Clustering and domain architecture analysis of the SET domain proteins in plants with emphasis on Oryza sativa reveals that in plants they can be assigned to eight groups

Abstract
SET(Su(var)3-9, E(z) and Trithorax) domain is an evolutionarily conserved motif present in chromosomal proteins modulating gene expression in eukaryotes. In animals, SET domain proteins can be classified into four groups exemplified by Drosophila proteins E(Z), TRX, ASH1 and SU(VAR)3-9. In Arabidopsis, previous classification paper got results consistent to this classification. However, via collecting protein sequences from SMART and ChromDB directly, we demonstrate that the previous classification in Arabidopsis may be oversimplified and there are four additional groups specific to plants. Additionally, we try to give a clue as to the function of their members in Oryza sativa for prospective research. Our analysis is based on the multiple sequence alignment of their SET domains. The approach, as outlined by the title, is clustering plus domain architecture analysis.

Introduction

Epigenetics' central view is that the phyiological template for all DNA-dependent pathways is chromatin, not simply the DNA sequence. Chromatin states (euchromatin or heterochromatin) is deeply connected with gene expression and to a large extent, the genome reprogramming, cell differentiation and organism development. Histone proteins play a fundamental role in chromosome organization. They are involved in the first order and likely the higher-order packaging and compaction of DNA in the form of nucleosome. Histones can alter the chromotin states in three principal ways [Iizuka et al., 2003]. First, the position or properties of nucleosomes can be remodeled by DNA-dependent ATPase complexes [Martens et al., 2003]. Second, the composition of nucleosomes can be modulated by replacing the major histone proteins with specialized variants [Santisteban et al., 2000; Smith et al., 2002; Ahmad et al., 2002]. Finally, the histones are subject to a rich variety of covalent modifications that regulate their function. Known modifications include acetylation, phosphorylation, methylation, ubiquitination and ADP-ribosylation [van Holde, 1988; Wolffe, 1998]. Acetylation [Roth et al., 2001] and methylation [Stallcup, 2001] have been linked mainly with transcriptional stimulation. Phosphorylation [Cheung et al., 2000] instead is a marker for activation of immediate early genes and a signal for mitotic chromatin condensation. Less is known about other modifications. In addition, the crosstalk between the different modification mechanisms reveals the combinatorial nature behind histone modifications, which is referred to as 'histone code' [Jenuwein et al., 2001a]. The requirement of H2B K(lysine)123 ubiquitination for H3K4 and H3K79 Methylation is just a glimpse of the crosstalk [Sun et al., 2002; Dover et al., 2002].

SET domain proteins are known to be involved in methylation of histones. As the recent discovery of the causal relationship between histone methylation and DNA methylation [Bird 2001; Tamaru et al., 2001], SET domain proteins have been the focus of the chromatin -based research. The recently established link between RNAi and heterochromatin or H3K9 methylation greatly highlights this field [Volpe et al., 2002; Hall et al., 2002; Mochizuki et al., 2002; Taverna et al., 2002]. Something needs to be pointed out is that not all histone methylation is connected with SET domain proteins. One example is that methylation of H3K79 which resides within the core domain not the tail is associated with Dot1, a protein containing no SET [Feng et al., 2002; van Leeuwen et al., 2002; Ng et al., 2002; Lacoste et al., 2002].

The goal of this paper is to systemize and comprehensively analyze currently available SET domain proteins in plants. The focus is their members in Oryza sativa. Previous work may help to reflect our work. In 1998, there are around 40 SET domain proteins. Based on the homology of their SET domains they were distributed into four families, SU(VAR)3-9, E(Z), ASH1 and TRITHORAX [Jenuwein et al., 1998]. In 2001, the number of SET domain proteins increases to about 300. Jenuwein in Vienna, Austria presented a paper focusing on histone methyltransferases dividing them into seven subclasses, SUV39H1, G9a, NSD1, HRX, ESET, EZH1 and RIZ [Jenuwein 2001b]. In Arabidopsis [Baumbusch et al., 2001], SET domain proteins were assigned to four classes consistent to [Jenuwein et al., 1998]. In 2002, SET

domain proteins were classified into four classes, SET1 (E(Z) & TRX), SET2 (ASH1), SUV39 and RIZ [Kouzarides 2002]. In 2003 (in the time of paper writing), SMART gives an output of 600 SET domain proteins among which 100 are from plants. This time, Jenuwein provides us an epigentics roadmap trying to assemble all information known about histone methylation [Lachner et al., 2003].

Materials and Methods

Sequences:
Totally, there're 87 SET domain proteins in the analysis with 69 from plants and 18 from others. Initially, seqences were fetched from SMART(http://smart.embl-heidelberg.de/). In the later phase, nineteen SDG proteins(Oryza sativa) were added from ChromDB(http://www.chromdb.org/). Below is a table illustrating the distribution of these proteins among the selected organisms(abbreviation in parenthesis).

	Organism	Number of Proteins	I	II	III	IV	V	VI	VII	VIII
Plants	Arabidopsis thaliana(At)	33	3	4	4	4	4	3	3	8
	Nicotiana tabacum(Nt)	2					1			1
	Oryza sativa(Os)	23	3	3	2	3	6	5	1	
	Pisum sativum(Ps)	1								1
	Spinacia oleracea(So)	2								2
	Zea mays(Zm)	8	1		2	1	3	1		
	Caenorhabditis elegans(Ce)	1		1						
	Drosophila melanogaster(Dm)	4	1	1	1	1				
	Homo sapiens(Hs)	9	1	1	1	2				
	Saccharomyces cerevisiae(Sc)	2		1	1					
	Schizosaccharomyces pombe(Sp)	1				1				
	Neurospora crassa(Nc)	1				1				

As mainly interested in Oryza sativa, we identified SET domain proteins of Oryza sativa from two sources, SMART and ChromDB with fourteen and nineteen proteins respectively. Ten proteins among them are overlapping, six identical and four highly similar(with asterisk).(See table below)

Chrmodb	SMART-pacc	SMART-pname
SDG704	Q9AT64	OsSET1*
SDG706	Q8S3S4	OspSETreg*
SDG709	Q8RUS3	OssSET1
SDG711	Q9FP07	Ospclf2
SDG714	Q8S1X3	OspSUVH4
SDG715	BAC56009	OsP0705A05.28
SDG716	Q8LN53	OsHypo*
SDG718	Q8LLD6	OsSET
SDG719	Q8S1J4	Os*
SDG701		
SDG702		
SDG703		
SDG705		
SDG707		
SDG708		
SDG710		
SDG712		
SDG713		
SDG717		
	Q8S7V1	Os57.7KDa
	Q8LJP4	OsP0489G09.11
	Q8GTZ5	OspTPR
	Q8H2G7	Ospclf

In the four highly similar pairs,
SDG704 = OsSET1 +Middle (1aa),

SDG706 = N-terminal (397aa) + OspSETreg,
SDG716 = OsHypo + C-terminal (54aa),
SDG719 + N-terminal(1a) = Os.

So we choose all 19 ChromDB proteins with addition of 4 SMART proteins to constitute the SET domain proteins of Oryza sativa.

Domain Analysis Tools:
The large volume requires us to adopt an automatic and batch way to process the sequence data. The platform is Debian Linux woody 3.0.

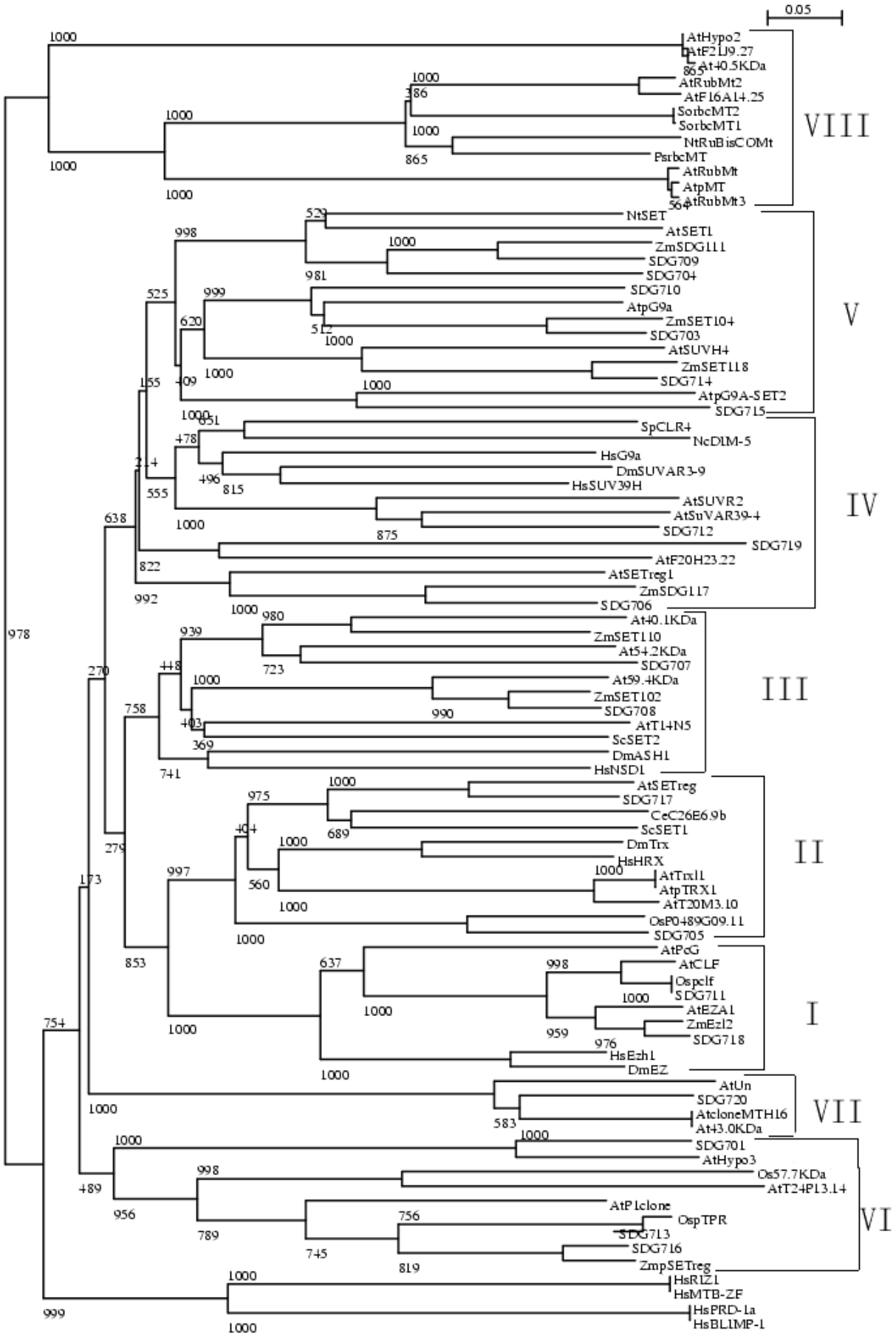
First, utilize PostgreSQL to build up a database called SETdb which contains two core tables, protein and dom_prot. The former is to hold the sequence data. The latter is to hold domain information corresponding to the sequences in the protein table.

Second, download the InterPro database and InterProScan search Package from EBI's ftp and run locally.

Third, write Python script to submit the local InterPro search results into dom_prot table.

Fourth, write programs to display the domains graphically. We used the lindna program from the EMBOSS package to draw domains initially. However , an interactive displaying program is much better. So we write a GNOME program to display domains in a SMART-like way through calling the bubble.pl perl script(downloadable from SMART) and can further display domain's location and full name when a specific domain-picture is pointed by the cursor.

Multiple alignments of protein sequencs were done with the ClustalX program(linux version 1.82).
Pfam (http://pfam.wustl.edu/hmmsearch.shtml/) and InterPro (http://www.ebi.ac.uk/interpro/) are the two main sources of speicfic domain function. Also, some recent literature may help to update the information from the two websites.



Results and Discussion

The 69 SET domain proteins of plants can be classified into eight groups based on their SET domains

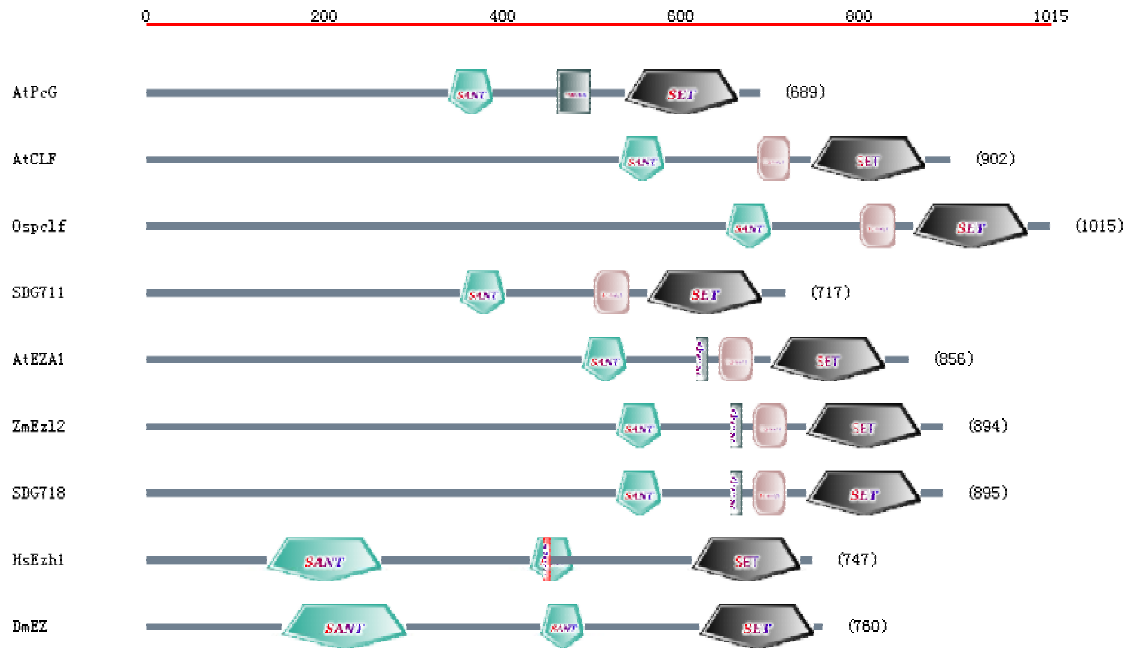
Sixty-nine SET domains of six plants' proteins were aligned with seventeen selected proteins from Caenorhabditis elegans, Drosophila melanogaster, Homo sapiens, Saccharomyces cerevisiae and Schizosaccharomyces pombe using ClustalX program. The six plants are Arabidopsis thaliana, Nicotiana tabacum, Oryza sativa, Pisum sativum, Spinacia oleracea and Zea mays. Except Oryza sativa, SET domain proteins of other plants are not complete. A tree based on the alignment of 69 plant SET domains and 17 such domains of proteins from other species was constructed by the neighbor-joining method, using ClustalX. Number of bootstrap trials is 1000 and random number generator seed is 111(all default values).

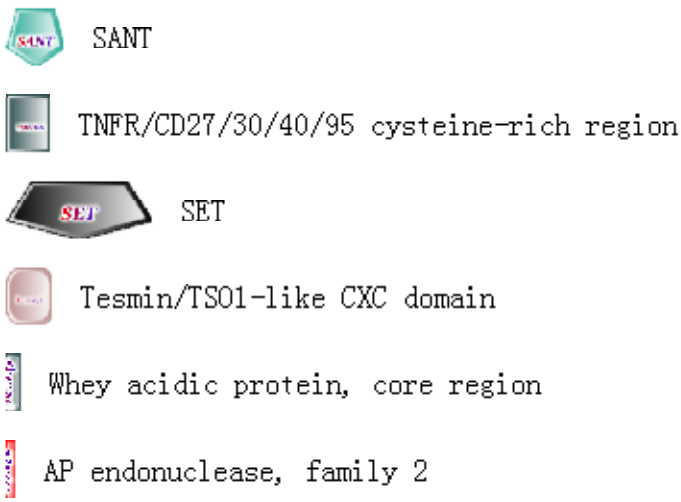
The previous 4 groups discussed by T.Jenuwein et. al. (1998) correspond to the first half of the 8 groups with E(Z) to group I, TRX to group II, ASH1 to group III and SU(VAR)3-9 to group IV. The second half of the 8 groups is specific to plants and no animals or bacteria included. (Table I)

E(Z): group I

Group	Dbsrc	Pname	Pacc
1	sptrembl	AtPcG(MEDEA)	O65312
1	sptrembl	AtCLF	O80455
1	sptremblnew	Ospclf	Q8H2G7
1	chromdb	SDG711	SDG000711
1	sptrembl	AtEZA1	Q9ZSM8
1	sptrembl	ZmEzl2	Q8S4P5
1	chromdb	SDG718	SDG000718
1	swissnew	HsEzh1	Q92800
1	sptrembl	DmEZ	Q9VTA3

The tree gives very good support for recognition of the E(Z)-like proeins of all species included, as a distinct group.





Important domains in the E(Z) group

SANT domain

It is also named Myb DNA-binding domain, which is encoded by the retroviral oncogene v-myb, and its cellular counterpart c-myb. This domain family specifically recognize the sequence YAAC(G/T)G [Borgmeyer et al., 1998, Stewart et al., 1996]. In myb, one of the most conserved regions consisting of three tandem repeats has been shown to be involved in DNA-binding [Sippel et al., 1987]. Interestingly, a recent paper reveals that the SANT domain is required for histone acetylation directed by SAGA complex in the yeast[Sterner et al., 2002].

Tesmin/TSO1-like CXC domain

This domain family includes proteins that have two copies of a cysteine rich motif as follows: C-X-C-X4-C-X3-YC-X-C-X6-C-X3-C-X-C-X2-C, which includes Tesmin (Q9Y4I5) [Sugihara et al., 1999] and TSO1 (Q9LE32) [Hauser et al., 2000]. This family is called a CXC domain in [Hauser et al., 2000].

Whey acidic protein, core region

A group of proteins containing 8 characteristically-spaced cysteine residues, which are involved in disulphide bond formation, have been termed '4-disulphide core' proteins [Sippel et al., 1982]. While the pattern of conserved cysteines suggests that the sequences may adopt a similar fold, the overall degree of sequence similarity is low (e.g. a few Pro and Gly residues are reasonably well conserved, as is the polar/acidic nature of residues between the third and fourth Cys, but otherwise there is little sequence conservation). The group of sequences that share this pattern include whey acidic protein (WAP) [Sippel et al., 1982]; elafin (an elastase-specific inhibitor from human skin) [Gregory et al., 1990]; WDNM1 protein (which is involved in the metastatic potential of adenocarcinomas in rats [Dear et al., 1988]; Kallmann syndrome protein [Bougueleret et al., 1991]; and caltrin-like protein II from guinea pig [Lardy 1990] (which inhibits calcium transport into spermatozoa).

Function of the E(Z) group members

AtCLF, the first SET domain protein described in plants, encodes a repressor of the floral homeotic gene *agamous* [Goodrich et al., 1997]. AtPcG(MEDEA) is involved in inhibition of endosperm development in the absence of fertilization [Grossniklaus et al., 1998; Kiyosue et al., 1999; Luo et al., 1999; Vielle et al., 1999]. DmEZ, a member of the polycomb proteins, restricts expression boundaries of HOM-C [Jones et al., 1990; Phillips et al., 1990] but has also been shown to be involved in repression of the early acting segmentation and gap genes [Moazed et al., 1992; McKeon et al., 1994; Pelegri et al., 1994].

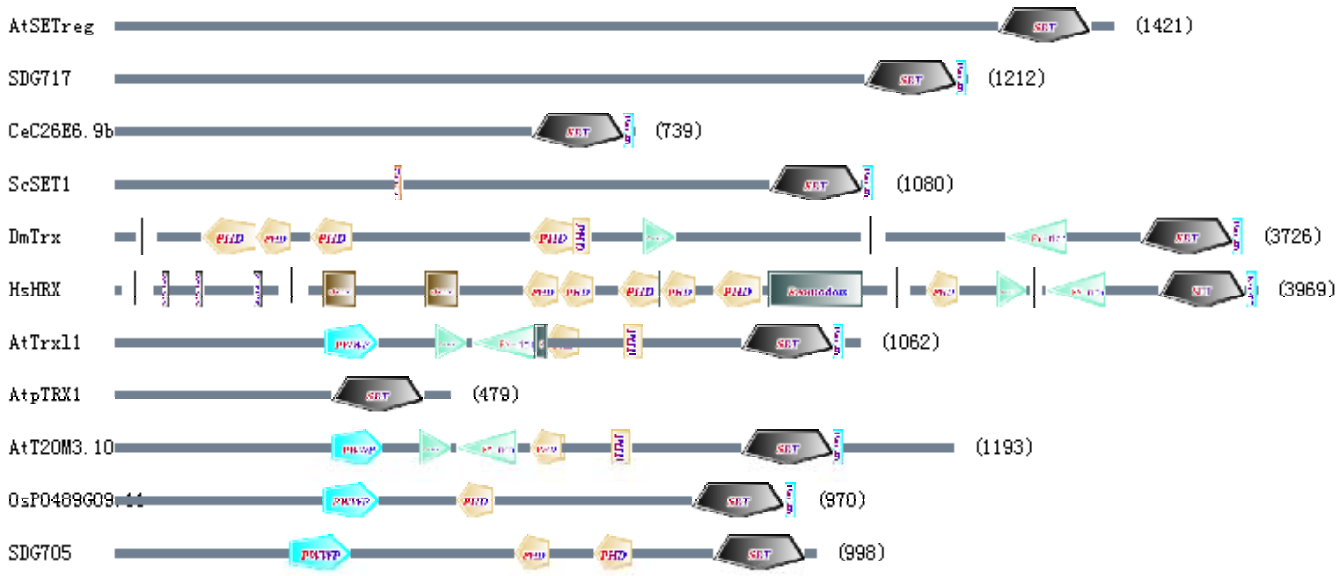
The above results demonstrate this group proteins to be repressive chromosomal proteins.

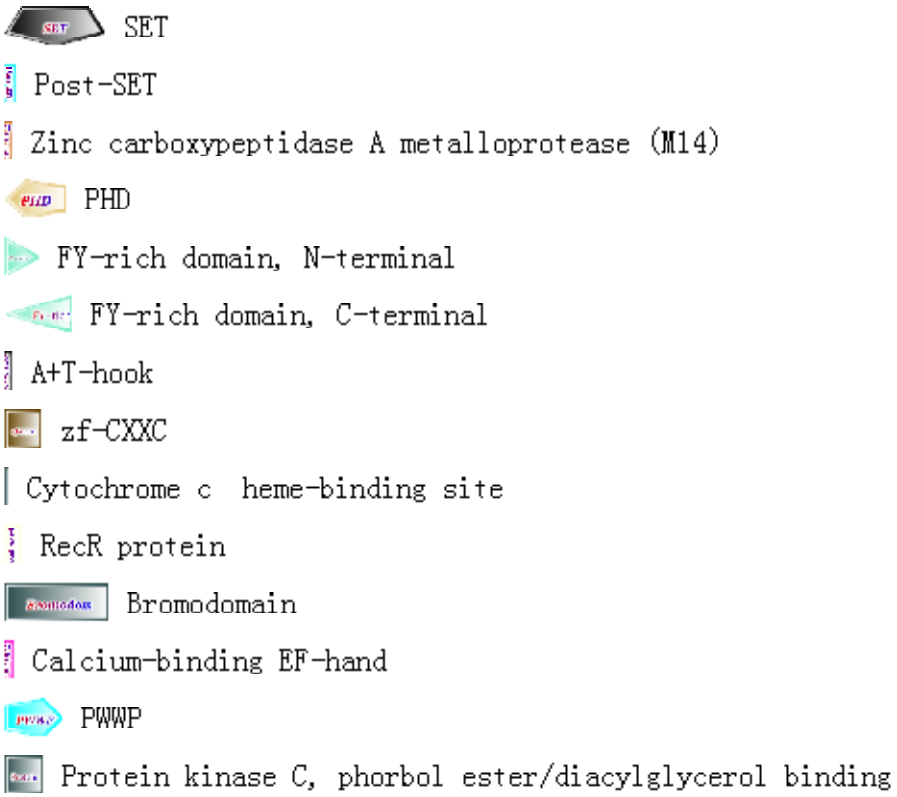
Function related to histone methylation
Currently, knowledge about histone methylation in this group is retricted in the animal field. DmEZ and HsEZH2 (HsEzh1's homologue) are involved in H3K9 and H3K27 Methylation [*review by Lachner et al., 2003*].

Putative function of the Os (Oryza sativa) proteins in the E(Z) group
Both the tree and domain architecture demonstrate that Ospclf and SDG711are most close to AtCLF, which suggest Ospclf and SDG711 may be involved in repressive function by H3K9, H3K27 methylation. The 'Whey acidic protein, core region' domain may make SDG718 to be distinct from the previous two proteins in speicfic function.

TRX: group II

Group	Dbsrc	Pname	Pacc
2	sptrembl	AtSETreg	Q9FIH7
2	chromdb	SDG717	SDG000717
2	sptrembl	CeC26E6.9b	Q18221
2	swissnew	ScSET1	P38827
2	sptrembl	DmTrx	Q9VFL1
2	sw	HsHRX	Q03164
2	sptrembl	AtTrxl1	Q9C5X4
2	sptremblnew	AtpTRX1	AAO22754
2	sptrembl	AtT20M3.10	Q9MA43
2	sptrembl	OsP0489G09.11	Q8LJP4
2	chromdb	SDG705	SDG000705





Important domains in the TRX group

PHD

The PHD finger, is a C4HC3 zinc-finger-like motif found in nuclear proteins thought to be involved in chromatin-mediated transcriptional regulation [Aasland et al. 1995]. The PHD finger motif is reminiscent of, but distinct from the C3HC4 type RING finger.

The function of this domain is not yet known but in analogy with the LIM domain it could be involved in protein-protein interaction and be important for the assembly or activity of multicomponent complexes involved in transcriptional activation or repression. PHD folds into an interleaved type of Zn-finger chelating 2 Zn ions in a similar manner to that of the RING and FYVE domains [Pascual et al., 2000].

PWWP

The PWWP domain is named after a conserved Pro-Trp-Trp-Pro motif. The function of the domain is currently unknown. The PWWP domain was first identified at the C-terminus of WHSC1, a gene mapping to the Wolf-Hirschhornsyndrome critical region [Stec et al., 1998]. It is present in proteins of nuclear origin and plays a role in cell growth and differentiation. Due to its position, the composition of amino acids close to the PWWP motif and the pattern of other domains it has been suggested that the domain is involved in protein-protein interactions [Stec et al., 2000].

Post-SET

This is a cysteine-rich motif following a subset of SET domains. Deletion of this region of SUV39H1 impairs its histone methyltransferase function.

FY-rich domain

The "FY-rich" domain consititutes two parts, N-terminal and C terminal, which are sometimes closely juxtaposed, but sometimes are far distant. It is of unknown function, but occurs frequently in chromatin-associated proteins like trithorax and its homologues.

Function of the TRX group members

DmTrx has been identified genetically as a positive regulator of homeotic genes [Mazo et al., 1990].

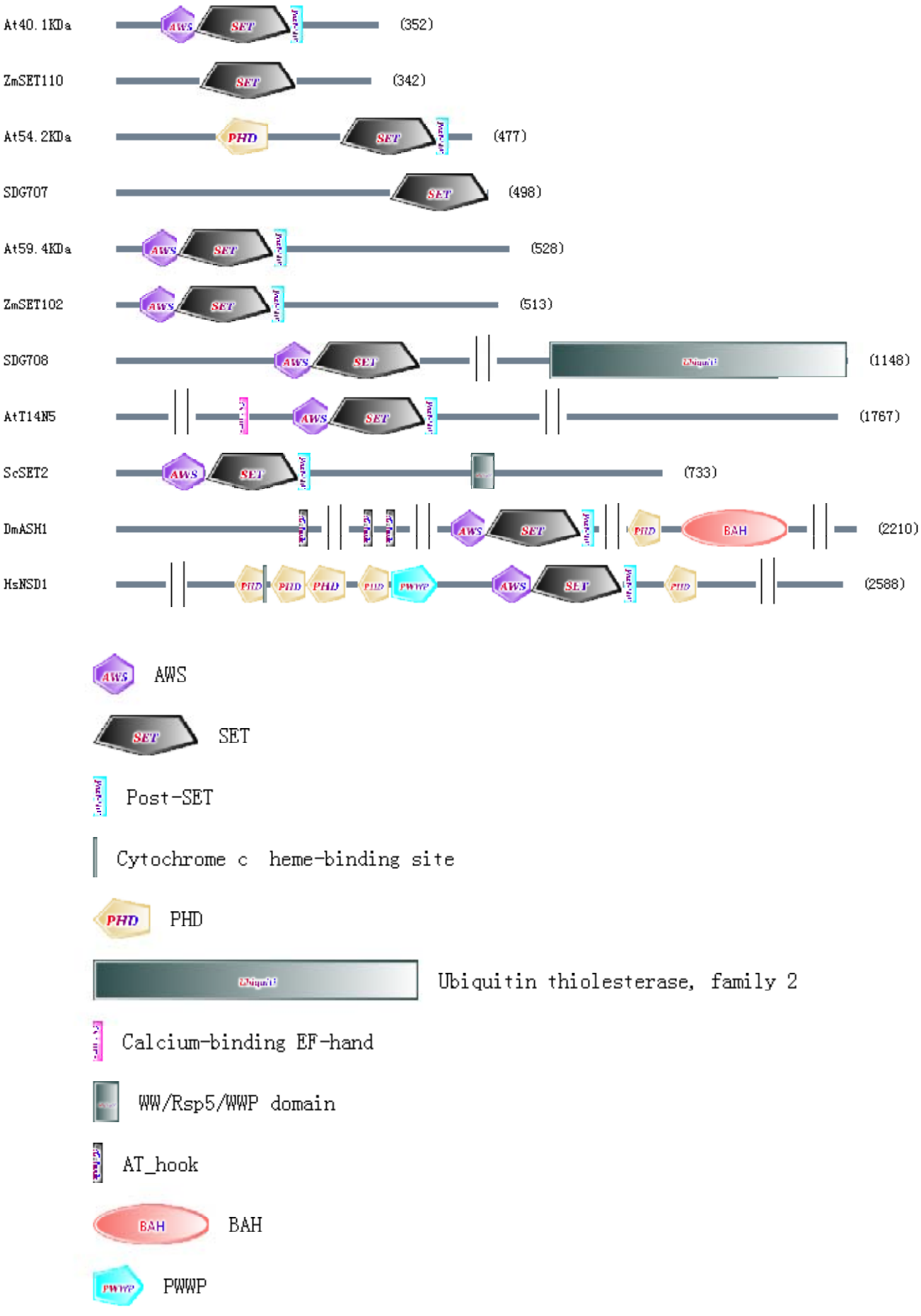
HsTRX (MLL, ALL-1) positively regulates the clustered Hox genes [Yu et al., 1995].
In a word, DmTrx and HsTRX are involved in trithorax activation [review by Lachner et al., 2003].
AtTrxl1 (ATX1), which is similar to the Drosophila trx, functions as an activator of homeotic genes by regulating floral organ development. The effects are specific: structurally and functionally related flower homeotic genes are under different control [Alvarez-Venegas et al., 2003].
ScSET1 is involved in rDNA silencing, telomeric silencing and transcriptional activation [review by Lachner et al., 2003].

Function related to histone methylation
ScSET1, DmTrx, HsHRX and AtTrxl1 are all involved in H3K4 methylation [review by Lachner et al., 2003]. Early experiments suggested that H3K4 methylation was linked with active genes. However, as depicted before, ScSET1 is involved in rDNA silencing, telomeric silencing. Recently, a study using antibodies that can distinguish between di- and tri-methylated H3 K4, revealed that tri-methylation is specific for the active state of transcription, whereas di-methylated K4 exist in both active and repressed genes [Santos-Rosa et al., 2002]. Thus the number of methyl groups in a modification, and not just the particular site, appears to play an important role in the functional consequences of histone methylation [Iizuka et al., 2003].

Putative function of the Os (Oryza sativa) proteins in the TRX group
The domain architecture of OsP0489G09.11 and SDG705 shows great similarity with AtTrxl1 and so these two probably function as activators of homeotic genes through H3K4 methylation. Lack of FY-rich domain should be considered for these two proteins as it is present in AtTrxl1. As to SDG717, it may also be involved in H3K4 methylation, but probably functions as a repressor in addition to an activator because it is more close to ScSET1.

ASH1: group III

Group	Dbsrc	Pname	Pacc
3	sptrembl	At40.1KDa	Q9M1X9
3	sptremblnew	ZmSET110	AAN41254
3	sptrembl	At54.2KDa	O65563
3	chromdb	SDG707	SDG000707
3	sptrembl	At59.4KDa	Q9SRE2
3	sptrembl	ZmSET102	Q8L819
3	chromdb	SDG708	SDG000708
3	sptrembl	AtT14N5	O80663
3	swissnew	ScSET2	P46995
3	sptrembl	DmASH1	Q24189
3	sptrembl	HsNSD1	O88491



Important domains in the ASH1 group

AWS

This domain, Associated With SET, of unknown function is found in eukaryotic proteins of unknown function. This domain, as the name suggests, is often found in association with the SET domain, suggesting a role in gene regulation by methylation of lysine residues in histones and other proteins.

Ubiquitin thiolesterase, family 2
This domain is also known as Ubiquitin carboxy-terminal esterase or Ubiquitin carboxy-terminal hydrolase.
Ubiquitin thiolesterases (EC: 3.1.2.15) (UCH) (deubiquitinating enzymes) are thiol proteases that recognize and hydrolyze the peptide bond at the C-terminal glycine of ubiquitin [Hauser et al., 1991; Pellman et al., 1998]. These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquinated proteins.
There are two distinct families of UCH. The second family consist of large proteins (800 to 2000 residues) that share two regions of similarity, a region that contains a conserved cysteine which is probably implicated in the catalytic mechanism and a region that contains two conserved histidines residues, one of which is also probably implicated in the catalytic mechanism.

Post-SET
(previously described in group II)

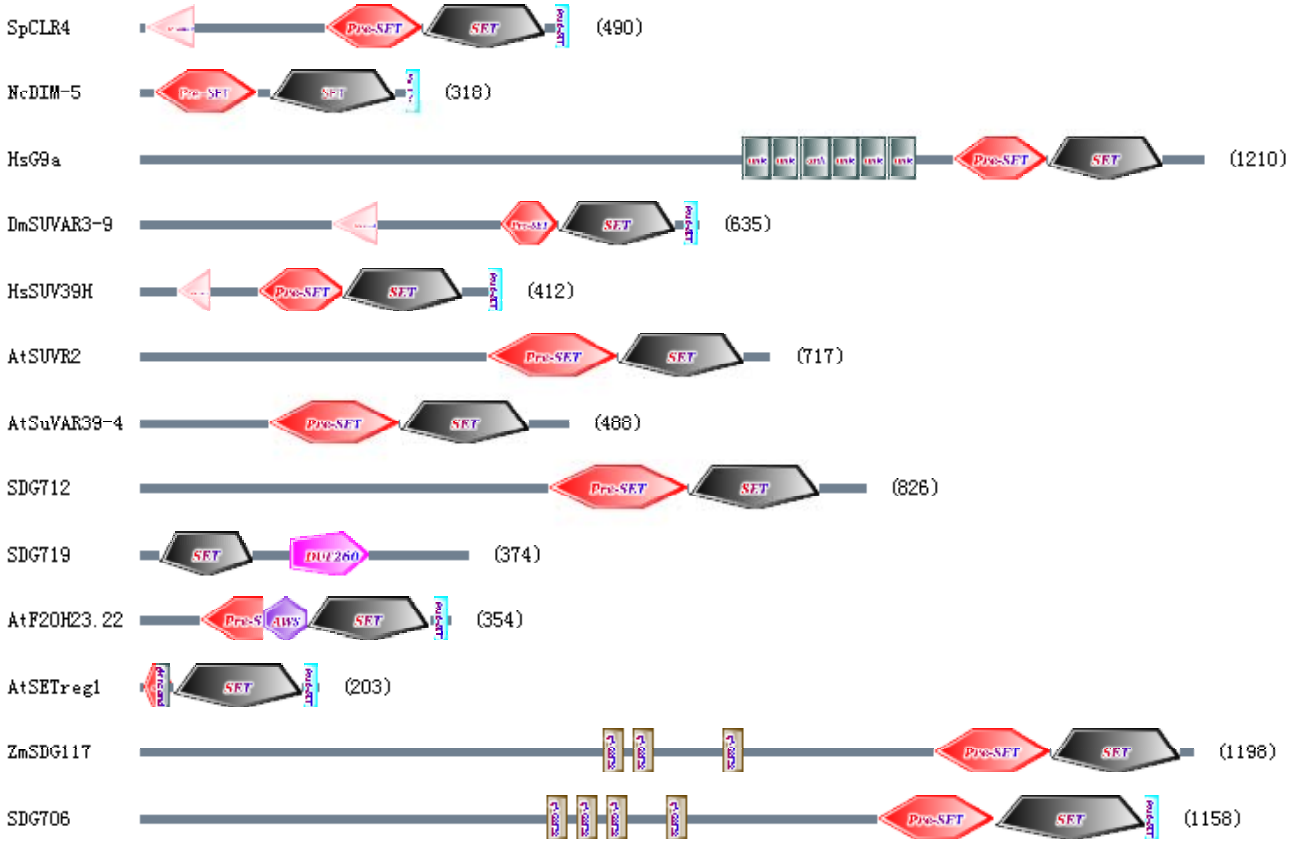
Function of the ASH1 group members
DmASH1 is involved in trithorax activation. ScSET2 is involved in gene repression [review by Lachner et al., 2003].

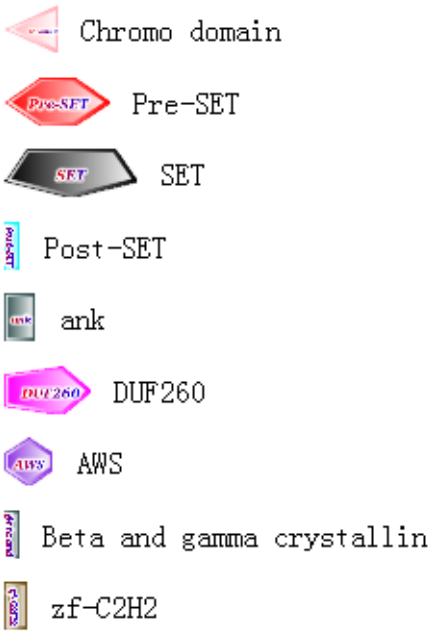
Function related to histone methylation
DmASH1 is involved in H3K4, H3K9 and H4K20 methylation. ScSET2 is involved in H3K36 methylation [review by Lachner et al., 2003].

Putative function of the Os (Oryza sativa) proteins in the ASH1 group
Function known in this group is quite diverse and restricted in non-plant field. Os proteins' function maybe easy to predict with any revelation in the plant field.

SU(VAR)3-9: group IV

Group	Dbsrc	Pname	Pacc
4	swissnew	SpCLR4	O60016
4	sptrembl	NcDIM-5	Q8X225
4	swissnew	HsG9a	Q96KQ7
4	swissnew	DmSUVAR3-9	P45975
4	sw	HsSUV39H	O43463
4	sptrembl	AtSUVR2	Q941L4
4	sptrembl	AtSuVAR39-4	Q8W595
4	chromdb	SDG712	SDG000712
4	chromdb	SDG719	SDG000719
4	sptrembl	AtF20H23.22	Q9SRV2
4	sptrembl	AtSETreg1	O64827
4	sptremblnew	ZmSDG117	AAO32935
4	chromdb	SDG706	SDG000706





Important domains in the Su(VAR)3-9 group

Pre-SET

This is a Cys-rich putative Zn2+-binding domain that occurs N-terminal to some SET domains. The function of this domain is unknown.

Chromo domain

The CHROMO (CHRromatin Organization MOdifier) domain [Gaunt et al., 1991; Koonin et al., 1995; Paro et al., 1990; Stewart et al., 1995] is a conserved region of around 60 amino acids, originally identified in Drosophila modifiers of variegation. These are proteins that alter the structure of chromatin to the condensed morphology of heterochromatin, a cytologically visible condition where gene expression is repressed. In one of these proteins, Polycomb, the chromo domain has been shown to be important for chromatin targeting. Proteins that contain a chromo domain appear to fall into 3 classes. The first class includes proteins having an N-terminal chromo domain followed by a region termed the chromo shadow domain [Stewart et al., 1995], eg. Drosophila and human heterochromatin protein Su(var)205 (HP1); and mammalian modifier 1 and modifier 2. The second class includes proteins with a single chromo domain, eg. Drosophila protein Polycomb (Pc); mammalian modifier 3; human Mi-2 autoantigenand and several yeast and Caenorhabditis elegans hypothetical proteins. In the third class paired tandem chromo domains are found, eg. in mammalian DNA-binding/helicase proteins CHD-1 to CHD-4 and yeast protein CHD1.

zf-C2H2

The C2H2 zinc finger is the classical zinc finger domain. The two conserved cysteines and histidines co-ordinate a zinc ion. The following pattern describes the zinc finger. #-X-C-X(1-5)-C-X3-#-X5-#-X2-H-X(3-6)-[H/C] Where X can be any amino acid, and numbers in brackets indicate the number of residues. The positions marked # are those that are important for the stable fold of the zinc finger. The final position can be either his or cys. The C2H2 zinc finger is composed of two short beta strands followed by an alpha helix. The amino terminal part of the helix binds the major groove in DNA binding zinc fingers.

Zinc finger domains are nucleic acid-binding protein structures first identified in the Xenopus transcription factor TFIIIA [Evans et al., 1988] . These domains have since been found in numerous nucleic acid-binding proteins. A zinc finger domain is composed of 25 to 30 amino-acid residues including 2 conserved Cys and 2 conserved His residues in a C-2-C-12-H-3-H type motif. The 12 residues separating the second Cys and the first His are mainly polar and basic, implicating this region in particular in nucleic acid binding. The zinc finger motif is an unusually small, self-folding domain in which Zn is a crucial component of its tertiary structure. All bind 1 atom of Zn in a tetrahedral array to yield a finger-like projection, which interacts with nucleotides in the major groove of the nucleic acid. The Zn binds to the conserved Cys and His residues. Fingers have been found to bind to about 5 base pairs of nucleic acid containing short runs of guanine residues. They have the ability to bind to both RNA and DNA, a versatility not demonstrated by the helix-turn-helix motif. The zinc finger may thus represent the original nucleic acid binding protein. It has also been suggested that a Zn-centred domain could be used in a protein interaction, e.g. in protein kinase C. Many classes of zinc fingers are characterized according to the number and positions of the histidine and cysteine residues involved in the zinc atom coordination. In the first class to be characterized, called C2H2, the first pair of zinc coordinating residues are cysteines, while the second pair are histidines.

Function of the Su(VAR)3-9 group members

SpCLR4 is involved in centromeric/mating-type silencing. NcDIM-5 is involved in DNA methylation. HsG9a is involved in transcriptional repression, while its murine homologue is involved in imprinting as well as transcriptional repression. HsSUV39H is involved in Rb-mediated silencing, while its murine homologues, Suv39h1 and Suv39h2 are involved in pericentric heterochromatin and DNA methylation. DmSUVAR3-9 is a dominant PEV modifier. [review by Lachner et al., 2003]

Function related to histone methylation

SpCLR4, NcDIM-5, HsG9a, DmSUVAR3-9 and HsSUV39H are all involved in H3K9 methylation [review by Lachner et al., 2003]. HsG9a's murine counterpart is also involved in H3K29 methylation [review by Lachner et al., 2003].

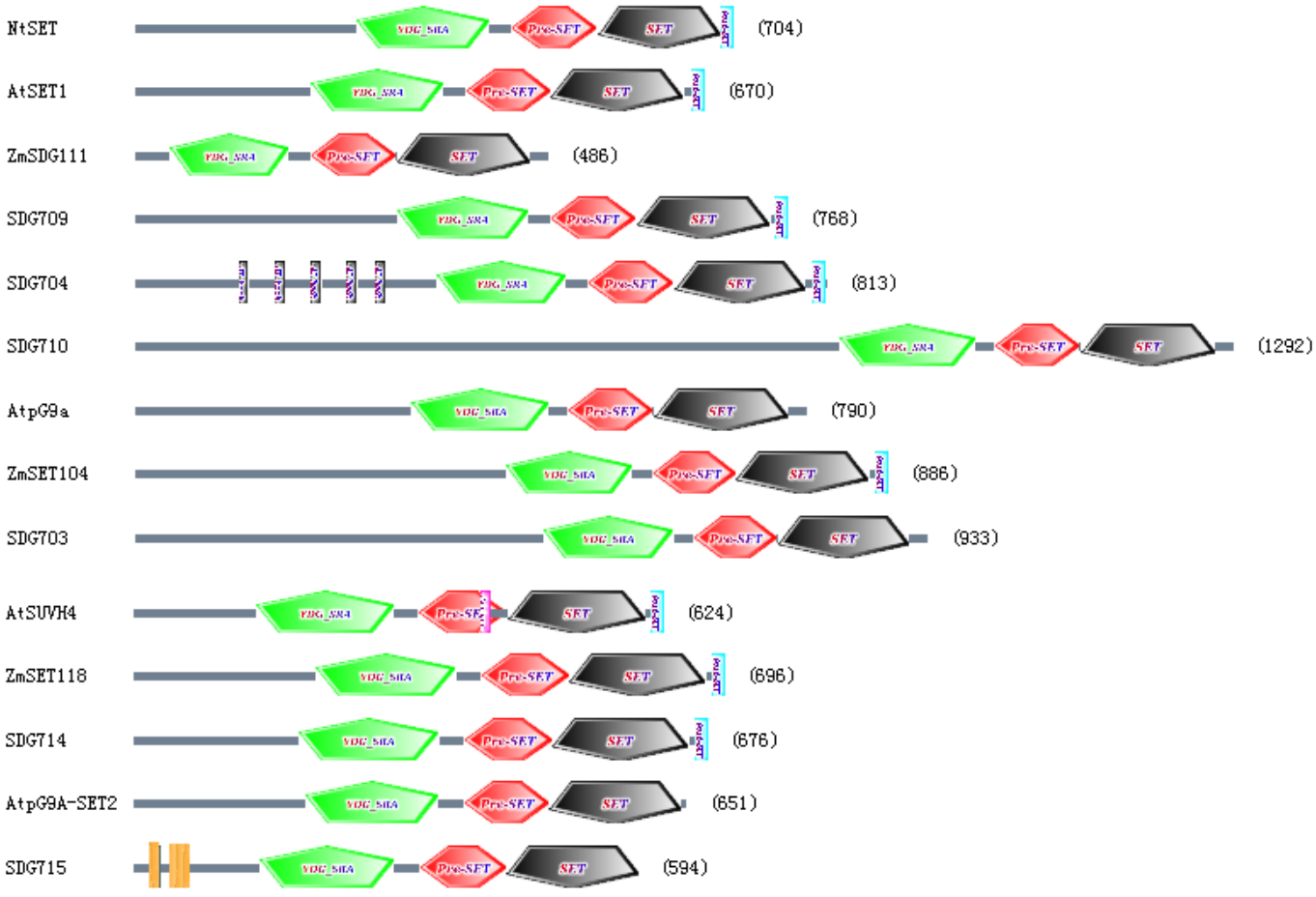
Putative function of the Os (Oryza sativa) proteins in the Su(VAR)3-9 group

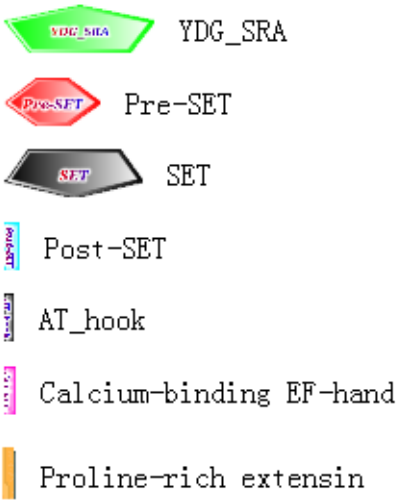
As the function consistency in this group, Os proteins (SDG712, SDG719 and SDG706) are probably involved in transcriptional repression through H3K9 methylation. However, the difference in the tree should be noticed; function-unknown proteins are not clustered in the same clade with the function-

known proteins.

Group V

Group	Dbsrc	Pname	Pacc
5	sptrembl	NtSET	Q93YF5
5	sptrembl	AtSET1	Q9FF80
5	sptremblnew	ZmSDG111	AAO32934
5	chromdb	SDG709	SDG000709
5	chromdb	SDG704	SDG000704
5	chromdb	SDG710	SDG000710
5	sptremblnew	AtpG9a	AAO22580
5	sptrembl	ZmSET104	Q8L820
5	chromdb	SDG703	SDG000703
5	sptrembl	AtSUVH4	Q9C5P3
5	sptrembl	ZmSET118	Q8L821
5	chromdb	SDG714	SDG000714
5	sptrembl	AtpG9A-SET2	O22781
5	chromdb	SDG715	SDG000715





Important domains in the group V

YDG_SRA

The function of this domain is unknown, it contains a conserved motif YDG after which it has been named [Baumbusch et al., 2001].

Domain in SET domain containing proteins. In mouse it is found in a nuclear protein associated with cell proliferation [Fujimori et al., 1998].

Pre-SET

(previously described in group IV)

Post-SET

(previously described in group II)

AT_hook

At hooks are DNA binding motifs with a preference for A/T rich regions.

High mobility group (HMG) proteins are a family of relatively low molecular weight non-histone components in chromatin. HMG-I and HMG-Y are proteins of about 100 amino acid residues which are produced by the alternative splicing of a single gene. HMG-I proteins bind preferentially to the minor groove of AT-rich regions in double-stranded DNA [Reeves et al., 1990; Friedmann et al., 1993]. It is suggested that these proteins could function in nucleosome phasing and in the 3' end processing of mRNA transcripts. They are also involved in the transcription regulation of genes containing, or in close proximity to, AT-rich regions. DNA-binding of these, and several related, proteins is effected by an 11-residue domain known as an AT-hook. Within known HMG-I proteins are found three highly conserved regions, closely related to the consensus sequence TPKRPRGRPKK. A synthetic oligopeptide with this sequence specifically binds to substrate DNA in a manner reminiscent of intact HMG-I proteins. Structure predictions suggest that the peptide has a secondary structure similar to the anti-tumour and anti-viral drugs netropsin and distamycin, and to the dye Hoechst 33258. These ligands, which also preferentially bind to AT-rich DNA, effectively compete with both the synthetic peptide and the HMG-I proteins for DNA binding. The peptide also contains novel structural features such as a predicted Asx bend, or 'hook', at its N-terminus, and laterally-projecting cationic Arg/Lys 'bristles', which may play a role in the binding of HMG-I proteins. The predicted peptide structure, the AT-hook, is a previously undescribed DNA-binding motif [Reeves et al., 1990].

Proline-rich extensin

This entry matches protein sequences characterised by multiple tandem pro-ser-rich repeats (often ser-

pro-pro-pro-pro penta- peptide repeats or variations of these), in which the proline residue is hydroxylated and then glycosylated. This pattern is found in a variety of proteins from prokaryotes and eukaryotes, including a group of extensins. Extensins are plant cell-wall proteins; they can account for up to 20% of the dry weight of the cell wall [Giralt et al., 1988]. They are highly-glycosylated, possibly reflecting their interactions with cell-wall carbohydrates. Amongst their functions is cell wall strengthening in response to mechanical stress (e.g., during attack by pests, plant-bending in the wind, etc.).

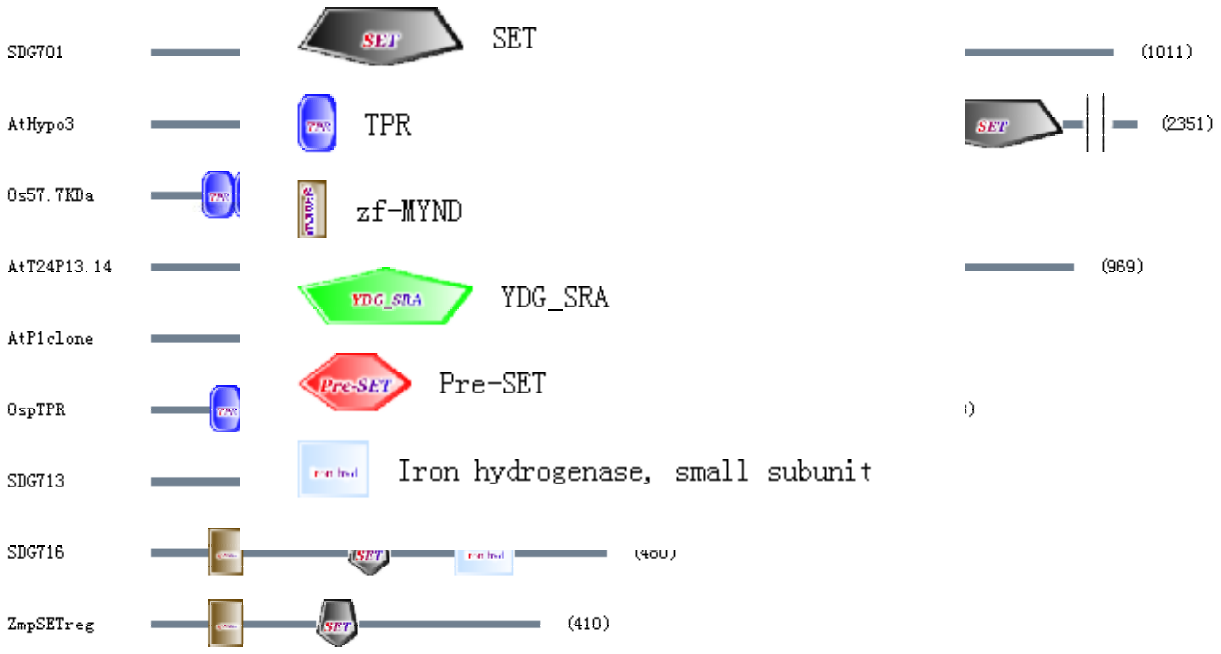
Function of the group V members
Transgenic experiment suggests NtSET may be involved in transcriptional repression of growth control genes through the formation of higher-order chromatin domains [Shen 2001].
AtSUVH4 is known as KRYPTONITE. It is involved in CpNpG DNA methylation through interaction of CMT3 with methylated chromatin [Jackson et al., 2002].

Function related to histone methylation
AtSUVH4 is involved in H3K9 methylation [Jackson et al., 2002].

Putative function of the Os (Oryza sativa) proteins in the group V
For the good support from the domain architecture of this group, Os proteins (SDG709, SDG704, SDG710, SDG703, SDG714, SDG715) are probably associated with DNA methylation through H3K9 methylation.

Group VI

Group	Dbsrc	Pname	Pacc
6	chromdb	SDG701	SDG000701
6	sptrembl	AtHypo3	O23372
6	sptrembl	Os57.7KDa	Q8S7V1
6	sptrembl	AtT24P13.14	Q9LQX6
6	sptrembl	AtP1clone	Q9LSY2
6	sptremblnew	OspTPR	Q8GTZ5
6	chromdb	SDG713	SDG000713
6	chromdb	SDG716	SDG000716
6	sptrembl	ZmpSETreg	Q8SA95



Important domains in the group VI

TPR

The tetratricopeptide repeat of typically 34 amino acids was first described in the yeast cell cycle regulator Cdc23p and later found to occur in a large number of proteins [Lamb et al., 1995; Das et al., 1998; Goebl et al., 1991]. A function for this repeat seems to be protein-protein interaction, but common features in the interaction partners have not been defined. It has been proposed that TPR proteins preferably interact with WD-40 repeat proteins, but in many instances several TPR-proteins seem to aggregate to multi-protein complexes. Prominent examples of TPR-proteins include, Cdc16p, Cdc23p and Cdc27p components of the cyclosome/APC, the Pex5p/Pas10p receptor for peroxisomal targeting signals, the Tom70p co-receptor for mitochondrial targeting signals, Ser/Thr phosphatase 5C and the p110 subunit of O-GlcNAc transferase.

zf-MYND

This domain is found in some suppressors of cell cycle entry [McGinnis et al., 1996;Blalock et al., 1998]. The MYND zinc finger (ZnF) domain is one of two domains in AML/ETO fusion protein required for repression of basal transcription from the multidrug resistance 1 (MDR-1) promoter. The other domain is a hydrophobic heptad repeat (HHR) motif [Hiebert et al., 1998]. The AML-1/ETO fusion protein is created by the (8;21) translocation, the second most frequent chromosomal abnormality associated with acute myeloid leukemia. In the fusion protein the AML-1 runt homology domain, which is responsible for DNA binding and CBF beta interaction is linked to ETO, a gene of unknown function [Hiebert et al., 1995].

Iron hydrogenase, small subunit

This family represents the small subunit of the Fe-only hydrogenases EC: 1.18.99.1. The subunit is comprised of alternating random coil and alpha helical structures that encompass the large subunit in a novel protein fold [Fontecilla-Camps et al., 1999].

Function of the group VI members

(not available)

Putative function of the Os (Oryza sativa) proteins in the group VI

(we don't know)

Group VII

Group	Dbsrc	Pname	Pacc
7	sptrembl	AtUn	Q9FNE9
7	chromdb	SDG720	SDG000720
7	sptrembl	AtcloneMTH16	Q9FXW6
7	sptrembl	At43.0KDa	Q9LXE2

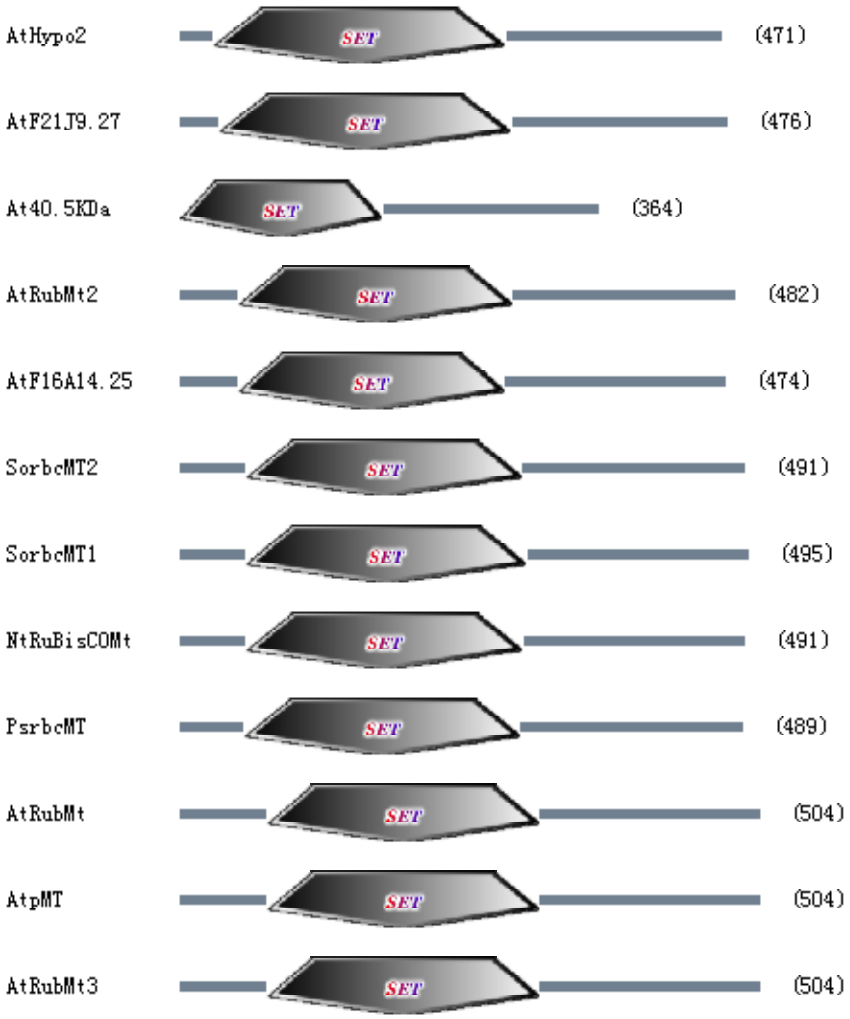
Important domains in the group VII
PHD
(previously described in group II)

Function of the group VII members
(not available)

Putative function of the Os (Oryza sativa) proteins in the group VII
(we don't know)

Group VIII

Group	Dbsrc	Pname	Pacc
8	sptrembl	AtHypo2	Q8L9I6
8	sptrembl	AtF21J9.27	Q9FYK3
8	sptrembl	At40.5KDa	Q94B58
8	swissnew	AtRubMt2	Q9XI84
8	sptrembl	AtF16A14.25	Q9LMF5
8	sptrembl	SorbcMT2	Q9TIM3
8	sptrembl	SorbcMT1	O80013
8	swissnew	NtRuBisCOMt	P94026
8	swissnew	PsrbcMT	Q43088
8	sptrembl	AtRubMt	Q8LF17
8	sptrembl	AtpMT	O65218
8	sptrembl	AtRubMt3	Q9S7D2



Important domains in the group VIII
Domain architecture of this group is very simple with only the SET domain present.

Function of the group VIII members
NtRuBisCOMt and PsrbcMT show ability to catalyze the posttranslational methylation of the epsilon-amino group of Lys-14 in the large subunit (LS) of Rubisco [Ying et al., 1996; Klein et al., 1995]. These two proteins with their counterparts in spinach, SorbcMT1, SorbcMT2, can catalyze the methylation of the alpha-amino group of the N-terminal methionine of the processed form of the small subunit (SS) of the Rubisco [Ying et al., 1999]. However SorbcMT1 and SorbcMT2 shows activity limited to methylation of the SS [Ying et al., 1999].

Putative function of the Os (Oryza sativa) proteins in the group VIII
Currently, there is no rice protein in this group for lack of the genome information.

References

Aasland R, Gibson TJ, Stewart AF.
The PHD finger: implications for chromatin-mediated transcriptional regulation.
Trends Biochem Sci 20:56-59. (1995)

Ahmad K, Henikoff S: The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. Mol Cell 9:1191-1200. (2002)

Alvarez-Venegas, R., Pien, S., Sadder, M., Witmer, X., Grossniklaus, U. and Avramova, Z.: ATX-1, an

Arabidopsis Homolog of Trithorax, Activates Flower Homeotic Genes. Current Biology, Vol. 13, 627-637, (2003)

Baumbusch LO, Thorstensen T, Krauss V, Fischer A, Naumann K, Assalkhou R, Schulz I, Reuter G, Aalen RB. The Arabidopsis thaliana genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. Nucleic Acids Res 29:4319-4333. (2001)

Bird, A.: Methylation Talk Between Histones and DNA. Science VOL 294: 2113-2115 (2001)

Blalock J.E. , Propst S.M. , Stone A.S. , Green M.M. , Ban E.M. , LeBoeuf R.D. , Tauber J.D. Molecular cloning, sequence analysis, expression, and tissue distribution of suppressin, a novel suppressor of cell cycle entry. J. Biol. Chem. 273: 361- 368 (1998)

Boehm S, Frishman D, Mewes HW. Variations of the C2H2 zinc finger motif in the yeast genome and classification of yeast zinc finger proteins. Nucleic Acids Res 25:2464-2469. (1997)

Borgmeyer, U., Biednkapp, H., Sippel, A.E., Klempnauer, K.-H. Viral myb oncogene encodes a sequence-specific DNA-binding activity. Nature 335: 835- 837 (1988)

Bougueleret L. , Caterina D. , Cohen D. , Le Paslier D. , Millasseau P. , Wunderle V. , Compain S. , Petit C. , Claverie J.-M. , Weissenbach J. , Levilliers J. , Lutfalla G. , Hardelin J.-P. , Legouis R. , Delemarre-Van De Waal H. The candidate gene for the X-linked Kallmann syndrome encodes a protein related to adhesion molecules. Cell 67: 423- 435 (1991)

Cheung, P., Allis, C. D. and Sassone-Corsi, P.: Signaling to chromatin through histone modifications. Cell 103, 263-271. (2000)

Das AK, Cohen PW, Barford D. The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. EMBO J 17:1192-1199. (1998)

Dear T.N. , Kefford R.F. , Ramshaw I.A. Differential expression of a novel gene, WDNM1, in non-metastatic rat mammary adenocarcinoma cells. Cancer Res. 48: 5203- 5209 (1988)

Dover J, Schneider J, Tawiah-Boateng MA, Wood A, Dean K, Johnston M, Shilatifard A: Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6. J Biol Chem 277:28368-28371.(2002)

Evans RM, Hollenberg SM. Zinc fingers: gilt by association. Cell Mar 11;52(5):783 (1988)

Feng, Q., Wang, H., Ng, H. H., Erdjument-Bromage, H., Tempst, P., Struhl, K. and Zhang, Y. Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. Curr. Biol. 12, 1052-1058. (2002)

Fontecilla-Camps J.C. , Piras C. , Hatchikian C.E. , Legrand P. , Nicolet Y. Desulfovibrio desulfuricans iron hydrogenase: the structure shows unusual coordination to an active site Fe binuclear center. Structure 7: 13- 23 (1999)

Friedmann M, Holth LT, Zoghbi HY, Reeves R. Organization, inducible-expression and chromosome localization of the human HMG-I(Y) nonhistone protein gene. Nucleic Acids Res Sep 11; 21(18): 4259-67 (1993)

Fujimori A, Matsuda Y, Takemoto Y, Hashimoto Y, Kubo E, Araki R, Fukumura R, Mita K, Tatsumi K, Muto M. Cloning and mapping of Np95 gene which encodes a novel nuclear protein associated with cell

proliferation. Mamm Genome Dec;9(12):1032-5 (1998)

Gaunt S.J. , James T.C. , Paro R. , Burton R.D. , Miller J.R. , Kothary R. , Pearce J. , Singh P.B. A sequence motif found in a Drosophila heterochromatin protein is conserved in animals and plants. Nucleic Acids Res. 19: 789- 794 (1991)

Giralt E. , Albericio F. , Perez-Grau L. , Stiefel V. , Puigdomenech P. , Ludevid M.D. , Ruiz-Avila L. Molecular cloning of cDNAs encoding a putative cell wall protein from Zea mays and immunological identification of related polypeptides. Plant Mol. Biol. 11: 483- 493 (1988)

Goebl M, Yanagida M. The TPR snap helix: a novel protein repeat motif from mitosis to transcription. Trends Biochem Sci May;16(5):173-7 (1991)

Goodrich, J., Puangsomlee, P., Martin, M., Long, D., Meyerowitz, E.M. and Coupland, G. A Polycomb-group gene regulates homeotic gene expression in Arabidopsis. Nature, 386, 44-51. (1997)

Gregory H. , Christophers E. , Young J.A. , Schroeder J.-M. , Wiedow O. Elafin - An elastase-specific inhibitor of human skin - Purification, characterisation, and complete amino acid sequence. J. Biol. Chem. 265: 14791- 14795 (1990)

Grossniklaus, U., Vielle-Calzada, J.-P., Hoeppner, M.A. and Gagliano, W.B. Maternal control of embryogenesis by MEDEA, a polycomb-group gene in Arabidopsis. Science, 280, 446-450. (1998)

Hall IM, Shankaranarayana GD, Noma K, Ayoub N, Cohen A, Grewal SI: Establishment and maintenance of a heterochromatin domain. Science 297:2232-2237. (2002)

Hauser, B. A., He, J. Q., Park, S. O. & Gasser, C. S. TSO1 is a novel protein that modulates cytokinesis and cell expansion in Arabidopsis. Development 127:2219-2226. (2000)

Hauser H.-P. , Seufert W. , Jentsch S. Genetic analysis of the ubiquitin system. Biochim. Biophys. Acta 1089: 127- 139 (1991)

Hiebert S.W. , Meyers S. , Lenny N. Functional domains of the t(8;21) fusion protein, AML-1/ETO. Oncogene 11: 1761- 1769 (1995)

Hiebert S.W. , Schuetz J. , Sun D. , Lutterbach B. The MYND motif is required for repression of basal transcription from the multidrug resistance 1 promoter by the t(8;21) fusion protein. Mol. Cell. Biol. 18: 3604- 3611 (1998)

Iizuka M. and Smith M. M. : Functional consequences of histone modifications. Current Opinion in Genetics & Development 13:154-160 (2003)

Jackson, J. P., Lindroth, A. M., Cao, X. and Jacobsen, S. E.: Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature 416, 556-560. (2002)

Jenuwein T. and Allis C. D.: Translating the Histone Code. SCIENCE. VOL 293:1074-1080. (2001a)

Jenuwein, T.: Re-SET-ting heterochromatin by histone methyltransferases. TRENDS in Cell Biology Vol.11 No.6: 266-273 (2001b)

Jenuwein, T., Laible, G., Dorn, R., Reuter, G.: SET domain proteins modulate chromatin domains in eu- and heterochromatin. Cell. Mol. Life Sci. 54, 80-93. (1998)

Jones R. S. and Gelbart W. M. Genetic analysis of the Enhancer of zeste locus and its role in gene regulation in Drosophila melanogaster. Genetics 126, 185-199 (1990)

Kinoshita, T., Yadegari, R., Harada, J.J., Goldberg, R.B. and Fischer, R.L. Imprinting of the MEDEA polycomb gene in the Arabidopsis endosperm. Plant Cell, 11, 1945-1952 (1999)

Klein,R.R.; Houtz,R.L.; Cloning and developmental expression of pea ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit N-methyltransferase. Plant Mol. Biol. 27:249 (1995)

Koonin E.V. , Lucchesis J.C. , Zhou S. The chromo superfamily - new members, duplication of the chromo domain and possible role in delivering transcription regulators to chromatin. Nucleic Acids Res. 23: 4229- 4233 (1995)

Kouzarides, T.: Histone methylation in transcriptional control. Current Opinion in Genetics & Development 12:198-209 (2002)

Lachner, M., O'Sullivan, R.J. and Jenuwein, T. : An epigenetic road map for histone lysine methylation. Journal of Cell Science 116, 2117-2124 (2003)

Lacoste N, Utley RT, Hunter JM, Poirier GG, Cote J: Disruptor of telomeric silencing-1 is a chromatin-specific histone H3 methyltransferase. J Biol Chem 277:30421-30424. (2002)

Lamb JR, Tugendreich S, Hieter P. Tetratrico peptide repeat interactions: to TPR or not to TPR? Trends Biochem Sci 20:257-259. (1995)

Lardy H.A. , San Agustin J. , Coronel C.E. Purification and structure of caltrin-like proteins from seminal vesicle of the guinea pig. J. Biol. Chem. 265: 6854- 6859 (1990)

Luo, M., Bilodeau, P., Koltunow, A., Dennis, E.S., Peacock, W.J. and Chaudhury, A.M. Genes controlling fertilization-independent seed development in Arabidopsis thaliana. Proc. Natl. Acad. Sci. USA, 96, 296-301 (1999)

Martens J.A. and Winston F. Recent advances in understanding chromatin remodeling by Swi/Snf complexes. Current Opinion in Genetics & Development 13:136-142 (2003)

Mazo, A., Huang, D.-H., Mozer, B., and Dawid, I. The trithorax gene, a trans-acting regulator of the bithorax complex in Drosophila encodes a protein with zinc-binding domains. Proc. Natl. Acad. Sci. USA 87, 2112-2116.(1990)

McGinnis W. , Gross C.T. DEAF-1, a novel protein that binds an essential region in a Deformed response element. EMBO J. 15: 1961- 1970 (1996)

McKeon J., Slade E., Sinclair D. A., Cheng N., Couling M. and Brock H. W. Mutations in some Polycomb group genes of Drosophila interfere with regulation of segmentation genes. Mol. Gen. Genet. 244: 474-483 (1994)

Moazed D. and O'Farrell P. H. Maintenance of the engrailed expression pattern by Polycomb group genes in Drosophila. Development 116: 805-810 (1992)

Mochizuki K, Fine NA, Fujisawa T, Gorovsky MA: Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in Tetrahymena. Cell 110:689-699. (2002)

Ng HH, Feng Q, Wang H, Erdjument-Bromage H, Tempst P, Zhang Y, Struhl K: Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein

association. Genes Dev 16:1518-1527. (2002)

Paro R. Imprinting a determined state into the chromatin of Drosophila. Trends Genet. 6: 416- 421 (1990)

Pascual J, Martinez-Yamout M, Dyson HJ, Wright PE. Structure of the PHD zinc finger from human williams-beuren syndrome transcription factor. J Mol Biol 304:723-729. (2000)

Pelegri F. and Lehmann R. A role of Polycomb group genes in the regulation of gap gene expression in Drosophila. Genetics 136: 1341-1353 (1994)

Pellman D. , D'andrea A. Deubiquitinating enzymes: a new class of biological regulators. Crit. Rev. Biochem. Mol. Biol. 33: 337- 352 (1998)

Phillips M. D. and Shearn A. Mutations in poly-combeotic, a Drosophila Polycomb-group gene, cause a wide range of maternal and zygotic phenotypes. Genetics 125:91-101 (1990)

Reeves R, Nissen MS. The A.T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. J Biol Chem May 25;265(15):8573-82 (1990)

Roth, S. Y., Denu, J. M. and Allis, C. D.: Histone acetyltransferases. Annu. Rev. Biochem. 70, 81-120. (2001)

Santisteban MS, Kalashnikova T, Smith MM: Histone H2A.Z regulates transcription and is partially redundant with nucleosome remodeling complexes. Cell 103:411-422. (2000)

Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NC, Schreiber SL, Mellor J, Kouzarides T: Active genes are tri-methylated at K4 of histone H3. Nature 419:407-411. (2002)

Shen WH: NtSET1, a member of a newly identi®ed subgroup of plant SET-domain-containing proteins, is chromatin-associated and its ectopic overexpression inhibits tobacco plant growth. The Plant Journal 28(4), 371-383 (2001)

Sippel, A.E., Klempnauer, K.-H. The highly conserved amino-terminal region of the protein encoded by the v-myb oncogene functions as a DNA-binding domain. EMBO J. 6: 2719- 2725 (1987)

Sippel A.E. , Hennighausen L.G. Mouse whey acidic protein is a novel member of the family of 4-disulfide core proteins. Nucleic Acids Res. 10: 2677- 2684 (1982)

Smith MM: Centromeres and variant histones: what, where, when and why? Curr Opin Cell Biol 2002, 14:279-285. (2002)

Stallcup, M. R.: Role of protein methylation in chromatin remodeling and transcriptional regulation. Oncogene 20, 3014-3020. (2001)

Stec I, Nagl SB, van Ommen GJ, den Dunnen JT. : The PWWP domain: a potential protein-protein interaction domain in nuclear proteins influencing differentiation? FEBS Lett May 4;473(1):1-5 (2000)

Stec I, Wright TJ, van Ommen GJB, de Boer PAJ, van Haeringen A, Moorman AFM, Altherr MR, den Dunnen JT. : WHSC1, a 90 kb SET domain-containing gene, expressed in early development and homologous to a Drosophila dysmorphy gene maps in the Wolf-Hirschhorn syndrome critical region and is fused to IgH in t(4;14) multiple myeloma. Hum Mol Genet 7:1071-1082 (1998)

Sterner, D. E., Wang, X., Bloom, M. H., Simon, G. M. & Berger, S. L. The SANT Domain of Ada2 Is Required for Normal Acetylation of Histones by the Yeast SAGA Complex. THE JOURNAL OF BIOLOGICAL CHEMISTRY Vol 227, No. 10:8178-8186 (2002)

Stewart A.F. , Aasland R. The chromo shadow domain, a second chromo domain in heterochromatin-binding protein 1, HP1. Nucleic Acids Res. 23: 3168- 3174 (1995)

Stewart, A.F., Aasland, R., Gibson, T. The SANT domain: a putative DNA-binding domain in the SWI-SNF and ADA complexes, the transcriptional co-repressor N-CoR and TFIIIB. Trends Biochem. Sci. 21: 87- 88 (1996)

Sugihara, T., Wadhwa, R., Kaul, S.C. & Mitsui Y. A novel testis-specific metallothionein-like protein, tesmin, is an early marker of male germ cell differentiation. Genomics 57:130-136. (1999)

Sun ZW, Allis CD: Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. Nature 418:104-108.(2002)

Tamaru, H. & Selker, E. U.: A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. Nature VOL 414: 277-283. (2001)

Taverna SD, Coyne RS, Allis CD: Methylation of histone H3 at lysine 9 targets programmed DNA elimination in Tetrahymena. Cell 110:701-711. (2002)

van Holde, K. E. Chromatin. New York: Springer Verlag. (1988)

van Leeuwen F, Gafken PR, Gottschling DE: Dot1p modulates silencing in yeast by methylation of the nucleosome core. Cell 109:745-756. (2002)

Vielle Calzada, J.P., Thomas, J., Spillane, C., Coluccio, A., Hoeppner, M.A. and Grossniklaus, U. Maintenance of genomic imprinting at the Arabidopsis MEDEA locus requires zygotic DDM1 activity. Genes Dev., 13, 2971-2982 (1999)

Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA: Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science 297:1833-1837. (2002)

Wolffe, A. P.: Chromatin: Structure and Function. San Diego: Academic Press. (1998)

Ying,Z.; Janney,N.; Houtz,R.L.; Organization and characterization of the ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit epsilon N-methyltransferase gene in tobacco. Plant Mol. Biol. 32:663 (1996)

Ying Z., Mulligan R.M., Janney N., Houtz R.L.: Rubisco small and large subunit N-methyltransferases. Bi- and Mono- functional methyltransferases that methylate the small and large subunits of rubisco. J. Biol. Chem. 274:36750-36756 (1999)

Yu, B.D., Hess, J.L., Horning, S.E., Brown, G.A., and Korsmeyer, S.J. Altered Hox expression and segmental identity in Mll-mutant mice. Nature 378, 505-508. (1995)