

RESEARCH ARTICLE

Open Access

Promoter conservation in HDACs points to functional implications



Toni A. Boltz^{1,6}, Sawsan Khuri² and Stefan Wuchty^{1,3,4,5*}

Abstract

Background: Histone deacetylases (HDACs) are the proteins responsible for removing the acetyl group from lysine residues of core histones in chromosomes, a crucial component of gene regulation. Eleven known HDACs exist in humans and most other vertebrates. While the basic function of HDACs has been well characterized and new discoveries are still being made, the transcriptional regulation of their corresponding genes is still poorly understood.

Results: Here, we conducted a computational analysis of the eleven HDAC promoter sequences in 25 vertebrate species to determine whether transcription factor binding sites (TFBSs) are conserved in HDAC evolution, and if so, whether they provide useful information about HDAC expression and function. Furthermore, we used tissue-specific information of transcription factors to investigate the potential expression patterns of HDACs in different human tissues based on their transcription factor binding sites. We found that the TFBS profiles of most of the HDACs were well conserved in closely related species for all HDAC promoters except HDAC7 and HDAC10. HDAC5 had particularly strong conservation across over half of the species studied, with nearly identical profiles in the primate species. Our comparisons of TFBSs with the tissue specific gene expression profiles of their corresponding TFs showed that most HDACs had the ability to be ubiquitously expressed. A few HDAC promoters exhibited the potential for preferential expression in certain tissues, most notably HDAC11 in gall bladder, while HDAC9 seemed to have less propensity for expression in the nervous system.

Conclusions: In general, we found evolutionary conservation in HDAC promoters that seems to be more prominent for the ubiquitously expressed HDACs. In turn, when conservation did not follow usual phylogeny, human TFBS patterns indicated possible functional relevance. While we found that HDACs appear to uniformly expressed, we confirm that the functional differences in HDACs may be less a matter of location of activity than a question of which proteins and which acetyl groups they may be acting on.

Keywords: HDAC, Promoter analysis, TFBS, Tissue specificity

Background

Histone deacetylases (HDACs) remove the acetyl group from lysine residues of the N-terminal tail of core histones, allowing the repression of transcription. These metal binding proteins are mostly active in large multiprotein complexes, and can also act on non-histone proteins. Human histone deacetylases require zinc, and have been grouped into different classes based on their sequence similarity to homologues they have evolved from in yeast. Class I HDACs (1, 2,

3, 8) are most similar to yeast RPD3 protein, while Class II HDACs (4, 5, 6, 7, 9, 10) are homologues of yeast HDA1 [1]. HDAC11 forms Class IV on its own, sharing features from both Class I and Class II enzymes [1], while the sirtuin enzymes, which require NAD⁺ for catalysis and were formerly categorized as Class III HDACs, have evolved independently.

The HDAC proteins have been well characterized, and the position of their active site(s), their genomic position and cellular localization are well established (Table 1). Their modes of action have been investigated extensively over the last two decades, with particular emphasis on HDAC inhibitors as possible drugs for use in cancer therapy [4, 5]. However, the high level of similarity

* Correspondence: wuchty@cs.miami.edu

¹Department of Computer Science, University of Miami, Coral Gables, FL, USA

³Department of Biology, University of Miami, Coral Gables, FL, USA

Full list of author information is available at the end of the article



Table 1 Genomic location, amino acid length of the main isoform, number of biochemically characterized active sites, and cellular localization of human histone deacetylases

	Chm	Protein length of main isoform (aa)	Number of active sites	Present in nucleus	Present in cytoplasm	
Class I	HDAC1	1p34.1	482	1	P	N (If present, results in axonal damage [2])
	HDAC2	6q21	488	1	P	N
	HDAC3	5q31.2	428	1	P	P
	HDAC8	Xq13	377	1	P	P [3]
Class IIA	HDAC4	2q37.2	1084	1	P	P
	HDAC5	17q21	1122	1	P	P
	HDAC7	12q13.1	952	1	P	P
	HDAC9	7p21.1	1011	1	P	P
Class IIB	HDAC6	Xp11.23	1215	2	P	P
	HDAC10	22q13.31	669	1	P	P
Class IV	HDAC11	3p25.1	347	1	P	P

Data from the NCBI Gene database, and localization data from [1], except where shown. P present, N Not present

between the HDACs and the seemingly interchangeable nature of their activity makes them a complex family of proteins that has proven difficult to fully decipher [6].

Early studies on HDAC evolution found evidence of an ancient family of proteins with de-acetylase activity [7]. At the time, research focused on phylogenetic studies of protein sequences for the characterization of vertebrate HDAC active domains and localization signals to infer functional overlap and clues of alternative functions [8]. It was quickly ascertained that HDAC1 and HDAC2 are closely related and work in concert most of the time, and more recent work confirms that one is not a direct substitute for the other [9]. HDAC3 is equally widely expressed, interacts with Class II HDACs, and affects a wide range of cellular processes [10–12].

HDAC8, while also considered a Class I HDAC, seems to have evolved from a separate, equally ancient lineage that works on multiple substrates and is involved in several pathways [13, 14]. This histone deacetylase has a particular structure/function conformation that is different from its human homologues [15] and a propensity for de-fatty-acetylation [16]. Such findings suggest that some of the functional differences among HDACs may be linked to the nature of the acetyl compounds that they remove from their substrate proteins, which has profound implications on how we view and investigate this enzyme family.

Traditionally considered as recruiters for Class I HDACs due to their low catalytic activity when compared to other histone deacetylases [17], Class II HDACs 4, 5, 7, and 9 are now known to be active in their own right, playing a central role in regulating gene expression relating to muscle development, tissue differentiation and other pathways [18]. Extensive research has shown that HDAC4 is involved in a myriad of roles [19, 20],

while HDAC5 is increasingly implicated in axon regeneration [21] and in cardiovascular contexts [6]. HDAC7 seems to play an important role for bone development [22] and in diabetes [23], while a flurry of recent articles have similarly associated HDAC9 with several disease pathways including various cancers and stroke [24–26].

HDACs 6 and 10 seem to have an interesting relationship and are often classified separately as Class IIB, as both have two highly similar catalytic domains, although the second domain of HDAC10 is considered inactive [1]. Comparative sequence analysis indicated that HDAC10 and HDAC6 may have shared a common ancestor at some point in vertebrate evolution [7], and ideas as to how both evolved separate functions are beginning to emerge. HDAC6 was the first histone deacetylase that was shown to work on a non-histone protein, tubulin, and is predominantly cytosolic [27], while recent work indicates that HDAC10 acts as a polyamine deacetylase [28]. Like other HDACs, both seem to be active in a variety of developmental and pathological contexts.

The only Class IV member, HDAC11, is arguably also the least well understood of the HDAC family. Recent reviews focus on its role in the immune system [29, 30], to the exclusion of other roles it may play that have not yet been discovered.

Given their vital and extensive roles in the regulation of gene expression and protein activity in eukaryotic genomes both in and out of the nucleus, relatively little is known about the regulation of HDAC expression. We expect that the transcription of HDACs does not differ markedly from other genes whose promoters are regulated by histone phosphorylation and acetylation [31], which they, as *the* histone deacetylases, are necessarily involved in [32].

Regulation of HDACs in cancer cells by the ubiquitous transcription factors Sp1 and Sp3 has been well investigated [33], and their expression profile has been studied in some disease cases [34, 35]. Furthermore, HDACs are subjected to the same array of post-translational modification as other proteins [36]. Increasing evidence is being accumulated about HDAC roles in development, housekeeping functions, and disease onset and progression, none the least in cancers. In plants, histone deacetylases have been shown to act on a wide array of molecules, including *N*-acetyleserotonin [37].

Despite our expanding knowledge, there seems to be an urgent need to elucidate their separate functions and intersections of function, and to better understand how their own expression is regulated. Turning to computational methods as potential guides to bench experimentation, we conducted an in-depth analysis of HDAC promoter sequences with two questions in mind: Are transcription factor binding sites (TFBSs) conserved in HDAC evolution, and if so, do they provide useful information about HDAC transcriptional regulation and HDAC function?

These questions were fueled by recent literature on the slow evolution of TFBSs [38, 39] and their potential use in highlighting gene expression patterns (reviewed in [40]). Given that there is no gold standard to assess methods for TF analysis [41], and as divergent as promoter sequences can be among closely related species (e.g. [42]) and among the promoter regions of closely related genes (e.g. [43]), there seems to be enough signal in them to imply functional relevance [39] which can then be confirmed by experimental data. Given the functional overlap that HDACs seem to have, and increasing evidence of their ubiquity in the human system, we were curious whether there were any signals in their promoters which could help deepen our understanding of this enzyme family.

Results

Evolutionary conservation of TFBSs in HDAC promoters

We found that human TFBS patterns in HDAC promoters are evolutionarily conserved across all HDACs, with only HDACs 5, 7 and 10 indicating unusual patterns of TFBS distribution along the promoter region. In Fig. 1 we present the HDAC1 promoter alignment as an example of the Genomatix output, showing promoter sequences aligned according to the quantitative phylogenetic distances between their TFBS patterns. Here, the TFBS patterns appeared in the predictable evolutionary groupings, with apes (*H. sapiens*, *G. gorilla* and *P. troglodytes*) and rodents (*R. norvegicus* and *M. musculus*) forming clades. We observed similar patterns in HDACs 2, 3, 4, 6, 8,

9 and in HDAC11 (Additional file 2: Figure S1, Additional file 3: Figure S2, Additional file 4: Figure S3, Additional file 5: Figure S4, Additional file 6: Figure S5, Additional file 7: Figure S6 and Additional file 8: Figure S7). HDACs 8 and 9 had evolutionarily conserved TFBS patterns in closely related organisms (Additional file 6: Figure S5, Additional file 7: Figure S6). Notably, a HDAC8 equivalent was absent in *P. troglodytes* while an HDAC9 equivalent was missing from *G. gorilla*.

The promoter of HDAC5 was conserved for the greatest number of species (Fig. 2), and had two fish species (*P. reticulata* and *D. rerio*) clustering close to two old world monkeys, while promoters of HDAC5 homologues from rat and mouse occupied different ends of the dendrogram. Promoter sequences of HDACs 7 and 10 also showed transcription factor binding site patterns where the classic phylogenetic lineages did not hold true. The TFBS profile of human HDAC7 (Fig. 3) appeared to be most similar to that of pig *S. scrofa*, green monkey *C. sabaeus* and rat *R. norvegicus*, while the TFBSs from other primate promoters appeared to follow different patterns. The promoter of human HDAC10 (Fig. 4) was most similar to that of rabbit, *O. cuniculus*, while the promoter sequences of other primate HDAC10s were more similar to that from horse, *E. caballus*.

TFBS patterns provide useful information on HDAC regulation and function

We analyzed the large amount of gene expression data available from mRNA studies [44] and found that most HDACs seemed to be able to be expressed in most tissues, albeit showing higher expression levels in some tissues than in others. For the most part, these expression profiles were based on experiments that were not targeting HDAC function per se, suggesting that drawing patterns of tissue specificity from these results may be challenging.

Using the previously gathered promoter motif data in a new context, we considered transcription factor binding site trimers that were present on HDAC promoters to denote expression in a given tissue. In particular, we used a non-exclusive set defined as TFs expressed in most tissues, and a preferentially expressed set of TFs that were more highly expressed in the given tissues when compared to other tissues found in Genomatix (as described in [Methods](#)). In particular, a trimer was only considered if all 3 units - a unit being the transcription factor whose binding site is present on the relevant HDAC promoter - were expressed within the given tissue, pointing to evidence that the corresponding HDAC is expressed in the underlying tissue. To help identify overarching patterns, we collapsed the 59 non-disease tissue specificity designations available into the 11

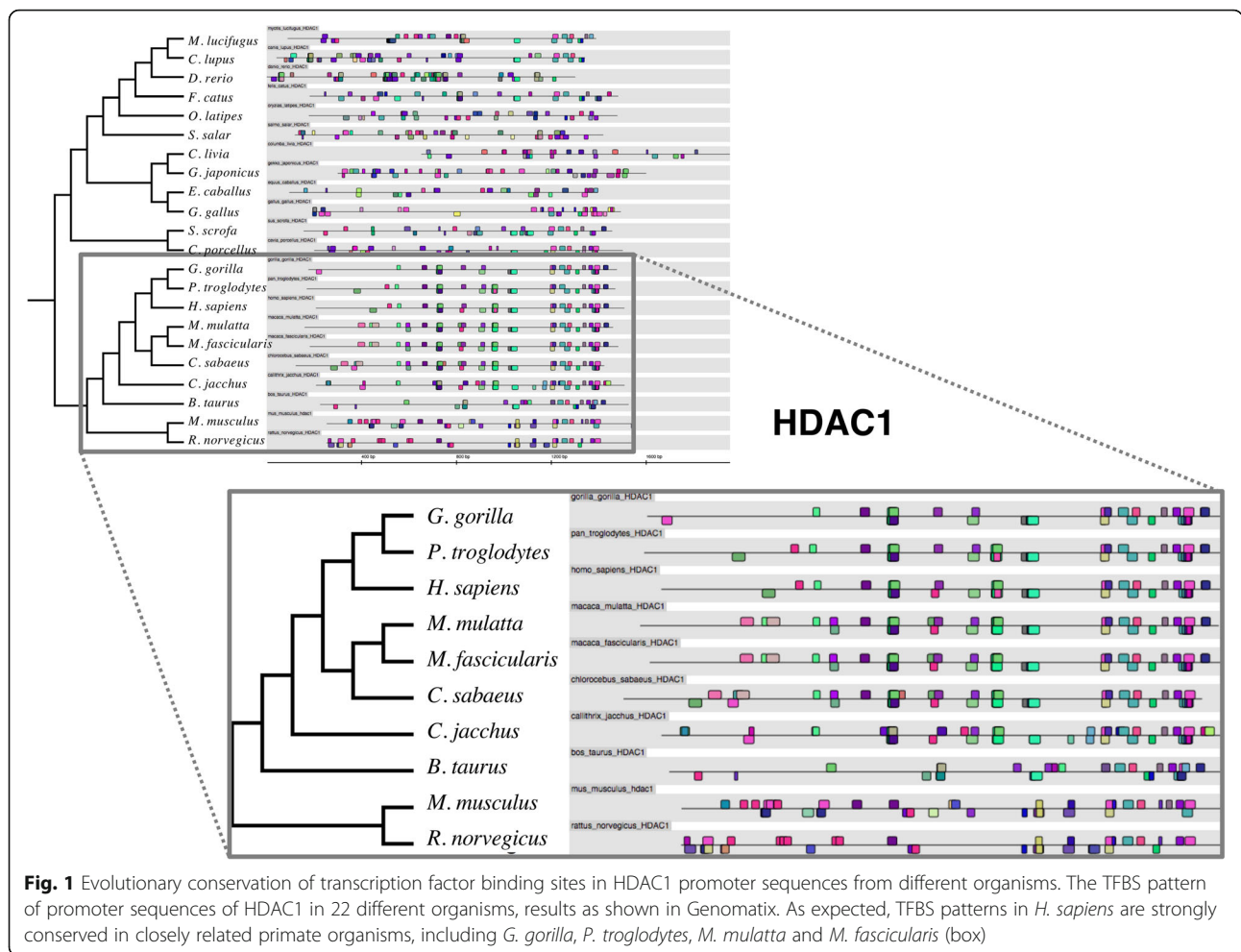


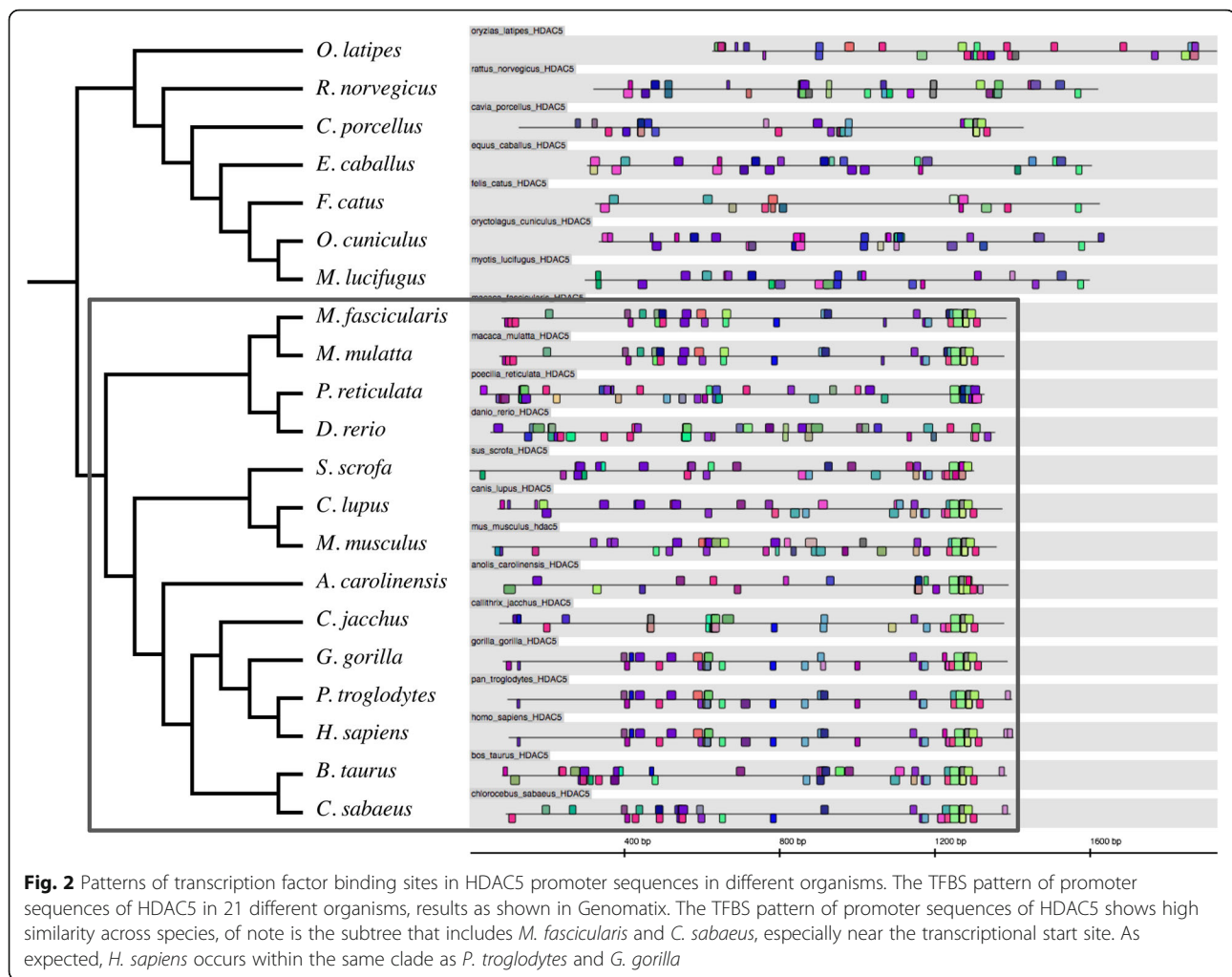
Fig. 1 Evolutionary conservation of transcription factor binding sites in HDAC1 promoter sequences from different organisms. The TFBS pattern of promoter sequences of HDAC1 in 22 different organisms, results as shown in Genomatix. As expected, TFBS patterns in *H. sapiens* are strongly conserved in closely related primate organisms, including *G. gorilla*, *P. troglodytes*, *M. mulatta* and *M. fascicularis* (box)

human biological systems (Additional file 1: Table S1), namely cardiovascular/hematopoietic, digestive, endocrine, excretory, immune/lymphatic, integumentary, muscular, nervous, respiratory, reproductive, and skeletal systems, with an additional designation for embryonic expression. In Table 2, we list the biological systems where we found at least one TFBS trimer that appeared in a human HDAC promoter sequence, and compared these results to previously reported tissue specificity of HDACs. Since we only used human data in this analysis, the absence of a TFBS trimer on an HDAC promoter could indicate that this HDAC is not expressed in the corresponding tissue, indicated in Table 2 using an “All Except” annotation.

Class I HDACs 1, 2, and 3 are particularly well-studied and known to be ubiquitously expressed [9–12], validating our approach. HDAC8, known to be active on multiple substrates [15] and many different proteins [48] is only reported to be highly expressed in smooth muscle, while our results suggest that HDAC8 is another ubiquitous HDAC (Table 2). In fact, when non-exclusive TFs were included in this

trimer analysis, almost all of the HDACs had fairly widespread tissue representation, with the exception of HDACs 7 and 11, which had fewer tissues represented in the results, suggesting a narrower expression range. When only preferentially expressed TFs were considered, HDAC3 had TFBS trimers from only 3 major systems such as cardiovascular/hematopoietic, immune/lymphatic, and muscular. HDACs 9 and 10 also had these same three major systems represented as TFBS trimers in their promoters, in addition to the embryonic system in HDAC9 and the endocrine system in HDAC10 promoters. The promoter sequence of HDAC11 had no trimers present when the preferential expression filter was applied, suggesting that its expression is governed only through non-exclusive TFs.

This analysis of trimers also indicated that there were higher instances of TFs across the HDACs that were expressed in the nervous, immune and endocrine systems. However, this observation may reflect the fact that there were disproportionately more TFs listed under these systems in the Genomatix annotation than for other systems,

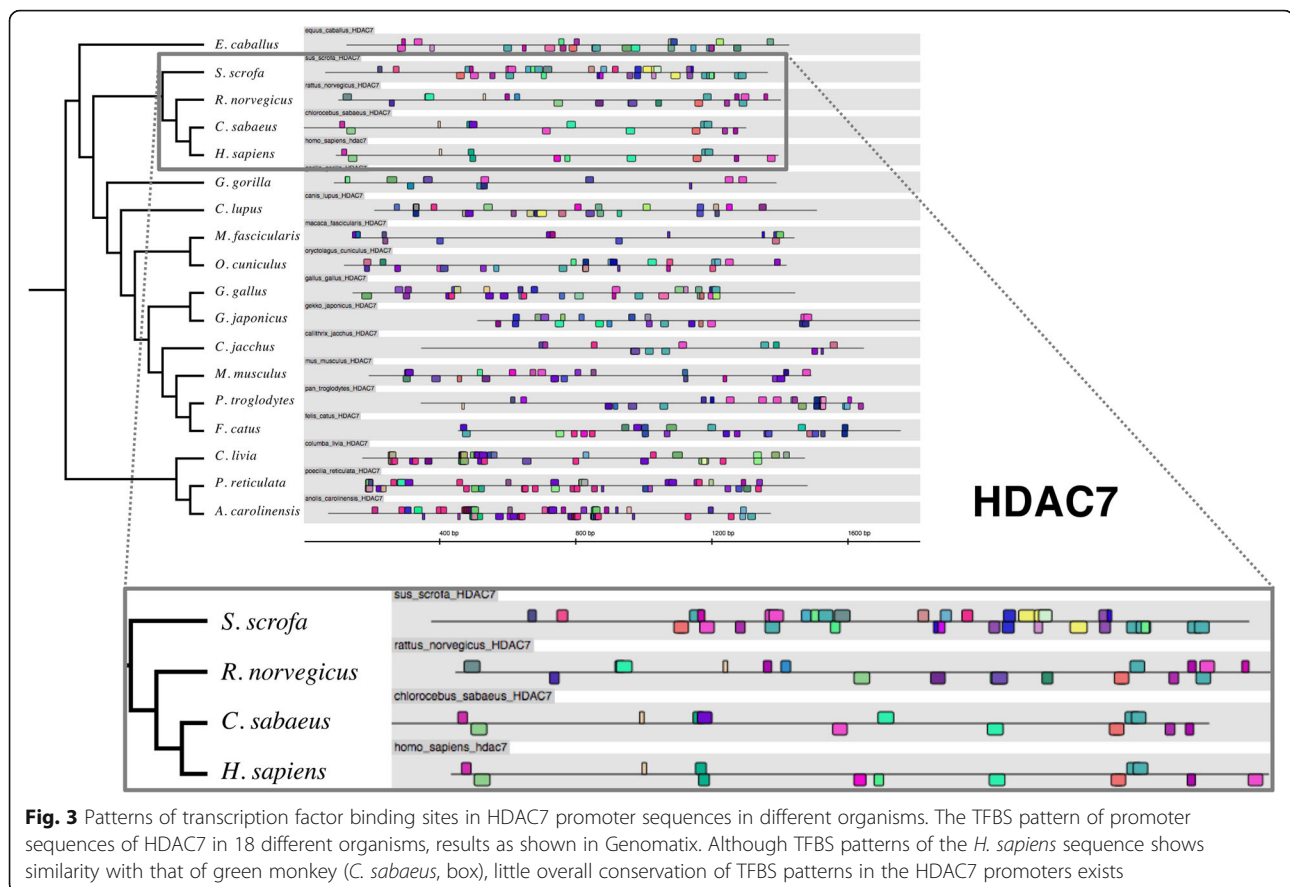


such as respiratory or integumentary. Therefore, we fine-tuned our approach, treating TFBSs individually and determining the observed numbers of binding sites of TFs that appear in HDAC promoter sequence and are expressed in specific tissues, and used a \log_2 -fold change to assess the significance of our findings (see [Methods](#)) when we compared observed to expected numbers. In the heatmap in Fig. 5a where we considered a non-exclusive set of TFs that were expressed in most tissues, we observed that HDACs 1,4,5,6 and 8 appear to be largely associated to the majority of 59 tissues. In turn, HDACs 2,3, 9,10 and 11 seem to be less likely expressed in most tissues. In particular, HDAC7 seems to be associated with expression in blood cells, and less in the embryonic system, while TFBSs for TFs associated with the nervous system seem to be underrepresented in the promoter of HDAC9. Furthermore, HDACs 6 and 8 appear strongly over-represented in thyroid gland, lung and cartilage. HDAC10 had a low score for TFBSs associated with muscle tissue, while HDAC11 had a high score for TFBSs associated with the gall bladder.

As for the heatmap in Fig. 5b, where we considered preferentially expressed set of TFs that were more highly expressed in given than other tissues we observed a high degree of paucity of binding sites on HDAC promoters. Specifically, we found enrichment for TFBSs associated with the nervous system in the HDAC3 promoter, and for those associated specifically with neuralgia and smooth muscle in the promoter of HDAC7. Similar to the trimer results, we observed fewer TFs associated with the nervous system on the promoter of HDAC9, while the high specificity for expression in gall bladder remained highlighted in the promoter region for HDAC11.

Discussion

In this work, we were curious whether there was any evolutionary conservation in HDAC promoter sequences, and if so what it can tell us about the transcriptional regulation and function of the eleven human HDACs. Our results confirmed that in general, there was evolutionary conservation in HDAC promoters, and in cases where this



conservation did not align with currently accepted phylogeny, the pattern of TFBS arrangement on human promoters showed some similarity with different species, indicating a possible functional relevance. Unusual patterns in genetic sequence phylogenies suggest dynamic and relatively recent changes in the evolution of such sequences, implying evolutionarily recent patterns in the way the corresponding HDACs are regulated [49]. In fact, promoter conservation among vertebrate species seems to be more prominent for the ubiquitously expressed HDACs, particularly for HDACs 1 and 2, suggesting that these have not undergone recent evolution, a hypothesis in line with literature on the evolution of so-called “house-keeping” and “essential” genes [50]. Those HDACs that exhibit an unusual pattern of TFBSs on their promoters seem to also have a propensity for expression in fewer tissues such as seen in our results of HDAC5 and their possible preferential association with the cardiovascular/hematopoietic, muscular, nervous and endocrine tissues (Table 2). The exceptional case in our results is HDAC11 that has a conserved promoter region which followed conventional species phylogenies. Yet, HDAC11 also exhibited a possible preferential association with expression in the gall bladder, a conclusion we recommend to be followed up with laboratory experimentation.

Overall, our results imply that most HDACs are able to be ubiquitously expressed. For example, our results of the HDAC8 promoter region concur with studies into the evolution of HDAC catalytic domains, highlighting the relatively recent functional evolution of HDAC8 [8]. In turn, these are validated by recent discoveries about HDAC8 function [15, 16, 51] which further indicate that the differences between the HDACs may not be one of location of expression as much as a structure/function difference in their catalytic process.

Early phylogenetic studies using HDAC protein sequences did not report differences between the catalytic domain of HDAC7 and the other Class II HDACs [8]. Our findings showed that there was little evolutionary conservation in the promoter sequence of HDAC7, and that it has a broad tissue specificity spanning most biological systems but perhaps not as strongly in embryonic tissues. Recent molecular investigations place HDAC7 in the endocrine and skeletal systems [23] as well as in the brain playing a key role in memory formation [52]. Like HDAC8, the regulation of HDAC7 seems to have evolved differently to its functional relevance, making it particularly interesting for further experimental investigations similar to those that have taken place in HDAC8 as detailed above.

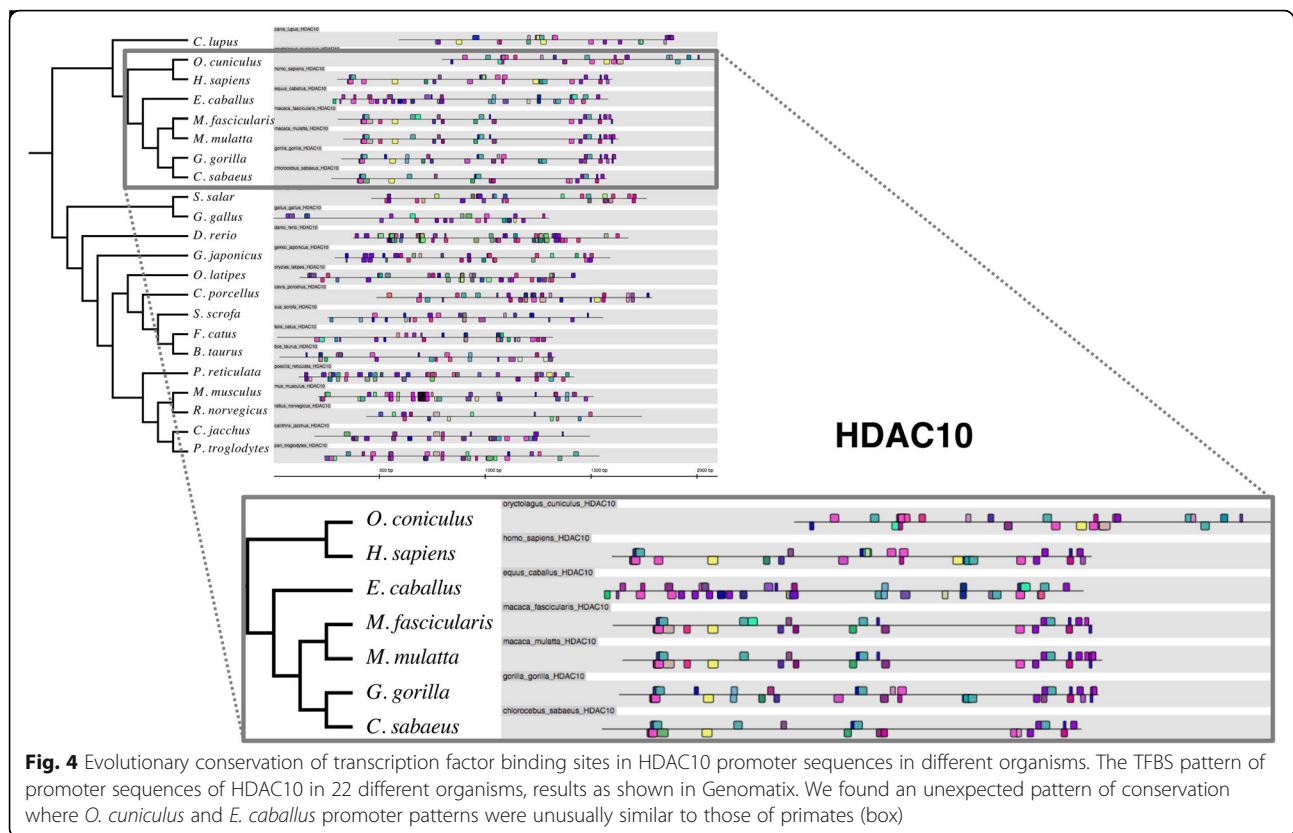


Fig. 4 Evolutionary conservation of transcription factor binding sites in HDAC10 promoter sequences in different organisms. The TFBS pattern of promoter sequences of HDAC10 in 22 different organisms, results as shown in Genomatix. We found an unexpected pattern of conservation where *O. coniculus* and *E. caballus* promoter patterns were unusually similar to those of primates (box)

Conclusions

Previous studies have shown that quantitative differences in transcription factor binding are observable even in closely related species, yet only a weak correlation is found between binding variation and regulatory function [53]. Furthermore, TFBSs have a high evolutionary turnover rate such that even closely related species may not have conserved binding sites on their promoters [54]. This may explain some of the differences we observed in TFBS pattern along the promoters of HDACs from different species. Given that low evolutionary conservation at the promoter level may not have a significant effect on gene expression [54, 55], we posit that our two-pronged approach points to new avenues for studying the regulation of HDAC expression. Since the HDACs themselves are heavily involved in gene expression, further studies into the transcriptomic levels of these genes may prove useful to compare to TFBS patterns reported here. This will allow us to infer how sensitive the regulation of regulatory proteins is with regards to differences at both the binding site and transcriptomic level.

Exploring the biochemical role of each of the HDAC homologues in the different species we tested would shed further light as to how these proteins have evolved, and why. Early thoughts about why there are eleven HDACs in the human system had focused on tissue or time specificity, and there is enough information about their

function now to add several layers of complexity on this question. Our results suggest that all HDACs are ubiquitously expressed, and that the differences between them rest in which acetyl group they remove from a protein, and which proteins they act on, instead of where they act. Given their role in gene expression and the impact that dysregulation of HDACs can have on the health of an organism, it is crucial that a comprehensive analysis of the biochemical roles and transcriptional regulation of these enzymes is performed, so that better targeted therapeutics can be identified. With HDAC inhibitors gaining traction in the treatment of various cancers and other diseases, the ability to fully understand their regulation and function, including through experimental promoter validation, remains a crucial research priority.

Methods

Transcription factor binding sites

To determine TFBSs in the promoters of human HDAC genes, we considered promoter sequences starting from 1,200 base pairs upstream to 100 base pairs downstream of the transcriptional start site (TSS) as designated in the National Center for Biotechnology Information's (NCBI) Nucleotide database. Research in yeast found that conservation of TFBSs is highest within 200 bp upstream of the TSS [42]. Furthermore, there is evidence of multiple TSSs

Table 2 Tissue and system specificity of histone deacetylases. Reported data is from cited references [5, 17, 45–47]. Predicted data is based on enrichment of TFBS trimers in the analysis described in the main text. Following TF groupings in Genomatix, *non-exclusive* refers to the presence of TFs that are active in most tissues, and *preferential* refers to TFs that are more highly expressed in these systems than in others. Dark bullet points denote predicted activity, outline bullet points denote a predicted lack of activity in these systems

		Reported tissue specificity	Predicted non-exclusive system specificity	Predicted preferential system specificity
Class I	HDAC1	Ubiquitous	All	All
	HDAC2	Ubiquitous	All Except o Respiratory	• Cardiovascular/Hematopoietic • Embryo • Endocrine • Immune/Lymphatic • Muscular • Nervous
	HDAC3	Ubiquitous	All Except o Integumentary o Respiratory	• Cardiovascular/Hematopoietic • Immune/Lymphatic • Muscular
	HDAC8	• Muscular (smooth muscle)	All	All
Class IIA	HDAC4	• Cardiovascular/Hematopoietic • Muscular (smooth muscle) • Nervous (brain) • Endocrine (liver)	All	All Except o Respiratory
	HDAC5	• Cardiovascular/Hematopoietic • Muscular (smooth muscle) • Nervous (brain) • Endocrine (liver)	All Except o Respiratory	All Except o Integumentary o Respiratory o Skeletal
	HDAC7	• Cardiovascular/Hematopoietic • Muscular (smooth muscle) • Endocrine (several tissues) • Skeletal	• Cardiovascular/Hematopoietic • Digestive • Endocrine • Excretory • Immune/Lymphatic • Muscle • Nervous	• Cardiovascular/Hematopoietic • Digestive • Endocrine • Excretory • Muscle • Nervous
	HDAC9	• Immune/Lymphatic • Muscular (smooth muscle) • Nervous (brain)	All Except o Digestive o Excretory o Integumentary o Respiratory	• Cardiovascular/Hematopoietic • Embryo • Immune/Lymphatic • Muscular
Class IIB	HDAC6	• Cardiovascular/Hematopoietic • Endocrine (liver, pancreas) • Nervous (brain)	All Except o Respiratory	All Except o Embryo o Respiratory
	HDAC10	• Endocrine (kidney, liver, spleen) • Excretory (kidney)	All Except o Integumentary o Respiratory o Skeletal	• Cardiovascular/Hematopoietic • Endocrine • Immune/Lymphatic • Muscular
Class IV	HDAC11	• Cardiovascular/Hematopoietic • Endocrine (kidney) • Excretory (kidney) • Muscular (smooth muscle) • Nervous (brain)	• Cardiovascular/Hematopoietic • Digestive • Endocrine • Excretory • Embryo • Immune/Lymphatic • Nervous • Reproductive	Not enough TF trimers for a signal

and alternative promoters per HDAC according to the Database for Transcriptional Start Sites (DBTSS) [56]. We therefore considered a range of 1,200 bp upstream of the NCBI's TSS to capture as many TFBS signals as might be present from possible alternative TSSs.

We extracted known and annotated HDAC sequences in 25 different species from the NCBI's Nucleotide

database. To check their similarity and possible kinship we aligned organism-specific HDACs with their human counterparts. In particular, we established 11 HDAC groups of sequences that are annotated according to their corresponding human HDAC. Although we only considered known organism-specific HDACs, we reviewed their annotations and assigned a given organism-specific HDAC to the corresponding group, if the corresponding

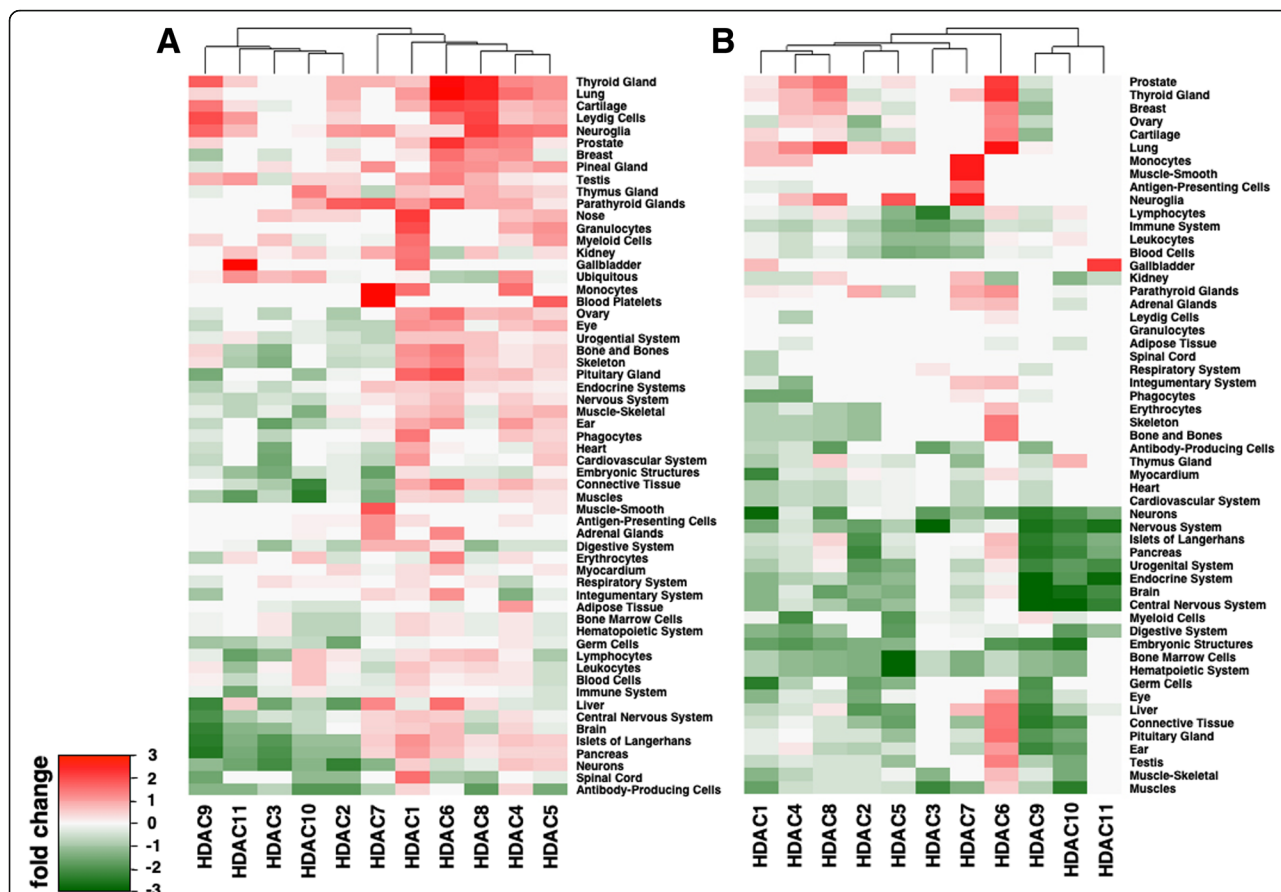


Fig. 5 Enrichment of tissue type TFBSs in each HDAC promoter. Showing the log-fold change of observed and expected TFBS frequencies in 59 different tissues we assessed the prevalence of TFBSs on HDAC promoters that correspond to TFs that are expressed in a tissue specific manner. In particular, we consider the enrichment/dilution of TFBSs through a (a) non-exclusive set defined as TFs which were expressed in most tissues, and (b) a preferentially expressed set of TFs that were more highly expressed in the given tissues when compared to other tissues

sequence was most similar to the underlying human HDAC. We then extracted their promoter sequences for our analysis using the Genomatix software suite (www.genomatix.com), as its transcription factor database has a taxonomically relevant classification system that was applicable to all the considered species. The computational detection of TFBS motifs was based on scanning these promoter sequences through position weight matrices of corresponding transcription factors with MatInspector as implemented in the Genomatix software suite [57, 58], that was also used to visualize TFBSs on the promoter sequences. We set a core similarity of 0.75 (maximum is 1.0) and a matrix similarity of the optimized value + 0.10 to find TFBSs. We used transcription factor motifs from transcription factor families that were found in either all species or only in vertebrates.

Similarity of TFBS profiles

Every promoter is initially represented by a sequence of TFBSs. We normalized the presence of 3-mers

(trimers) of TFBSs by $p(\alpha_1, \alpha_2, \alpha_3) = \frac{f(\alpha_1, \alpha_2, \alpha_3)}{L-2}$ where L is the number of binding sites on the promoter, and α_i refers to a particular transcription factor. Randomness in this data was reduced via the corresponding 2-mers and 1-mers through

$$p^0(\alpha_1, \alpha_2, \alpha_3) = \frac{p(\alpha_1, \alpha_2)p(\alpha_2, \alpha_3)}{p(\alpha_2)}$$

In a promoter sequence we determined the occurrence of a 3-mer m of TFBSs as

$$m(\alpha_1, \alpha_2, \alpha_3) = \begin{cases} \frac{p(\alpha_1, \alpha_2, \alpha_3) - p^0(\alpha_1, \alpha_2, \alpha_3)}{p^0(\alpha_1, \alpha_2, \alpha_3)} & \text{if } p^0 \neq 0 \\ 0 & \text{if } p^0 = 0 \end{cases}$$

As a consequence, each promoter sequence was represented as a profile of trimers.

Comparing pairs of trimer profiles of TFBSs between species, we defined a distance between promoter sequences M and N as the cosine distance

between profiles of transcription factor binding sites as

$$D(M, N) = 1 - \frac{\sum_{i=1}^k m_i \times n_i}{\sqrt{\sum_{i=1}^k m_i^2} \sqrt{\sum_{i=1}^k n_i^2}}.$$

This similarity measure was used to determine all pairwise distances between promoter profiles of TFBSs between species. Distance matrices were used to reconstruct the trees using the neighbor-joining algorithm as implemented in the DendroPy Phylogenetic Computing Library [59]. The resulting dendrograms were visualized using Fig-Tree, a freely available web-based software tool (<http://tree.bio.ed.ac.uk/software/figtree/>).

Tissue specific TFBSs

The Genomatix database [57, 58] was again used to determine the names and descriptions of the transcription factors as well as their recorded tissue-specific expression. According to Genomatix, “the tissue associations of matrix families are determined by automatic evaluation of all PubMed abstracts (co-citations of transcription factors and tissues) and subsequent manual curation.” Specifically, we considered a non-exclusive set defined as TFs expressed in most tissues, and a preferentially expressed set of TFs that were more highly expressed in the given tissues when compared to other tissues. Only human promoters were used for tissue specificity analysis, due to availability of data.

As a general expected value for transcription factors that appear on a given promoter p and are expressed in a tissue t , $E_{p,t}$, we defined $E_{p,t} = \frac{x_t}{|\cup_t x_t|} n_p$, where x_t is the number of transcription factors that are expressed in tissue t , while $|\cup_t x_t|$ is the total number of transcription factors in all tissues, and n_p is the number of transcription factor binding sites in the underlying promoters sequence p . We utilized this background distribution to determine the enrichment of a promoter sequence p in a tissue t , defined as $f_{p,t} = \frac{O_{p,t}}{E_{p,t}}$, where $O_{p,t}$ is the observed number of transcription factor binding sites that appear in promoter sequence p and are expressed in tissue t . Given the difficulty of assessing significance in this domain, and the lack of direct activity data of TFs on HDAC promoters [41], we considered the \log_2 -fold change of observed and expected TFBS frequencies $fc_{p,t} = \log_2 \frac{O_{p,t}}{E_{p,t}}$ which allowed us to assess the prevalence of expressed TFBSs that occur in a given HDAC promoter sequence in a given tissue. As a consequence, a promoter sequence appears enriched in a given tissue if $fc_{p,t} > 1$ and diluted if $fc_{p,t} < -1$.

Additional files

Additional file 1 : Table S1 Tissue to System classification. (DOCX 17 kb)

Additional file 2 : Figure S1 Evolutionary conservation of transcription factor binding sites in HDAC2 promoter sequences in different organisms. (DOCX 3802 kb)

Additional file 3 : Figure S2 Evolutionary conservation of transcription factor binding sites in HDAC3 promoter sequences in different organisms. (DOCX 3534 kb)

Additional file 4 : Figure S3 Evolutionary conservation of transcription factor binding sites in HDAC4 promoter sequences in different organisms. (DOCX 3746 kb)

Additional file 5 : Figure S4 Evolutionary conservation of transcription factor binding sites in HDAC6 promoter sequences in different organisms. (DOCX 4280 kb)

Additional file 6 : Figure S5 Evolutionary conservation of transcription factor binding sites in HDAC8 promoter sequences in different organisms. (DOCX 3673 kb)

Additional file 7 : Figure S6 Evolutionary conservation of transcription factor binding sites in HDAC9 promoter sequences in different organisms. (DOCX 3052 kb)

Additional file 8 : Figure S7 Evolutionary conservation of transcription factor binding sites in HDAC11 promoter sequences in different organisms. (DOCX 3544 kb)

Abbreviations

HDAC: Histone deacetylases; TFBS: Transcription factor binding site

Acknowledgements

Not applicable.

Authors' contributions

SW and SK conceived the study. TB and SW analyzed data. TB, SW and SK wrote the paper. All authors read and approved the final manuscript.

Funding

The study was supported by start-up funds of the Univ. of Miami.

Availability of data and materials

All used data was obtained from the databases referenced. Tissue specific and transcription factor binding site data was obtained from Genomatix (www.genomatix.com).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Computer Science, University of Miami, Coral Gables, FL, USA. ²University of Exeter College of Medicine and Health, Exeter, UK. ³Department of Biology, University of Miami, Coral Gables, FL, USA. ⁴Center of Computational Science, University of Miami, Coral Gables, FL, USA. ⁵Sylvester Comprehensive Cancer Center, University of Miami, Miami, FL, USA. ⁶Present address: University of California, Los Angeles, Los Angeles, CA, USA.

Received: 25 October 2018 Accepted: 12 July 2019

Published online: 27 July 2019

References

- de Ruijter AJ, van Gennip AH, Caron HN, Kemp S, van Kuilenburg AB. Histone deacetylases (HDACs): characterization of the classical HDAC family. *Biochem J.* 2003;370(Pt 3):737–49.

2. Kim JY, Shen S, Dietz K, He Y, Howell O, Reynolds R, Casaccia P. HDAC1 nuclear export induced by pathological conditions is essential for the onset of axonal damage. *Nat Neurosci*. 2010;13(2):180–9.
3. Li J, Chen S, Cleary RA, Wang R, Gannon OJ, Seto E, Tang DD. Histone deacetylase 8 regulates cortactin deacetylation and contraction in smooth muscle tissues. *Am J Physiol Cell Physiol*. 2014;307(3):C288–95.
4. Ceccacci E, Minucci S. Inhibition of histone deacetylases in cancer therapy: lessons from leukaemia. *Br J Cancer*. 2016;114(6):605–11.
5. Witt O, Deubzer HE, Milde T, Oehme I. HDAC family: what are the cancer relevant targets? *Cancer Lett*. 2009;277(1):8–21.
6. Guise AJ, Cristea IM. Approaches for studying the subcellular localization, interactions, and regulation of histone deacetylase 5 (HDAC5). *Methods Mol Biol*. 2016;1436:47–84.
7. Leipe DD, Landsman D. Histone deacetylases, acetoin utilization proteins and acetylpolymine amidohydrolases are members of an ancient protein superfamily. *Nucleic Acids Res*. 1997;25(18):3693–7.
8. Gregoret IV, Lee YM, Goodson HV. Molecular evolution of the histone deacetylase family: functional implications of phylogenetic analysis. *J Mol Biol*. 2004;338(1):17–31.
9. Dovey OM, Foster CT, Cowley SM. Histone deacetylase 1 (HDAC1), but not HDAC2, controls embryonic stem cell differentiation. *Proc Natl Acad Sci U S A*. 2010;107(18):8242–7.
10. Choi HK, Choi Y, Kang H, Lim EJ, Park SY, Lee HS, Park JM, Moon J, Kim YJ, Choi I, et al. PINK1 positively regulates HDAC3 to suppress dopaminergic neuronal cell death. *Hum Mol Genet*. 2015;24(4):1127–41.
11. Emmett MJ, Lim HW, Jager J, Richter HJ, Adlanmerini M, Peed LC, Briggs ER, Steger DJ, Ma T, Sims CA, et al. Histone deacetylase 3 prepares brown adipose tissue for acute thermogenic challenge. *Nature*. 2017;546(7659):544–8.
12. Janardhan HP, Milstone ZJ, Shin M, Lawson ND, Keane JF Jr, Trivedi CM. Hdac3 regulates lymphovenous and lymphatic valve formation. *J Clin Invest*. 2017;127(11):4193–206.
13. Hua WK, Qi J, Cai Q, Carnahan E, Ayala Ramirez M, Li L, Marcucci G, Kuo YH. HDAC8 regulates long-term hematopoietic stem-cell maintenance under stress by modulating p53 activity. *Blood*. 2017;130(24):2619–30.
14. Tian Y, Wong VW, Wong GL, Yang W, Sun H, Shen J, Tong JH, Go MY, Cheung YS, Lai PB, et al. Histone deacetylase HDAC8 promotes insulin resistance and beta-catenin activation in NAFLD-associated hepatocellular carcinoma. *Cancer Res*. 2015;75(22):4803–16.
15. Marek M, Shaik TB, Heimburg T, Chakrabarti A, Lancelot J, Ramos-Morales E, Da Veiga C, Kalinin D, Melesina J, Robaa D, et al. Characterization of histone deacetylase 8 (HDAC8) selective inhibition reveals specific active site structural and functional determinants. *J Med Chem*. 2018;61(22):10000–16.
16. Aramsangtienchai P, Spiegelman NA, He B, Miller SP, Dai L, Zhao Y, Lin H. HDAC8 catalyzes the hydrolysis of long chain fatty acyl lysine. *ACS Chem Biol*. 2016;11(10):2685–92.
17. Parra M. Class IIa HDACs - new insights into their functions in physiology and pathology. *FEBS J*. 2015;282(9):1736–44.
18. Seto E, Yoshida M. Erasers of histone acetylation: the histone deacetylase enzymes. *Cold Spring Harb Perspect Biol*. 2014;6(4):a018713.
19. Clocchiatti A, Di Giorgio E, Demarchi F, Brancolini C. Beside the MEF2 axis: unconventional functions of HDAC4. *Cell Signal*. 2013;25(1):269–76.
20. Fitzsimons HL. The class IIa histone deacetylase HDAC4 and neuronal function: nuclear nuisance and cytoplasmic stalwart? *Neurobiol Learn Mem*. 2015;123:149–58.
21. Walker RF, Sparks DL, Slevin J, Rush ME. Temporal effects of norepinephrine on pineal serotonin in vitro. *J Pineal Res*. 1986;3(1):33–40.
22. Bradley EW, Carpio LR, Olson EN, Westendorf JJ. Histone deacetylase 7 (Hdac7) suppresses chondrocyte proliferation and beta-catenin activity during endochondral ossification. *J Biol Chem*. 2015;290(1):118–26.
23. Daneshpajoo M, Bacos K, Bysani M, Bagge A, Ottosson Laakso E, Vikman P, Eliasson L, Mulder H, Ling C. HDAC7 is overexpressed in human diabetic islets and impairs insulin secretion in rat islets and clonal beta cells. *Diabetologia*. 2017;60(1):116–25.
24. Gil VS, Bhagat G, Howell L, Zhang J, Kim CH, Stengel S, Vega F, Zelent A, Petrie K. Deregulated expression of HDAC9 in B cells promotes development of lymphoproliferative disease and lymphoma in mice. *Dis Model Mech*. 2016;9(12):1483–95.
25. Qingxu G, Yan Z, Jiannan X, Yunlong L. Association between the gene polymorphisms of HDAC9 and the risk of atherosclerosis and ischemic stroke. *Pathol Oncol Res*. 2016;22(1):103–7.
26. Jamiruddin MR, Kaitsuka T, Hakim F, Fujimura A, Wei FY, Saitoh H, Tomizawa K. HDAC9 regulates the alternative lengthening of telomere (ALT) pathway via the formation of ALT-associated PML bodies. *Biochem Biophys Res Commun*. 2016;481(1–2):25–30.
27. Hubbert C, Guardiola A, Shao R, Kawaguchi Y, Ito A, Nixon A, Yoshida M, Wang XF, Yao TP. HDAC6 is a microtubule-associated deacetylase. *Nature*. 2002;417(6887):455–8.
28. Hai Y, Shinsky SA, Porter NJ, Christianson DW. Histone deacetylase 10 structure and molecular function as a polyamine deacetylase. *Nat Commun*. 2017;8:15368.
29. Chen J, Sahakian E, Powers J, Lienlaf M, Perez-Villarreal P, Knox T, Villagra A. Functional analysis of histone deacetylase 11 (HDAC11). *Methods Mol Biol*. 2016;1436:147–65.
30. Yanginlar C, Logie C. HDAC11 is a regulator of diverse immune functions. *Biochim Biophys Acta*. 2018;1861(1):54–9.
31. Hauser C, Schuettengruber B, Bartl S, Lagger G, Seiser C. Activation of the mouse histone deacetylase 1 gene by cooperative histone phosphorylation and acetylation. *Mol Cell Biol*. 2002;22(22):7820–30.
32. Schuettengruber B, Simboeck E, Khier H, Seiser C. Autoregulation of mouse histone deacetylase 1 expression. *Mol Cell Biol*. 2003;23(19):6993–7004.
33. Yang H, Salz T, Zajac-Kaye M, Liao D, Huang S, Qiu Y. Overexpression of histone deacetylases in cancer cells is controlled by interplay of transcription factors and epigenetic modulators. *FASEB J*. 2014;28(10):4265–79.
34. Gunawardhana LP, Gibson PG, Simpson JL, Powell H, Baines KJ. Activity and expression of histone acetylases and deacetylases in inflammatory phenotypes of asthma. *Clin Exp Allergy*. 2014;44(1):47–57.
35. Giaginis C, Alexandrou P, Delladetsima I, Giannopoulou I, Patsouris E, Theocharis S. Clinical significance of histone deacetylase (HDAC)-1, HDAC-2, HDAC-4, and HDAC-6 expression in human malignant and benign thyroid lesions. *Tumour Biol*. 2014;35(1):61–71.
36. Segre CV, Chiocia S. Regulating the regulators: the post-translational code of class I HDAC1 and HDAC2. *J Biomed Biotechnol*. 2011;2011:690848.
37. Lee K, Lee HY, Back K. Rice histone deacetylase 10 and Arabidopsis histone deacetylase 14 genes encode N-acetylserotonin deacetylase, which catalyzes conversion of N-acetylserotonin into serotonin, a reverse reaction for melatonin biosynthesis in plants. *J Pineal Res*. 2018;64(2):e12460.
38. Nitta KR, Jolma A, Yin Y, Morgunova E, Kivioja T, Akhtar J, Hens K, Toivonen J, Deplancke B, Furlong EE, et al. Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. *Elife*. 2015;4:e04837.
39. Thompson D, Regev A, Roy S. Comparative analysis of gene regulatory networks: from network reconstruction to evolution. *Annu Rev Cell Dev Biol*. 2015;31:399–428.
40. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR, Weirauch MT. The human transcription factors. *Cell*. 2018;172(4):650–65.
41. Berchtold E, Csaba G, Zimmer R. Evaluating transcription factor activity changes by scoring unexplained target genes in expression data. *PLoS One*. 2016;11(10):e0164513.
42. Schaeffe B, Wang TY, Wang CY, Li WH. Gains and losses of transcription factor binding sites in *Saccharomyces cerevisiae* and *Saccharomyces paradoxus*. *Genome Biol Evol*. 2015;7(8):2245–57.
43. Barrière A, Gordon KL, Ruvinsky I. Coevolution within and between regulatory loci can preserve promoter function despite evolutionary rate acceleration. *PLoS Genet*. 2012;8(9):e1002961.
44. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015;347(6220):1260419.
45. Dequiedt F, Kasler H, Fischle W, Kiermer V, Weinstein M, Herndier BG, Verdin E. HDAC7, a thymus-specific class II histone deacetylase, regulates Nur77 transcription and TCR-mediated apoptosis. *Immunity*. 2003;18(5):687–98.
46. Waltregny D, Glenisson W, Tran SL, North BJ, Verdin E, Colige A, Castronovo V. Histone deacetylase HDAC8 associates with smooth muscle alpha-actin and is essential for smooth muscle cell contractility. *FASEB J*. 2005;19(8):966–8.
47. Mihaylova MM, Vasquez DS, Ravnskjaer K, Denechaud PD, Yu RT, Alvarez JG, Downes M, Evans RM, Montminy M, Shaw RJ. Class IIa histone deacetylases are hormone-activated regulators of FOXO and mammalian glucose homeostasis. *Cell*. 2011;145(4):607–21.
48. Banerjee S, Adhikari N, Amin SA, Jha T. Histone deacetylase 8 (HDAC8) and its inhibitors with selectivity to other isoforms: an overview. *Eur J Med Chem*. 2019;164:214–40.

49. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, GA MV. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467(7319):1061–73.
50. Lv W, Zheng J, Luan M, Shi M, Zhu H, Zhang M, Lv H, Shang Z, Duan L, Zhang R, et al. Comparing the evolutionary conservation between human essential genes, human orthologs of mouse essential genes and human housekeeping genes. *Brief Bioinform*. 2015;16(6):922–31.
51. Chakrabarti A, Oehme I, Witt O, Oliveira G, Sippl W, Romier C, Pierce RJ, Jung M. HDAC8: a multifaceted target for therapeutic interventions. *Trends Pharmacol Sci*. 2015;36(7):481–92.
52. Jing X, Sui WH, Wang S, Xu XF, Yuan RR, Chen XR, Ma HX, Zhu YX, Sun JK, Yi F, et al. HDAC7 ubiquitination by the E3 ligase CBX4 is involved in contextual fear conditioning memory formation. *J Neurosci*. 2017; 37(14):3848–63.
53. Bradley RK, Li XY, Trapnell C, Davidson S, Pachter L, Chu HC, Tonkin LA, Biggin MD, Eisen MB. Binding site turnover produces pervasive quantitative changes in transcription factor binding between closely related *Drosophila* species. *PLoS Biol*. 2010;8(3):e1000343.
54. Doniger SW, Fay JC. Frequent gain and loss of functional transcription factor binding sites. *PLoS Comput Biol*. 2007;3(5):e99.
55. Romero IG, Ruvinsky I, Gilad Y. Comparative studies of gene expression and the evolution of gene regulation. *Nat Rev Genet*. 2012;13(7):505–16.
56. Suzuki A, Kawano S, Mitsuyama T, Suyama M, Kanai Y, Shirahige K, Sasaki H, Tokunaga K, Tsuchihara K, Sugano S, et al. DBTSS/DBKERO for integrated analysis of transcriptional regulation. *Nucleic Acids Res*. 2018; 46(D1):D229–38.
57. Quandt K, Frech K, Karas H, Wingender E, Werner T. MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res*. 1995;23(23):4878–84.
58. Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, Frisch M, Bayerlein M, Werner T. MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics*. 2005;21(13):2933–42.
59. Sukumaran J, Holder MT. DendroPy: a Python library for phylogenetic computing. *Bioinformatics*. 2010;26(12):1569–71.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

