Chromatin-based transcriptional punctuation

Paul B. Talbert1 and Steven Henikoff
Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA

The long polycistronic transcription units of trypanosomes do not appear to be demarcated by the usual DNA motifs that punctuate transcription in familiar eukaryotes. In this issue of *Genes & Development*, Siegel and colleagues (pp. 1063–1076) describe a system for the demarcation of transposonspse transcription units based on the deposition and turnover of histone variants rather than on the binding of transcription factors. Replication-independent incorporation of histone variants and destabilization of nucleosomes is an emerging theme at promoters of more familiar eukaryotes, and it now appears that this system is an evolutionarily conserved mode of transcriptional punctuation.

Eukaryotes have evolved remarkably diverse cellular forms encoded by genomes that are regulated by a common set of chromatin proteins. Studies that catalog the diversity of mechanisms for eukaryotic gene expression focus on only a few well-studied model organisms, and “from yeast to man” is a phrase that is often used to indicate generality in eukaryotic biology. However, our favorite model organisms thrive in man-made environments, and only when a “freaky” eukaryote has an impact on human health or wealth does it become a popular organism for intensive study. In this regard, trypanosomes belong to a highly distinctive eukaryotic lineage that has become a model organism because of its enormous impact on human health. They are well-known to us because of the human misery they cause as blood-borne parasites, including *Trypanosoma brucei*, which causes African sleeping sickness that is transmitted by tsetse flies, and others that cause Chagas’ disease and leishmaniasis.

The intensive study of transposon biology aimed at dealing with these scourges has also uncovered remarkable features of their genomic organization and chromatin biology. Notably, gene transcription is polycistronic, with 5’ and 3’ ends that appear to lack the usual sequence motifs for initiation and termination of transcription, which suggests a very different strategy for transcriptional regulation from what is found in more familiar organisms, including animals, fungi, and plants. These observations have led to the notion that a different system is responsible for transcriptional regulation (Clayton 2002).

As described by Siegel et al. [2009] in this issue of *Genes & Development*, the punctuation of transcription appears to be governed by the incorporation of histone variants. This discovery has important implications for more familiar organisms, where histone variant incorporation has taken a back seat to DNA-binding proteins in specifying where transcription starts and stops.

**Trypanosome histone variants and transcription**

Eukaryotic transcription takes place in the context of nucleosomes that wrap DNA, thereby achieving a greater compaction of the genome while restricting access of the transcriptional machinery. The bulk of nucleosomes are composed of octamers of two molecules each of the four “canonical” core histones—H3, H4, H2B, and H2A—which are among the most highly conserved proteins in eukaryotes. Each of these histones, however, also exists in variant forms that can carry out specialized functions. Functionally distinct variants are uncommon for H4 and H2B but are universal, or nearly so, for H3 and H2A (Malik and Henikoff 2003). Considerable work in yeast and multicellular eukaryotes in recent years has established the common functions that are mediated by nucleosomes bearing H3 and H2A variants. H3.3 mediates nucleosome replacement and marks active chromatin (Waterborg 1993; Ahmad and Henikoff 2002), CenH3 functions in chromosome segregation (Amor et al. 2004), H2A.X participates in DNA repair (Loizou et al. 2006), and H2AZ is associated with promoters (Raisner et al. 2005) and maintains accessible chromatin that prevents silencing in yeast (Meneghini et al. 2003) and DNA methylation in plants (Zilberman et al. 2008). Histone variants in unicellular protists, however, do not fit so neatly into these common functional categories. Where has this been more of a puzzle than in trypanosomes? Trypanosomes have two forms of each of the four histones, all of which are among the most highly conserved proteins in the canonical histones of plants and animals (Alsford and Horn 2004; Siegel et al. 2009).

*Keywords: Trypanosoma brucei; histone variants; histone modification; genome-wide ChIP-seq; transcription initiation; transcription termination*

[Corresponding author.
E-MAIL ptalbert@fhcrc.org; FAX (206) 667-5889.
Article is online at http://www.genesdev.org/cgi/doi/10.1101/gad.1806409.
Freely available online through the *Genes & Development* Open Access option.*]
Talbert and Henikoff

The two forms of H2A in T. brucei are markedly divergent from other H2As but appear to correspond to the canonical H2A and its universal variant, H2AZ. Previous work has shown that T. brucei H2AZ is coincident in distribution and specifically coimmunoprecipitates with the H2B variant H2BV, with which it is presumed to dimerize (Lowell et al. 2005). H2BV has only 38% identity to its more conventional H2B counterpart and, together with H2AZ, it is required for viability, indicating that H2AZ/H2BV nucleosomes have a function distinct from H2A/H2B nucleosomes.

Siegel et al. [2009] describe a previously uncharacterized variant of H4 in T. brucei, designated H4V, which is 85% identical to its H4 counterpart. The unusual occurrence of variants of H4 and H2B in trypomonomes is striking, but no less so than the complement of only two forms of H3 variants, both highly divergent from canonical H3. Multicellular eukaryotes commonly have three forms of H3. In addition to the canonical form that is incorporated exclusively during replication, a nearly identical variant known as H3.3 incorporates in replacement nucleosomes that are assembled throughout the cell cycle, including during replication [Ahmad and Henikoff 2002]. H3.3 is incorporated after transcription [Schwarz and Ahmad 2005] and is enriched at the 5' end of genes and also in regulatory elements, where nucleosome remodeling complexes may be responsible for its turnover [Henikoff 2008]. Although the replication-dependent H3 is commonly called canonical, ascomycotes such as yeast and some unicellular organisms such as the algae Cyanidioschyzon and Chlamydomonas have only a single form of packaging H3 that corresponds functionally to replication-independent H3.3. Replication-specific forms of canonical H3 appear to have evolved recurrently in eukaryotic evolution, presumably by divergence from H3.3-like forms [Waterborg and Robertson 1996; Malik and Henikoff 2003]. A third near-universal but much less conserved H3 variant, called CENP-A or CenH3, specifies the centromere and is essential for kinetochore assembly [Amor et al. 2004]. The two T. brucei H3s, designated H3 and H3V, are ~60% identical. Neither variant corresponds to replication-coupled H3 nor, surprisingly, to CenH3, which is missing from trypanosome genomes [Lowell and Cross 2004].

Trypanosomes are highly unusual for eukaryotes in having polycistronic transcription units of up to 100 genes, similar to bacterial operons except that cotranscribed genes are not generally functionally related [Fig. 1]. mRNAs are separated post-transcriptionally by coupled reactions that polyadenylate and splice a 39-nucleotide [nt] leader onto every mRNA [Liang et al. 2003]. Neighboring transcription units can be either convergent or divergent, and the regions between them are known as strand switch regions (SSRs). Nuclear run-on assays in the related trypanosomatids Leishmania major and Trypanosoma cruzi, which are highly syntenic with T. brucei [El-Sayed et al. 2005], showed that RNA polymerase II [pol II] transcription initiates at divergent SSRs and terminates at convergent SSRs [Martinez-Calvillo et al. 2003, 2004; Respuelta et al. 2008]. Trypanosome genomes, however, lack recognizable pol II promoter elements [Clayton 2002]. General transcription factors that act on protein-encoding genes have been difficult to find, and most known families of locus-specific transcription factors appear to be absent from the genome [Iyer et al. 2008].

In order to gain insight into whether chromatin structure and histone variants play a role in transcription initiation, Siegel et al. [2009] used chromatin immunoprecipitation (ChIP) to tagged or untagged histone variants or modifications followed by Solexa sequencing of the nucleosomal DNA [ChIP-seq] to map the locations of histone variants on the assembled genome sequence. Using an antibody to H4 acetylated on Lys 10 [H4K10ac], they found twin peaks of H4K10ac enrichment at every SSR with divergently oriented transcription units, and no peaks at convergent SSRs. They also found 61 single peaks not at SSRs, most of which were downstream from pol III-transcribed tRNA genes that interrupted what otherwise appeared to be a single polycistronic pol II transcription unit. All but three other tRNA genes were found in convergent SSRs, leading Siegel et al. [2009] to propose that all peaks of H4K10ac enrichment correspond to transcription start sites either at divergent SSRs or between adjacent co-oriented transcription units, many of which terminate at sites associated with pol III-transcribed tRNA genes. Comparing the positions of these putative transcription start sites between procyclic and bloodstream forms of the parasite, they found only one unique peak in each form. Antibodies against tagged H2AZ and H2BV gave peaks of enrichment coincident with anti-H4K10ac, suggesting that transcription start sites are characterized by nucleosomes containing H2AZ, H2BV, and H4K10ac. Previous studies showed that H2BV-containing nucleosomes are enriched in tri methylated H3K4 [Mandava et al. 2008] and that divergent SSRs are also enriched with acetylated H3 [Respuelta et al. 2008].

Acetylation of histone tails has been thought to increase nucleosome mobility by neutralizing the charged
lysines that help mediate interactions with DNA and other histones (Waterborg 2002). Nucleosomes containing H2AZ have also been shown to be less stable than H2A nucleosomes when in nucleosome core particles with H3.3 (Jin and Felsenfeld 2007). This raises the possibility that less stable nucleosomes at transcription start sites of trypanosomes facilitate transcription initiation. To test this possibility, Siegel et al. (2009) used antibodies against equivalently tagged versions of H2AZ and H2A or H2BV and H2B and compared the ability of each tagged histone to communoprecipitate H3 and H4 after washing chromatin pellets in increasing concentrations of salt. They found significantly less association of H3 and H4 with tagged H2AZ and H2BV, consistent with H2AZ/H2BV nucleosomes being less stable.

Besides modifying nucleosome mobility, acetylated lysines on histones can serve as binding sites for bromodomain proteins. Siegel et al. (2009) investigated such proteins and found one, BDF3, that colocalized cytologically with H4K10ac. ChIP-seq profiles showed that BDF3 overlapped H4K10ac peaks and was concentrated over the 5’ ends of transcription units. Siegel et al. (2009) also analyzed the DNA sequences of the putative transcription start sites and found them enriched in runs of nine to 15 consecutive guanines.

The distribution of tagged versions of variants H3V and H4V was also investigated. H3V was known previously to be localized at telomeric/subtelomeric sites (Lowell and Cross 2004). ChIP-seq confirmed this distribution, but also revealed tagged H3V at convergent SSRs and upstream of single non-SSR H4K10ac peaks, the sites of presumed pol II transcription termination. Tagged H4V coincided with H3V at presumed termination sites, but showed less enrichment over telomeric and subtelomeric regions.

Siegel et al. (2009) propose a model of transcriptional initiation in which H3K4 methylation is needed for H4K10 acetylation, which would serve as a binding site for BDF3. In turn, BDF3 would recruit transcription factors or chromatin remodeling complexes involved in incorporating H2AZ/H2BV dimers into nucleosomes. The less stable H2AZ/H2BV/H3K4me/H4K10ac nucleosomes are proposed to be permissive for polymerase binding and transcription initiation. The nearly complete acetylation of H4K10 at all putative transcription start sites argues against a role in regulating transcription levels. Instead, mRNA levels appear to be regulated post-transcriptionally (Clayton 2002). Runs of Gs may confer directionality on transcription, either through direct effects on polymerase processivity or through DNA-binding factors. H2A/H2B/H3V/H4V nucleosomes mark sites of pol II transcription termination and often coincide with tRNA genes. Yeast tRNA genes are clustered and their subnuclear localization can inhibit pol II transcription (Kendall et al. 2000), and Siegel et al. (2009) suggest that a similar clustering process in T. brucei could lead to an island of pol III transcription from which pol II is effectively excluded. An alternative model is that H3V/H4V-containing nucleosomes have increased stability and are not easily traversed by pol II. Perhaps these nucleosomes are incorporated following pol III transcription. These and other ideas are ripe for testing, starting with careful transcript mapping and further investigation of the structures and properties of variant nucleosomes.

Evolution of eukaryotic transcriptional punctuation

Are trypanosomes just evolutionary freaks, or do they offer a window into the evolution of transcription and histone variants in eukaryotes? Phylogenetic trees of small subunit ribosomal RNA initially led to the view that trypanosomes, their euglenid relatives, and a handful of protists lacking mitochondria are very early-branching eukaryotes, in contrast to the “crown” eukaryotes that include plants, animals, fungi, and numerous other protist groups (Sogin and Silberman 1998). However, early branching in these trees has been claimed to be an artifact of systematic errors in tree construction methods, especially long branch attraction, and one “early-branching” group, the Microsporidia, was shown to be related to fungi (Hirt et al. 1999). Improved methods and additional data have not succeeded in relating other early-branching groups to “crown” groups, but most other early-branching eukaryotes including trypanosomes have been proposed to belong to the Excavata, a group defined by cytoskeletal features mostly related to the flagellar apparatus, and by molecular phylogenies (Simpson 2003). The monophyly of excavates has been difficult to verify [Simpson 2003; Yoon et al. 2008] but appears to be gaining support [Rodriguez-Ezpeleta et al. 2007; Hampel et al. 2009]. The position of excavates such as trypanosomes relative to the root of the eukaryotic tree and whether they are truly early branching, however, remains controversial [Rodriguez-Ezpeleta et al. 2007; Hampel et al. 2009] and, therefore, so is the notion that their transcriptional peculiarities, including the lack of recognizable promoter elements and transcription initiation factors, might preserve ancestral characteristics.

Transcriptional regulation in other putative excavates such as the diplomonad Giardia lamblia and the parasalid Trichomonas vaginalis differs from trypanosomes, lacking polycistronic messages and trans-spliced leaders, but has other unusual, putatively primitive features (Vanacova et al. 2003). Whereas metazoans have heterogeneous promoters that combine diverse genespecific enhancers with common shared promoter elements such as the TATA-box and initiator element, Giardia has simple compact promoters generally occupying <70 base pairs [bp] and including a short upstream TATA-like element and an initiator element, both of which lack any highly conserved sequence beyond AT richness (Vanacova et al. 2003; Teodorovic et al. 2007). These promoters function bidirectionally, with the TATA-like and initiator elements apparently switching roles, producing an abundance of noncoding antisense transcripts. Transcripts also can be produced from AT-rich patches [cryptic promoters] throughout the genome [Teodorovic et al. 2007], suggesting a rather promiscuous mode of transcriptional initiation. The three Giardia RNA polymerases have most of the standard subunits...
found in polymerases from crown eukaryotes but lack those subunits specific to just one of the polymerases and have only four of 12 general transcription initiation factors, which suggests that the missing factors evolved after the divergence of \textit{Giardia} from the lineage of the crown group [Best et al. 2004]. \textit{Trichomonas} has promoters with a simple highly conserved initiator element and additional small upstream gene-specific elements [Liston and Johnson 1999; Vanacova et al. 2003]. The initiator sequence is bound by a unique protein, IBP39, that interacts with the C-terminal domain (CTD) of the largest subunit of pol II [Lau et al. 2006]. IBP39 is without known homologs in other eukaryotes, suggesting that \textit{Trichomonas} also diverged from the main line of eukaryotic descent before general transcription factors became sophisticated and fixed in their roles. Consistent with this scenario, the DNA-binding domain of IBP39 has given rise to a family of >100 putative transcription factors in \textit{Trichomonas} [Iyer et al. 2008].

In plants, animals, and fungi, the CTD of the largest subunit of pol II is composed of numerous heptad repeats of YSPTSPS and functions in several processes in transcription, including transcriptional elongation and premRNA processing. These heptad repeats are absent from the CTDs of \textit{Giardia}, \textit{Trichomonas}, and trypanosomes and are reduced and divergent in a fourth putative excvate, the free-living heterolobosean amoeboflagellate \textit{Naegleria gruberi} [Dacks et al. 2002, Stiller and Hall 2002]. Investigation into the evolution of the pol II largest subunit (RPB1) led to the proposal that heptad repeats may have been ancestral for eukaryotes but did not become essential and fixed in sequence until a later time, when increased sophistication of CTD functions led to more intricate patterns of gene expression and splicing that made possible a “CTD” clade of developmentally complex eukaryotes [Stiller and Hall 2002].

Histone variants reconsidered

Siegel et al. [2009] provide evidence that the seemingly unregulated transcription of trypanosomes is directed by histone variants, and that transcription initiation is effected by unstable nucleosomes containing H2AZ/H2BV and acetylated H4, and terminated by nucleosomes containing H3Vand H4V. Do other early-branching excavates also employ unusual histone variants to regulate transcription? The answer is not known, but it is interesting to note that \textit{Giardia}, \textit{Trichomonas}, and \textit{Naegleria} all encode divergent H3 variants in addition to somewhat more conventional H3s and putative or confirmed [Dawson et al. 2007] CenH3s. In \textit{Giardia}, the variant H3B is found in numerous punctate dots on metaphase chromosomes, and does not much overlap signal from antibodies to mono- and dimethylated H3K4, which are also present in relatively few punctate dots [Dawson et al. 2007]. This pattern does not immediately suggest a general role in gene transcription, and H3B was proposed to have a role in nonpericentric heterochromatin. ChIP-seq technology would likely clarify its role. \textit{Giardia} has only a single form of H2A that does not closely resemble canonical H2A or H2AZ but must have the general packaging function of H2A and also has the peptide sequence necessary to serve as an H2A.X (Malik and Henikoff 2003). In yeast, promoters are marked by nucleosome depletion around the transcription initiation site that is bound by a Myb-related DNA-binding protein and by flanking H2AZ nucleosomes [Raisner et al. 2005]. In \textit{Giardia}, promoters occupy <70 bp, and the average space between genes is only 372 bp, or about two nucleosomes. Are these intergenic spaces kept depleted of nucleosomes by DNA-binding proteins or, if not, how are the nucleosomes mobilized to allow transcription?

The apparent punctuation of transcription by histone variant incorporation and turnover seen in trypanosomes parallels an emerging theme in transcriptional regulation of more familiar eukaryotes, where these processes occur in the context of regulation of transcription by nonhistone DNA-binding proteins. DNA elements can bind proteins or position nucleosomes, histone variants can produce nucleosomes of different stabilities, and modifications to histone tails can further stabilize or destabilize nucleosomes both by affecting their interaction with DNA and by binding proteins that help anchor or remodel nucleosomes [Henikoff 2008]. This toolkit of nuclear proteins has been employed differently to regulate transcription in different organisms, and some processes, such as turnover of nucleosomes at regulatory sequences upstream of promoters, have become necessary innovations as transcription itself evolved in complexity. Lacking sophisticated regulatory sequences, trypanosomes appear to have innovated the deployment of histone variants to achieve transcriptional regulation. Early-branching eukaryotes, even if some of them turn out to be derived rather than primitive, offer a window into the evolutionary plasticity and constancy of methods of achieving gene expression in the context of nucleosome packaging. To the extent that they reflect ancestral conditions, they also give clues to the sequence of events that gave rise to the sophisticated gene regulation in multicellular organisms today.

References


Transcriptional punctuation by chromatin


Chromatin-based transcriptional punctuation

Paul B. Talbert and Steven Henikoff

*Genes Dev.* 2009, 23:

Access the most recent version at doi:10.1101/gad.1806409

---

**Related Content**

Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*

T. Nicolai Siegel, Doeke R. Hekstra, Louise E. Kemp, et al.

*Genes Dev.* May, 2009 23: 1063-1076

**References**

This article cites 40 articles, 17 of which can be accessed free at:

http://genesdev.cshlp.org/content/23/9/1037.full.html#ref-list-1

Articles cited in:

http://genesdev.cshlp.org/content/23/9/1037.full.html#related-urls

**License**

Freely available online through the *Genes & Development* Open Access option.

**Email Alerting Service**

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.