

## In Silico Study of Cis-acting Elements of Histone Deacetylase gene Family of *Homo sapiens* & *Mus musculus*

<sup>1</sup>Raghvendra Singh, <sup>2</sup>Swati Singh, <sup>3</sup>P. N. Pandey,

<sup>1</sup>Center of Bioinformatics, IIDS, Nehru Science Center, University of Allahabad, Allahabad, India

<sup>2</sup>Center of Bioinformatics, IIDS, Nehru Science Center, University of Allahabad, Allahabad, India.

<sup>3</sup>Department of Mathematics, University of Allahabad, Allahabad, India

---

**ABSTRACT :** HDAC is a well known member of regulatory proteins that binds to chromatin and negatively regulates the expression of genes. These genes are under regulation of cis-acting elements, present in upstream as well as downstream regions, that provide platforms for different trans factors to bind and regulate expression of corresponding HDAC genes. Each HDAC has approximately 9 cis acting regulatory elements. Several potential cis-acting elements in up/down stream regions of HDAC genes were discovered through in silico biological methods. Transcription factors that could bind to these elements have also been identified and their association with cancer ascertained. Distribution and comparative conservation profile of human and mouse of HDAC gene family have been acknowledged. In this study we have shown the mode of HDAC action on the basis of binding of various regulatory elements. Deregulation of predicted TFs: n-MYC, Nkx, S8, Sox-5, S8, FREAC, GATA and SP play significant role in cancer.

**KEYWORDS:** HDAC, deactivation of genes, cis-acting elements, Transcription Factor (TF), Cancer

---

### I. INTRODUCTION.

HDACs are members of an ancient enzyme family found in plants, animals, and fungi, as well as archaeobacteria and eubacteria (Leipe DD & Landsman D 1997). Phylogenetic study of HDAC gene family (Ivan V Gregoretti *et.al.* 2004) reveals HDAC members are divided into 4 classes on the basis sequence homology score. Class I contain, HDAC1, 2, 3 and 8. Class II is further subdivided into class IIa (HDAC4, 5, 7 and 9) and class IIb (HDAC6 and HDAC10) and class IV contain HDAC11. Class I deacetylases, HDAC1 and HDAC2 are found in the ubiquitously expressed mSin3A, NURD/Mi2/NRD and CoREST corepressor complexes (Yang XJ, Seto E. 2003). HDAC3 associates to and is activated by SMRT and NCoR co-repressors that play an important role in the regulation of gene expression by nuclear hormone receptors (Yang XJ, Seto E. 2003). HDAC3 also deacetylate non-histone substrates, such as the RelA subunit of NF-κB, thereby affecting its stability and DNA-binding properties (Chen LF, *et.al.* 2001). HDAC8 play important role in muscle contractibility and linked with cancer activity (Waltregny D, *et al.* 2005). The class IIa HDACs (subtypes 4, 5, 7 and 9) are involve in tissue-specific expression and stimulus-dependent nucleo-cytoplasmic shuttling. In cytoplasm they are targeted by several kind of kinases (Verdin E, *et al* 2003). Inside the nucleus they associate with transcription factors, like MEF and Runx families, and regulate differentiation and cellular hypertrophy in muscle and cartilage tissues (Zhang CL, *et al.* 2002). Class IIb (HDAC6 and HDAC10) have a duplication of their catalytic domains, but the second catalytic domain is thought to be dysfunctional in HDAC10 (Hubbert C, *et al.* 2002). HDAC6 is the only deacetylase known to act on tubulin, its deacetylation leads for disposal of misfolded proteins in aggresomes (Kawaguchi Y *et al.* 2003) class IV (HDAC11), function of HDAC11 it act novel drug target in carcinoma.

Last 10 decade research on HDAC concludes that, deregulation of all the class members of HDAC lead in cancer. So it is important to explore mode of regulation and cis acting elements that regulate the synthesis of class of HDAC gene members. Cis regulatory elements are often spread over regions spanning thousands of base pairs around (up and down stream regions) the targeted gene in multi-cellular eukaryotes. Computational predictions have been successfully applied for telling potential regulatory regions for further experimental analysis; (Aparicio S *et.al.*1995; Bagheri-Fam S *et.al.* 2001; Christensen TH *et.al.*1993; Gumucio DL *et.al.* 1992). Transcriptional regulation, in particular modeling and prediction of TFBS, is one of the most studied problems in computational biology (Wasserman WW and Krivan W (2003), Stormo GD 2000) Reliable profile-based methods and model frameworks have been developed over the years which accurately describe the DNA-binding specificity of a TF (Stormo GD 2000). In this study Analysis of cis-acting elements in HDAC gene family of *Homo sapiens* and *Mus musculus*, greatly enhance our understanding the expression of HDAC genes which epigenetically control expression of genes which are vital force of developmental process.

The search for putative cis-acting elements in the HDAC gene family of *Homo sapiens* and *Mus musculus* would lead to a better understanding of the oncogenic pathways that activate this gene. In this study we analyzed the up and down stream regions of HDAC gene family of *H. sapiens* and *M. musculus* using computational DNA pattern recognition methods. We report here that several potential cis-regulatory motifs were identified in this way.

## II. MATERIALS AND METHODS

**Sequence retrieval :** The genomic sequences of *Homo sapiens* and *Mus musculus* HDAC gene (table 1) have been collected from DBTSS <http://dbtss.hgc.jp/index.html> (Yamashita R et.al. 2011). In this study, 1000 base pairs upstream and downstream from the position of TSS (Transcription Start Site) have been retrieved. These sequences are further subjected to analysis of cis acting elements.

**Transcription factor binding sites of HDAC gene family of *H. sapiens* and *M. musculus* and repeat analysis :** Selected sequence, length 1kb of upstream and 1kb of downstream from TSS position are most probable region to locate the cis acting elements and TF binding site. All the fetched sequences of HDAC gene members are further used for prediction of cis-acting elements by using CONSITE (Albin Sandelin *et.al.* 2004) (<http://asp.iu.uib.no:8090/cgi-bin/CONSITE/consite/>). ConSite is a web-based tool for predicting cis-regulatory elements in genomic sequences. Predictions are based on the integration of binding site with high-quality transcription factor models and cross-species comparison filtering (phylogenetic footprinting). Results produced by ConSite, are consist of number of steps. In brief, the program (i) aligns input promoter sequences by ORCA aligner (Arenillas and Wasserman, unpublished), (ii) calculates the degree of conservation in the alignment, (iii) scans the sequences of a set of TF binding profile models, (iv) performs filtering on the initial sets of sites using phylogenetic footprinting and (v) presents the results in user-selected output formats.

## III. RESULT

Selected upstream and downstream sequences of HDAC gene family have been analyzed by CONSITE tool for prediction of cis-acting elements and corresponding transcription factors. A comparative illustration of cis acting elements present in human and mouse HDAC gene family has been shown in Table 2 and table 3. There are four grouped classes of HDAC members (11 variants in human and mouse) and Class I (HDAC1, HDAC2, HDAC3 and HDAC8) has 21 unique cis acting elements in *H. sapience* whereas 26 unique cis acting element present in *M. musculus*. Class IIa (HDAC4, HDAC5, HDAC7 and HDAC9) has 26 and 28 unique cis acting elements in *H. sapience* and *M. musculus* respectively and Class IIb (HDAC6 and HDAC10) has 12 and 11 unique cis acting elements in *H. sapience* and *M. musculus* respectively. Finally in Class IV there is only one member HDAC 11 which has 8 and 2 unique cis acting elements in *H. sapience* and *M. musculus* respectively. Hence, we compared and found in all different classes that there is no common cis acting elements between HDAC members of Human and mouse. Cis acting element distribution in human and mouse HDAC gene family is also graphically represented together in Fig.7 where HDAC family members are shown on x-axis is and number of cis acting elements on y-axis. It was found that for HDAC1 there are 6 cis acting elements in human and 9 cis acting elements in mouse while for HDAC2 a total of 8 and 9 cis acting elements are present in human and mouse respectively. Furthermore, the occurrence of cis acting elements for HDAC3 is 8 and 5, for HDAC4 is 5 and 6, for HDAC5 8 is 9, for HDAC6 is 4 and 7 , for HDAC7 is 9 and 6, for HDAC8 is 7 and 8, for HDAC9 is 4 and 7 cis acting elements, for HDAC10 is 8 and 4 while for HDAC11 it is 8 and 2 in human and mouse respectively. A comparative conservation profile of human and mouse of HDAC gene family has been shown in Fig1 - Fig 6, where x-axis represents nucleotide positions and y-axis has conservation cut of predicting cis acting elements and it also has analogous transcription factor. Green line illustrates upstream alignment whereas Cyan line in Figures demonstrates the alignment of down stream sequences of both human and mouse in HDAC gene family.

## IV. DISCUSSION

The use of In Silico Biology methods represent new, faster, cost-efficient techniques to identify novel regulatory elements that provide areas for more in depth in vitro investigations to confirm their functional effects. The main focus has been on analysis of cis-acting elements in upstream and downstream regions of HDAC gene family of *H. sapience* and *M. musculus* where the function of predicted cis-acting elements present in *H. sapience* and *M. musculus* in HDAC gene family were elucidated animal cis-acting database. Abnormality in expression and mutation in binding sites of predicted putative transcription factors play significance role in cancer and other syndroms. Predicted transcription factor, GATA1-3 belongs to the GATA family of transcription factor. GATA1 directly contributes to the silencing of genes associated with cellular proliferation such as Kit, Myc, and Myb, which are proto-oncogenes [22] Failure to silence these genes might be expected to

result in hyperproliferation of immature erythroid cells. It regulates luminal epithelial cell differentiation in the mammary gland, study suggested that changes in the structure or expression GATA-2 uncover broader role in human diseases [23] Another major predicted transcription factor is SP-1 belongs to class SP transcription factor, it play important role in regulating cancer cell metabolism [24]. There are other TFs like n-MYC, Nkx, S8, Sox-5, S8, and FREAC profoundly signature their binding site in upstream regions of HDAC gene family in both human and mouse.

## V. CONCUSSION

In silico study was done to determine the functional motifs at sequential and structural level in members HDAC gene family. The cis-acting elements study in 1000bps upstream and 1000 bps downstream from TSS of HDAC gene family has been investigated to elucidate the functional role and behavior of each members of this family.

## REFERENCES

- [1] Aparicio S, Morrison A, Gould A, Gilthorpe J, Chaudhuri C, Rigby P, Krumlauf R and Brenner S (1995) Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, *Fugu rubripes*. *Proc. Natl Acad. Sci.* 92:1684–1688.
- [2] Bagheri-Fam S, Ferraz C, Demaille J, Scherer G and Pfeifer D (2001) Comparative genomics of the SOX9 region in *Homo sapiens* and *Fugu rubripes*: conservation of short regulatory sequence elements within large intergenic regions. *Genomics*, 78:73–82.
- [3] Christensen TH, Prentice H, Gahlmann R and Kedes L (1993) Regulation of the *Homo sapiens* cardiac/slow-twitch troponin C gene by multiple, cooperative, cell-type-specific, and MyoD-responsive elements. *Mol. Cell Biol.* 13:6752–6765.
- [4] Gumucio DL, Heilstedt-Williamson H, Gray TA, Tarle SA, Shelton DA, Tagle DA, Slightom JL, Goodman M and Collins FS (1992) Phylogenetic footprinting reveals a nuclear protein which binds to silencer sequences in the *Homo sapiens* gamma and epsilon globin genes. *Mol. Cell Biol.* 12 4919–4929.
- [5] Wasserman WW and Krivan W (2003) In silico identification of metazoan transcriptional regulatory regions. *Naturwissenschaften*, 90:156–166.
- [6] Stormo G D 2000 DNA binding sites: representation and discovery. *Bioinformatics* 16 16–23.
- [7] Kim D. Pruitt, Tatiana Tatusova, and Donna R 2007 Maglott/NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins *Nucleic Acids Res.* 35 D61–D65.
- [8] Thompson JD, Higgins DG and Gibson TJ. 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22 4673-80.
- [9] Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S 2011 MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* 28 2731-2739.
- [10] Yamashita R, Sathira NP, Kanai A, Tanimoto K, Arauchi T, Tanaka Y, Hashimoto S, Sugano S, Nakai K, Suzuki Y. 2011 Genome-wide characterization of transcriptional start sites in *Homo sapiens* s by integrative transcriptome analysis. *Genome Res.* 5 775-789.
- [11] Albin Sandelin, Wasserman WW and Boris Lenhard 2004 ConSite: web-based prediction of regulatory elements using cross-species comparison. *Nucleic Acids Res* 32 W249–W252.
- [12] Ivan V Gregoret, Yun-Mi Lee and Holly V Goodson 2004 Molecular Evolution of the Histone Deacetylase Family: Functional Implications of Phylogenetic Analysis. *Journal of Molecular biology* 1 17–31.
- [13] Wasserman, W.W. and Krivan, W. (2003) In silico identification of metazoan transcriptional regulatory regions. *Naturwissenschaften*, 90, 156–166.
- [14] Stormo, G.D. (2000) DNA binding sites: representation and discovery. *Bioinformatics*, 16, 16–23.
- [15] Yang XJ, Seto E. Collaborative spirit of histone deacetylases in regulating chromatin structure and gene expression. *Curr Opin Genet Dev* 2003; 13:143–153.
- [16] Chen LF, Fischle W, Verdin E, et al. Duration of nuclear NF-kappaB action regulated by reversible acetylation. *Science* 2001; 293:1653–1657.
- [17] Waltregny D, Glenisson W, Tran SL, et al. Histone deacetylase HDAC8 associates with smooth muscle alpha-actin and is essential for smooth muscle cell contractility. *FASEB J* 2005; 19:966–968.
- [18] Verdin E, Dequiedt F, Kasler HG. Class II histone deacetylases: versatile regulators. *Trends Genet* 2003; 19:286–293.
- [19] Zhang CL, McKinsey TA, Chang S, et al. Class II histone deacetylases act as signal-responsive repressors of cardiac hypertrophy. *Cell* 2002; 110:479–488.
- [20] Hubbert C, Guardiola A, Shao R, et al. HDAC6 is a microtubule-associated deacetylase. *Nature* 2002; 417:455–458.
- [21] Kawaguchi Y, Kovacs JJ, McLaurin A, et al. The deacetylase HDAC6 regulates aggresome formation and cell viability in response to misfolded protein stress. *Cell* 2003; 115:727–738.
- [22] Munugalavadla V, Dore LC, Tan BL, et al. Repression of c-kit and its downstream substrates by GATA-1 inhibits cell proliferation during erythroid maturation. *Mol Cell Biol.* 2005; 25:6747–59.
- [23] Cooper G, editor. *Oncogenes*. 2nd ed. London, UK: Jones and Bartlett Learning; 1995.
- [24] Archer MC, Role of Sp Transcription Factors in the Regulation of Cancer Cell Metabolism; *Genes & Cancer* 2011 2: 712-719.

Table 1: RefSeq accession ids of *M. musculus* and *H. sapiens* HDAC gene members.

Gene Name	Mus musculus RefSeq gene id	Homo sapiens RefSeq gene id
HDAC1	NM_008228	NM_004964
HDAC10	NM_199198	NM_032019
HDAC11	NM_144919	NM_024827

HDAC2	NM_008229	NM_001527
HDAC3	NM_010411	NM_003883
HDAC4	NM_207225	NM_006037
HDAC5	NM_010412	NM_005474
HDAC6	NM_010413	NM_006044
HDAC7	NM_019572	NM_015401
HDAC8	NM_027382	NM_018486
HDAC9	NM_024124	NM_178425

Table2: Predicted Cis acting elements, corresponding transcription factors and corresponding start and end position of cis acting elements in *H. sapience*

Gene name	Putative Cis-regulatory Elements	start Position	End Position	Corresponding Transcription factors
HDAC1	AGATAG	849	854	GATA-3
	CGGAAG	661	666	SPI-1
	ACCATC	234	239	Yin-Yang
	TATACATA	6	13	FREAC-7
	GGATA	-111	-107	GATA-2
	GCCATC	-524	-519	Yin-Yang
HDAC2	GGGAAG	487	492	SPI-1
	TGGGGA	432	437	MZF_1-4
	GGGAAG	430	435	SPI-1
	TGGGGA	180	185	MZF_1-4
	GGATA	32	36	GATA-2
	AGAGGAA	-86	-80	SPI-B
	CGGAAG	-597	-592	SPI-1
CACGTGG	-639	-633	USF	
HDAC3	GCCCCGGGC	550	559	AP2alpha
	GGATA	532	536	GATA-2
	GGGAAG	503	458	SPI-1
	GGGAAG	144	149	SPI-1
	GGATA	-336	-332	GATA-2
	GGATA	-626	-621	GATA-2
	GGATA	-821	-817	GATA-2
GGATA	-860	-855	GATA-2	
HDAC4	ACCATC	787	792	Yin-Yang
	AGAGGAA	376	382	SPI-B
	ATTGTGGTT	-232	-224	AML-1
	AGATAG	-513	-508	GATA-3
	CACGTGG	-527	-521	USF
HDAC5	TGGGGA	413	418	MZF1-4
	GGGAAG:	342	247	SPI-1
	CTATTTATAG	-1	8	MEF2
	GGGAAG	-71	-66	SPI-1
	ACCATC	-464	-459	Yin-Yang
	ACCATC	-517	-512	Yin-Yang
	AGATAG	-653	-648	GATA-3
GGGAAG	-947	-942	SPI-1	
HDAC6	TGGGGA	430	435	MZF_1-4
	AGATAG	394	399	GATA-3
	ACCATC	-244	-239	Yin-Yang
	TGGGGA	-705	-700	MZF_1-4
HDAC7	GCCATC	999	1009	Yin-Yang
	TGGGGA	918	923	MZF_1-4

	GGGAAG	563	568	SPI-1
	GGGAAG	537	542	SPI-1
	CGGAAG	503	508	SPI-1
	GTAAACAT	36	43	FREAC-4
	ACCATC	-59	-54	Yin-Yang
	TGGGGA	-117	-112	MZF_1-4
	AGAGGAA	-924	-918	SPI-B
HDAC8	GGATA	565	569	GATA-2
	GGGAAG	463	468	SPI-1
	ACCATC	220	225	Yin-Yang
	AGAGGAA	-650	-644	SPI-B
	GCCATC	-758	-753	Yin-Yang
	GTAAACAAT	-779	-2558	SRY
	GTAAACAA	-779	-772	FREAC-4
HDAC9	AGAGGAA	184	190	SPI-B
	ATTGTGGTT	90	98	AML-1
	AGATAG	-272	-267	GATA-3
	GGATA	-321	-317	GATA-2
HDAC10	GGGAAG	748	753	SPI-1
	GCCATC	397	402	Yin-Yang
	TGGGGA	349	354	MZF_1-4
	GGATA	-136	-132	GATA-2
	GCCCCGGGG	-240	-232	AP2alpha
	TGGGGA	-532	-527	MZF_1-4
	GGGAAG	-554	-549	SPI-1
	GGGAAG	-657	-652	SPI-1
HDAC11	TGGGGA	683	688	MZF_1-4
	GGGAAG	606	611	SPI-1
	TGGGGA	-82	-77	MZF_1-4
	GCCATC	-638	-633	Yin-Yang
	AGATAG	-678	-673	GATA-3
	GGATA	-716	-712	GATA-2
	TGGGGA	-761	-756	MZF_1-4
	GCCATC	-953	-948	Yin-Yang

Table3: Predicted Cis acting elements, corresponding transcription factors and corresponding start and end position of cis acting elements in *M. musculus*.

Gene name	Putative Cis-regulatory Elements	Start Position	End Position	Corresponding Transcription factors
HDAC1	AGATGG	986	991	GATA-1
	AGATGG	533	538	GATA-1
	AATTA	409	413	S8
	AATTA	-257	-253	S8
	AGATGG	-277	-272	GATA-1
	AGATGG	-314	-309	GATA-1
	AATTA	-382	-378	S8
	TGCGTG	-808	-803	Ahr-ARNT
	CGATGG	-953	-948	GATA-1
HDAC2	TTAATTG	986	992	Nkx
	TTAATTG	618	624	Nkx
	AATTA	523	527	S8
	AATTA	106	110	S8
	AATTA	40	44	S8
	AAACAAT	-191	-185	Sox-5
	CACGTG	-696	-691	ARNT
	CACGTG	-696	-691	n-MYC
	AGATGG	-797	-792	GATA-1

HDAC3	AGATGG TGCGTG TGCGTG AGATGG GGATGG	914 387 -159 -692 -888	919 392 -154 -687 -883	GATA-1 Ahr-ARNT Ahr-ARNT GATA-1 GATA-1
HDAC4	AATTA AATTA AATTA AGATGG GGATGG GGVAGTTG	679 422 359 -156 -458 -992	683 426 363 -151 -453 -985	S8 S8 S8 GATA-1 GATA-1 c-MYB_1
HDAC5	AGATGG AGATGG CACGTG CACGTG GGATGG AGATGG AGATGG GGATGG TAACAAT	939 758 692 692 648 409 192 -298 -940	944 762 697 697 653 413 196 -293 -934	GATA-1 S8 ARNT n-MYC GATA-1 S8 S8 GATA-1 Sox-5
HDAC6	TGCGTG TGCGTG AGATGG GGATGG AGATGG CGATGG AGATGG	788 784 443 103 -56 -629 -916	793 789 448 108 -51 -624 -911	Ahr-ARNT Ahr-ARNT GATA-1 GATA-1 GATA-1 GATA-1 GATA-1
HDAC7	AATTA CACGTG AATTA CGATGG CACGTG CACGTG	172 135 94 -148 -518 -518	176 140 98 -143 -513 -513	S8 ARNT S8 GATA-1 ARNT n-MYC
HDAC8	AATTA AATTA AAACAAT AATTA AATTA AATTA AGATGG CGATGG	868 502 85 -279 -441 -501 -635 -930	872 506 91 -275 -437 -497 -630 -925	S8 s8 Sox-5 S8 ss8 S8 GATA-1 GATA-1
HDAC9	AATTA AATTA AATTA AGATGG AATTA AATTA AAACAAT	900 820 318 -197 -266 -756 -904	904 824 322 -192 -262 -752 -898	S8 S8 S8 GATA-1 S8 S8 Sox-5
HDAC10	GGATGG AATTA CACGTG CACGTG	-333 -449 -656 -656	-328 -445 -651 -651	GATA-1 S8 ARNT n-MYC
HDAC11	TAACAAT GGATGG	676 -321	682 -316	Sox-5 GATA-1

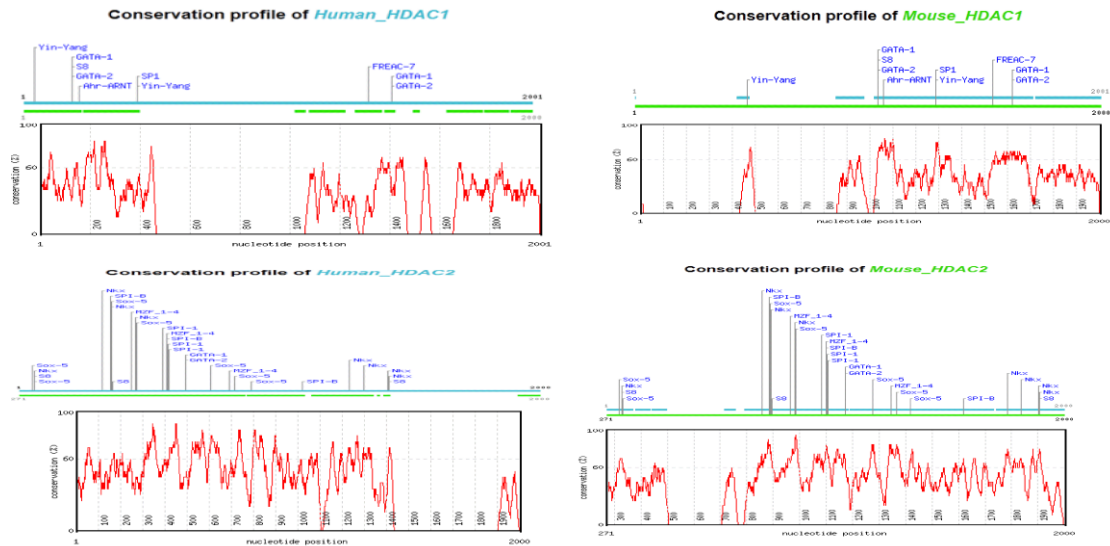
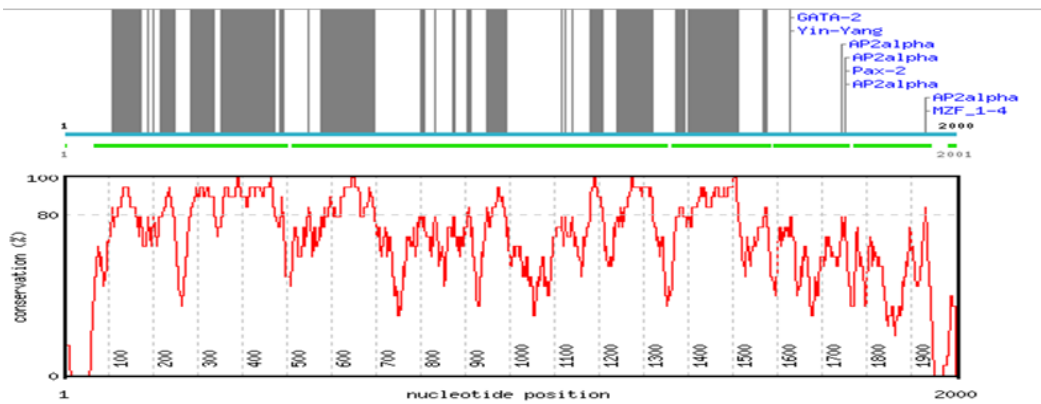


Figure1: Comparative conservation profile of human and mouse HDAC1 and HDAC2. Green and cyan line represent the alignment of up and down stream sequence both human and mouse HDAC1 and HDAC2 respectively. In graph x-axis represent nucleotide positions and y-axis represent conservation cut of predicting cis acting elements.

Human *hdac3*



Mouse *hdac3*

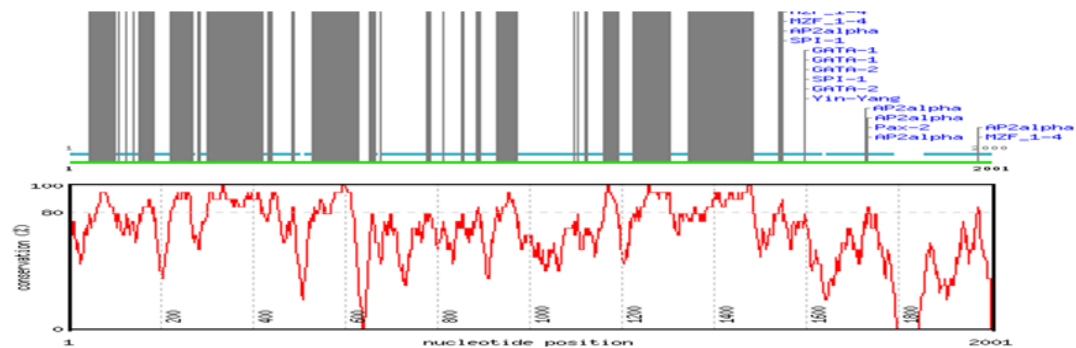


Figure2: Comparative conservation profile of human and mouse HDAC3. Green and cyan line represent the alignment of up and down stream sequence both human and mouse HDAC3. In graph x-axis represent nucleotide positions and y-axis represent conservation cut of predicting cis acting elements.

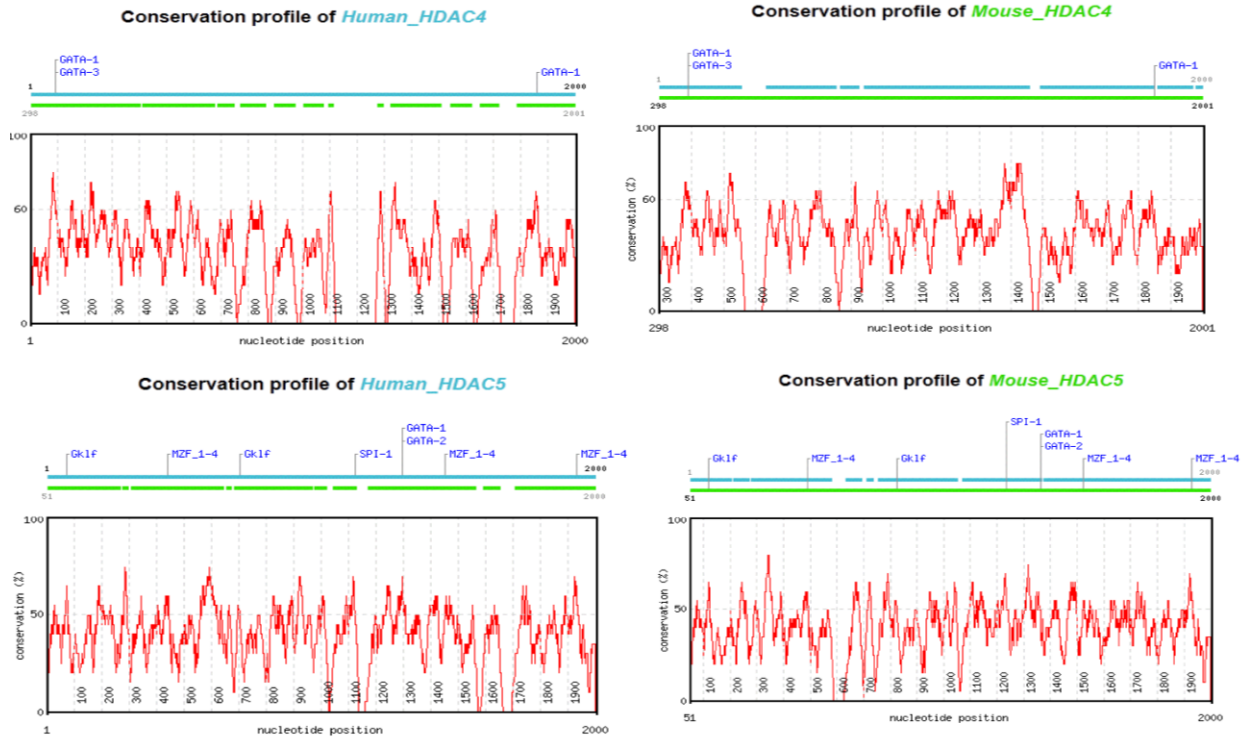


Figure3: Comparative conservation profile of human and mouse HDAC4 and HDAC5. green and cyan line represent the alignment of up and down stream sequence both human and mouse HDAC5 and HDAC5 respectively. In graph x-axis represent nucleotide positions and y-axis represent conservation cut of predicting cis acting elements.

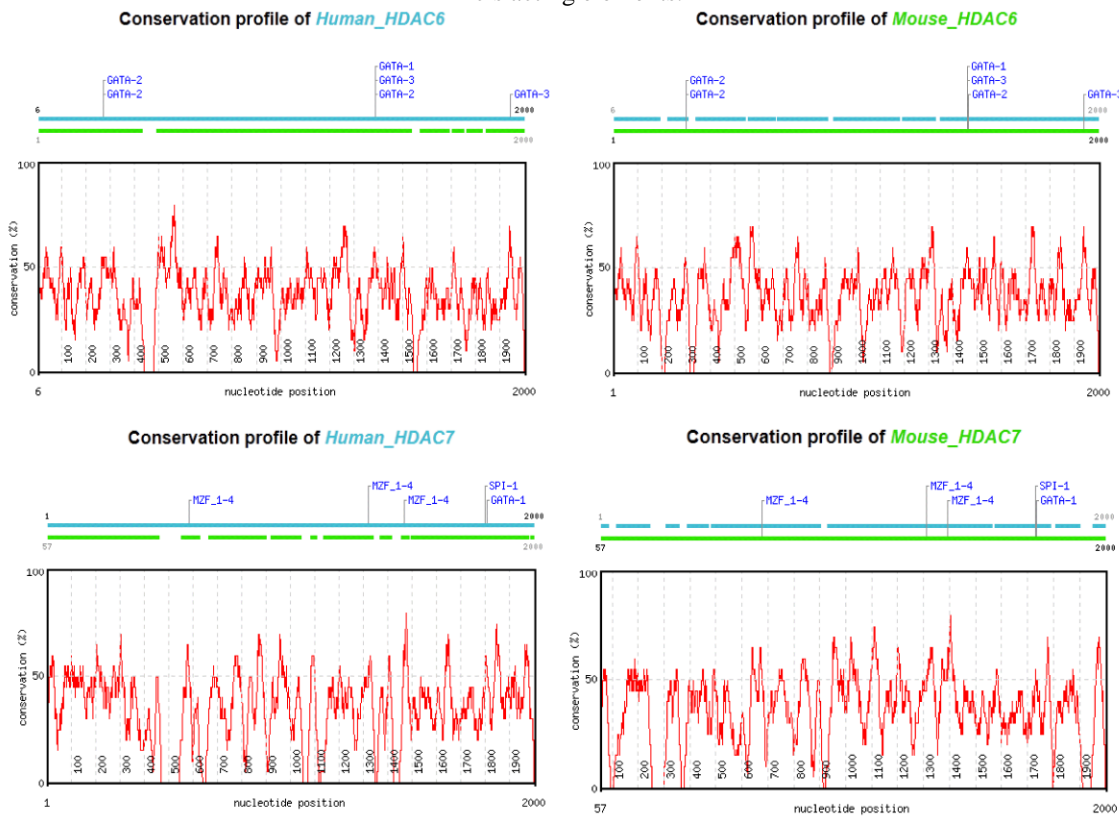


Figure4:



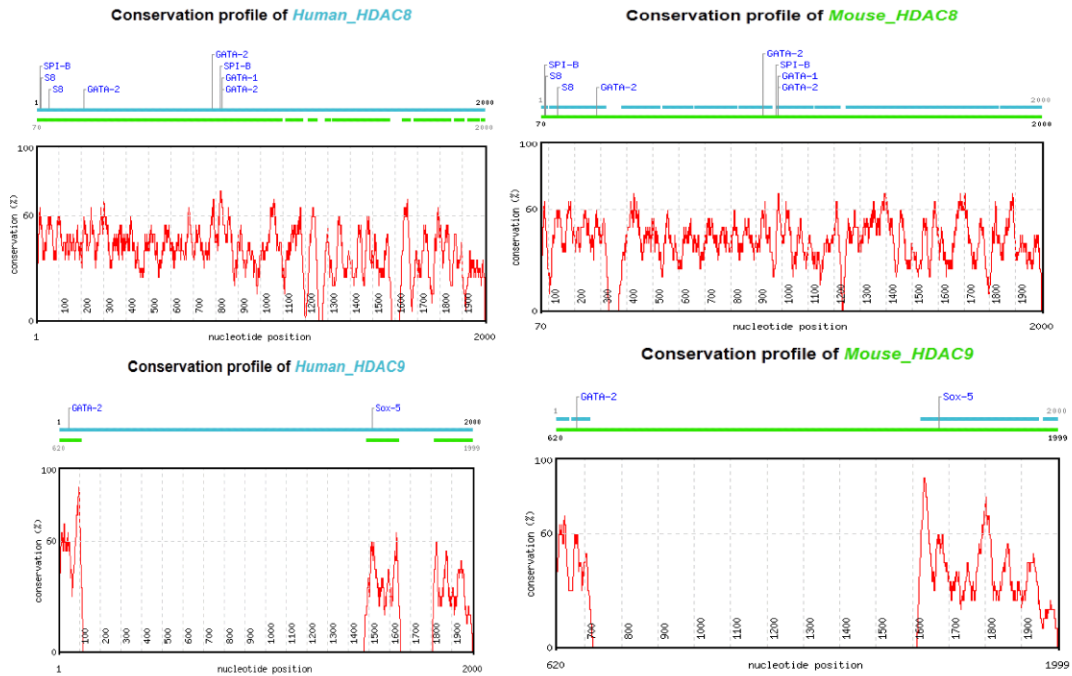


Figure5: Comparative conservation profile of human and mouse HDAC8 and HDAC9. green and cyan line represent the alignment of up and down stream sequence both human and mouse HDAC8 and HDAC9 respectively. In graph x-axis represent nucleotide positions and y-axis represent conservation cut of predicting cis acting elements.

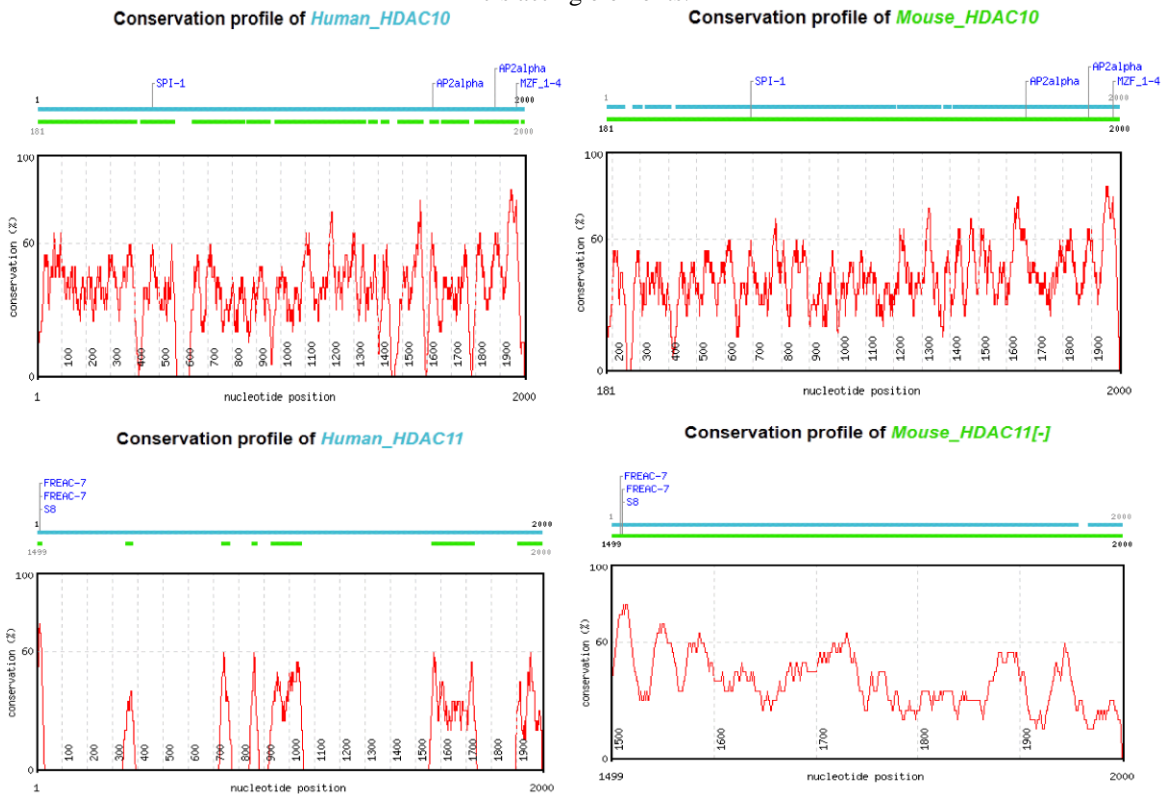


Figure6: Comparative conservation profile of human and mouse HDAC10 and HDAC11. green and cyan line represent the alignment of up and down stream sequence both human and mouse HDAC10 and HDAC11 respectively. In graph x-axis represent nucleotide positions and y-axis represent conservation cut of predicting cis acting elements.

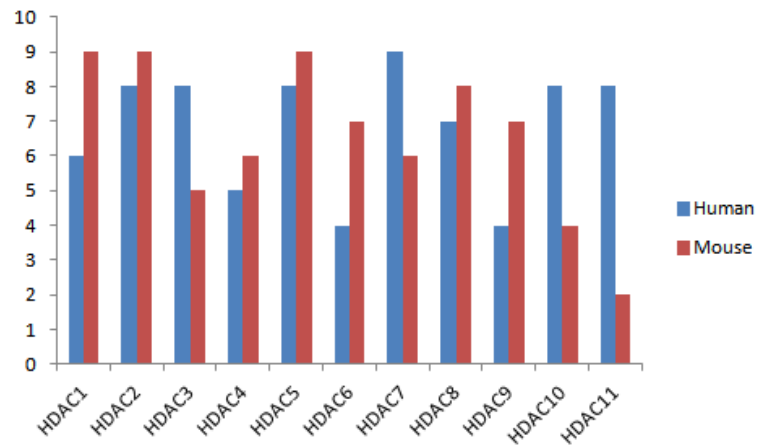


Figure:7 Distribution of cis acting elements in human and mouse HDAC gene family.