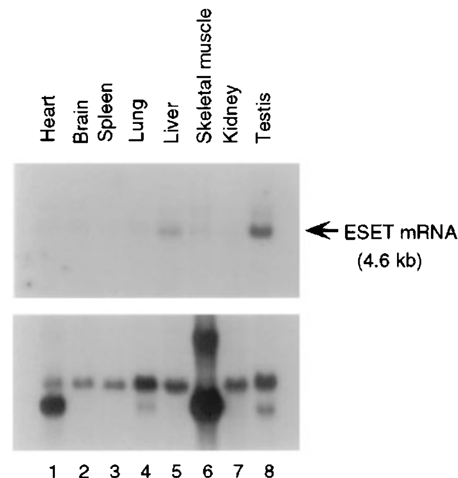


Figure 2 *In vitro* translation and expression of ESET. (**a**) The TNT Coupled Reticulocyte Lysate Systems (Promega) were used in the *in vitro* translation of ESET cDNA. The protein products, generated from empty pBS phagemid (lane 1) and pBS-ESET (lane 2), were separated on a 6% SDS–PAGE. Positions of protein markers are labeled at left and ESET protein is indicated by an arrow. (**b**) A mouse multi-tissue Northern blot was purchased from Clontech and hybridized to a ^{32}P-labeled ESET cDNA probe (top panel) or to a β-actin DNA probe (bottom panel). The 4.6 kb ESET mRNA transcript is indicated by an arrow

used as a template for *in vitro* transcription and translation. While the empty pBS phagemid failed to generate any labeled protein product (Figure 2a, lane 1), *in vitro* transcription and translation of pBS-ESET phagemid generated a prominent protein band of approximately 180 kD on SDS–PAGE (Figure 2a, lane 2). The discrepancy between this apparent molecular weight and the calculated molecular weight is likely due to post-translational modification such as protein phosphorylation.

To examine the expression pattern of the *ESET* gene in various tissues, a mouse multi-tissue Northern blot was hybridized to a ^{32}P-labeled ESET cDNA probe. The 4.6 kb ESET mRNA transcript is most prominent in mouse liver and testis (Figure 2b, lanes 5 and 8, top panel). Interestingly, the human *SETDB1* gene also appears to be highly expressed in testis although expression is detectable in all the tissues examined (http://www.kazusa.or.jp/huge/gfimage/northern/html/KIAA0067.html).

To verify the yeast two-hybrid interaction between ESET and ERG, GST 'pull-down' assays were performed. ^{35}S-labeled ERG was transcribed and translated *in vitro* from the pSG5-FL-ERG (Figure 3a, lane 1) and used to bind the GST-ESET(1–167). After three washes with binding buffer, the bound ^{35}S-labeled ERG was resolved by SDS–PAGE. Results shown in Figure 3a (lane 4) indicate that ESET interacts with ERG. However, under the same conditions, the ^{35}S-labeled ERG failed to bind to the negative controls (Figure 3a, lanes 2 and 3).

The finding that ERG interacts with ESET in the yeast two-hybrid screen and in the GST 'pull-down' assay suggested that these two proteins might also associate with each other in mammalian cells. To demonstrate their association *in vivo*, plasmids expressing Flag-tagged ERG and Myc-tagged ESET were co-transfected into 293T cells, and cell lysates from the co-transfected cells (Figure 3b, lane 1) were used for immunoprecipitation. A mouse monoclonal anti-Myc antibody co-immunoprecipitated the Myc-ESET along with appreciable amount of Flag-ERG (Figure 3b, lane 2), whereas a control mouse IgG failed to bring down either epitope-tagged protein (Figure 3b, lane 3).

To provide further evidence supporting the *in vivo* association between ESET and ERG, a rabbit polyclonal antibody against the N-terminal 167 amino acids of ESET was generated. This antibody recognizes both mouse ESET and human SETDB1 due to a high degree of sequence homology between these two proteins (Figure 3C, lanes 1 and 2). In a previous study, the human K562 leukemia cells were transduced with a retrovirus that expresses Flag-tagged ERG protein at the physiological level (Yang *et al*., 2000). Western blotting analysis of lysate from these K562 cells detected both retroviral Flag-ERG and endogenous SETDB1 (Figure 3d, lane 1). When endogenous SETDB1 protein was immunoprecipitated from these K562 cells with affinity-purified anti-ESET antibody, retroviral Flag-ERG was detectable from the anti-ESET immunoprecipitate (Figure 3d, lane 2). The SETDB1 and Flag-ERG association is specific as Flag-ERG was absent in the control immunoprecipitate obtained using an unrelated rabbit polyclonal antibody (Figure 3d, lane 3).

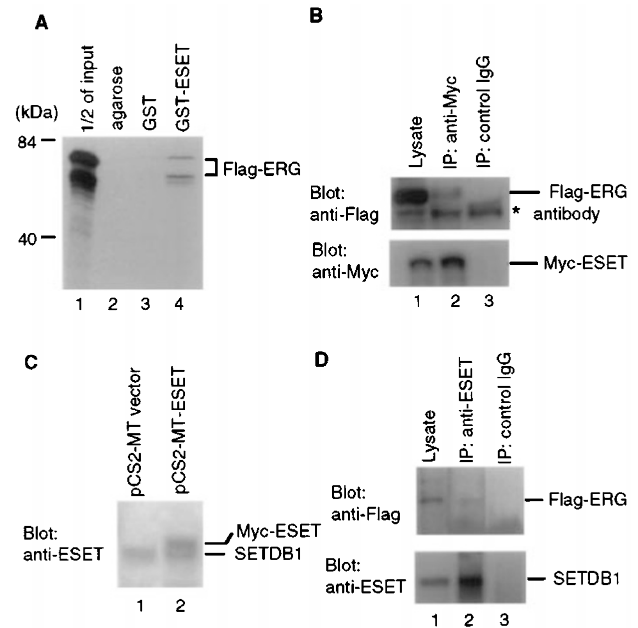
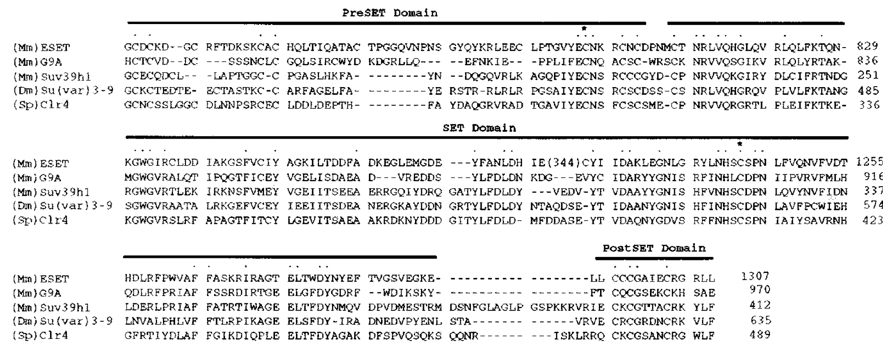


Figure 3 *In vitro* and *in vivo* Association of ERG and ESET. (**a**) Full-length ERG-2 was cloned in-frame into the *Eco*RI-*Sma*I sites of pSG5-FL vector for *in vitro* translation of Flag-tagged ERG protein. The Glutathione S-transferase (GST) and GST-ESET fusion protein were expressed from plasmid pGEX4TK and pGEX-ESET (with a DNA insert corresponding to amino acids 1–167 of ESET) in *E. coli* as previously described (Wu *et al*., 1996). ^{35}S-labeled Flag-ERG (lane 1) was used to bind glutathione-agarose (lane 2), GST-agarose (lane 3) and GST-ESET-agarose (lane 4). The bound ^{35}S-labeled ERG was eluted, subjected to SDS–PAGE, and visualized by autoradiography. (**b**) Plasmids expressing Flag-tagged ERG and Myc-tagged ESET were co-transfected into 293T cells. The cell lysate (lane 1), the immunoprecipitates using the 9E10 anti-Myc antibody (lane 2), or a control IgG (lane 3) were separated by SDS–PAGE then blotted with the M2 anti-Flag (top panel) or the 9E10 anti-Myc (bottom panel) antibodies. (**c**) 293T cells transfected with empty pCS2-MT vector (lane 1) or pCS2-MT-ESET (lane 2) was blotted with a rabbit polyclonal anti-ESET antibody. Positions of Myc-ESET and endogenous SETDB1 are indicated. (**d**) K562 lysate expressing retroviral Flag-ERG (lane 1) was immunoprecipitated with affinity-purified anti-ESET (lane 2) or a control IgG (lane 3). The samples were blotted with M2 anti-Flag (top panel) or rabbit anti-ESET (bottom panel)

The association of ESET (or SETDB1) with ERG may affect the functions of both proteins. One of the potential activities of ESET is suggested by the presence of the SET domain. In this regard, it was demonstrated recently that a SET domain-containing protein, the mammalian homologue of the *Drosophila* SU(VAR)3-9 protein, is a site-specific histone H3 methyltransferase (HMTase) (Rea *et al*., 2000). The G9A protein, another SET domain-containing molecule, has also been reported to function as a histone H3 methyltransferase (Tachibana *et al*., 2001). Interestingly, all the invariant amino acid residues within the SET domains and their adjacent catalytically

A



B

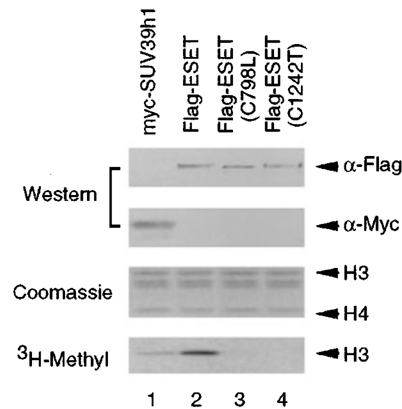


Figure 4 H3-specific histone methyltransferase assay with ESET protein. (**a**) Alignment of the mouse ESET (AF091628) with other selected SET domain containing proteins. Included in this alignment is the mouse G9A (AF109906), the HMTase Suv39h1 from mouse (AF019969), *Drosophila* (X80070), and *S. pombe* (AF061854). The preSET, SET, and postSET domains are indicated. Residues that are invariant in this alignment are indicated by a dot above their positions. The two cysteines mutated in the mutants are indicated by '*'. Numbers correspond to the amino acid of respective proteins. Note that the SET domain of ESET is divided by a 344 amino acid insertion, indicated by (), at amino acid 878. (**b**) The SUV39h1 control (lane 1), wild-type ESET (lane 2) and mutant ESET (lanes 3 and 4) were immunoprecipitated and analysed by Western (top two panels). For HMTase activity assays, immunoprecipitates were incubated with HeLa core histone octamers and S-adenosyl-L-[*methyl*-^{3}H]methionine (15 Ci/mmmol; NEN Life Science Products) at 30°C for 1 h. The histone substrates were separated in an 18% SDS–PAGE. Western indicates equal amounts of ESET protein was used in the HMTase assays. Coomassie indicates equal amounts of core histones in each reaction. Fluorogram indicates only H3 is methylated and the two mutated cysteines are critical for the HMTase activity (bottom two panels)

important cysteine rich regions (preSET and postSET domains) of Suv39h1 and G9A are conserved in the ESET protein (Figure 4a). This high degree of conservation raised the possibility that ESET is also a HMTase. To test this possibility, a plasmid encoding Flag-tagged ESET was transiently expressed in 293T cells and immunoprecipitated with antibodies against Flag. The immunoprecipitates were divided into two parts for Western blotting and HMTase activity assay. The previously described *myc*-tagged SUV39h1 (Rea *et al*., 2000) was used as a positive control in the methylation assay. The HMTase substrates in the assay were HeLa histone octamers consisting of histone H2A, H2B, H3 and H4 subunits. Results shown in Figure 4b indicate that ESET, like SUV39h1, can specifically methylate histone H3 while inactive toward histone H2A, H2B and H4 (Figure 4b, lanes 1 and 2). To demonstrate that ESET, not its associated proteins are responsible for the enzymatic activity, two highly conserved cysteines at 798 and 1242 in preSET domain and SET domains of ESET proteins were mutated and analysed for HMTase activity. Both mutants (C798L and C1242T) were expressed at levels similar to that of wild-type ESET, however neither mutant retained the HMTase activity (Figure 4b, lanes 3 and 4).

These studies indicate that mouse ESET (or its human homologue SETDB1) is an ERG-associated protein, and that ESET protein functions as a histone H3-specific methyltransferase. Our findings raise the intriguing possibility that the ETS-related transcription factor ERG may participate in transcriptional regulation, at least in part, through the recruitment of ESET histone methyltransferase and subsequent modification of local chromatin structure. Methylation of histones on both lysine and arginine residues has been well documented (Strahl and Allis, 2000), and this mod-

ification can create binding sites for other proteins (Lachner *et al*., 2001; Bannister *et al*., 2001), leading to either gene activation or gene silencing (Chen *et al*., 1999). In addition, protein methylation can also affect cellular functions such as RNA processing and receptor-mediated signaling (Aletta *et al*., 1998). While the *in vivo* targets of the ESET methyltransferase are likely to include histone H3, the functional consequences of ESET (or SETDB1) association with ERG are currently under investigation.

Acknowledgments
We thank Dr Thomas Jenuwein for the *myc*-SUV39h1 plasmid. This work is supported in part by funds from the Department of Orthopedics and Sports Medicine at the University of Washington and by a New Investigator Grant from the Leukemia Research Foundation (to L Yang). Y Zhang is a V-foundation scholar and work in his lab is supported by NIH (GM63067-01) and ACS (RSG-00-351-01-GMC).

References

Aletta JM, Cimato TR and Ettinger MJ. (1998). *Trends Biochem. Sci.,* **23,** 89–91.
Bird AP and Wolffe AP. (1999). *Cell,* **99,** 451–454.
Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC and Kouzarides T. (2001). *Nature,* **410,** 120–124.
Buhler D, Raker V, Luhrmann R and Fischer U. (1999). *Hum. Mol. Genet.,* **8,** 2351–2357.
Chen D, Ma H, Hong H, Koh SS, Huang SM, Schurter BT, Aswad DW and Stallcup MR. (1999). *Science,* **284,** 2174–2177.
Dunn T, Praissman L, Hagag N and Viola MV. (1994). *Cancer Genet Cytogenet.,* **76,** 19–22.
Giovannini M, Biegel JA, Serra M, Wang JY, Wei YH, Nycum L, Emanuel BS and Evans GA. (1994). *J. Clin. Invest.,* **94,** 489–496.
Hahn SL, Wasylyk B, Criqui-Filipe P and Criqui P. (1997). *Oncogene,* **15,** 1489–1495.
Harte PJ, Wu W, Carrasquillo MM and Matera AG. (1999). *Cytogenet. Cell. Genet.,* **84,** 83–86.
Ichikawa H, Shimizu K, Hayashi Y and Ohki M. (1994). *Cancer Res.,* **54,** 2865–2868.
Im YH, Kim HT, Lee C, Poulin D, Welford S, Sorensen PH, Denny CT and Kim SJ. (2000). *Cancer Res.,* **60,** 1536–1540.
Kozak M. (1991). *J. Biol. Chem.,* **266,** 19867–19870.
Lachner M, O'Carroll D, Rea S, Mechtler K and Jenuwein T. (2001). *Nature,* **410,** 116–120.
Leprince D, Gegonne A, Coll J, de Taisne C, Schneeberger A, Lagrou C and Stehelin D. (1983). *Nature,* **306,** 395–397.
Li R, Pei H and Watson DK. (2000). *Oncogene,* **19,** 6514–6523.
Nomura N, Nagase T, Miyajima N, Sazuka T, Tanaka A, Sato S, Seki N, Kawarabayasi Y, Ishikawa K and Tabata S. (1994). *DNA Res.,* **1,** 223–229.
Nunn MF, Seeburg PH, Moscovici C and Duesberg PH. (1983). *Nature,* **306,** 391–395.
Ponting CP. (1997). *Trends Biochem. Sci.,* **22,** 51–52.
Prasad DD, Ouchida M, Lee L, Rao VN and Reddy ES. (1994). *Oncogene,* **9,** 3717–3729.
Rao VN, Papas TS and Reddy ES. (1987). *Science,* **237,** 635–639.
Raouf A and Seth A. (2000). *Oncogene,* **19,** 6455–6463.
Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, Opravil S, Mechtler K, Ponting CP, Allis CD and Jenuwein T. (2000). *Nature,* **406,** 593–599.
Sorensen PH, Lessnick SL, Lopez-Terrada D, Liu XF, Triche TJ and Denny CT. (1994). *Nat. Genet.,* **6,** 146–151.
Strahl BD and Allis CD. (2000). *Nature,* **403,** 41–45.
Tachibana M, Sugimoto K, Fukushima T and Shinkai Y. (2001). *J. Biol. Chem.,* **20,** 20.
Trojanowska M. (2000). *Oncogene,* **19,** 6464–6471.
Tschiersch B, Hofmann A, Krauss V, Dorn R, Korge G and Reuter G. (1994). *EMBO J.,* **13,** 3822–3831.
Wu DY, Kalpana GV, Goff SP and Schubach WH. (1996). *J. Virol.,* **70,** 6020–6028.
Yang L, Embree LJ and Hickstein DD. (2000). *Mol. Cell. Biol.,* **20,** 3345–3354.
Zucman J, Melot T, Desmaze C, Ghysdael J, Plougastel B, Peter M, Zucker JM, Triche TJ, Sheer D and Turc-Carel C. (1993). *EMBO J.,* **12,** 4481–4487.