UNIVERSITÉ DE STRASBOURG

igbmc institut de génétique et de biologie moléculaire et cellulaire

Thèse présentée pour obtenir le grade de

Docteur de l'Université de Strasbourg

Discipline: Sciences du vivant

Aspects Moléculaires et Cellulaires de la Biologie

Par

KREBS Arnaud

Caractérisation fonctionnelle des complexes contenant GCN5

Functional characterization of GCN5 containing complexes

Soutenue publiquement le 29 septembre 2010

Membres du jury:

- Directeur de thèse: Dr Làszlò Tora
- Rapporteur interne : Dr Florence Cammas
- Rapporteur externe : Dr Matthieu Gérard
- Rapporteur externe : Dr Mueller Ferenc

Abstract

In higher eukaryotes, RNA Polymerase II transcription is a central cellular process allowing the spatio-temporal expression of a particular set of genes contained in a given genome. Accurately regulated transcription is a prerequisite for many key aspects of the life of an organism. Thus, transcription is tightly controlled by multiple regulatory layers.

GCN5 is a member of the Histone Acetyl Transferase (HAT) family that are part of the chromatin modifier complexes. These complexes tune transcription by acting on chromatin states through the addition of post-translational modifications on histones. HATs deposit acetyl groups on histone tails that is believed to favour chromatin opening and positively influence transcription initiation. GCN5 is contained in two different muti-subunit macromolecular complexes named ATAC (Ada-Two-A-Containing) and SAGA (Spt-Ada-Gcn5-Acetyl-transferase) that modulate its functional specificity. While the composition and the *in vitro* activity of these two complexes were intensively studied, little is known about the function of these complexes *in vivo*.

During my thesis, I aimed to better characterise the mode of action of GCN5 containing complexes (GCC) by using recent technology developments. First through the systematic analysis of the *in vivo* interacting partners of SAGA and ATAC by sensitive mass spectrometry, I identified a stable functional interaction between ATAC and the Mediator, another coactivator complex. Then, I showed that this interaction is dependent upon a molecular bridge that allows the formation of the ATAC-Mediator structure in defined cell contexts. By using genome wide mapping techniques, I could show that this particular complex is recruited to regulate a set of ncRNA genes.

Second, I analysed genome wide binding maps of ATAC produced in different cell types by chromatin immunoprecipitation coupled with high throughput sequencing (ChIP-seq). By crossing these data with existing information on chromatin marking, I could show that ATAC binds both active promoter and active enhancer elements. Moreover, I demonstrate that while the ATAC binding at promoters is generally invariant across cell lines, the binding at enhancer is highly variable. This suggests that cell specific programs controlled by ATAC are mainly regulated at the enhancer level.

Finally, along the studies I developed innovative bioinformatics approaches to facilitate the interpretation of genome wide mapping datasets that I combined in two softwares, which are freely available for the scientific community.

Acknowledgments

I would like to thank the members of the Jury Dr Gerard Matthieu, Dr Müller Ferenc, Dr Cammas Florence for accepting to evaluate my work.

Particular thanks go to Làszlò Tora that additionally to a fruitful project provided me constant unconditional support, exceptional freedom and exceptional working conditions. I am especially grateful to him for giving me the possibility to make a break in the research project to acquire additional competences that I anticipate to be highly valuable for me in the future. Moreover, I own him a number of epi-scientific advices on multiple aspects of the scientific life that I progressively realize to be essential.

I am of course very grateful to all the Tora's lab member who created such a friendly but prolific scientific environment. Additional thanks go to Eli that additionally to a friendly bench neighbour was a great in & out of lab adviser. I am also grateful for the similar kind of support that was constantly provided by Didier that allowed me to go up in various aspects of science and rock based activities. I particularly appreciated Didier's constant availability; his key critical reviewing that allowed me to raise up my standards; his pertinent and broad scientific vision that guided me in a couple of "middle of the desert" situations. I want also to particularly thank Maria Elena for enriching scientific discussion, creating a stimulating environment of scientific excellence, work reviewing and key advices on career management.

I am grateful to all the people that by their work made these studies possible. I would like to mention: Andrée Dierich that thought me all the ES cells manipulation techniques; Jeroen Demmers for performing all the MS analysis and key experimental suggestions; Pascal Eberlin and Gilles Duval for respectively the synthesis of the peptides and the production of excellent antibodies; the organizers of the excellent "Informatics in Biology" course in Pasteur Paris; Bernard Jost, Stephanie Legras and Serge Vicaire for the high throughput sequencing.

I would like to thank people for past and present scientific collaborations: Fred who supported the bioinformatics work in the background. Tao for incredibly fast implementation and design of great bioinformatics tools, Celine that gives a new breathe to the seqMINER project.

Additionally, I would like to thank the neighbouring perturbating elements of the IGBMC that made these years a pleasant period.

Special thank goes to Emilie for out of the lab constant and invaluable support.

Table of contents

PREFACE

Eukaryotic gene expression is a complex stepwise mechanism that requires several multi-subunit cellular machines. In higher eukaryotes, RNA Polymerase II transcription is a central cellular process allowing the spatio-temporal control of expression of particular gene programs out of the whole set of transcription units contained in a genome. Accurately regulated transcription is a prerequisite for many key aspects of the life of an organism. For example, gene regulation gives control to an organism over its structure and function, and is the basis for cellular differentiation, morphogenesis and the adaptability. Through evolution, transcription regulation mechanisms gradually increased in complexity to become sophisticated multilayered regulatory networks in higher eukaryotes.

In the initial models proposed to describe transcriptional regulatory mechanisms, interaction between regulatory proteins (transcription factors) and DNA had a central role in the establishment of active transcription states. This DNA centred vision, inherited from knowledge in prokaryotes, was challenged thanks to several technical breakthroughs in the last decades. The completion of the sequencing of the human genome quickly followed by other species revealed that DNA regulatory elements appeared to be largely degenerated in eukaryotes and their presence seems, unlike in prokaryote, to be rather the exception than the rule in the typical promoter in higher eukaryotes. Additionally, an increasing number of evidences revealed that the presence of chromatin in eukaryotes was not only a static structure required to better compact larger genomes within the nucleus, but rather a dynamic regulator having a key influence on all DNA related processes including transcription. Thus, contemporary models tend to lower the importance of DNA regulatory elements and to give to chromatin a preponderant role in the regulation of transcription related processes.

The present work aims to study one particular member of the large panel of transcriptional regulatory players used by cells to tune their transcription at proper levels. More precisely, my work focused on the study of GCN5, a member of the Histone Acetyl Transferase (HAT) family that is part of the chromatin modifiers complexes. These complexes tune transcription by acting on chromatin states through the addition of post-translational modifications on histones In the present study, I will present new insight on the function of these complexes at the light of post-genomic era technologies.

INTRODUCTION

In this chapter, I will first extensively review the general mechanistic concepts that are thought to rule transcription. I will detail the role of different elements of the regulatory system including, DNA regulatory sequences, protein regulators as well as transcription mechanisms. Then I will discuss emerging concepts regarding transcription regulation that arises from a recent technology breakthrough that allow asking biological questions at the scale of the genome. I will particularly detail the increasing role given to chromatin in the recent models describing transcriptional mechanisms and the interplay of this new regulatory layer with the transcription machinery. In the final introductory part, I will present the current knowledge regarding the main topic of this study that is HAT complexes containing GCN5. I will review in detail the knowledge about the composition of these complexes, their mechanistic action and more globally their known function in transcriptional regulation.

First Part: Core RNA Pol II transcriptional mechanisms

Transcription is a central cellular process through which a gene is enzymatically copied by a DNA dependant RNA polymerase to produce a complementary RNA. Eukaryotic transcription requires several steps including promoter recognition, pre-initiation complex (PIC) assembly, promoter escape, elongation and termination. In most of the generally admitted transcriptional regulatory model, transcription initiation is assumed to be the rate limiting step of the process. Recent evidences have demonstrated the existence of regulatory mechanisms acting after PIC formation. However, the importance of these mechanisms relative to pre-initiation steps in the global transcription activation process remains elusive. Thus the present review of the mechanisms controlling RNA Polymerase II (Pol II) activity will extensively describe the knowledge on the regulatory steps that precedes transcription initiation. I will first present the DNA regulatory sequences known to play a role in transcription, then, I will detail the nature, the role of the main actors of transcription initiation. Finally, I will discuss new concepts that recently emerged from studies performed at the scale of the genome and that challenges existing transcriptional regulatory models.

I. DNA regulatory elements

At the DNA level, genes transcribed by Pol II are typically regulated by (i) common core-promoter elements that are recognized by general transcription initiation factors (GTFs) and (ii) gene-specific DNA elements that are recognized by regulatory factors, which in turn modulate the function of the general initiation factors. (iii) various distal regulatory elements including enhancers and insulators, that tune transcription through long distance interactions.

A. Core-promoter elements

Core-promoter elements are defined as minimal DNA elements that are necessary and sufficient for accurate transcription initiation by Pol II in reconstituted cell-free systems. These elements are recognized most often by a specific GTF. In the subsequent section, I will briefly review the characteristics of each element found in mRNA genes promoters, that are also summarized in Table 1. Additionaly, I wil detail the elements that differentiate promoters of snRNA genes from mRNA genes.

DNA element	Acronym meaning	Description	Bound protein
TATA	--	Located 25-30 bp upstream the TSS, site of PIC assembly. <10% of human gene contain a TATA box	TBP
BRE	TFIIB Recongnition Element	Two distinct motifs flank the site where the TATA box typically resides, helps to orient the PIC	TFIIB
Inr	Initiator	Immediately adjacent to the TSS, can accurately direct initiation alone or with the TATA box	TAF1,TAF2
MTE	Motif Ten Element	Located 20bp downstream of TSS, functions with Inr to enhance transcription, can substitute for TATA	NA
DPE	Downstream Promoter Element	Located 30bp downstream of the TSS in *Drosophila*	TAF6-TAF9
enhancer		Distant element recognized by activators to stimulate transcription by recruitment of coactivators	
insulator		Distant elements recognized by insulator proteins that block activation/repression signals from spreading along the genome	

Table 1 : Summary of the common DNA elements regulating Pol II transcription
(Adapted from (Venters and Pugh 2009b))

a. TATA box

The TATA box, a short sequence upstream of the initiation start site, was the first core promoter element identified in eukaryotic protein-coding genes. The discovery of the TATA box in 1979 emerged from a comparison of the 5' flanking sequences in a number of *Drosophila*, mammalian, and viral protein-coding genes (Breathnach and Chambon 1981). Following the early studies, it was speculated that the TATA box might be strictly conserved and essential for transcription initiation from all protein-coding genes from yeast to human. However, with the release of the complete sequence of several higher eukaryote genomes, this assumption was infirmed. Two recent database analyses of human genes revealed that TATA boxes were present in only 10 to 20% of all promoters examined (Gershenzon and Ioshikhes 2005; Cooper, Trinklein et al. 2006).

The TATA-binding protein (TBP) a subunit of the GTF called TFIID, was identified to be the protein binding the TATA box (Dynlacht, Hoey et al. 1991; Nikolov, Hu et al. 1992). TBP was later found to be also a component of distinct complexes that contribute to transcription

initiation by RNA polymerases I and III (Hernandez 1993; Goodrich and Tjian 1994). Consensus sequences for TATA function and TBP binding have been difficult to define. A binding site selection analysis identified the sequence 5'-TATATAAG-3' as the optimal TBP recognition sequence (Wong and Bateman 1994). However, several other studies revealed that a wide variety of A/T-rich sequences can function as potential TATA boxes and can interact with TBP (Singer, Wobbe et al. 1990).

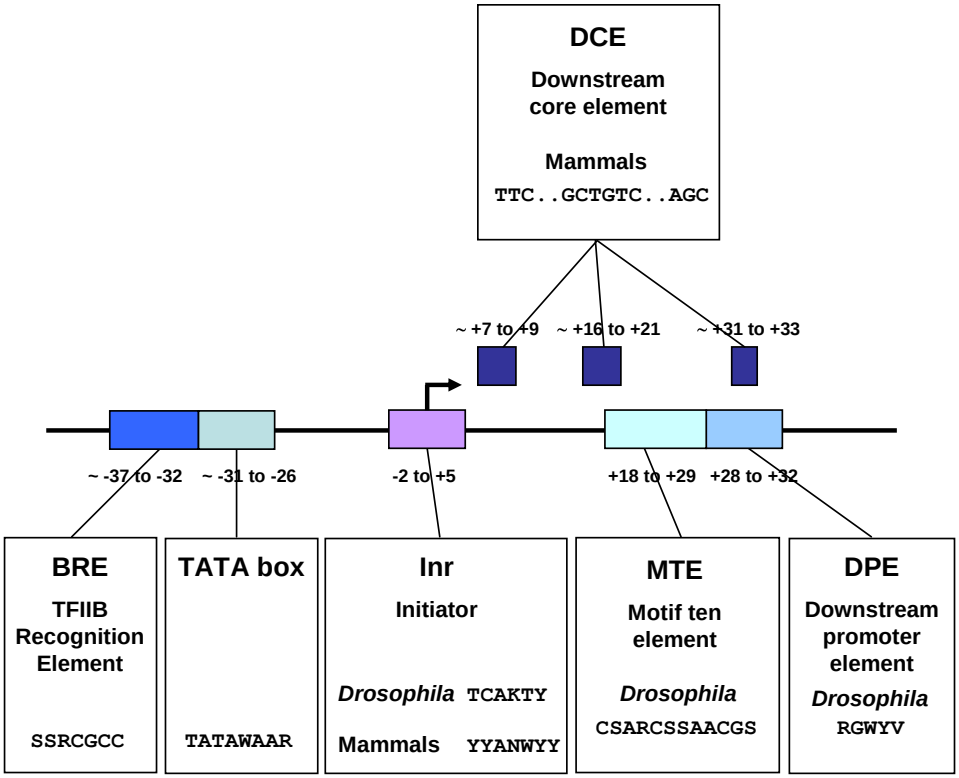


Figure 1: Core promoters motifs.
This diagram depicts some of the sequence elements that can contribute to basal transcription from a core promoter. Each of these sequence motifs is found in only a subset of core promoters. A particular core promoter may contain some, all, or none of these elements. The TATA box can function in the absence of TFIIB recognition Element (BRE), Initiator (Inr) and Downstream Promoter Element (DPE) motifs. In contrast, the DPE motif requires the presence of an Inr. The BRE is located immediately upstream of a subset of TATA box motifs. The DPE consensus was determined with *Drosophila* core promoters. The Inr consensus is shown for both mammals and *Drosophila*. Motif Ten Element (MTE) and DCE have been recently identified in *Drosophila* and b-globin gene, respectively. For description purposes, the consensus sequences are presented as International Union for Pure and Applied Chemistry (IUPAC) consensus symbols: N, any nucleotide; R, A or G (purine); S, C or G; V, A or C or T; W, A or T; Y, C or T (pyrimidine); K, G or T. Adapted from (Smale and Kadonaga 2003).

b. Initiator element

The initiator element (Inr) contains within itself the transcription start site (+1) and spreads from -2 to +5 positions (Figure 1). The Inr is necessary and sufficient for accurate transcription *in vitro* and *in vivo* (Smale and Baltimore 1989). The *Drosophila* and mammalian Inr consensus sequences determined by database analysis are similar but not identical: Py C A(+1) N T/A Py Py in mammals and T C A(+1) G/T T Py in *Drosophila*. Similar to the TATA box, the Inr element was not present in all protein-coding genes: the *Drosophila* consensus, or a sequence containing one mismatch, was present in 67.3% of 1941 core promoters (Ohler, Liao et al. 2002). The prevalence of the Inr in mammalian promoters has not been determined using rigorous methods similar to those used to calculate its prevalence in *Drosophila*. Studies of TATA-Inr spacing have shown that the two elements act synergistically when separated by 25-30 bp but act independently when separated by more than 30 bp (Ohler and Wassarman; O'Shea-Greenfield and Smale 1992).

Although the Inr element is not symmetrical and therefore has the potential to dictate the direction of transcription, its contribution to directionality from a promoter appears minimal (O'Shea-Greenfield and Smale 1992). Several studies demonstrated that the Inr element is recognized by the TFIID complex (Wang and Van Dyke 1993; Bellorini, Dantonel et al. 1996). Moreover, Verrijzer et al. have detected stable Inr binding by a complex consisting of two TBP-associated factors (TAFs), TAF1 and TAF2 (Chalkley and Verrijzer 1999). The domains of TAF1 and TAF2 that are responsible for Inr recognition have not been determined. However, their involvement is consistent with functional studies showing that a trimeric TBP-TAF1-TAF2 complex is sufficient for Inr activity in reconstituted transcription assays (Verrijzer, Chen et al. 1995).

c. Downstream promoter element (DPE)

The downstream promoter element (DPE) was identified as a downstream core promoter motif that is required for the binding of purified TFIID to a subset of TATA-less promoters (reviewed in (Kadonaga 2002). The DPE acts in conjunction with the Inr, and the core sequence of the DPE is located at precisely +28 to +32 relative to the +1 nucleotide in the Inr motif (Kutach and Kadonaga 2000) (Figure 1). The DPE is conserved from *Drosophila* to humans and is typically, but not exclusively, found in TATA-less promoters

(Kutach and Kadonaga 2000). In a *Drosophila* database of 205 core promoters, it was estimated that about 29% contain a TATA-box and no DPE motifs, 26% possess a DPE but no TATA, 14% contain both TATA and DPE motifs, and 13% do not appear to contain either a TATA or a DPE (Kutach and Kadonaga 2000). The frequency of occurrence of the DPE motif in the human genome has not been determined yet.

d. TFIIB recognition element (BRE)

The TFIIB recognition element (BRE) is another well-characterized element in the core promoters of protein-coding genes that is recognized by a factor other than TFIID. Compelling evidence that TFIIB interacts with DNA in a sequence-specific manner emerged from the analysis of the T6 promoter from the archaeal *Sulfolobus shibatae* virus (Qureshi and Jackson 1998). Archaeal TBP and TFB (the archaeal homolog of TFIIB) bound cooperatively to the promoter when both the TATA box and upstream element were present. A binding site selection analysis revealed that, in the presence of TBP and TFB, purines were strongly preferred 3 and 6 bp upstream the TATA box. A parallel study with human TFIIB established the existence of a eukaryotic BRE that prefers a 7-bp sequence: G/C G/C G/A C G C C (Lagrange, Kapanidis et al. 1998) (Figure 1). Although the interaction between the archaeal TFB and BRE clearly enhances the assembly of a pre-initiation complex and transcription initiation, the function of the human TFIIB-BRE interaction appears to be very different. The BRE was reported to be a repressor of basal transcription *in vitro* with crude nuclear extracts as well as *in vivo* in transfection assays (Evans, Fairley et al. 2001). These results suggest that the function of the BRE may have expanded during evolution. In the archaea, it stimulates promoter activity, but in eukaryotes, it also represses transcription.

e. Motif ten element (MTE)

The motif ten element (MTE) was identified as an overrepresented sequence motif in a computational analysis of nearly 2000 *Drosophila* core promoters (Ohler, Liao et al. 2002). This sequence, motif 10, is located downstream of the transcription start site, from +18 to +29 relative to the A_{+1} position in the Inr. The consensus sequence is CSARCSSAACGS (Figure 1). Subsequent studies have revealed that the motif 10 sequence contains a new core promoter element termed MTE that can promote transcription by RNA Pol II in conjunction with the Inr but independently of the TATA box or DPE (Lim, Santoso et al. 2004).

f. Downstream core element (DCE)

The downstream core element (DCE) was first identified in the human adult B-globin promoter (Lewis, Kim et al. 2000). The existence of this element was indicated by two mutations at +22 and +33 downstream of the B-globin transcriptional start site in humans with thalassemia. The DCE is a tripartite element that spreads approximately from +7 to +33 positions, downstream the transcription start site (Figure 1). The DCE functions in concert with the B-globin CATA box and Inr element, as well as in a heterologous, TATA-less context. DCE mutants show a reduced affinity for TFIID, indicating that TFIID makes sequence-specific contacts to the DCE and that TFIID binding is necessary for DCE function (Lewis, Kim et al. 2000). More recently, the TAF1 subunit of TFIID was reported to interact with the DCE in a sequence-dependent manner (Lee, Gershenzon et al. 2005).

g. Element specific for snRNA genes promoters (PSE, DSE)

The human small nuclear RNA (snRNA) genes, which encode snRNAs that are involved in RNA processing reactions such as mRNA splicing, serve as prototypes for a family of genes whose promoters are characterized by the presence of a proximal sequence element (PSE) and a distal sequence element (DSE) (Hernandez 2001). The human pol II snRNA core promoters contain only one essential element, the PSE, whereas the pol III snRNA core promoters consist of two elements, the PSE and a TATA box located at a fixed distance downstream. Both the DSE and the PSE can be interchanged between pol II and III snRNA promoters with no effect on RNA polymerase specificity, which is determined by the presence or absence of the TATA box.

The factor binding to the PSE element has been characterized in the human system and is known as $SNAP_c$. It is a complex containing five types of subunits, SNAP190, SNAP50 (PTFβ), SNAP45 (PTFδ), SNAP43 (PTFγ), and SNAP19 (Mittal, Ma et al. 1999; Ma and Hernandez 2001), that is required to initiate transcription from snRNA promoters. The precise set of factors binding to PSE is still unknown but the presence of these element has been shown to enhance transcription from snRNA genes.

B. Distal elements

Distal elements are gene-specific DNA elements recognized by regulatory factors that allow the modulation of the function of GTFs.

1. Enhancer

Enhancers were originally identified as *cis*-acting DNA sequences that increase transcription in a manner that is independent of their orientation and distance relative to the transcription start site. Although most enhancers are located tens of kilobases away, some have been found at distances of up to a megabase from the gene they regulate (Nobrega, Ovcharenko et al. 2003; Qin, Kong et al. 2004). Enhancer elements, therefore, have the potential to activate a number of neighbouring genes over a large chromosomal region (Figure 2). Enhancers are bound by transcriptional activators/co-activators prior to promoter activation. Some enhancers are specific for promoters that contain either DPE or TATA box elements (Butler and Kadonaga 2001), whereas some others are regulated by promoter-proximal tethering elements or through promoter competition (Calhoun, Stathopoulos et al. 2002). However, most enhancers appear to be promiscuous, thus, the action of enhancers *in vivo* must be restricted in order to prevent the activation of non-target genes within their long reach.

2. Insulator

Insulators or boundary elements are another class of regulatory sequences that possess the ability to protect genes from inappropriate signals emanating from their surrounding environment. Insulators protect an expressing gene from its surroundings in two ways. First, insulators protect genes is by acting as “barriers” (Sun and Elgin 1999) that prevent the spreading of nearby condensed chromatin that might otherwise silence expression (Figure 2). The second way is by blocking the action of a distal enhancer on a promoter (Kellum and Schedl 1991; Geyer and Corces 1992). Enhancer blocking occurs if the insulator is located between the enhancer and the promoter (Figure 2). Such activity can prevent an enhancer from activating expression of an adjacent gene from which it is blocked, while leaving it free to stimulate expression of genes located on its unblocked side. Recently a catalogue of insulator elements has been defined based on binding of CTCF a repressive protein known to

bind these elements (Kim, Abdullaev et al. 2007). This study allowed to define a consensus motif for these elements (CCACCAGGGG) (Kim, Abdullaev et al. 2007).

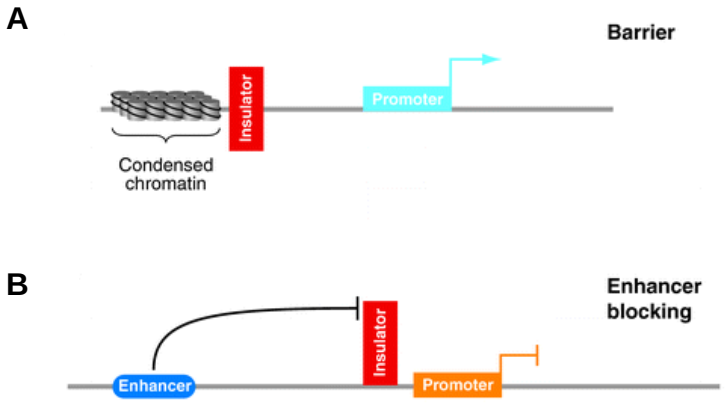


Figure 2 Insulators block enhancer and silencer elements in a position-dependent manner.
A, Barrier elements block the linear spread of silenced chromatin protecting the reporter gene from silencing. B, Enhancer-blocking elements interfere with enhanced transcription when placed between an enhancer element and the promoter. From (Valenzuela and Kamakaka 2006).

II. Actors of Pol II transcription initiation

A. RNA Polymerase II (Pol II)

RNA polymerase II is the central enzyme that catalyses DNA-directed mRNA synthesis during transcription of protein-coding genes. RNA polymerase is conserved in all living organisms. Thus, bacterial RNA polymerase, archaeal RNA polymerase and eukaryotic RNA polymerases I, II and III are members of a conserved protein family, termed the "multisubunit RNA polymerase family" (Table 2). In eukaryotes, whereas RNA polymerases I and II synthesize ribosomal and mainly mRNA respectively, RNA polymerase III transcribes small RNAs, including transfer RNAs, 5S ribosomal RNA and U6 small nuclear RNA.

Eukaryotes					
Pol I	Pol II	Pol III	Archaea	Bacteria	Class[1]
A190	Rpb1	C160	A' + A''	β'	Core
A135	Rpb2	C128	B (B' + B'')	β	Core
AC40	Rpb3	AC40	D	α	Core
AC19	Rpb11	AC19	L	α	Core
Rpb6	Rpb6	Rpb6	K	ω	Core and common
Rpb5	Rpb5	Rpb5	H	—	Common
Rpb8	Rpb8	Rpb8	—	—	Common
Rpb10	Rpb10	Rpb10	N	—	Common
Rpb12	Rpb12	Rpb12	P	—	Common
A12.2	Rpb9	C11	X	—	Unclear
A14	Rpb4	C17	F	—	Rpb4/7
A43	Rpb7	C25	E	—	Rpb4/7
A34.5	—	C82	+1 other	—	Specific
A49		C34			Specific
		C31			Specific

Table 2 : RNA Polymerase subunits.
Core: sequence partially homologous in all RNA polymerases. Common: shared by all eukaryotic RNA polymerases, Rpb4/7: Rpb4/7 heterodimer and its structural counterparts. Unclear: it is unclear if A12.2 and C11 are true Rpb9 homologs. It appears that the C-terminal domain of the Pol II subunit C11 is functionally and structurally homologous to the Pol II transcript cleavage factor TFIIS.

1. Composition and structure of Pol II

Purified by conventional column chromatography, the eukaryotic RNA Polymerase II consists of a 10-subunit catalytic core, which alone is capable of elongating the RNA transcript, and a complex of two subunits, Rpb4/7, that is required for transcription initiation. Structures of the complete yeast 12-subunit Pol II have been determined by X-ray analysis (Armache, Kettenberger et al. 2003; Bushnell and Kornberg 2003), as well as the structure of yeast Pol II in complex with the elongation factor TFIIS (Kettenberger, Armache et al. 2003). Pol II is an asymmetric and large multiprotein complex with a total molecular weight of 0.5 MDa. Five Pol II subunits, Rpb1, Rpb2, Rpb3, Rpb6 and Rpb11 show sequence and structural similarity in all cellular RNA polymerases and are referred to as the "core" subunits (Table 2). Rpb6, and four other subunits, Rpb5, Rpb8, Rpb10 and Rpb12, are shared between the three eukaryotic RNA polymerases I, II and III, and are referred to as the "common" subunits. The

10-subunit Pol II core comprises the core and common subunits and in addition, subunit Rpb9.

2. The large subunit carboxy terminal domain of Pol II

Rpb1, the largest subunit of Pol II, has an unusual C-terminal domain (CTD) composed of tandemly repeated copies of a serine-rich heptapeptide with consensus sequence Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7 (reviewed in (Young 1991). The CTD consensus sequence is repeated 26 times in yeast Pol II (Allison, Moyle et al. 1985) and 32, 45 and 52 times in the enzymes from *Caenorhabditis* (Bird and Riddle 1989), *Drosophila* (Zehring, Lee et al. 1988) and mammalian cells (Corden, Cadena et al. 1985), respectively. The CTD is essential for Pol II function; deletion mutations that remove most or all of the CTD are lethal in yeast (Nonet, Sweetser et al. 1987), *Drosophila* (Zehring, Lee et al. 1988) and mouse (Bartolomei, Halden et al. 1988) cells.

In vivo, Serines 2 and 5 (Ser2 and Ser5) were identified as major phosphorylation sites, and multiple functions for these modifications have been elucidated. More recently, phosphorylation of serine 7 (Ser7) and other covalent modifications have been described. Different phosphorylation states predominate at each stage of transcription, and each preferentially binds a distinct set of factors. For example, the Pol II phosphorylated on its Ser2 is mainly associated with initiation and is found at the promoter while Ser5 phosphorylation is associated with elongation and is found on the Pol II found the body of the gene.

These dynamic interactions provide a means for coupling and coordinating specific stages of transcription with other events necessary for proper gene expression (reviewed in (Buratowski 2009)). Briefly, the CTD serves as a flexible binding scaffold for numerous nuclear factors; the phosphorylation patterns of the CTD repeats during the transcription cycle determine which factors bind to it (Phatnani and Greenleaf 2006). Additionally, the CTD is required for efficient capping, splicing, cleavage and polyadenylation of mRNAs *in vivo*, and recruits RNA processing factors *in vitro* (Hirose and Manley 2000; Proudfoot, Furger et al. 2002; Buratowski 2009). Also coupled to transcription are chromatin remodeling and modification, DNA repair, mRNA packaging, RNA editing and nuclear mRNA export (Reed 2003; Sims, Mandal et al. 2004; Ares and Proudfoot 2005; Venters and Pugh 2009b).

B. General Transcription Factors (GTF)

General transcription factor	Subunits	Properties
TFIIA	37 kDa (α) 19 kDa (β) 13 kDa (γ)	Stabilization of TBP binding ; stabilization of TAF-DNA interactions ; antirepression functions
TFIIB	35 kDa	Pol II-TFIIF recruitment ; start site selection by Pol II
TFIID	38 kDa (TBP) 250 kDa (TAF1) 150 kDa (TAF2) 140 kDa (TAF3) 135 kDa (TAF4) 100 kDa (TAF5) 80 kDa (TAF6) 55 kDa (TAF7) 55 kDa (TAF8) 31 kDa (TAF9) 30 kDa (TAF10) 28 kDa (TAF11) 20 kDa (TAF12) 18 kDa (TAF13)	Core promoter recognition (TATA box) ; TFIIB recruitment ; TAFs function as co-activators
TFIIE	56 kDa 34 kDa	TFIIH recruitment ; modulation of TFIIH helicase, ATPase and kinase activities ; direct enhancement of promoter melting; Pol II elongation stimulation
TFIIF	58 kDa (RAP74) 26 kDa (RAP30)	Promoter targeting of Pol II ; destabilization of non-specific Pol II-DNA interactions; TFIIE and TFIIH recruitment into the PIC
TFIIH	89 kDa (XPB) 80 kDa (XPD) 62 kDa 52 kDa 44 kDa 40 kDa (cdk7/MO15) 38 kDa (cyclin H) 34 kDa 32 kDa (MAT1) 8 kDa	Promoter melting using helicase activity; promoter clearance by CTD phosphorylation; DNA repair

Table 3 : Subunit compositions and properties of the human GTFs.

GTFs have been defined biochemically as a set of factors essential for accurate transcription initiation at a TATA box-containing viral promoter *in vitro* (Orphanides, Lagrange et al. 1996; Roeder 2003). They have been named TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH (where TF stands for transcription factor) according to their chromatographic elution profiles and order of discovery (Matsui, Segall et al. 1980) (Table 3). The general model assumes that they are generically recruited to every Pol II promoters in order to allow transcription.

a. TFIID

TFIID is the GTF capable to recognize TATA elements in a variety of promoters (Parker and Topol 1984; Nakajima, Horikoshi et al. 1988). It is composed of the TATA-binding protein (TBP) and up to 14 TBP-associated factors (TAFs) (Table 3).

- TBP

The TBP subunit of TFIID is responsible for TATA box recognition. TBPs have been identified in a wide variety of organisms, emphasizing the universal requirement for this protein. TBP is a universal eukaryotic transcription factor (Sharp 1992; Hernandez 1993). TBPs range in size from 22 kDa (*Arabidopsis*) to 38 kDa (*Drosophila* and human) and can be divided into two structural domains. The conserved C-terminal domain consist of two imperfect direct repeats, can direct efficient and specific transcription initiation *in vitro* when combined with the remaining GTFs and Pol II (Horikoshi, Yamamoto et al. 1990; Peterson, Tanese et al. 1990), and is sufficient for formation of the complete TFIID complex (Zhou, Boyer et al. 1993). The N-terminus of TBP is divergent among species and its function is less clear.

- TAFs

In addition to TBP, TFIID contains 13-14 additional subunits – the TBP-associated factors (TAFs) – ranging in size from 20 to 250 kDa (Tora 2002) (Table 3). The TAFs are a phylogenetically conserved set of proteins identified in humans, *Drosophila* and yeast. The TAFs are required for activator-dependent transcriptional stimulation in human and *Drosophila* systems in vitro. Certain TAFs interact with activators to facilitate PIC formation

(Hoffmann, Oelgeschlager et al. 1997; Albright and Tjian 2000) and transcription reinitiation (Ainbinder, Revach et al. 2002).

TAFs also have a role in recognition and binding to core promoter elements. DNase I footprinting revealed direct contact of TAFs with sequences upstream and downstream of the TATA box (Sawadogo and Roeder 1985; Emanuel and Gilmour 1993; Kaufmann and Smale 1994). TAF1-TAF2 binds the Initiator element (Chalkley and Verrijzer 1999), multiple TAFs were crosslinked to the adenovirus major late promoter (Oelgeschlager, Chiang et al. 1996), and *Drosophila melanogaster* TAF6 (dmTAF6) and dmTAF9 were crosslinked to the downstream promoter element (Burke and Kadonaga 1997). Interestingly, some TAFs are specific to TFIID-specific and others are shared with the co-activator complex SAGA (Grant, Schieltz et al. 1998; Ogryzko, Kotani et al. 1998; Brand, Yamamoto et al. 1999b; Saurin, Shao et al. 2001).

One common feature found in 9 out of the 14 TAFs is the histone fold motif (Gangloff, Romier et al. 2001; Cler, Papai et al. 2009). This motif has been established as an essential protein-protein interaction domain that facilitates assembly of TFIID in a manner analogous to that for histones (Hoffmann, Chiang et al. 1996). TAF6 and TAF9 are structurally related to histones H4 and H3, respectively (Xie, Kokubo et al. 1996). TAF12-TAF4 and TAF11-TAF13 contains an H2-like domain, and interact with each other via the histone fold domain (HFD) (Gangloff, Werten et al. 2000; Guermah, Tao et al. 2001; Selleck, Howley et al. 2001). In vitro *Saccharomyces cerevisiae* TAF6 (scTAF6)-scTAF9 can assemble with scTAF12-scTAF4 to form a histone octamer-like structure (Selleck, Howley et al. 2001). The nucleosome-like interaction of TFIID with DNA and the presence of histone fold TAFs within this complex have led to the proposal that a nucleosome-like octamer within TFIID may be involved in direct DNA binding (Hoffmann, Chiang et al. 1996). However, the issue has remained elusive and, to date, it is unknown whether the HFDs of TAFs are involved in core promoter binding by TFIID (Cler, Papai et al. 2009).

b. TFIIA

cDNAs encoding TFIIA subunits have been isolated from yeast (Ranish, Lane et al. 1992), human (DeJong and Roeder 1993) and *Drosophila* (Yokomori, Admon et al. 1993). Human and *Drosophila* TFIIA consist of three subunits of 37 kDa (α subunit), 19 kDa (β subunit) and 13 kDa (γ subunit). The α and β subunits are the products of a single gene and

are produced by proteolytic cleavage of TFIIAαβ into TFIIAα and TFIIAβ subunits (De Jong and Roeder 1993).

The precise role of TFIIA in transcription initiation has been the subject of much controversy, because the requirement for TFIIA in reconstituted transcription varies from system to system. Generally, reactions reconstituted with recombinant TBP do not require TFIIA (Cortes, Flores et al. 1992), while reactions with purified TFIID are stimulated by TFIIA (Ozer, Moore et al. 1994). Much of this variability can be attributed to the ability of TFIIA to relieve the repressive effects of certain factors (including NC2/Dr1 (Inostroza, Mermelstein et al. 1992), topoisomerase I (Merino, Madden et al. 1993) and MOT1 (Auble, Hansen et al. 1994)) that may be present in cruder systems, a process known as antirepression.

c. TFIIB

Human TFIIB exists as a single polypeptide of 35 kDa (Ha, Lane et al. 1991). Homologs have been identified in *Drosophila* (Yamashita, Wada et al. 1992), yeast (Pinto, Ware et al. 1992) and archaeabacteria (Ouzounis and Sander 1992). The primary structure of TFIIB reveals some interesting features. The N-terminus of the protein contains a cysteine-rich sequence that forms a zinc-ribbon domain (Zhu, Zeng et al. 1996). Most of the remainder of the protein consists of two imperfect direct repeats of ~75 amino acids. Protease digestion of TFIIB defined the N- and C-terminal regions to be separate domains (Barberis, Muller et al. 1993).

The X-ray crystal structure of a complex of the C-terminal domain of TFIIB (TFIIBc) with TBP and TATA box fragment has shown non-sequence specific contacts of TFIIBc with the DNA upstream and downstream of the TATA box, as well as sequence-specific interactions with an element (BRE) immediately upstream (Lagrange, Kapanidis et al. 1998; Littlefield, Korkhin et al. 1999; Tsai and Sigler 2000). More recently, the structure of TFIIB in a complex with RNA Pol II was determined and revealed three features crucial for transcription initiation (Figure 3) (Bushnell, Westover et al. 2004). First, the N-terminal zinc ribbon of TFIIB contacts the “dock” domain of RNA Pol II, near the path of RNA exit from a transcribing enzyme. Second, the “finger” domain of TFIIB is inserted into the polymerase active center. Third, the C-terminal domain of TFIIB, whose interaction with both RNA Pol II and a TBP-promoter DNA complex orients the DNA for unwinding and transcription.

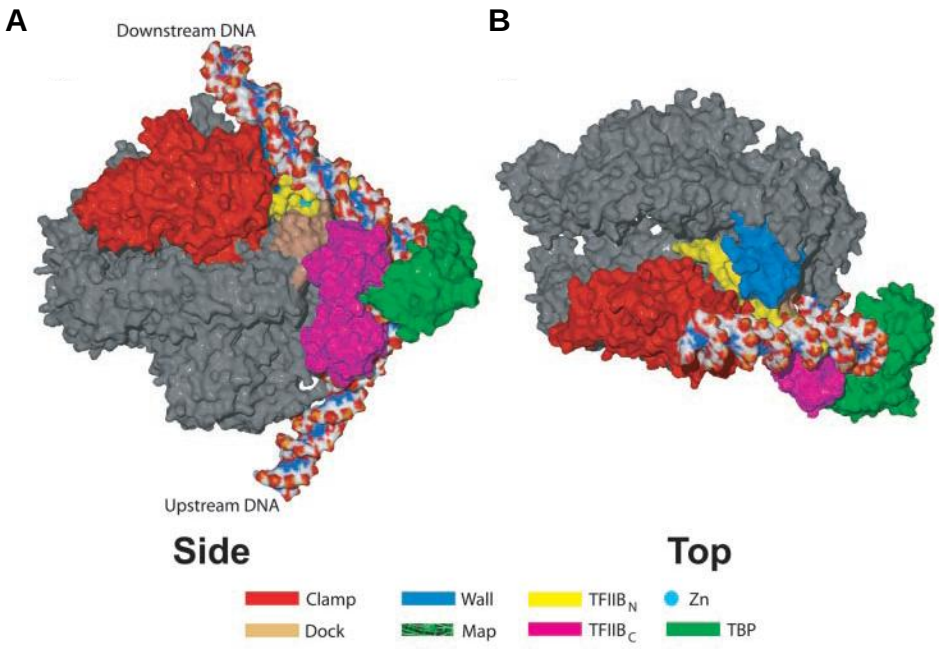


Figure 3 TFIIB in the Pol II-TFIIB complex and the inferred locations of TBP and promoter DNA in a transcription initiation complex.
A. Side view of the Pol II-TFIIB structure with a published model of a TFIIBc-TBP-TATA box DNA complex docked by least-squares alignement to the observed TFIIBc helix and with the TATA box DNA extended by the addition of 20 bp of B-form DNA at both ends. The colour code is shown at the bottom. B. Top view of the model in A. From (Bushnell, Westover et al. 2004).

The primary role of TFIIB is to physically link TFIID with the Pol II/TFIIF complex at the promoter. Consistent with this role, TFIIB contains binding sites for the TFIID-DNA complex (Buratowski, Hahn et al. 1989), TFIIF and Pol II (Ha, Roberts et al. 1993). Tethering Pol II and TFIIF to the promoter is not the only role of TFIIB in initiation. Mutations in the *S. cerevisiae* TFIIB gene (SUA7) can dramatically alter the position at which Pol II begins transcription (Pinto, Ware et al. 1992). A similar alteration in start sites accompanies mutations in genes encoding subunits of *S. cerevisiae* Pol II (Hekmatpanah and Young 1991). Thus, the interaction between TFIIB and Pol II is crucial for specifying the start site of transcription.

d. TFIIF

cDNAs encoding TFIIF subunits have been isolated from human (Sopta, Burton et al. 1989), *Drosophila* (Kephart, Price et al. 1993) and yeast (Henry, Campbell et al. 1994). Human TFIIF is a heterotetrameric factor consisting of 26 kDa (RAP30) and 58 kDa (RAP74)

subunits. Some of the functions of TFIIF can be accomplished by RAP30 alone. RAP30 can deliver Pol II to the promoter to support transcription initiation in the absence of RAP74. RAP74 is required for stimulation of the elongation rate of Pol II (Chang, Kostrub et al. 1993). In addition, the RAP30-RAP74 complex can remove Pol II from nonspecific DNA sites, whereas RAP30 alone can only prevent it from binding to these sites (Conaway and Conaway 1990; Killeen, Coulombe et al. 1992). RAP74 is heavily phosphorylated *in vivo*. Both initiation and elongation activities of TFIIF are stimulated by phosphorylation (Kitajima, Chibazakura et al. 1994). TFIIF is also critical for tight wrapping of DNA, possibly inducing torsional tension in the DNA, thereby facilitating promoter melting. Mutants in TFIIF that display transcription defects are also defective in DNA wrapping (Coulombe and Burton 1999).

e. TFIIH

TFIIH subunits have been progressively cloned and sequenced during the last two decades (i.e. (Gerard, Fischer et al. 1991; Fischer, Gerard et al. 1992)). Ten subunits are currently characterized, ranging from 8 to 89 kDa (Table 3). TFIIH can be separated into two subcomplexes, core TFIIH and a separable kinase/cyclin subcomplex (Feaver, Henry et al. 1997). TFIIH is known to possess several enzymatic activities (Svejstrup, Vichi et al. 1996). The p89/XPB and p80/XPD subunits are a 3'-5' and 5'-3' ATP-dependent helicases, respectively (Schaeffer, Roy et al. 1993; Schaeffer, Moncollin et al. 1994). The XPB ATP-dependent DNA helicase plays a major role in promoter opening (Holstege, van der Vliet et al. 1996; Kim, Ebright et al. 2000). Cyclin-dependent kinase 7 (cdk7), cyclin H and MAT1 form a ternary complex with a cdk-activating kinase (CAK) activity (Feaver, Svejstrup et al. 1994). Cdk7 was shown to phosphorylate the C-terminal repeats of the largest Pol II subunit (CTD) (Makela, Parvin et al. 1995), a modification that affects the elongation phase and the production of long run-off transcripts rather the initiation step (Akoulitchev, Makela et al. 1995). In addition to its role in transcription, TFIIH is involved in the nucleotide excision repair (NER) mechanism (Vermeulen, van Vuuren et al. 1994; Wang, Buratowski et al. 1995). Besides its role in the recognition of the lesion together with the two NER repair factors XPA and XPC (de Laat, Jaspers et al. 1999), TFIIH participates in the unwinding reaction that involves both helicases in order to allow DNA excision on both sides of the lesion (Evans, Moggs et al. 1997).

The structure of the human TFIIH complex has been determined by electron crystallography at 3.8 nm (Schultz, Fribourg et al. 2000) and 13 Å resolution (Chang and Kornberg 2000), but docking to the initiation complex model is not possible with currently available information. However, TFIIH can be placed approximately (Figure 4), on the basis of interaction between its p62 subunit and the largest subunit of TFIIE, whose locations within the structures revealed by EM have been inferred from other evidences.

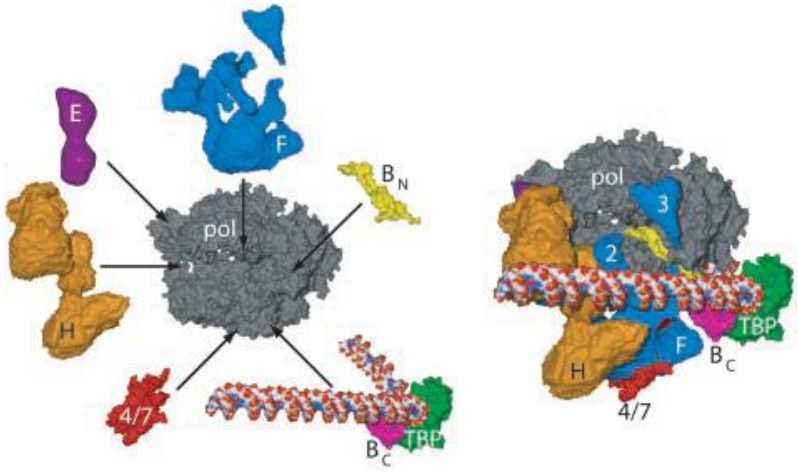


Figure 4 : Model of the RNA Pol II transcription initiation complex.
Electron microscopy structures of TFIIE, TFIIF and TFIIH are added to the X-ray structure obtained for the TFIIB-Pol II complex. An exploded view is shown at the left, to reveal the individual structures and their relationship to the Pol II structure more clearly. Regions of the second largest subunit of TFIIF corresponding to domains 2 and 3 of bacterial sigma factor are indicated (« 2 » and « 3 »). From (Bushnell, Westover et al. 2004).

f. TFIIE

cDNAs encoding TFIIE subunits have been isolated from human (Purrello, Di Pietro et al. 1994), *Drosophila* (Wang, Hansen et al. 1997) and yeast (Feaver, Henry et al. 1994). TFIIE is composed of two highly charged subunits with a molecular mass of 56 kDa (TFIIEα) and 34 kDa (TFIIEβ) (Ohkuma, Sumimoto et al. 1990) (Table 3). TFIIE enters the preinitiation complex after Pol II and interacts directly with the unphosphorylated form of Pol II, with TFIIB and with both subunits of TFIIF (Orphanides, Lagrange et al. 1996; Roeder 1996). TFIIE is required for the recruitment of TFIIH and for the regulation of its kinase and helicase activities (Lu, Zawel et al. 1992; Ohkuma and Roeder 1994). The TFIIE α-subunit is required for the specific interaction with TFIIH. The central core region of the β-subunit that

binds to double-stranded DNA is essential for basal and activated transcription, and a C-terminal half contains two basic stretches that interact with TFIIB, TFIIFβ and single-stranded DNA (Okamoto, Yamamoto et al. 1998). Protein-DNA crosslinking experiments showed that both TFIIE subunits bind to the start site of the promoter between positions -10 and +10 (Kim, Ebright et al. 2000).

g. Transcriptional machinery assembly

An ordered assembly of the transcription PIC was originally proposed on the basis of the formation of active transcription complexes *in vitro* (Buratowski 1994). Steps leading to Pol II transcription defined biochemically include: (1) the binding of TBP to the TATA element (Burley and Roeder 1996); (2) the binding of TFIIA to this complex (Buratowski, Hahn et al. 1989) through direct contacts both with TBP and with upstream DNA sequences (Geiger, Hahn et al. 1996; Tan, Hunziker et al. 1996); (3) the binding of TFIIB through direct interactions with TBP and with DNA sequences both upstream and downstream of the TATA element (Nikolov, Chen et al. 1995) that stabilizes TBP-TATA interactions; (4) the binding of a pre-formed TFIIF-Pol II complex, through direct interactions of TFIIB with both TFIIF and Pol II (Leuther, Bushnell et al. 1996); (5) the TFIIE binding through direct interactions with Pol II and potentially TFIIF and TBP (Zawel and Reinberg 1993; Maxon, Goodrich et al. 1994) that recruits TFIIH to complete the assembly of the PIC.

However, it is now clear that the core Pol II initiation machinery is more elaborate than previously anticipated and other transcriptional co-regulatory complexes described later participate also in this process. When combined with the profusion of additional co-factors observed to interact with the core machinery and required to regulate activated transcription, the assembly that may represent an initiation complex could be extraordinary large.

C. Coactivator complexes

As shown in the previous chapter, Pol II transcription machinery is by itself quite complicated, comprising numerous GTFs that are sufficient to promote efficient basal transcription *in vitro.* This indicates that the machinery possesses on its own an innate ability to efficiently carry out transcription. Thus the requirement of additional factors to control transcription was not initially suspected. In the last decades, coactivator complexes have emerged as central players in the transcription regulation process. They are defined as factors

that have the ability to enhance transcription that are recruited through DNA binding activators (or transcription factor), but are unable to bind DNA themselves. In the present section, I will give an overview on the function of the different complexes known to act as coactivators in eukaryotic transcription processes.

1. Mediator

The Mediator complex was initially indentified in yeast as an activity that helped to stimulate activator-dependent transcriptional activity in reconstituted transcription reactions (Fondell, Ge et al. 1996). Since its original description in yeast, Mediator has been isolated from mammalian cells using diverse approaches, showing a ~30 subunits composition partially conserved from yeast to humans (Table 4) (reviewed in (Conaway, Sato et al. 2005)). Structural studies by EM of the Mediator and its interaction with the RNA Pol II were conducted. They showed that upon incubation with RNA polymerase II, Mediator undergoes a large-scale conformational change and adopts an extended conformation in which separate structural/functional domains appear. This led to the proposal of a model of organisation for the mediator with four functional modules (head, middle, tail, kinase) (Asturias, Jiang et al. 1999).

The exact mechanisms by which mammalian Mediator complexes control mRNA synthesis have to date not been firmly established. However, substantial evidence argues that Mediator activates transcription, at least in part, via direct interactions with DNA-binding transcriptional activators bound at enhancers, with Pol II and, most likely, with one or more of the general initiation factors bound at the core promoter (Conaway, Sato et al. 2005). For example, the mammalian MED1 protein has an essential role in transcriptional activation by the large family of nuclear receptors. MED1 binds directly to liganded nuclear receptors, and depletion of MED1 specifically disrupts nuclear-receptor function (Yuan, Ito et al. 1998; Ito, Yuan et al. 2000).

The role of each subunit of the complex is not yet clear, however, studies with particular focuses have helped to assign functional categories to some of them. For example many subunits have been shown to mediate direct interaction with multiple transcription factors (MED1, 14, 15, 17, 23, 25, 29) (Conaway, Sato et al. 2005). These multiple associations are thought to be part of the basis of one of the main function of the complex that is to mediate molecular interactions between transcriptional activators. Another central

activity of the complex is thought to be mediated by the kinase module that has the ability to negatively regulate transcription by phosphorylating and inactivate one or more of the general initiation factors. Consistent with this possibility, CDK8–Cyclin C can phosphorylate the Cyclin H subunit of mammalian TFIIH, blocking the activity of the TFIIH CTD kinase and inhibiting TFIIH activity in transcription (Akoulitchev, Chuikov et al. 2000).

KINASE	CDK8
	CDK11
	CYCLINC
	MED12
	MED13
HEAD	MED6
	MED8
	MED11
	MED17
	MED18
	MED19
	MED20
	MED21
	MED22
	MED28
	MED29
	MED30
MIDDLE	MED1
	MED4
	MED7
	MED9
	MED10
	MED21
TAIL	MED14
	MED15
	MED16
UNASSIGNED	MED23
	MED24
	MED25
	MED26
	MED27
	MED31

Table 4 : Subunit composition of the Human mediator complex.
The mediator complex exerts a modular organization with four sub-modules (kinase, head, middle and tail). (Adapted from (Conaway, Sato et al. 2005))

On the first glance, the antagonistic effects on transcriptional activation of the mediator complex are somewhat surprising. Interestingly, intensive biochemical characterisation of the complex by purification using different procedures revealed that in higher eukaryotes, at least two mutually exclusive forms of the mediator co-exist (Conaway, Sato et al. 2005; Paoletti, Parmely et al. 2006), (i) a repressive (named TRAP) form containing the kinase module but devoted of the MED26 subunit and (ii) an active form (named PC2) (Kretzschmar, Stelzer et al. 1994), containing MED26 but where the kinase (or part of it) is lost. Based on these observations, a model was proposed suggesting that the two forms of the complex would act stepwise in the transcriptional activation process (Figure 5) (Malik and Roeder 2005). It is proposed that the repressive Mediator is initially recruited to the promoter. This could be followed by facilitated entry of Pol II into the PIC, forming a transient intermediate. At some point, possibly with the addition of MED26, Mediator could lose the kinase module and be converted to an active form. Finally, the newly generated (or activated) mediator could act upon the pre-assembled PIC either to facilitate the first initiation event or a subsequent step as part of the re-initiation scaffold.

Thus, in current models, the Mediator coactivator effects are proposed to be composite of multiple events. First, in conjunction with the activator, Mediator promotes PIC assembly. Second, it stimulates the basal activity of the PIC and likely favours Pol II loading on the DNA.

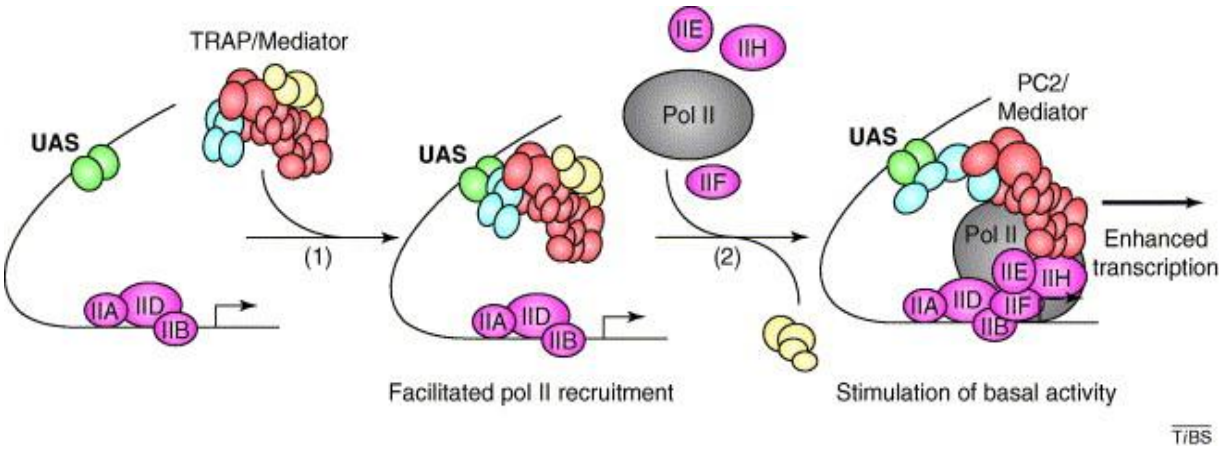


Figure 5 : Model for activated transcription involving interconversion of Mediator forms.
In this model, transcriptional activators bind to their target sites Upstream Activating Sequence (UAS) and recruit the repressive form of Mediator **(1)**. The Mediator could go on to

facilitate, in turn, the recruitment of Pol II into the nascent PIC **(2)**. This could be accompanied by loss of the kinase module (yellow) and MED26 integration coupled with transcription initiation. The activating form of the Mediator that would be generated could exert stimulatory effects on the PIC. (adapted from (Malik and Roeder 2005))

2. ATP-dependent chromatin remodelling complexes

Remodeling complexes were defined as complexes that process a DNA-stimulated ATPase activity and can destabilize histone DNA interactions in reconstituted nucleosomes in an ATP-dependent manner. The first remodelling activity identified was SWI/SNF (SWItch/Sucrose NonFermentable) in yeast (Cote, Quinn et al. 1994). SWI/SNF function was first linked to chromatin when extragenic suppressors of yeast swi/snf mutants were found to encode histone proteins or other components of chromatin (Winston and Carlson 1992). Subsequently, activities of yeast and human SWI/SNF complexes were found to alter nucleosome structure and facilitate transcription factor binding in an ATP-dependent reaction (Cote, Quinn et al. 1994; Imbalzano, Kwon et al. 1994).

Since then several proteins containing an ATP-dependent remodelling activity were identified. So far, all of these factors have been found to be contained in multi-subunit complexes that harbour an ATPase subunit, which belongs to the SWI/SNF superfamily. ATP-dependant remodelling factors were later sub-divided into four separated classes based on the presence of structural domains found on their ATPase subunit: SWI/SNF, ISWI, CHD and INO80 classes (Figure 6). Despite their structural similarities, the different members of the remodelling complexes family appears to act with different mechanisms on the chromatin structure affecting transcription in a positive or negative manner (Figure 7).

a. SWI/SNF family

The SWI/SNF family of chromatin remodelers, which also includes the RSC (Remodels Structure of Chromatin) complex, is generally viewed as a positive regulator of transcription (Angus-Hill, Schlichter et al. 2001). Location profiling by ChIP-chip finds RSC at the promoters of several hundred genes (Damelin, Simon et al. 2002). This observation suggests a role for these complexes in transcription initiation and nucleosome organization. Consistent SWI/SNF, in particular, might have additional functions in transcription elongation (Schwabish and Struhl 2007).

The presence of a bromodomain, that mediates recognition of acetyl groups, in the SWI/SNF family of chromatin remodelers suggested an involvement of histone acetylation in their recruitment (Figure 6). Several studies have demonstrated a direct functional link between HATs and SWI/SNF family remodelling complexes. For example, *in vitro* transcription studies show that nucleosome acetylation by NuA4/TIP60 stimulates the activity of RSC (Carey *et al.*, 2006). Moreover, SWI/SNF has been shown to be dependent on SAGA mediated histone acetylation to displace nucleosomes (Chandy, Gutierrez et al. 2006). Histone acetylation favours but is likely not sufficient to recruit remodelling complexes to promoters. It has been shown that both RSC and SWI/SNF are also directed to some promoters through interactions with activators (Cosma, Tanaka et al. 1999; Yudkovsky, Logie et al. 1999).

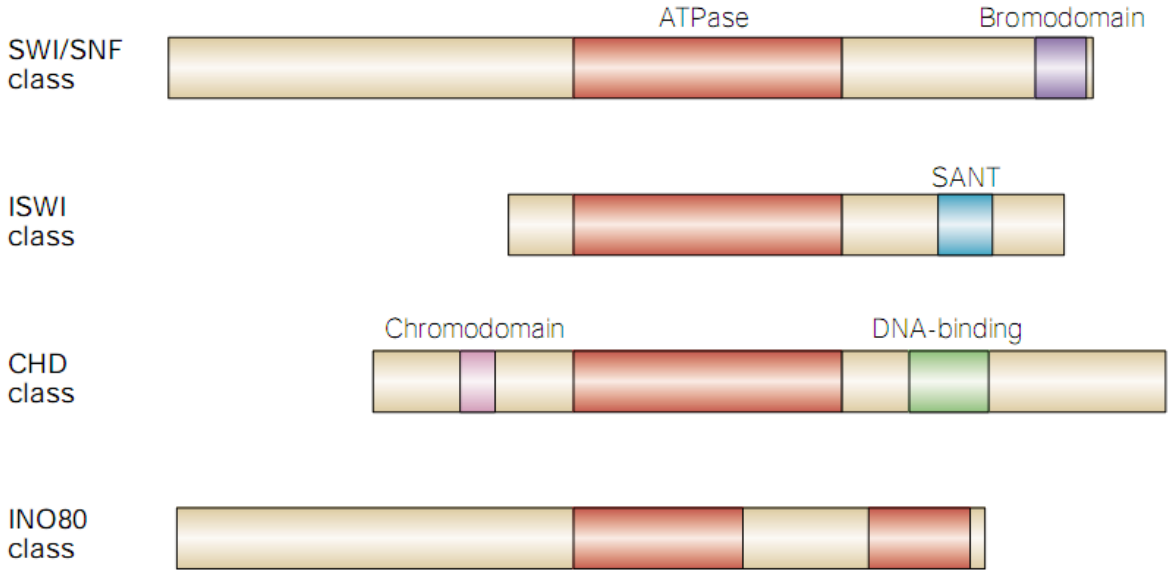


Figure 6 : Structure of four classes of ATPase in ATP-dependent chromatin-remodelling factors.
ATPase domains are shown in red. Signature motifs in each class — the bromodomain (dark blue), the SANT domain (light blue), the chromodomain (pink) and a putative DNA-binding domain (green). (Adapted from (Tsukiyama 2002))

b. ISWI family

In contrast to the SWI/SNF and INO80 families, the Imitation SWItch (ISWI) family of remodelers tends to negatively regulate transcription. For example, genome-wide expression profiling and DNase I sensitivity studies in *Saccharomyces* found that the ISW2 complex in concert with the histone deacetylase Rpd3 (Reduced Potassium Dependency 3)

represses meiotic genes by creating a repressive nucleosome arrangement (Fazzio, Kooperberg et al. 2001).

c. CHD family

The defining feature of the CHD family is that it contains a chromodomain, which binds to methylated lysines. Indeed, it was shown that Chd1 interacts with methylated H3K4 in vitro (Pray-Grant, Daniel et al. 2005), thus suggesting a role in active transcription mediation. However, the *in vivo* significance of Chd1 binding to H3K4me remains elusive.

The precise function of the CHD (Chromatin organization modifier, Helicase, and DNA-binding domains) family of chromatin remodelers is the least understood of the four families of remodelers. The major observation to date that gives an insight on its function was made using expression profiling in a Chd1 yeast mutant. It has shown that very few genes were affected by Chd1 disruption (Tran, Steger et al. 2000), suggesting that Chd1 loss might be compensated by functional redundancy with other remodelers or that it is targeted to few genes.

d. INO80/SWR1 family

The INO80/SWR1 (INOsitol requiring 80/Sick With Rat8 ts 1) family of remodelers is unique in that it contains ATPase domains. The INO80 complex likely plays a broader role in genome regulation than many other remodelling complexes. It was shown to participate in various processes including transcription activation, DNA repair, and resolving stalled replication forks (Shen, Mizuguchi et al. 2000).

For example, the SWR-C/SWR1 complex is a chromatin remodeler that alters the composition but not the position of nucleosomes. The SWR1 complex was shown to replace H2A with its variant H2A.Z in promoter nucleosomes (Guillemette, Bataille et al. 2005; Zhang, Roberts et al. 2005). This links these complexes with transcriptional activation since H2A.Z variants are thought to promote transcription by destabilizing nucleosomes.

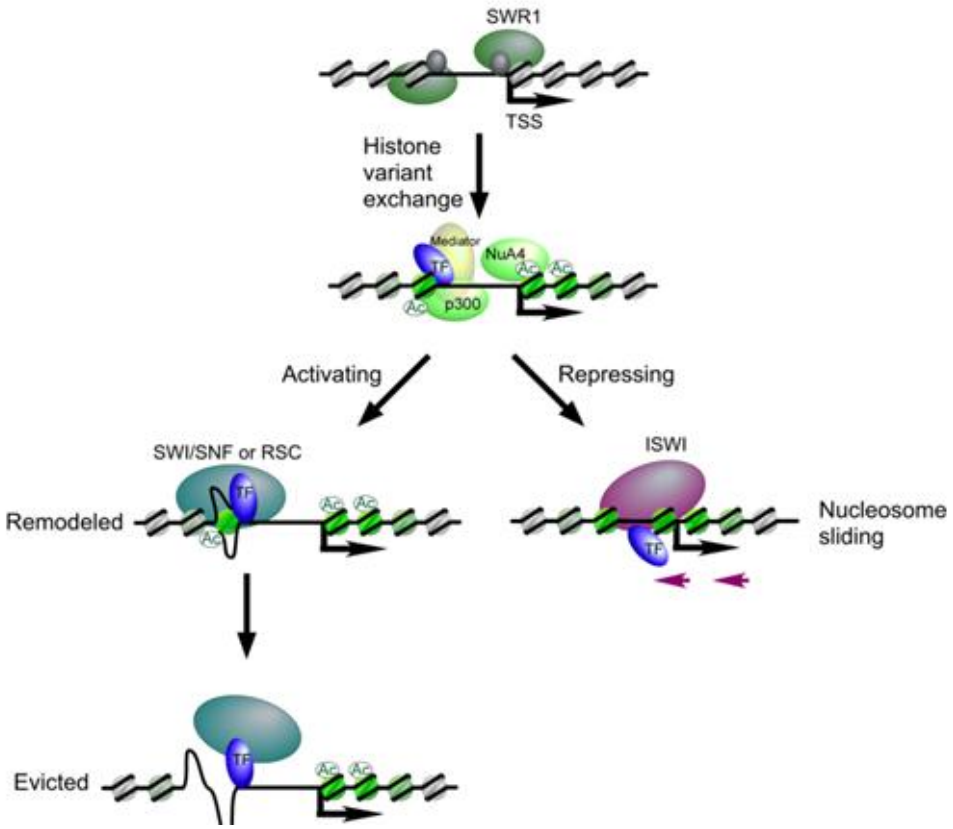


Figure 7: Different modes of action for chromatin remodeler families.
Chromatin remodeler families act differently on the structure of the chromatin. In the activating pathway, H2A.Z nucleosomes are incorporated by SWR complexes, SWI/SNF, RSC open the chromatin at active regions. Some of this may be facilitated by histone acetylation (SAGA and NuA4/TIP60). In the repressing pathway, ISWI complexes in concert with HDACs close the chromatin. (Adapted from (Venters and Pugh 2009b))

3. Histone acetyl transferases (HATs)

p55 was the first enzyme isolated (from *Tetrahymena*) bearing an acetyl transferase (AT) activity and shown to target histone proteins (Brownell and Allis 1995). This enzyme was then shown to be the homologue of the eukaryotic General Control Nonderepressible 5 (GCN5). This enzyme was the funding member of a new class of histone modifying enzymes, the Histone Acetyl Transferases (HATs) that were later shown to play a role in transcriptional regulation through chromatin. Since then many HATs were identified and a classification was proposed based on the structure of their catalytic site. In higher eukaryotes, three enzymatic families have been identified, GNAT, MYST, p300/CBP. Most of the HATs have been shown to be subunits of larger complexes (with the notable exception of CBP/p300).

Knowledge gained from *in vitro* transcription systems as well as ChIP assays allowed to narrow down the general function of HATs. They are principally thought to acts as transcriptional co-activators (reviewed in (Roth, Denu et al. 2001)). Mechanistically, histone acetylation is believed to increase the decompaction of chromatin, which in turn may increase the accessibility of factors that promote transcription (Krajewski and Becker 1998b; Akhtar and Becker 2000; Sterner and Berger 2000a; Shogren-Knaak and Peterson 2006). More precisely, the current vision is that HATs are recruited through the transcriptional activation process by DNA binding activators, leading to histone acetylation that acts both on the opening of the chromatin structure and as a signature for recognition by other activator complexes (i.e. recruitment of SWI/SNF remodelling complexes family).

The involvement of HATs in other processes than gene specific activation has been recently described. The MYST family complex Male specific Lethal (MSL) (containing the MOF HAT subunit) has been shown to globally enhance transcription on the X sex chromosome but not the autosomes in the process of dosage compensation in flies (Morales, Straub et al. 2004; Kind, Vaquerizas et al. 2008). The precise mechanism of this process is not completely understood yet, but suggests that HATs not only stimulate expression though the classical gene specific pathways, but may play key roles in more complex regulatory phenomena.

While the general role of the HATs in transcription is relatively clear, the need for such a large number of specific enzymes is not yet understood. This issue of functional specificity is exhaustively reviewed and discussed in the Results section of this manuscript (Anamika K, Krebs A et al, submitted) page 88). Furthermore, knowledge on GCN5-containing complexes is extensively reviewed in a separate chapter of this manuscript (page 48)

4. Model for the coordinated action of coactivator complexes

Even though many aspects on the mode of action of coactivators remains elusive (targeting specificity, collaborative action...), models have been proposed to describe their action in an integrative manner. Figure 8 describes a stepwise model proposed by (Naar, Ryu et al. 1998) to try to explain how known coactivtor families could act cooperatively to stimulate transcription initiation.

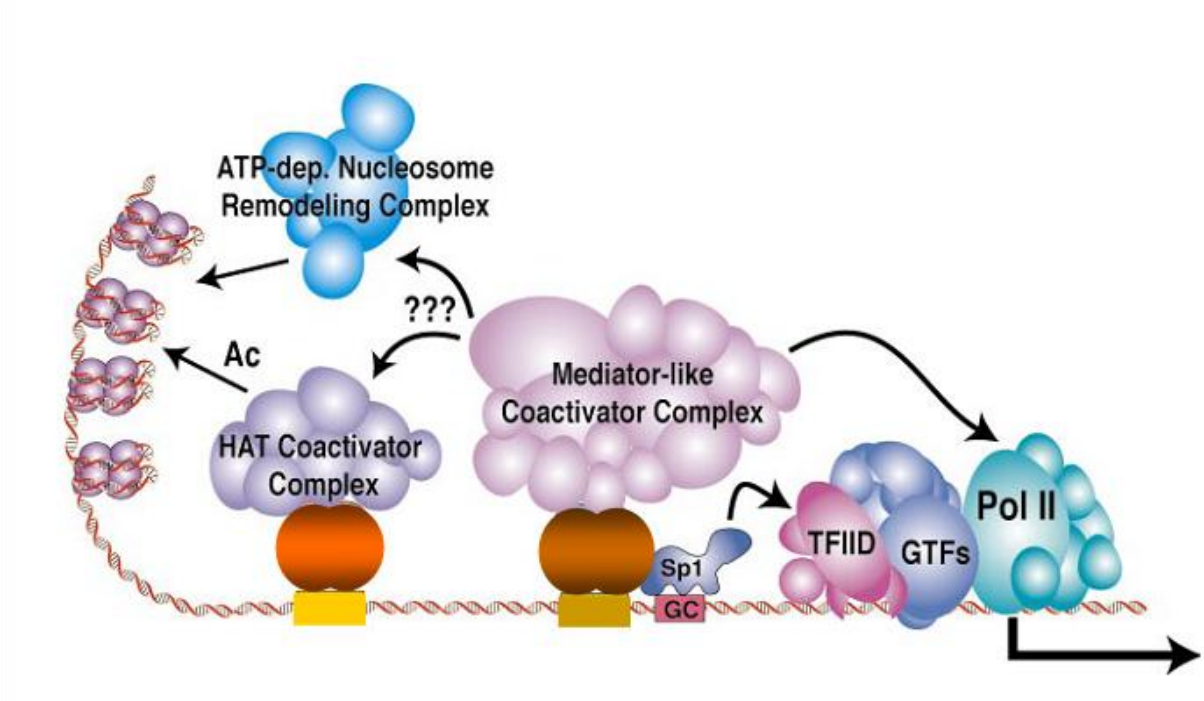


Figure 8 : A model of some of the different activities displayed by distinct types of transcriptional coactivator complexes in regulating gene activity.
ATP-dependent nucleosome remodelling complexes (light blue), which may or may not be directly recruited by sequence-specific activators, help activators and the transcriptional apparatus to gain access to their cognate enhancer/promoter sequences by mobilizing nucleosomes. Histone acetyltransferase (HAT) containing coactivator complexes (light purple) have been shown to be recruited by activators (orange circles) bound to a specific DNA sequence (light orange rectangle) to specific target genes. HATs may acetylate nucleosomes to counter repressive effects of local chromatin structure. HAT complexes could also acetylate other target polypeptides, such as the activators themselves. Mediator-type coactivator complexes (lavender) may be directed to select genes by interaction with activators (brown circles) bound to a specific DNA sequence (light brown rectangle), and to mediate recruitment of RNA polymerase II (Pol II, green) and, possibly, chromatin-directed activities. Other activators, (dark blue) may bind to TAF components of TFIID (light red) and serve to recruit this coactivator complex to nucleate the formation of a pre-initiation complex of general transcription factors (GTFs, blue) and Pol II, which can then proceed to initiate gene transcription.

III. Perspectives from genome wide studies on transcriptional activation process

Over the past fifteen years, several major technical advances have allowed the research community to reconsider the way to address questions concerning the analysis of transcription regulation. The completion of the Human Genome Project in 2003 has provided a road map for large-scale interrogation of gene functions and expression regulation. The single gene scale approach has progressively been replaced and/or complemented by systematic studies at the scale of the whole genome. These studies are progressively building the global rules defining transcription as a complex regulatory system. The classical approach, studying molecular mechanisms at a single gene scale and extrapolate the results to the whole genome, has been inverted. Now, the observations are directly made at the scale of the genome to try to deduce mechanistic rules applicable at the single gene level. Furthermore, these new approaches allowed the establishment of more complex regulatory rules by constructing interconnected networks and hubs. In the present section, I will review through several key examples, the new insight gained by the appearance of genome scale investigation techniques in the understanding of general transcription mechanisms and their integration to complex regulatory models. First I will discuss how these techniques were used to de novo annotate genomic elements. Second, I will present the evidences that led to attribute an increasingly important role of chromatin in transcription regulation. Third, I will describe how new mechanistic concepts could emerge out of global analyses of transcriptional actors positioning and the interplay between chromatin modifications and transcription states.

A. Annotation of the genome

1. Annotation of functional elements

In order to have a more comprehensive view of the human genome, several steps of annotation of the genomic landscape were initiated. To define where transcription starts exactly from the transcription units several genome-wide sequencing-based high-throughput methods that require reliable isolation of full-length cDNAs, sequencing of their 5' ends and mapping of the sequence to a completed genomic DNA sequence were used. The sequencing stage can use the 5' ends of cloned full-length cDNA libraries (so-called 5' ESTs), short tags

derived from 5' ends of capped RNAs (cap analysis of gene expression, CAGE) and 5'-SAGE (serial analysis of gene expression), or tags derived from 5'–3' ends (so-called paired-end tags) (reviewed in (Sandelin, Carninci et al. 2007). This intensive mapping allowed to predict the location of the transcription start sites (TSSs) and to define the promoter regions of most of the expressed genes (Bajic, Tan et al. 2006).

In addition to these sequencing-based methods, another, albeit less precise, approach to localize promoters is to use a genome wide location analysis (GWLA) identifying the genomic binding sites for the general transcription machinery associated with transcription start sites. By using the ChIP-on-chip technique with specific antibodies against subunits of different components of the basal transcription machinery (e.g. TFIID or RNA Polymerase II (Pol II)) that bind to the core promoters of different genes, many active promoters were mapped genome wide from different cell lines (Kim, Barrera et al. 2005a) demonstrating the validity of the method.

This technical breakthrough was part of a larger ambitious consortium project named **ENC**yclopedia **O**f **D**NA **e**lements (ENCODE). The goal of ENCODE is to identify all functional DNA elements in the human genome using a large set of high throughput techniques. The pilot phase (from 2003 to 2007) focused on 1% of randomly and non-randomly selected regions of the genome in order to develop the methodology for the analysis at a reasonable scale. During the first phase of this analysis, the project could not only map previously known, as well as novel functional elements using heterogeneous datasets, but also give new insights on the epigenetic signature of those elements (Birney, Stamatoyannopoulos et al. 2007). For example, the study by (Birney, Stamatoyannopoulos et al. 2007) shows that chromatin accessibility (presence of DNase I hypersensitivity sites (DHS)) and specific histone modification patterns are highly predictive for both the presence and the activity of a TSS. By contrast, combination of DHS, with another set of histone marks at regions distal from detected TSSs, could also define regulatory regions (i.e. insulator or enhancer binding sites). One can anticipate that the completion of the production phase of the ENCODE project (launched in 2007), targeting the whole genome, will contribute to transform the human genomic landscape from a plain base pair sequence to a fully annotated regulatory catalogue.

2. Pol II promoter characterisation

In order to be able to understand Pol II transcription regulation, a genome wide characterisation of core promoter composition in higher eukaryotes is a prerequisite. Using the genome wide promoter mapping data recently generated (discussed earlier), some general composition and regulation rules have emerged. In the generally accepted model, a promoter used to be defined by a TSS and several regulatory elements located in its close vicinity. Among those elements, the canonical TATA box was thought to be the main marker defining a promoter.

Interestingly, it rapidly became clear from the first genome wide studies that a vast majority of mammalian promoters were lacking the TATA box. Furthermore, the CAGE-based approaches indicated that most of the mouse and human promoters lack a distinctive sharp TSS, but rather harbour a broad array of closely located TSSs over 50-100 bp (Carninci, Kasukawa et al. 2005; Bajic, Tan et al. 2006; Carninci, Sandelin et al. 2006). These observations provide the basis for a new system of promoter classification based on the TSS distribution ("sharp" or "broad"). It has been observed that the presence of a TATA box is more often associated with promoters that have a single, sharply defined TSS. On the contrary, CpG islands have been shown to be overrepresented in the "broad" promoter categories (Sandelin, Carninci et al. 2007). Recently, some functional relevance of this new classification has emerged from a broad tissue GWLA of Pol II. This study demonstrates that sharp, TATA containing promoters are primarily used for tissue-specific expression, whereas broad, CpG island containing promoters are generally associated with ubiquitously expressed genes (Barrera, Li et al. 2008).

Altogether, the genome wide approaches generated novel datasets that helped to redefine the promoter nomenclature established by single gene scale studies and opened new perspectives in the understanding of the transcription regulation mechanisms at a larger scale.

B. An increasing role for chromatin in transcriptional regulation

1. Nucleosome positioning

In comparison with prokaryotes, the presence of chromatin in eukaryotes adds a layer of complexity to the transcriptional regulation mechanisms. Before being able to get access to a particular region of DNA to transcribe it, the transcription machinery has to overcome the nucleosome barrier. Two models have been proposed to explain the functional relationship between the transcription machinery and the chromatin. One possibility is that transcription factor binding at promoters acts through nucleosomes to activate or repress gene expression (Li, Carey et al. 2007). However, studies of the *S. cerevisiae PHO5* promoter suggested that nucleosomes are simply evicted from promoters, and the resulting naked DNA would allow transcription factors to gain access to their binding sites and for the basal transcriptional machinery to assemble (Boeger, Griesenbeck et al. 2003; Reinke and Horz 2003). These somewhat contradictory models illustrate the importance of associating the study of transcription regulation processes with the study of chromatin structure dynamics at regulatory regions.

Since the discovery of sites that exert a high sensibility to DNAse treatment in chromatin in the early eighties DNAse Hypersensitivity Sites (DHSs), the prevalent hypothesis was that those regions would correspond to nucleosome free regions. However, until improvements in the investigation methodology came about, it could not be demonstrated that the DHSs were due to the absence of nucleosomes in these regions. Recently, several genome wide studies addressed this question at different scales, significantly improving the knowledge on the chromatin structure.

The first strategy, focused on 480kbp of the *S. cerevisiae* genome at a very high resolution (20bp tiled DNA microarrays) (Yuan, Liu et al. 2005) whereas the other study covers the whole genome at a low resolution (1kb) (Bernstein, Liu et al. 2004). The use of very high resolution arrays led to the concept of constitutive nucleosome free region (NFR) spreading on a ~150bp region immediately upstream of the TSS, flanked on both sides by well positioned nucleosomes, that are deacetylated and enriched in H2A.Z (a histone variant of H2A) (Yuan, Liu et al. 2005) (Figure 11). Such NFRs were observed for most of the genes studied, with no correlation with their relative expression status. NFRs were associated in this

study with multiple stretches of poly-A or poly-T known to be associated with nucleosome instability. From this data, the authors concluded that NFRs are intrinsic components of promoter regions and thus, do not appear due to active removal of nucleosomes during transcription initiation. On the other hand, when studied at the whole genome scale and compared to transcription levels, the presence of nucleosome depleted regions seemed to correlate mainly with active regulatory elements (Bernstein, Liu et al. 2004). This apparent contradiction could be reconciled, however if upon gene induction regions flanking the fixed constitutive NFRs would lose additional nucleosomes. This model was further supported by a study where nucleosome positioning was investigated at a single DNA molecule level (Gal-Yam, Jeong et al. 2006). The authors could show for a specific gene that a constitutive NFR is observed at the promoter region and that upon induction, additional nucleosomes are lost downstream, but not upstream of the constitutive NFR. The recent extension of nucleosome mapping to regions distant from the TSS by the use of genome wide ChIP-seq revealed that NFRs are not restricted to TSSs, but are also present at the 3' end of most *S. cerevisiae* genes (Mavrich, Ioshikhes et al. 2008; Shivaswamy, Bhinge et al. 2008). The location of those NFRs coincides with the polyadenylation sites, suggesting that the 3'NFRs might be involved in transcription termination.

The conservation of NFRs in higher eukaryote genomes has been addressed by several studies. Initial observations were made over limited regions of the human genome, but at a high resolution, assessing the presence of NFRs at promoters of human cells (Heintzman, Stuart et al. 2007; Ozsolak, Song et al. 2007). These studies associated the presence of NFRs with the presence of a pre-initiation complex and could not detect NFRs at promoters of silenced genes. More recently, a study covering the whole genome using high throughput sequencing, generated genome-wide maps of nucleosome position in both resting and activated human T cells (Schones, Cui et al. 2008). The authors found, similar to the above described *S. cerevisiae* data, that in human cells nucleosomes are highly phased and NFRs are formed around the TSS of expressed genes. However the nucleosome phasing disappears for silent genes. Furthermore, by comparing the high resolution nucleosome position map with previously generated histone modification mark maps, these studies attributed particular modification patterns (mainly histone H3K4me3 and H2A.Z) to the highly positioned NFR flanking nucleosomes (see Figure 9). Altogether, these observations show an evolutionarily constrained role for this phenomenon in transcriptional regulation. Furthermore, these results

support the model in which the formation and the maintenance of NFR in higher eukaryotes are dependent upon downstream gene activation.

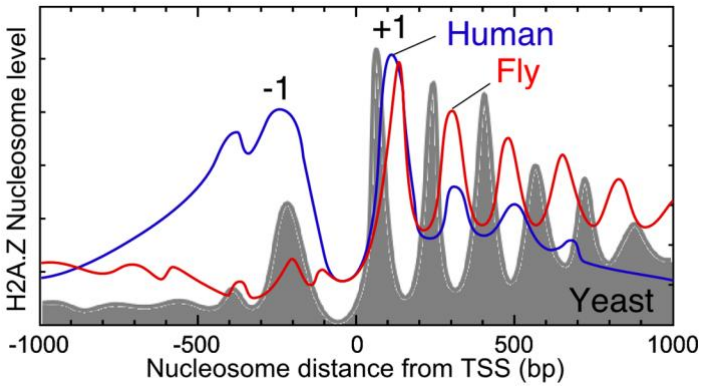


Figure 9 : The organization of nucleosomes throughout the genome.
The promoters of most genes reside in an open chromatin state in which they are competent to undergo activation. Frequency distributions for nucleosomes relative to the transcriptional start site (TSS) of all genes in budding yeast, fly, and human CD4+ T cells determined by ChIP-seq (Albert et al., 2007; Mavrich et al., 2008b; Schones et al., 2008). (Adapted from (Venters and Pugh 2009b))

As mentioned above, the chromatin can be considered as a barrier preventing the transcription machinery to access the DNA. The deciphering of the entire regulation of different chromatin states and their functional consequences will constitute a major step in the understanding of the transcription regulation processes.

2. Role of histone post translational modifications

The presence of NFR is not the only chromatin feature of regulatory regions. The systematic mapping of histone modifications by the high throughput ChIP techniques helped in defining the chromatin signatures associated with DNA regulatory regions. The individual patterns and the transcriptional state associated with the most characterised modifications have been reviewed in (Li, Carey et al. 2007) and shown in Figure 10. This shows that a high correlation is found at the genome scale between a particular marks, a region in the gene structure and a transcriptional state. This argues for an important functional role of chromatin modification in transcriptional regulation.

The integration of the data from these studies allowed creating combinatorial maps of these histone modifications, which defined special regulatory regions and thus, could also be used to predict such regions. The early data integration efforts, focused on a limited number

of marks, allowed to define simple signatures sufficient to characterise higher eukaryotic promoter and enhancer regions (e.g. (Heintzman, Stuart et al. 2007),(Heintzman, Hon et al. 2009)). According to these studies, active promoters were associated with H3K4 tri-methylation , histone H3-H4 global acetylation and H2A.Z enrichment. Similarly, enhancer regions were associated with histone H3K4 mono-methylation and H3K27 acetylation, but not histone H3K4 tri-methylation (Heintzman, Hon et al. 2009). It was also observed that NFR was a marker shared by both promoter and enhancers regions (Heintzman, Stuart et al. 2007). Despite the fact that those correlations could robustly be used for predicting promoter or enhancer regions, the limited set of markers considered in these studies did not reflect the entire complexity of all the chromatin marks, which have been implicated in potential transcriptional regulatory networks.

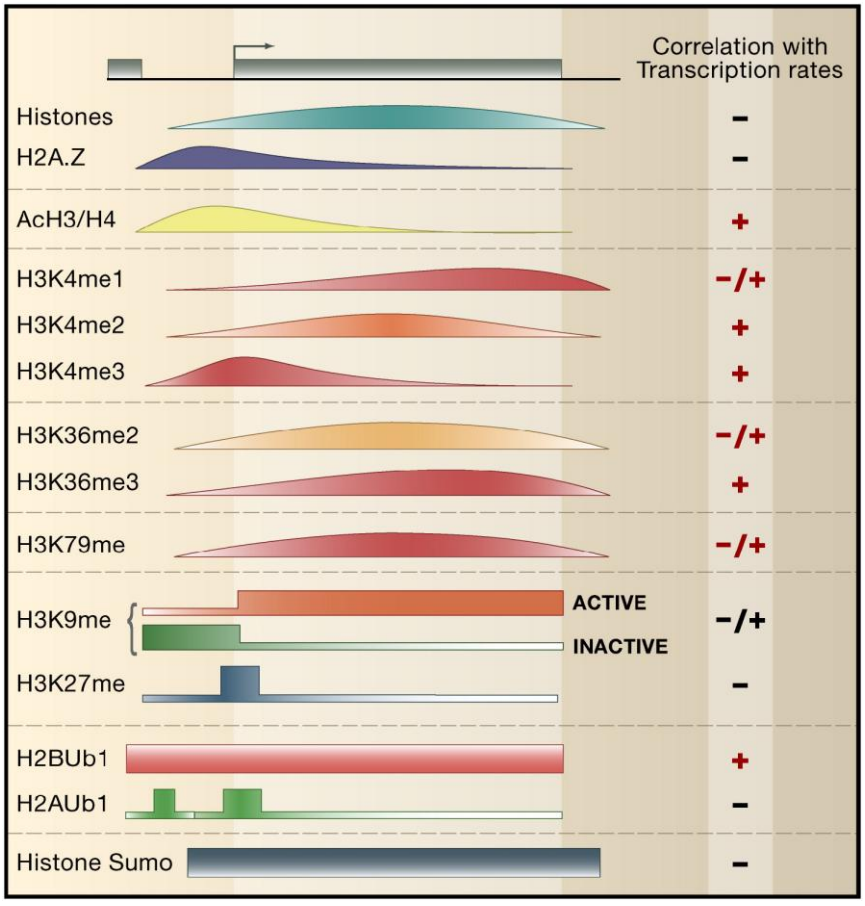


Figure 10 : Genome-wide distribution pattern of histone Modifications from a transcription perspective
The distribution of histones and their modifications are mapped on an arbitrary gene relative to its promoter, ORF, and 3'. The curves represent the patterns that are determined via genome-wide approaches. (Adapted from (Li, Carey et al. 2007)).

Recently, two systematic studies using ChIP-seq approaches have mapped 39 acetylation, 20 histone methylation marks and the enrichment of H2A.Z variant genome wide in human lymphocytes (Barski, Cuddapah et al. 2007; Wang, Zang et al. 2008). Out of the 4339 combinatorial patterns possible, only a small fraction was indeed observed at promoter regions. The prevalent patterns associated with promoters and enhancer regions were defined and classified in three expression level categories (Wang, Zang et al. 2008). Despite the complexity of the combinations observed, some modification patterns correlate very well with a type of regulatory region. Table 5 displays the marks most frequently observed at regulatory regions according to gene expressions.

	Methylation	**Acetylation**	**Histone Variant**	**Expression level**
Promoter I	H3K27me3, H3K4me1/2/3, H3K9me1,	excluded	H2A.Z	low
Promoter II	H3K36me3	H4K16Ac	H2A.Z	intermediate
Promoter III	H3K4me3, H2BK5me1, H4K20me1, H3K79me1/2/3	H4K16Ac	H2A.Z	high
Enhancer	H3K4me1/2, H3K9me1	H3K18Ac, H3K27Ac	H2A.Z	-

Table 5 : Histone modification patterns associated with transcription regulatory elements.
The table summarizes the most frequently observed histone modifications detected on or around regulatory elements (promoters and enhancers), classified according to the expression levels of the given genes localized in the close vicinity of the regulatory regions (based on (Heintzman, Stuart et al. 2007; Wang, Zang et al. 2008; Heintzman, Hon et al. 2009)

Interestingly, these combinations of histone marks define flexible transcription regulatory rules in a complex cellular system that can likely be partially extrapolated to other tissues and organisms. However, it becomes clear from several other studies that additional specific rules may govern each cellular system. For example, a ChIP-on-chip study against histone methylation marks H3K4me3 and H3K27me3 in mouse embryonic stem cells (mESCs) observed a specific pattern of modifications in this cell type (Bernstein, Mikkelsen et al. 2006). H3K4me3 is generally associated with active promoters (Santos-Rosa, Schneider et al. 2003; Pray-Grant, Daniel et al. 2005) and H3K27me3 with repressed transcription (Francis, Kingston et al. 2004; Ringrose, Ehret et al. 2004). One would expect that the distribution of these antagonistic marks would not be overlapping. However, in the undifferentiated mESCs, the combination of H3K4me3 and H3K27me3 marks is present simultaneously at some promoters. It appeared that this bivalent marking is present at the promoters of genes coding for transcription factors, which will be involved at a subsequent stage in cellular differentiation, but have to be repressed to keep the stem cell state. The authors then hypothesized that this bivalent mark is restricted to mESCs and to those genes, which will change their expression during differentiation. However, this hypothesis was contradicted by a study analysing the change of histone methylation marks in a model where mESCs were differentiated to neurons (Mohn, Weber et al. 2008). Similar to what was observed previously, authors observed that, the promoters harbouring bivalents chromatin domains in undifferentiated mESCs loose this particular bivalent mark during the differentiation process. However, they also observed that the bivalent marks were not restricted to mESCs, but during differentiation novel sets of promoters appeared harbouring the bivalent chromatin marks. This example illustrates the complexity of generalizing regulation rules deduced from a study in a particular cellular system.

C. New transcriptional regulatory concepts

Transcription regulation processes are known to be complex multi-step processes. Most of the mechanistic knowledge accumulated by classical approaches is issued from *in vitro* or *in vivo* studies on inducible gene models. Most of the studies performed in lower or higher eukaryotes supported a model in which transcription factors and cofactors are sequentially recruited, finally leading to the pre-initiation complex (PIC) formation, the loading of Pol II on the DNA and the production of a functional pre-mRNA transcript. Furthermore, the canonical model assumes that the transcription initiation is the rate-limiting

step in efficient transcript production. However, the systematic characterisation of Pol II location sites on the genome of a given cell type and the mapping of the other actors of the transcription initiation process has not only improved our view of transcription regulation processes, but at the same time also indicated that the transcription initiation process, when studied at a genome wide level, is much more complex than originally thought. In the next couple of small sub-chapters I will try to illustrate through several striking examples, how high throughput approaches brought new insights in our understanding of transcription regulation.

1. Mechanisms of action of the transcriptional machinery

To date, most of the GWLA studies performed against the transcription initiation machinery in higher eukaryotes were done either on selected regions at high resolution (e.g. ENCODE regions), or at relatively low resolution (~0,5kb) due to the size of the studied genomes. At low resolution, the precise localisation of the position of the investigated factors relative to the promoter of the genes and the surrounding chromatin structure remains limited. However, already at this low resolution a global view emerged that started to give information about the general behaviour of those factors. For example, the GWLA of p300 (a general co-activator harbouring a histone acetyltransferase activity (HAT)), together with TAF1 (a TFIID subunit) and Pol II (Heintzman, Stuart et al. 2007), suggest a physical interaction between enhancer and active promoter regions distant from several kilobases (Szutorisz, Dillon et al. 2005; Heintzman, Stuart et al. 2007). This observation would confirm the model of action in which enhancer regions are thought to be bridged to promoters by the action of activators and chromatin remodelling complexes resulting in the increase of the promoter activity (Figure 11) (Blackwood and Kadonaga 1998). More recently, the enhancer function has been further investigated at the genome wide scale in different cell types (Heintzman, Hon et al. 2009). Interestingly, this study shows that p300-bound enhancers are not common among different cell types. This observation supports the concept in which tissue specific activity is mainly controlled at enhancer level rather than at the promoter level.

In contrast to higher eukaryotes, studies in yeast were performed at high resolution, allowing the definition of promoter chromatin architecture (see above). More recently, the majority of the transcription actors (activators, co-activators, PIC subunits and Pol II) could be mapped around the promoters in *S. cerevisiae* (Venters and Pugh 2009a). It appears in this

study that, at least for lower eukaryotes all the transcription actors are bound within the NFR. The PIC is bound immediately upstream of the TSS at the location of the TATA box. In the middle of the NFR region (-100bp from the NFR border), co-activators, such as the mediator complex, and the Pol II are bound. Finally, other co-activators such as the HAT complex (SAGA) or ATP-dependent remodelling complexes are bound at the 5' limit of the NFR . This mapping gives some new organisational insights of promoter function and potential role of each member of the machinery regarding to its location. For example, it shows that the active removal of the 5' NFR nucleosome is necessary prior Pol II docking to the promoter. These results further explain the observed location of remodelling complexes at this precise location.

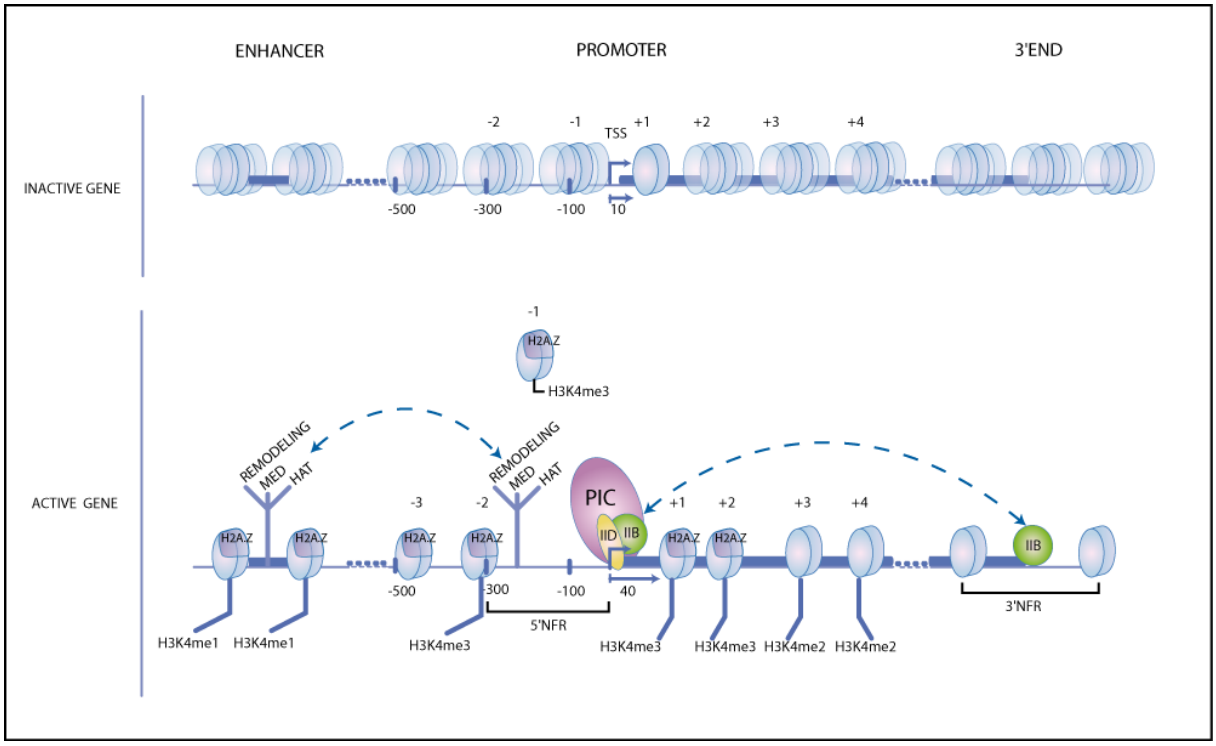


Figure 11 : Structural and mechanistic changes in the chromatin organization of the Pol II regulatory elements upon gene activation.
The figure gives a projection of the structural organisation of enhancer, promoter and transcription termination (3'END) regions in transcriptionally silent (top panel) and active mode (bottom panel). Upon gene activation, the position of the fuzzy histones around the transcription start sites (TSS) get fixed, the -1 nucleosome is evicted, to form the 5' nucleosome free region (NFR). A similar positioning of the histones occurs also for histones flanking the enhancer and the 3'NFR regions. The nucleosomes located outside of the represented regions remain fuzzy after activation (not represented). Several histone modification and histone variants characterise each region upon activation. A looping is believed to happen between enhancer and promoter regions allowing the activity of co-activator complexes (e.g. Mediator, HATs, ATP-dependent remodelling complexes). The cylinders are representing nucleosomes with fixed (single non-transparent cylinder) and fuzzy positioning (three transparent cylinder array). Several histone modifications and variants

characteristic of a particular localisation or activation state are symbolised (e.g. H2A.Z, methylation marks). Members of the general transcription machinery are symbolised either by colour circles (PIC: pre-initiation complex; IID: TFIID; IIB: TFIIB) or by their acronyms (REMODELLING: chromatin remodelling complexes; MED: mediator complex; HAT: histone acetyl-transferase complexes). The dotted arrows show potential long distance physical interactions within the different regulatory DNA elements (looping).

2. Non productive PIC formation

Expression of genes was long thought to be regulated primarily at the level of RNA polymerase II recruitment to gene promoter regions, and the few genes that did not fit this paradigm were regarded as exceptions. The comparison of the genome wide mapping by ChIP-on-chip of the pre-initiation complex subunit TAF1 in human fibroblast with the rate of gene expression surprisingly showed that for ~20% of the human genes a PIC is observed, but no transcript can be detected (Kim, Barrera et al. 2005b). This result may indicate that another mechanism of regulation is responsible for the expression of these genes at a post transcription initiation step. This concept was further developed by a genome wide study on hESC and differentiated lymphocytes, showing that most of the silent genes in these cell types indeed show PIC formation in a non productive manner (Guenther, Levine et al. 2007). Moreover, recent genome-wide analyses of Pol II distribution in *Drosophila* (Zeitlinger, Stark et al. 2007) and mammalian systems have indicated that a large number of genes might be regulated at a step subsequent to Pol II recruitment, during early transcription elongation. At these genes, Pol II begins transcription but stalls after synthesizing a short RNA. The release of this engaged Pol II from the promoter-proximal region may be the rate limiting step for transcription. From these genome wide studies it seems that Pol II stalling at promoter-proximal is prevalent on genes involved in development and response to stimuli, suggesting that Pol II stalling during early elongation plays important roles in rapid and precise control of gene expression (reviewed in (Nechaev and Adelman 2008). Moreover, c-Myc was recently shown to be a key player of the general mechanism that release Pol II from pausing (Rahl, Lin et al. 2010). These results further confirm that regulation at post initiation steps is a general mechanism widespread in metazoan organisms.

3. Transient and permanent regulation of expression

Several processes can be involved in relieving the repression that prevents the formation of an active PIC and thus inhibits active transcription. Amongst those, the transition

from a silenced chromatin barrier state to an activated one, the active recruitment of the components of the PIC and the ability to initiate a productive elongation are thought to be major regulatory mechanisms. The relative contributions of those three phenomena to achieve gene silencing have been recently investigated by comparing cell lines at different phases of the cell cycle and during differentiation (from mESC to embryonic bodies) (Komashko, Acevedo et al. 2008). The authors first isolated subsets of genes that are silent or have a very low expression level in the different conditions tested. Then, using ChIP-on-chip, they could investigate the presence of chromatin repressive marks (H3K9me3, H3K27me3), Pol II and DNA methylation at the corresponding promoters. They could distinguish between at least two different mechanisms explaining the lack of gene expression. Interestingly, it seems that repressive mechanisms used by the cell during the cell cycle and differentiation are different. Along the cell cycle, cells achieve transient gene regulation by changes in the recruitment of the PIC, presumably through the action of different transcription factors. In the contrary, during differentiation changes in different chromatin states seem to establish permanent changes in expression of genes. This result illustrates how different transcription regulation strategies can be used by the cells to achieve transient or permanent changes in their expression patterns.

The use of systematic approaches from different systems will help to define the epigenetic landscape that interacts with the transcription machinery and better understand the gene expression regulation processes

Second part: GCN5 containing complexes (GCC)

The protein p55 was the first enzyme isolated (from *Tetrahymena*) bearing an acetyl transferase (AT) activity (Brownell and Allis 1995). It was then shown to be the homologue of the eukaryotic General Control Nonderepressible 5 (GCN5). It rapidly appeared that GCN5 similarly to most of the HATs is included in multi-subunits macromolecular complexes *in vivo*. GCN5 was demonstrated to be included in two distinct functional complexes named SAGA (Spt-Ada-Gcn5-Acetyl transferase) and ATAC (Ada Two A Containing). Here I will review in detail the current knowledge on GCC. First I will discuss the protein composition as well as the structural features of these complexes. Then I will present the catalytic spectra of these HAT complexes, such as the specificity of GCC for histone residues as well as for non histone targets. Finally, I will present the current models describing the mechanistic of action of GCC.

I. Structural organisation of the GCN5 and PCAF proteins

At the primary sequence level, GCN5 is remarkably conserved throughout the evolution. Particularly, the bromodomain and the catalytic domain (AT domain) show a very high degree of conservation. While in most metazoan genomes, only one copy of the GCN5 gene is found, in vertebrates, GCN5 is encoded by two paralogous genes, that share a high degree of sequence homology (>70%) (reviewed in (Nagy and Tora 2007)) named GCN5 (or KAT2A) and PCAF (or KAT2B). All known metazoan GCN5 homologues can be divided in two parts: the N-terminal half of the molecule that seems to be metazoan specific, and the C-terminal half, which is highly homologous to the shorter yeast protein (Figure 12). The N-terminal half contains the so-called PCAF homology domain, and the C-terminal half contains two other conserved domains: the AT domain and the bromodomain.

The structure of the AT domain of *Tetrahymena* GCN5 (tGCN5) bound to its physiologically ligands, coenzyme A (CoA) and a histone H3 peptide has been solved (Rojas, Trievel et al. 1999). This study together with the one of the PCAF AT domain (Clements, Rojas et al. 1999), revealed that the central core region of these domains mediate acetylCoA

binding and catalysis, while the N- and C-terminal regions of the AT domains contain a related scaffold that seem to mediate histone substrate specificity.

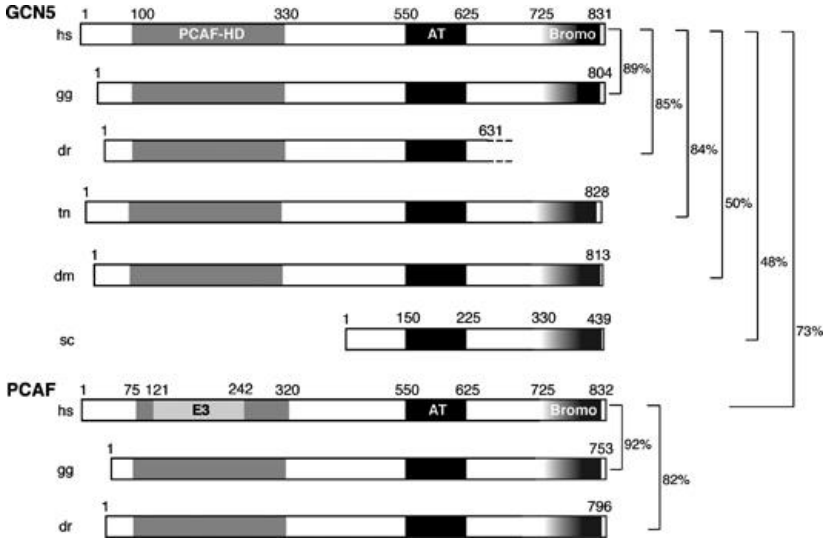


Figure 12 Overall structures of the GCN5 and PCAF enzymes in vertebrates, *Drosophila* and yeast.
Schematic representation and domain organization of the GCN5 and PCAF proteins from human (hs; *Homo sapiens*), chicken (gg; *Gallus gallus*), zebrafish (dr; *Danio rerio*), pufferfish (tn; *Tetraodon nigroviridis*), *Drosophila melanogaster* (dm) and yeast (sc; *Saccharomyces cerevisiae*) are shown. The PCAF homology domain (PCAF-HD) is shown in grey, the AT domain is shown in black,the bromo domain (Bromo) is shaded and the ubiquitin E3 ligase domain (E3) of PCAF is indicated in light grey. The numbers over the boxes indicate amino-acid positions. The identity between the different factors is indicated in percentage (%) on the right of the horizontal lines, representing the pair wise comparisons. AT, acetyl transferase. (Adapted from (Nagy and Tora 2007))

II. Sub-unit composition and ultra-structure of GCCs

Extensive efforts were spend in the last decade in order to define the subunit composition and the macromolecular organisation of these HAT complexes, by the use of biochemistry and electron microscopy (EM), respectively.

A. Subunit composition: SAGA & ATAC

1. The SAGA (Spt-Ada-Gcn5-Acetyl transferase) complex (also known as TFTC, STAGA)

In metazoans, the first GCN5-containing complex purified from human cells was SAGA. This human complex was purified using either double immunoprecipitations (IP) with antibodies recognizing endogenous factors, (Wieczorek, Brand et al. 1998; Brand, Yamamoto et al. 1999b), or overexpressing a putative subunit with a tag and then using the tag to purify the complexes (Martinez, Kundu et al. 1998; Ogryzko, Kotani et al. 1998). Following its identification, the characterization of the subunit composition of the human complex (Table 6) indicated that it is similar to the previously identified yeast SAGA complex (for complete review see (Nagy and Tora 2007)).

This complex contains either GCN5 or PCAF as catalytic HAT subunit (Table 6). In addition, it contains a set of TBP-associated factors (TAFs) and several human homologues of proteins, earlier identified in yeast screening for being necessary for either correct initiation site selection by Pol II (Spt proteins (Winston and Sudarsanam 1998) or for transcriptional activation (the Ada group of proteins (Roberts and Winston 1996). These human complexes also contain a 400 kDa protein, TRRAP that was originally isolated as a Myc-associated transcription co-activator (McMahon, Van Buskirk et al. 1998). Additional proteomic investigations of the yeast SAGA complex (Sanders, Jennings et al. 2002; Powell, Weaver et al. 2004) and later of the human SAGA complex (Zhao, Lang et al. 2008) allowed the identification of four novel SAGA-associated factors (ySus1/hENY2, ySgf73/hATXN7, ySgf29 and Sgf11/hATXN7L3) together with yUbp8/hUSP22, a ubiquitin-specific protease component. These four sub-units were shown to form a functional module, named de-ubiquitination module (dUB), which brings a second enzymatic activity within the SAGA complex.

2. The ATAC (Ada Two A Containing) complex

More recently, alternative purification strategies revealed a second GCN5/PCAF containing macro-molecular complex named ATAC, whose existence is exclusive of higher eukaryotes. The complex was first purified from *Drosophila* cells (Suganuma, Gutierrez et al.

2008) showing homologies in the subunit composition with SAGA since additionally to GCN5, the ADA3 and SGF29 sub-units are shared between the two complexes (Table 6). However, it also showed that most subunits such as ZZZ3, YEATS2, MBIP, NC2β and WDR5 were specific for ATAC suggesting functional differences between the two complexes. The presence of ATAC2 (CRSP2BP) that contains a domain with a putative HAT activity, would suggest that ATAC could be a double HAT containing complex. The presence of a similar ATAC complex in mammals was assessed by a series of studies (Nagy, Riss et al.; Wang, Faiola et al. 2008) confirming the subunit composition of the complex (Table 6).

SAGA	ATAC
GCN5 (or PCAF)	GCN5 (or PCAF)
-	ATAC2 (CSRP2BP)
ADA1	-
ADA2b	-
-	ADA2a
ADA3	ADA3
SPT3	-
SPT7L	-
SPT20 (p38IP or FAM48)	-
TAF5L	-
TAF6L	-
TAF9 (and TAF9b)	-
TAF10	-
TAF12	-
ATXN7L3	-
USP22	-
ENY2	-
ATXN7	
SGF29	SGF29
TRRAP	-
-	ZZZ3
-	Yeats2
-	MBIP
-	NC2beta
-	WDR5

Table 6 : Comparative summary of the sub-unit composition of SAGA and ATAC identified in human cells.
Sub-units were classified according the module they belong to in the complex: HATs (purple); Adaptator-ADA (orange); Spt (green); TAF (blue); de-Ubiquitination (brown).

B. SAGA structure: a functional modularity

Insights into the organisation of yeast and human SAGA complexes were gained by resolving low-resolution three-dimensional structure by Electron Microscopy (EM) (Brand, Leurent et al. 1999; Wu, Ruhlmann et al. 2004). In these studies, negatively stained SAGA yielded a three-dimensional model at 30 Å resolution, which revealed that both yeast and human SAGA have an elongated shape of 160 Å 270 Å in size (Figure 13). The complexes consisted of five modular substructures of 70–100 Å in diameter (called domains) separated by solvent-accessible grooves (see Figure 13).

An overlay of the yeast and human SAGA structures indicated, in good agreement with their similar subunit composition, a high degree of structural conservation in size and shape. This is therefore expected that the metazoan subunits will position to similar domains as described in the yeast complex. The distinct domains of SAGA were revealed by mapping its subunits by immuno-EM methods (Wu, Ruhlmann et al. 2004). The domain I of SAGA which contains yTra1(TRRAP) seems to represent the activator interaction surface. Domains II, III and IV contain several histone fold-containing TAFs and TAF5, which might play an architectural role. In domain III, the two bromodomain-containing subunits, GCN5 and Spt7, were detected (Figure 13a and b). Domain V, for which the structure seems to be more flexible, contains ySpt3, ySpt20 and probably ySpt8. Thus, the EM studies suggested a modular organisation within the complex with the possible existence of sub-modules dedicated to a particular function.

More detailed *in vitro* studies further described the inner organization of the complex and the role of different proteins within particular sub-modules (acetylation, de-ubiquitination). For example, (Balasubramanian, Pray-Grant et al. 2002) and (Gamper, Kim et al. 2009) showed that the ADA protein family (ADA2b-ADA3 in SAGA) directly interact with GCN5. Moreover they demonstrated that they participate in the catalytic activation of GCN5 within the SAGA complex. They have therefore argued for the existence of an acetylation catalytic module in the complex. It is interesting to note that the authors report a similar organisation for the ATAC complex with an ADA2a-ADA3-GCN5 module. However, they could not observe an *in vitro* catalytic activation in that case (Gamper, Kim et al. 2009).

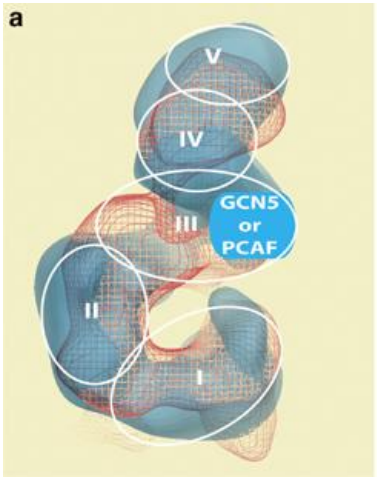


Figure 13 : The overall three-dimensional structure of the yeast and human SAGA complexes is evolutionarily conserved and display a modular organisation.
(a) The low-resolution three-dimensional structure of the yeast SAGA and human TFTC was elucidated via EM methods. Image reconstruction yielded a three-dimensional model at 30 Å resolution, which revealed that both yeast and human SAGA complexes have an evolutionarily well-conserved structure. Alignment and superposition of the yeast (blue) with that of human (red) SAGA models is shown and the five modular domains of the complexes as defined by(Wu, Ruhlmann et al. 2004) are indicated with white circles. The theoretical position of GCN5 (or PCAF) in the superposition is indicated. (b) The different subunits of SAGA, which were identified in the distinct domains (Wu, Ruhlmann et al. 2004), are summarized. (Adapted from (Nagy and Tora 2007))

Similar reconstitution studies were performed for another subgroup of proteins (namely USP22, ATXN7L3, ATXN7 and ENY2) that were postulated to be functionally linked in yeast based on their loss of integration in SAGA upon genetic deletion of one of the partner (Ingvarsdottir, Krogan et al. 2005; Lee, Florens et al. 2005). Using recombinant proteins, (Zhao, Lang et al. 2008) demonstrate that mammalian ATXN7L3 mediates the interaction between this subgroup of proteins within the complex. Interestingly, this group includes the enzyme USP22 that process a histone ubiquitin protease activity (H2A-Ub, H2B-Ub). Moreover, they show that ATXNL3 is required for both the catalytic activation of the enzyme and its integration o to the SAGA complex. More recently, high resolution crystallographic structures of this tetra protein sub-complex were obtained (Kohler, Zimmerman et al. 2010; Samara, Datta et al. 2010). These two studies could mechanistically dissect the structural implications of sgf73 (the yeast homolog ATXN7L3) in its central role within this module of SAGA. Altogether, these data led to the proposal of the existence of a second catalytic module within SAGA dedicated to the de-ubiquitination (dUB) activity.

Taken together, the EM structural determination and the in vitro reconstruction studies, demonstrates that the organisation of SAGA is highly modular containing interconnected sub-modules harbouring a particular set of functions. At present, less is known about the structure of the ATAC-type complexes, which will need to be determined in the future. However, by analogy with SAGA, since ATAC contains a HAT module composed by GCN5 and ADA3/ADA2a (instead of ADA2b in SAGA), one can anticipate that a similar modular organisation could be observed in ATAC with at least one functional module.

III. Catalytic spectra of GCCs

As discussed earlier, multiple catalytic activities such as HAT and dUB, have been described for SAGA and ATAC. Numerous studies were then focused on the identification of modification targets of these complexes. Two types of studies were conducted, the ones identifying the acetylated residues on histone targets and the ones identifying non-histonal targets.

A. Histone substrates

1. Acetylation (Ac)

Two major methodological approaches were used to study SAGA/ATAC histone acetylation spectra. First, by using recombinant free enzyme or immunopurified native complexes, the acetylation on different substrates can be tested *in vitro* (*in vitro* synthesized peptides, purified histones, reconstituted nucleosomal arrays). Second, by performing Knock Out (KO) or Knock Down (KD) of the targeted enzymes and monitoring the effect on global or loci specific acetylation using specific antibodies. Numerous studies used one or the other approach, which combined together allowed to define a consensus acetylation spectra for SAGA and ATAC (reviewed in Table 7).

		Histone Lysine	H3				H4				H2A		H2B				references
			9	14	18	23	5	8	12	16	5	9	5	12	15	20	
	molecular complex	HAT Family															
GCN5-KAT2A/ PCAF - KAT2B	SAGA	GNAT															(Nagy, Riss et al.; Brand, Yamamoto et al. 1999b),
	ATAC						?		?	?							(Nagy, Riss et al.; Wang, Faiola et al. 2008; Guelman, Kozuka et al. 2009)
	Free																(Nagy, Riss et al.; Brand, Yamamoto et al. 1999b; Nagy, Riss et al. 2009a)

Table 7 : Histone acetylation specificity of SAGA and ATAC.
Review of the substrate specificity described in mammalian systems, for SAGA and ATAC enzymes either as a free protein or within their respective macromolecular complexes. Dark grey boxes represent lysine residues highly acetylated; light grey boxes represent residues where weak acetylation activity has been observed and white boxes represent residues where no acetylation was detected. ND represents residues that have not been tested for acetylation. Question marks represent residues for which contradictory data have been reported.

These data shows that SAGA and ATAC have partially overlapping but different spectra of acetylation. SAGA preferentially acetylates histone H3 on K14-K9 which was also reported for ATAC in mammalian systems. Additionally, ATAC was shown to acetylate residues (K5, K12 and K16) on histone H4 in *Drosophila* (Guelman, Suganuma et al. 2006) and mouse cells (Guelman, Kozuka et al. 2009), mediated through its second HAT sub-unit ATAC2. However, this result could not be confirmed in human (Nagy, Riss et al. 2009a) suggesting that ATAC2 activity might have been lost through evolution. Another interesting observation is that free GCN5 is unable to acetylate histones (Brand, Yamamoto et al. 1999b) requiring its integration in SAGA/ATAC to gain full activity. This result highlights the

importance of studying HAT enzymes within their native macromolecular complexes, since their acteylation capability depends on their intergration within these.

2. De-ubiquitination (dUB)

As discussed previously, SAGA contains a de-ubiquitination activity that is absent in ATAC. To date, this activity has been mostly tested in *in vitro* dUB assays (Kohler, Zimmerman et al.; Samara, Datta et al.; Zhang, Varthi et al. 2008; Zhao, Lang et al. 2008). In terms of specificity, the SAGA dUB activity preferentially targets mono ubiquitination of histone H2BK120 and in less extend histone H2AK119 (Zhao, Lang et al. 2008). Moreover, similarly to GCN5 that requires its ADA partners within the complexes to exert full activity, USP22 needs its dUB module partners and integration into SAGA to be fully active (Kohler, Zimmerman et al. 2010).

B. Non histone substrates

During the last years, the research on the SAGA/ATAC complexes was focused on defining extensively their specificity over histone targets. However, a couple of isolated studies raised evidences that these complexes can also target non-histone targets. Here I will present a non-exhaustive review of the most documented examples of non histone acetylation by GCN5 containing complexes.

One of the best studied examples is the transcription factor MyoD, a key regulator of myogenic differentiation. PCAF was shown to directly acetylate MyoD and consequently to be required for MyoD targeted gene activation (Sartorelli, Puri et al. 1999; Dilworth, Seaver et al. 2004). However, these studies did not clearly distinguish between SAGA or ATAC and it is not clear which complex functions in MyoD regulation.

More recently, a study from our laboratory (Orpinell, Fournier et al. 2010) showed that GCN5 within ATAC but not SAGA was able to acetylate the cell cycle regulator CyclinA. The study latter demonstrates that ATAC mediated acetylation regulates Cyclin A stability and by consequence the downstream cell cycle progression under its control. This study gave for the first time an indication on the GCN5 non-histonal acetylation spectra within the ATAC complex.

These two representative examples clearly show that GCN5-conatining complexes can have key regulatory roles through targeting non histone targets. However, many efforts will be needed in the future to define the complete acetylation spectra of SAGA/ATAC. Particularly, one can expect that the recent apparition of protein spotted microarrays (Lin, Lu et al. 2009) will allow large scale identification of protein substrate targets for HATs acetylation. Furthermore possible SAGA mediated ubiquitination/dUB pathways regulating non histone targets remain to be studied in the future.

IV. Mechanistic

While description of the GCC composition and the characterization of their enzymatic activities have been thoroughly studied over the last decade, little is known about their mechanistic mode of action. Early studies used either *in vitro* transcription systems or reporter assays to monitor the effect of the GCC on transcription (i.e. (Hardy, Brand et al. 2002)). This allowed to categorize these complexes as transcriptional coactivators complexes. However, these studies monitor the final effect on transcription without understanding the mechanistic behind (i.e. how these complexes acts in the activation pathway *in vivo*). In the present section I will review the current knowledge on how GCC are recruited into the transcription activation network.

A. Mechanistic models for GCC role in transcription activation

The current model of action of GCC (particularly studied for SAGA in yeast models) emerged from multiple independent reports of direct physical interactions between DNA binding transcriptional activators and SAGA (i.e. nuclear receptors (Yanagisawa, Kitagawa et al. 2002), c-Myc (McMahon, Wood et al. 2000), E2F...) (see (Nagy and Tora 2007) for full review). These activators would recruit the GCC onto the promoters of their target genes where they would participate in transcriptional activation (Carrozza, Utley et al. 2003).

One of the best described examples is the interaction between the transcriptional activator, c-Myc and SAGA. c-Myc was shown to recruit SAGA via direct physical interactions of its N-terminal activation/transformation domain and the TRRAP subunit of SAGA (McMahon, Wood et al. 2000; Liu, Tesfai et al. 2003). The authors could then show that the downstream c-Myc activation process was dependent on the GCN5 HAT activity. Later reports showed

that in this activation process, c-Myc bound to SAGA helps for the recruitment of the Mediator complex to the promoter via direct interaction (Liu, Vorontchikhina et al. 2008). Interestingly, the c-Myc/SAGA dependant activation has been also reported to be important on a set of Pol III driven genes (Kenneth, Ramsbottom et al. 2007), suggesting that GCC activation may not be restricted to Pol II transcribed genes.

The recruitment of GCC was also shown to be facilitated by the chromatin environment itself. Once targeted, SAGA acetylates histone H3 in the vicinity of the promoter. Histone acetylation by SAGA was shown to stabilize its own binding. Similar observation was done for H3K10 phosphorylation that enhanced targeting of GCN5 (Lo, Duggan et al. 2001). The binding to acetylated histones requires the bromodomains of GCN5, which bind to acetyl-lysine. Since bromodomains are found in other chromatin-modifying complexes and TFIID, acetylation might stabilize promoter interactions of several complexes involved in transcription activation. For example, it was shown that SAGA mediated acetylation promotes the recruitment of the chromatin-remodelling complex SWI/SNF (Syntichaki, Topalidou et al. 2000).

The current model of action of GCC based on these observations and the one made in the context of other transcriptional activators pathways would be that (Figure 14): GCC are recruited by a particular activator and then acts (i) through their enzymatic activities that both reinforce their binding and allow the recruitment of other coactivators (ii) acts possibly as a binding platform favouring the synergic recruitment of other coactivators.

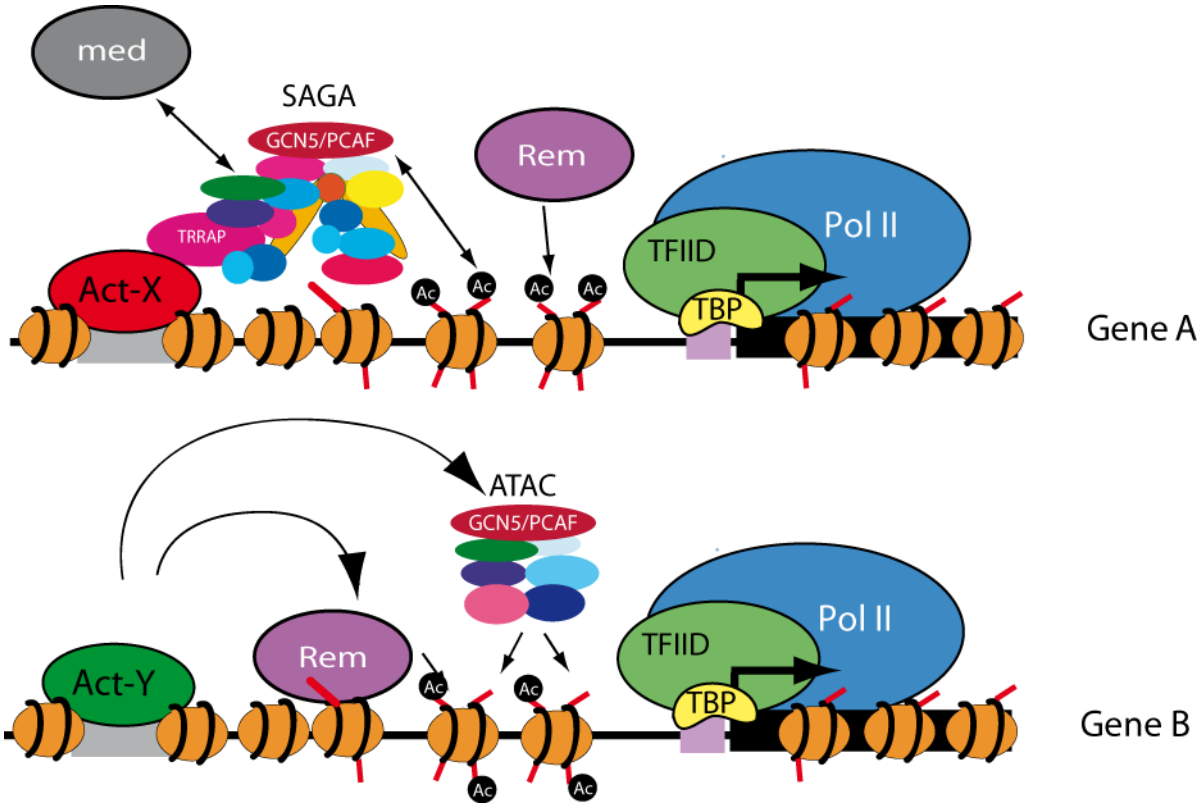


Figure 14 : Mechanistic models for the role of GCC in transcriptional activation
The GCC are recruited by DNA binding transcriptional activators. They can act both through their enzymatic activities and as a molecular bridge. GCC acetylates the chromatin leading to both the recruitment of other coactivators (i.e. remodelling complexes) and a stabilisation of their own binding. Moreover, the acetylation promotes the decompation of the chromatin favouring the binding of other transcriptional players. GCC can also favour the recruitment of others regulators through direct protein interactions (i.e. Mediator, TBP).

B. SAGA/ATAC recruitment specificity

Two recent parallel studies from our laboratory investigated the existence of specificity in the recruitment of SAGA and ATAC using separate stress induced gene activation pathways as a model (Nagy, Riss et al.; Nagy, Riss et al. 2009c). Interestingly the results show that while SAGA is recruited to the promoter of ER stress-induced genes and is necessary for their activation, ATAC is not recruited to these promoters (Nagy, Riss et al. 2009c). On the other hand, ATAC but not SAGA is recruited on the promoter of Na-arsenite stress-induced genes and is necessary during the activation process (Nagy, Riss et al.). Thus taken together, these results argue for a separation in the gene networks regulated by one or the other complexes. In the future, studies based on genome wide determination of the location of these complexes will help to better understand the functional differences between the two complexes.

C. Binding profile of the GCC

1. GCN5 binds promoters and ORFs (?)

Recent technological advances allowed limited but significant insight on the binding profile of GCC over the genome. In the present section, I will review the different GCC genome wide location analysis studies published to date highlighting the convergences, putative contradictions and limitations regarding the technology used.

The earliest study was performed in *Saccharomyces cerevisiae*, analysing the presence of yGcn5 over promoters by ChIP-chip on promoter arrays (Robert, Pokholok et al. 2004). The major conclusion was that yGcn5 was present at promoters at a level correlating with the expression level of the bound gene. Surprisingly, a similar observation was obtained by performing similar experiments on arrays representing open reading frames (ORFs) (Rosaleny, Ruiz-Garcia et al. 2007). Taken together, these data suggest that in *S. cerevisiae*, Gcn5 is generally recruited to both promoters and ORFs depending on the expression level of the gene. More recently, a study performed using ChIP-chip high resolution tiling arrays in *Schizosaccharomyces pombe* concluded that Gcn5 was mainly recruited to ORFs in a gene activity manner (Johnsson, Durand-Dubief et al. 2009) (Figure 15a) suggesting that GCN5, likely through SAGA, would be mainly involved in transcriptional elongation in yeast.

Interestingly, a study in human lymphoblastic cells mapped the binding of hGCN5/PCAF genome wide by ChIP-sequencing (Wang, Zang et al. 2009). From these results, it is clear that in human cells, GCN5 is mainly bound to promoters and has weak or no binding in the coding region (Figure 15b). Taken together, this would suggest strong functional differences in the mode of action of GCC in yeast compared to humans. GCN5 would be mainly recruited to ORFs in lower eukaryotes regulating transcription at the elongation stage while it would be mainly recruited to promoters in higher eukaryotes presumably regulating transcription at the initiation stage.

However, one should be cautious while interpreting these data that were obtained using emerging technologies for which the data analysis background is weak. Nevertheless, one can hope that more GCC mapping studies will help understanding the mechanistic of these complexes and the functional differences between SAGA/ATAC.

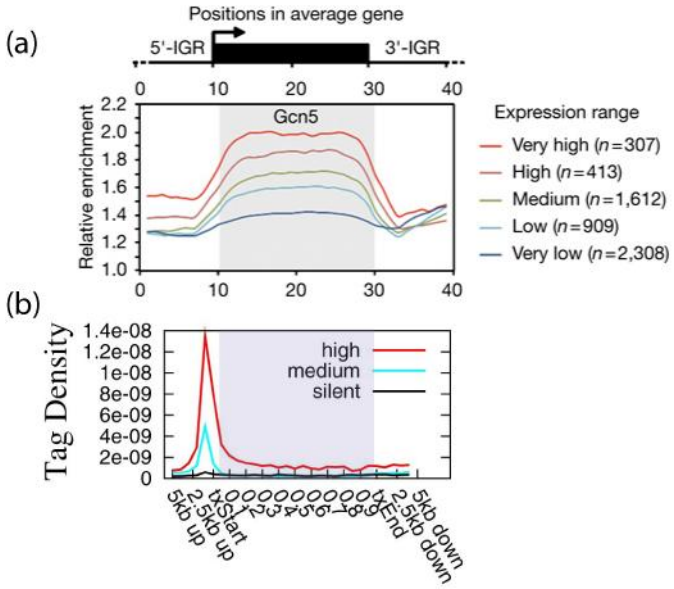


Figure 15 : Comparison of the GCN5 average gene profiles observed in yeast and human.

Genes were ranked and classified according to their expression level in the tested system. Average signal density was calculated at regular interval along promoter (5'IGR: Inter-Genic Region), gene body (light blue box) and termination region (3'IGR). Average signal density for each category was plotted (a) Average gene profile obtained using yGcn5 ChiP-chip data from S. pombe (adapted from (Johnsson, Durand-Dubief et al. 2009)) (b) Average gene profiles obtained using GCN5 ChIP-seq data from human lympoblast cells (adapted from (Wang, Zang et al. 2009)).

2. GCC binding on enhancers

Recent genome wide studies proposed that enhancer elements could be defined based on a couple of genomic criteria including particular specific histone modifications (presence of H3K4me1) and binding of a co-activator complex (the HAT p300/CBP) (Heintzman, Stuart et al. 2007; Heintzman, Hon et al. 2009). Since GCC are also HATs, one can speculate that they might also bind enhancer elements. Here, I will review the line of evidences that could suggest that GCC not only bind promoters but may also be recruited to distant regulatory elements. The largest pool of data abounding in this direction comes from ChIP studies in yeast that showed that SAGA was often recruited to upstream activating sequences (UAS). For example, on the well described HO model gene, it was clearly showed that SAGA binds stepwise to the UAS prior to the promoter during gene activation (Takahata, Yu et al. 2009). This proves that GCC can act at distance of the promoter. However, it is not clear to which extend the observations made on UAS can be extrapolated to enhancers that have different properties (longer range interaction, chromatin modification profile). In higher eukaryote, the complete GCN5 binding map in resting T cells was recently published (Wang, Zang et al. 2009). In this study, the authors mention that they could not attribute a promoter at all GCN5 binding events, meaning that a proportion of the binding sites are located distant from annotated transcripts. It suggest that GCC could bind distal enhancer elements, but additional studies rigorously investigating the nature of these events would be needed to affirm that they are true enhancer elements. Thus the binding of GCC to distal enhancer elements is likely happening in higher eukaryotes but remains to be strictly demonstrated.

RESULTS

Rational for the study

As illustrated in the Introduction part, the biochemical characterization of the GCN5 Containing Complexes (GCC) has been extensively conducted in the last decade. Thus the sub-unit compositions as well as the enzymatic activities of the complexes are now well defined. However, many aspects regarding how these complexes acts functionally in transcription are not or poorly understood.

The ambition of the present work was to gain more insight on the mechanistic of action of GCC by addressing key questions in higher eukaryotic cellular models. In this aim, two major research axes were used: (i) we did characterize the interaction partners of GCC to better understand their regulatory network (ii) we established the genome wide maps (GWM) for several members of the GCC to determine the loci regulated by the GCC. The ultimate goal of this work was to integrate these data to define some of the regulatory rules that govern the action of these complexes at the scale of the genome.

One of the strategic postulates of this study was to follow up recent technical development in the post-genomic era (High sensitivity proteomics, ChIP-sequencing) as a guiding lane in order to bring original perspectives on the function of GCC. Thus in parallel to the biological results obtained several innovative bioinformatics analysis methods were developed and implemented for distribution to the scientific community. Thus in this section, I will separately present primary research results followed by the methodological results.

In the upcoming section, I will present (i) a study that explores the ineractome of GCC in mouse embryonic stem cells. A specific and stable association of the ATAC complex with the Mediator coactivator complex is detected. Moreover, using ChIP-seq, a subset of target genes of this newly identified meta-coactivator complex (MECO) are identified. (ii) A study that establish the GWM of ATAC in different cells lines. The genomic context of ATAC binding is analyzed, showing binding to promoter and enhancer elements. Moreover, the behaviour of ATAC through variations between cell types is analyzed. (iii) A data mining study that explores the functional specificity within the HAT family through the bioinformatics analysis of published genome wide maps of several HATs. (iv) Presentation of

a tool named GPAT that allow annotating GWM experiments with neighbouring genomic features. (v) Presentation of a bioinformatics tool named seqMINER that allow comprehensive analysis of GWM experiments through multiple data comparisons methods.

I. ATAC and MEDIATOR form a stable complex and regulates a set of noncoding RNA (Krebs *et al*, 2010 EMBO R)

EMBO reports

scientific report

ATAC and Mediator coactivators form a stable complex and regulate a set of non-coding RNA genes

Arnaud R. Krebs[1], Jeroen Demmers[2], Krishanpal Karmodiya[1], Nan-Chi Chang[3w], Alice Chien Chang[4] & La´szló Tora[1+]

[1]Department of Functional Genomics, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR 7104, INSERM U 964, Universite´ de Strasbourg, Illkirch, France, [2]Proteomics Center, Erasmus University Medical Centre, Rotterdam, The Netherlands, [3]Institute of Microbiology and Immunology, and [4]Institute of Neuroscience, School of Life Science, National Yang-Ming University, Taipei, Taiwan

The Ada-Two-A-containing (ATAC) histone acetyltransferase and Mediator coactivator complexes regulate independent and distinct steps during transcription initiation and elongation. Here, we report the identification of a new stable molecular assembly formed between the ATAC and Mediator complexes in mouse embryonic stem cells. Moreover, we identify leucine zipper motif-containing protein 1 as a subunit of this meta-coactivator complex (MECO). Finally, we demonstrate that the MECO regulates a subset of RNA polymerase II-transcribed non-coding RNA genes. Our findings establish that transcription coactivator complexes can form stable subcomplexes to facilitate their combined actions on specific target genes.

Keywords: GCN5; PCAF; SAGA; LUZP1; snRNA biogenesis

EMBO reports advance online publication 28 May 2010; doi:10.1038/embor.2010.75

INTRODUCTION

Gene expression is a tightly regulated process. Initiation of transcription by RNA polymerase II (Pol II) is believed to be the outcome of a sequence of events beginning with the binding of specific activators to their cognate binding sites. This initial step triggers the recruitment of coactivator complexes and general transcription factors at promoters to allow the loading of Pol II into the preinitiation complex (PIC) to achieve transcription initiation. In this process, coactivators have crucial roles through their enzymatic and non-enzymatic functions. For example, they allow chromatin modification and opening, as well as enhancement of the formation of initiation complexes (Thomas & Chiang, 2006).

The histone acetyltransferases (HATs) general control of amino-acid synthesis 5 (GCN5 or KAT2A) and p300/CBP-associated factor (PCAF or KAT2B) are encoded by paralogue genes and share 70% sequence identity. GCN5 and PCAF are mutually exclusive subunits of two functionally distinct, but related, multisubunit coactivator complexes: Spt–Ada–Gcn5 acetyltransferase (SAGA) and Ada-Two-A-containing (ATAC; Wang et al, 2008; Nagy et al, 2010). These complexes have been shown to differentially regulate both locus-specific gene expression and global chromatin structure through their enzymatic activities (HAT activity and histone deubiquitination).

The Mediator complex (MED) is another multisubunit key transcription coactivator complex. Even though the complete mechanism by which this complex controls transcription is not fully understood, the MED is believed mainly to be required to mediate molecular bridges among activators, coactivators and general transcription factors to enhance or repress the formation of PIC leading to transcription initiation (Malik & Roeder, 2005). The existence of at least two functional forms of the MED complex has been reported. In the presence of the MED kinase module, the MED (known as TRAP) represses transcription. By contrast, when the kinase module is lost and the subunit MED26 is instead incorporated to the complex, the MED (known as PC2) activates transcription (Malik & Roeder, 2005; Paoletti et al, 2006).

To gain more insight into the molecular mechanisms of ATAC function, we carried out a comprehensive analysis of the ATAC interactome. We showed that in certain cell types, including mouse embryonic stem cells (mESCs), the ATAC and MED complexes form a highly stable meta-coactivator complex (MECO) and that the MECO associates and regulates transcription of a subset of Pol II-transcribed non-coding RNA (ncRNA) genes.

[1]Department of Functional Genomics, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Parc d'Innovation, 1 rue Laurent Fries, CNRS UMR 7104, INSERM U 964, Universite´ de Strasbourg, BP 10142, Illkirch Cedex 67404, France
[2]Proteomics Center, Erasmus University Medical Centre, Faculty of Medicine, Dr. Molenwaterplein 50, 3015 GE, Rotterdam, The Netherlands
[3]Institute of Microbiology and Immunology, and
[4]Institute of Neuroscience, School of Life Science, National Yang-Ming University, 155, Linong Street, Taipei 112, Taiwan
[w]Present address: Department of Medicine, Mackay Medical College, Taipei County, 252 WuHsing Street, Taiwan
[+]Corresponding author. Tel: þ 33 388 653 444; Fax: þ 33 388 653 201; E-mail: laszlo@igbmc.fr

Received 14 January 2010; revised 20 April 2010; accepted 26 April 2010; published online 28 May 2010

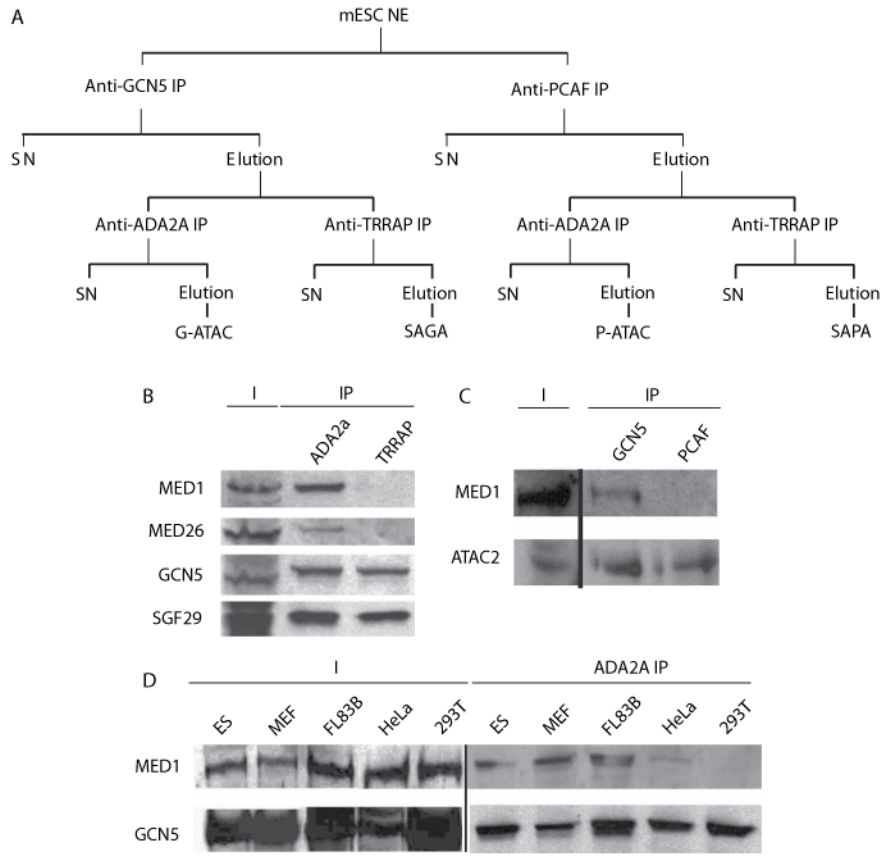


Fig 1 | Ada-Two-A-containing and the Mediator complex interact stably in a cell-specific manner. (A) Schematic representation of the ATAC and SAGA tandem immunoprecipitation (IP). First, GCN5- (G-) or PCAF- (P-) containing complexes were isolated by the corresponding IPs followed by peptide competition elution. From these pools ATAC (a-ADA2a IP) or SAGA (a-TRRAP IP) complexes were isolated, respectively. (B) Equivalent amounts of purified ATAC (ADA2a IP) and SAGA (TRRAP IP) complexes (normalized for their GCN5 content) were analysed by western blot against ATAC (GCN5, SGF29) and MED subunits (MED1, MED26). Input (I) and immunoprecipitated fractions are shown. (C) Equivalent amounts of GCN5- and PCAF-containing complexes (purified by either an a-GCN5 IP or an a-PCAF IP, respectively, and normalized for their ATAC2 content) were analysed by western blot with the indicated antibodies. (D) Cell extracts were prepared from the indicated mouse (ES, MEF and FL83B) and human (HeLa and 293T) cell lines. From these, ATAC complexes were prepared by a single a-ADA2a IP and the complexes analysed by western blot with the indicated antibodies. ATAC, Ada-Two-A-containing; ES, embryonic stem; GCN5, general control of amino-acid synthesis 5; mESC, mouse embryonic stem cell; MED, mediator complex; MEF, mouse embryonic fibroblast; NE, nuclear extract; PCAF, p300/CBP-associated factor; SAGA, Spt–Ada–Gcn5 acetyltransferase; SAPA, Spt–Ada-PCAF-acetyltransferase; SGF29, SAGA-associated factor 29; SN, supernatant.

RESULTS AND DISCUSSION

ATAC interacts stably with the active Mediator complex

To identify partners of the GCN5- (G-) and PCAF- (P-) containing complexes in mESC, we used a tandem immunoprecipitation (IP) strategy. First, complexes containing either one or the other HAT paralogues were separated. Then the ATAC and SAGA complexes were separated successively by immunoprecipitations using antibodies specific to each type of complex (Fig 1A). At this stage, the bead–antibody-bound complexes were washed with 500 mM KCl containing IP buffer. The four preparations were analysed by mass spectrometry (MS) to identify potential interacting partners of the complexes in mESCs. In the four complex preparations, all the known core subunits of the SAGA and ATAC complexes were identified with high confidence (Table 1; supplementary Fig S1A online). Unexpectedly, in the G-ATAC preparation, 19 subunits of the MED were also identified (Table 1). By contrast, no MED subunits were identified in the different SAGA immunoprecipitations or in the P-ATAC preparation (Table 1; supplementary Fig S1A online; data not shown). This observation indicates a new specific interaction between G-ATAC and the MED complexes in mESCs. Remarkably, whereas the MED subunit MED26 was identified with good confidence, MED kinases CDK8 or CDK11 were not detectable in our preparation. These data, together with the fact that we also identified subunits of Pol II (Table 1), indicate strongly that the active PC2-Pol II form of the MED is the one that associates preferentially with the GCN5-containing ATAC.

To validate further the G-ATAC–MED interaction and the lack of interaction between the SAGA and MED complexes, we separated equivalent amounts (normalized to their GCN5 content) of ATAC (ADA2a IP) and SAGA (TRRAP IP) complexes by sodium dodecyl sulphate–polyacrylamide gel electrophoresis and tested

scientific report

Table 1 | Results of mass spectrometry analysis of GCN5- and PCAF-ATAC showing the percentage of coverage and the number of unique peptides found for each subunit

Uniprot ID	GCN5-ATAC		PCAF-ATAC	
	Coverage (%)	Unique peptides	Coverage (%)	Unique peptides
ATAC core				
YETS2	67.4	64	47.3	41
ZZZ3	56.5	36	24.3	17
SGF29	53.6	13	45.7	8
MBIP1	46.3	14	36.7	11
WDR5	44	10	40.7	8
TAD2L	42.9	15	28.4	9
TAD3L	36.1	11	16.4	5
CSR2B	27.9	20	14.8	11
GCNL2	23.4	19	—	—
PCAF	—	—	21.2	14
NC2B	4	1	4	1
Mediator				
MED27	31.5	5	—	—
MED15	21.6	12	—	—
MED28	21.3	2	—	—
MED20	20.8	4	—	—
MED17	19.1	9	—	—
MED31	18.3	2	—	—
MED4	15.6	2	—	—
MED14	15	13	—	—
MED21	13.9	1	—	—
MED26	12.9	5	—	—
MED8	12.3	2	—	—
MED18	8.7	2	—	—
MED6	8.1	2	—	—
MED24	7.1	5	—	—
MED12	6.1	8	—	—
MED16	6	4	—	—
MED23	5.7	6	—	—
MED1	3.3	4	—	—
MED13	1.2	2	—	—
RNA Pol II				
RPB1	4.8	7	—	—
RPB2	2.9	3	—	—

ATAC, Ada-Two-A-containing; GCN5, general control of amino-acid synthesis 5; PCAF, p300/CBP-associated factor; RPB, RNA polymerase B subunit.

for the presence of MED subunits in these immunopurified complexes by western blot. In agreement with the MS data, MED1 was detected in ATAC, but not in SAGA immunoprecipitations (Fig 1B). The association between G-ATAC and MED was also confirmed using different antibodies specific to either ATAC (a-ATAC2) or MED (a-MED1). These immunoprecipitations further confirmed the results obtained with the a-ADA2a IP (supplementary Fig S1B,C online). Moreover, the G-ATAC–MED interaction was not mediated by DNA because when the nuclear extracts were treated with either DNAse I or ethidium bromide we obtained the same co-immunoprecitation results as before (supplementary Fig S2 online). When the presence of MED subunits was investigated in the GCN5- and PCAF-containing complexes by loading equivalent amounts of purified G-ATAC and P-ATAC (normalized for ATAC2; Fig 1C), the MED1 subunit was detected only in the GCN5-containing complexes. This result confirms our MS data and indicates, for the first time, a functional difference between G-ATAC and P-ATAC. Moreover, this indicates that GCN5 probably has a structural role in the formation of the interaction(s) between ATAC and the MED complexes. However, as GCN5 is shared between SAGA and ATAC, further ATAC-specific subunits and new MECO subunits could also be involved in the interaction (see below).

Next, we analysed the G-ATAC–MED stability by increasingly concentrated salt washes of the bead–antibody-bound complexes. This experiment indicated that G-ATAC–MED interaction is extremely stable as the a-ADA2a IP co-purified MED1 up to 1 M KCl washes (Fig 2A). This result indicates that, in contrast to the previously reported low-stability interaction between hSAGA and MED (Liu et al, 2008), the G-ATAC–MED complex seems to be a very stable MECO structure formed by the assembly of ATAC and MED in vivo.

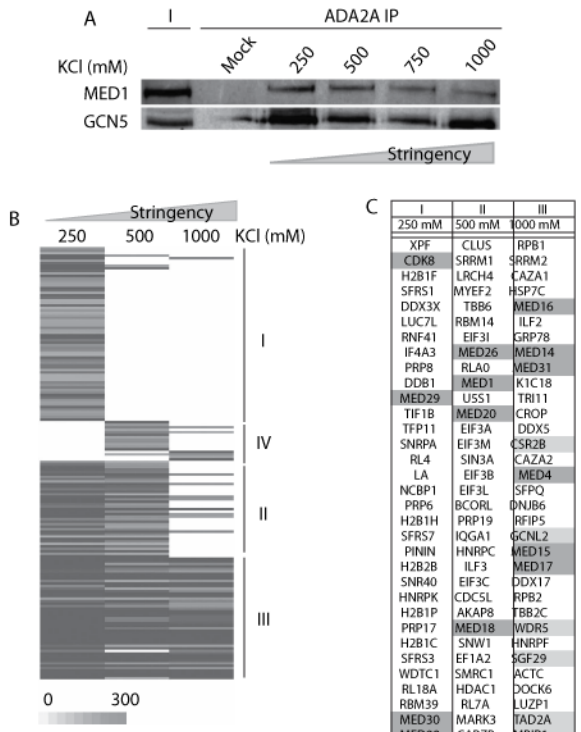
Altogether, the above data show that in mESCs, G-ATAC, but not SAGA nor P-ATAC, specifically co-purifies with a fraction of the PC2 MED26-containing MED.

The ATAC–MED interaction is not restricted to mESC

To analyse whether this newly identified ATAC–MED complex is formed specifically in mESCs because of their particular pluripotent state, or whether this complex can be formed in other cell types, we repeated the ATAC immunopurification in different mouse cell lines representing different stages of differentiation, and in human cell lines from which ATAC was originally purified (HeLa and HEK293Tl; Wang et al, 2008; Nagy et al, 2010). Interestingly, we observe that this interaction can be found in cell lines other than mESC, as MED1 also co-purifies with GCN5 in fibroblastic lineages (MEF, FL83B). In agreement with previous reports, this interaction could not be observed in HEK293T cells and was very weak in HeLa cells. These data show that the association of the ATAC and the MED is not restricted to pluripotent cells, and that its abundance varies in the different cell lines tested, indicating that the formation of this MECO complex is dependent on the cellular context. This result also indicates that a bridging factor might exist between the two complexes in some cell types, but not in others.

Candidates to form the bridge between the two complexes

To isolate candidate bridging factors between ATAC and MED, we analysed by MS different ATAC preparations in which the



I	II	III
250 mM	500 mM	1000 mM
XPF	CLUS	RPB1
CDK8	SRRM1	SRRM2
H2B1F	LRCH4	CAZA1
SFRS1	MYEF2	HSP7C
DDX3X	TBB6	MED16
LUC7L	RBM14	ILF2
RNF41	EIF3I	GRP78
IF4A3	MED26	MED14
PRP8	RLA0	MED31
DDB1	MED1	K1C18
MED29	U5S1	TRI11
TIF1B	MED20	CROP
TFP11	EIF3A	DDX5
SNRPA	EIF3M	CSR2B
RL4	SIN3A	CAZA2
LA	EIF3B	MED4
NCBP1	EIF3L	SFPQ
PRP6	BCORL	DNJB6
H2B1H	PRP19	RFIP5
SFRS7	IQGA1	GCNL2
PININ	HNRPC	MED15
H2B2B	ILF3	MED17
SNR40	EIF3C	DDX17
HNRPK	CDC5L	RPB2
H2B1P	AKAP8	TBB2C
PRP17	MED18	WDR5
H2B1C	SNW1	HNRPF
SFRS3	EF1A2	SGF29
WDTC1	SMRC1	ACTC
RL18A	HDAC1	DOCK6
RBM39	RL7A	LUZP1
MED30	MARK3	TAD2A
MED28	CAPZB	MBIP1
RL12	MATR3	MED27
PSPC1	PABP1	RPB3
OGT1	EMSY	HNRPM
TRA2B	K2C8	MARK2
PCM1	SHKB1	MED24
TRA2A	SMRD1	ACTS
RS18		YETS2
RALY		TBB2B
NONO		MED23
H2B1B		ZZZ3
PR40A		MED6
BAT2L		MED12
PLRG1		K0802
MGN		TBA1B
SFRS4		MED13
SYF1		XPF
H2B1M		
KV3AG		
NC2B		
H2B1K		
MED8		
PUF60		
MED25		
ACINU		
RL24		
SK2L2		
DHX15		
MD13L		
RNPS1		
K2C72		
VIME		
SFRS5		
MGNR		
FUSIP		
EIF3H		
CRNL1		
K2C79		

b Fig 2 | Identification of candidates for constituting the Ada-Two-A-containing–Mediator molecular bridge. (A) ATAC complexes were immunoprecipitated with an a-ADA2a antibody and the bead–antibody-bound complexes were washed with the indicated salt-containing immunoprecipitation (IP) buffer (from 250 to 1000 mM) three times before elution. The purified complexes were then analysed by western blot with the indicated antibodies. (B) ADA2a-containing complexes were prepared as in Fig 1A except that the bead–antibody-bound complexes were washed with 250, 500 and 1000 mM KCl-containing buffers. Proteins in these three ATAC complex preparations were analysed by mass spectrometry. The heat-map represents the clustering analysis of the proteins identified in the ATAC purifications under different salt stringencies. Proteins were grouped according to their enrichment score variation (from 0 to 300 MASCOT unit) in the different conditions. Three groups representing gradual stability in the complex were isolated: I lost at 500 mM; II lost at 1000 mM; III stable in all conditions tested; VI only present at high stringencies. (C) Table presenting the Uniprot ID for the proteins for cluster categories I–III. Light grey denotes ATAC subunits and medium grey MED subunits. ATAC, Ada-Two-A-containing; GCN5, general control of amino-acid synthesis 5; MED, mediator complex.

bead–antibody-bound ATAC complexes were washed with increasing salt concentrations. As the G-ATAC–MED complex is stable up to 1 M KCl washes (Fig 2A), we prepared G-ATAC complexes at three selected washing stringencies (250, 500 and 1000 mM KCl) and analysed by MS the associated proteins in each condition. Next, all the identified proteins in each individual list were clustered according to their enrichment scores, to group proteins behaving similarly in the three different conditions (Fig 2B). Four distinct groups of proteins could be isolated. Group III contained proteins observed at all stringencies, which were also unaffected by the 1 M KCl wash. Groups I–II and IV represented groups of proteins that disappear (groups I–II) or appear (group IV) with the increasing stringency of washes used. Remarkably, we observed that almost all ATAC subunits are highly stable and are in group III (Fig 2C), except the 10 kDa NC2 b subunit. Furthermore, most MED subunits were found in group III, meaning that the interaction between ATAC and MED is maintained at all stringencies. Consequently, group III is also likely to contain factors that are potentially involved in the molecular bridge that could have a role in the strong association of G-ATAC with MED. By excluding proteins that are also found in control immunoprecipitations (a-GST) and in other non-ATAC preparations from group III, we could isolate a restricted subset of candidates (supplementary Table S2 online), including the leucine zipper motif-containing protein 1 (LUZP1; Fig 2C; Sun et al, 1996). As LUZP1 presented the highest enrichment score among the candidates (supplementary Table S2 online), we further characterized the function of this protein in the formation of MECO.

LUZP1 is part of the bridge between ATAC and Mediator

To verify whether LUZP1 is a bona fide subunit of the MECO, we compared the ATAC and MED contents of a-GCN5 and a-LUZP1 immunoprecipitations (Fig 3A). We detected LUZP1 in the GCN5 IP and more importantly, we could detect both GCN5 and MED1 in the a-LUZP1 IP, showing that LUZP1 is associated with the newly identified ATAC–MED meta-complex.

To determine whether LUZP1 has a role in the establishment of the interaction between ATAC and MED, we purified the ATAC complex from mESCs derived from Luzp1 knockout mice (Lee et al, 2001) and determined the quantity of the MED remaining associated with ATAC (Fig 3B). In Luzp1 $^{-/-}$ mESCs, we observed a marked (B 50%) decrease in the interaction of MED with ATAC (Fig 3C). These data indicate that LUZP1 facilitates the formation of the molecular bridge between ATAC and MED, but also that further factors or post-translational modifications could be required for this interaction.

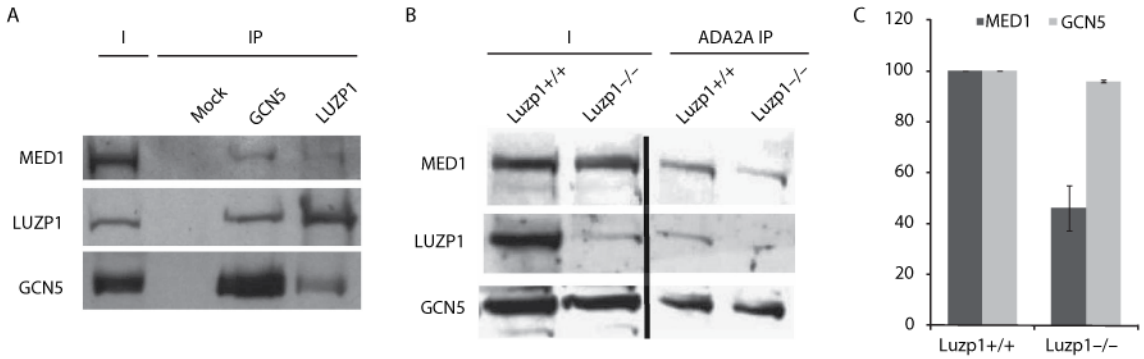


Fig 3 | Leucine zipper motif-containing protein 1 contributes to the molecular bridge between Ada-Two-A-containing and Mediator complexes. (A) Proteins were immunoprecipitated with antibodies raised against GCN5 or LUZP1 from embryonic stem cell extracts and immunoprecipitated complexes were analysed by western blot using antibodies against ATAC (GCN5) and MED (MED1) subunits as well as LUZP1. Input (I) and immunoprecipitation (IP) fractions are shown. (B) Whole-cell extracts were prepared from either Luzp1 (þ /þ) or Luzp1 (−/−) cells. From these extracts ATAC complexes were immunoprecipitated with a-ADA2a antibody. Complexes were analysed by western blot with the indicated antibodies. (C) The quantity of MED1 and GCN5 in Luzp1 (−/−) cells was determined relative to the quantity detected in Luzp1 (þ /þ) cells using the Quantity One software. Mean and range in the value over two biological replicates were calculated. ATAC, Ada-Two-A-containing; GCN5, general control of amino-acid synthesis 5; LUZP1, leucine zipper motif-containing protein 1; MED, mediator complex.

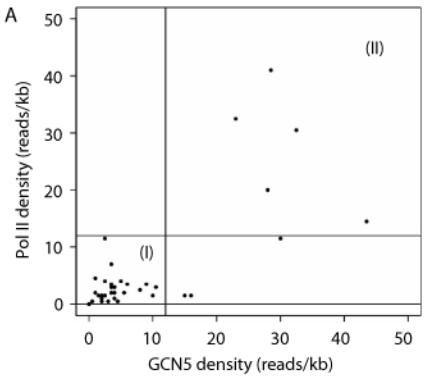
MECO is required for the activity set of ncRNA genes

Two lines of evidence indicate that the newly identified MECO would enhance transcription. First, the subunit composition of the ATAC-associated MED corresponds to the active form of the MED (MED26-containing, CDK-free, Pol II-associated). Second, ATAC is a HAT complex, known to be involved in transcription stimulation of certain stress-regulated genes (Nagy et al, 2010). To gain more insight into the in vivo function of the ATAC–MED complex, we undertook chromatin immunoprecipitation against LUZP1 and GCN5 followed by high-throughput sequencing (ChIP-seq) to determine the common loci to which the LUZP1-containing ATAC–MED complex binds. After bioinformatics analysis, 46 LUZP1 binding sites were identified with high confidence, which most probably represented only the most significant LUZP1 binding events. This low number might be because of the dynamic behaviour of LUZP1 in vivo limiting its crosslinking to chromatin. When comparing these LUZP1 binding sites with those of the genome-wide GCN5 density map obtained after ChIP-seq using mESCs, we observed that most LUZP1 sites lack GCN5, indicating that LUZP1 could also be detected on the genome in an ATAC–MED-independent manner (group I, Fig 4A). When the genome-wide LUZP1 binding sites were compared with those bound by GCN5 and Pol II (available from mESCs for Pol II; Mikkelsen et al, 2007), we identified a number of sites that were bound by all three factors (group II, Fig 4A). In agreement with the three ChIP-seq data sets, our ChIP-quantitative PCR (CRIP-qPCR) validation indicated that these sites were bound by LUZP1, GCN5, Pol II and MED1 (Fig 4C). Surprisingly, we observed that all the identified, bound genes belong to the family of genes expressing unspliced ncRNAs (Fig 4B), most of them being small nuclear RNA (snRNA) genes. To investigate how general is the recruitment of MECO subunits on snRNA promoters, we systematically analysed GCN5, Pol II and LUZP1 levels at all mouse snRNA promoters. We observed that the relevant GCN5 levels are detected systematically at active Pol II-transcribed snRNA promoters, indicating that the regulation of the expression of snRNAs by MECO would be a widespread mechanism (supplementary Fig S3 online).

To characterize the function of MECO at the identified ncRNA promoters, we first tested their expression in Luzp1 $^{-/-}$ ES cell lines. Under these conditions no marked decrease in the tested ncRNA gene expression levels was observed (supplementary Fig S5 online). We hypothesize that the partial effect of LUZP1 knockout on MECO formation is not sufficient to reduce the expression of these genes. Next, we carried out a series of knockdowns against catalytic (GCN5) and structural subunits (ATAC2 and MED1) of ATAC and MED complexes and measured the changes in the expression levels of the four MECO target genes by quantitative PCR after reverse transcription of RNA (RT–qPCR). The downregulation of the three MECO subunits was around 60% (supplementary Fig S4 online). Under these conditions we observed a marked decrease (between 20–60%) in the relative expression levels of the target genes (snRNA and long ncRNA; Fig 4D). This result indicates that MECO is a main regulator of the tested ncRNAs.

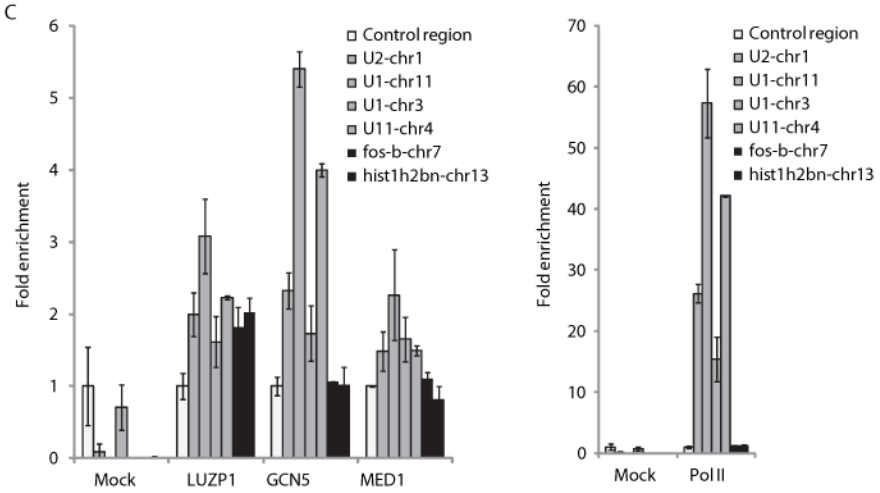
Interestingly, transcriptional mechanisms of these snRNAs are known to be different from those of messenger RNAs (Jawdekar & Henry, 2008). Although the precise mechanistic action of this bridged ATAC–MED MECO and its integration into the snRNA biosynthesis mechanisms need to be studied further on the newly identified genomic loci, our results clearly indicate a new role for the identified MECO complex in the expression of these ncRNA transcripts.

We have shown the existence of a new MECO formed in cells by the stable association of the G-ATAC and the active PC2 form of the MED complexes. The integrity of this complex depends partly on the presence of LUZP1 in MECO and this complex is required for the active transcription of a particular class of non-coding transcripts. Our findings support the new idea that coactivator networks are functionally interconnected and, under particular conditions, can be modulated in stable meta structures to potentiate their actions at particular sites on the genome.



B

Genomic coordinate	Gene name	Category
chr11: 87236290–87236891	U1	snRNA
chr19: 5800961–5801470	MALAT-1	ncRNA
chr4: 131825690–131826167	U11	snRNA
chr3: 96132459–96132952	U1	snRNA
chr1: 72301807–72301995	U2	snRNA



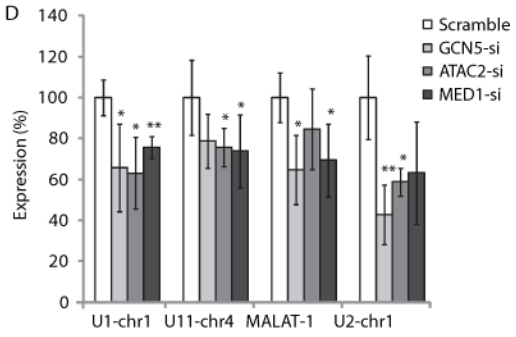


Fig 4 | The Ada-Two-A-containing–Mediator meta-complex is bound to transcriptionally active non-coding RNA loci together with leucine zipper motif-containing protein 1 in mouse embryonic stem cells. (A) Two genome-wide ChIP-seq experiments were carried out in mESC using both LUZP1 and GCN5 polyclonal antibodies. The tag density over a 2 kb region around the identified LUZP1 binding sites was then compared with those of GCN5 or Pol II. The plot is delimited by bars representing cut-off(15 reads/kb as a cut-offvalue) used in further subdivision of the data. (I) Loci bound only by LUZP1; (II) loci bound by LUZP1, GCN5 and Pol II. (B) Analysis of the functional category of the transcripts neighbouring the group II loci (± 500 bp). (C) Validation of the ChIP-seq data by ChIP-qPCR quantification of LUZP1–MECO binding on control loci (white bars) and loci from category I (LUZP1 only, black bars) and II (LUZP1–MECO, grey bars) using the indicated antibodies. (D) Quantification of non-coding gene expression changes on MECO subunit knockdowns. Measurements of the efficiencies of siRNA knockdowns against MECO subunits are presented in supplementary Fig S4 online. The relative levels of expression of MECO target genes after knockdown of GCN5 (light grey bars), ATAC2 (medium grey bars) and MED1 (dark grey bars) by siRNA relative to scramble siRNA (white bars) were quantified by reverse transcription–qPCR. Mean and standard deviation over three biological replicates were calculated.P-values were calculated using an unpairedt-test for triplicates (*Po 0.05; **Po 0.01). ATAC, Ada-Two-A-containing; IP, immunoprecipitation; ChIP-qPCR, chromatin immunoprecipitation-quantitative PCR; ChIP-seq, chromatin immunoprecipitation-coupled high-throughput sequencing; chr, chromosome; GCN5, general control of amino-acid synthesis 5; LUZP, leucine zipper motif-containing protein 1; MECO, meta-coactivator complex; MED, mediator complex; mESC, mouse embryonic stem cell; ncRNA, non-coding RNA; siRNA, small interfering RNA; snRNA, small nuclear RNA.

Moreover, it indicates that ncRNA genes have different needs for chromatin-modifying coactivator assemblies than have the broadly studied protein-coding genes.

METHODS

Protein immunopurification and MS analysis. Detailed IP procedure can be found in supplementary methods online. For MS analysis, mESC nuclear extracts were subjected to tandem IP followed by peptide competition elution. Retrieved purified complexes were concentrated and loaded on acrylamide gels. Gel lanes were cut into slices and each was washed and trypsinized. Proteolytic peptides were then collected and analysed by nanoflow liquid chromatography–tandem mass spectrometry.

ChIP-seq. ChIP was carried out as described previously (Nagy et al, 2010). For ChIP-seq analysis, 300 mg of chromatin (DNA) was incubated with GST (mock), LUZP1 and GCN5 antibodies. Retrieved purified DNA was sequenced using Illumina Genome Analyzer II following the manufacturer's recommendations.

Peptides and antibodies. Previously described a-GCN5 (2676), a-PCAF (2760), a-ADA2a (2AD2A1), a-TRRAP (2TRA1B3), a-ATAC2 (2734) and a-GST (TFII1D10; Nagy et al, 2010); a-mLUZP1 (Sun et al, 1996); and commercial MED1 (sc-5334) and MED26 (sc-48776) were used for IP, western blotting and ChIP experiments, respectively.

Bioinformatics analysis. Detailed bioinformatics procedures can be found in the supplementary methods online, and see supplementary Table S1 online for quality control. Briefly, LUZP1 enrichment clusters were detected by using model-based analysis of ChIP-seq (MACS; Zhang et al, 2008). After repeat masking, peaks were ranked and a cut-offwas determined according to ChIP-qPCR validation results. Loci were annotated using a genomic position annotation tool (Krebs et al, 2008). Average read densities were collected in GCN5 and Pol II (GSE12241; Mikkelsen et al, 2007) raw data sets and compared with a mock ChIP data set.

Supplementary information is available at EMBO reports online (http://www.emboreports.org).

Note added in proof. The raw and processed ChIP-seq datasets have been deposited in the Gene Expression Omnibus (GEO) under the accession number GSE21717.

ACKNOWLEDGEMENTS

We are grateful to M. Orpinell and D. Devys for critical reading of the paper, M. Meinsterernst for antibodies, R. Poot and M. Ballarino for advice, A. Dierich for help in mESC culture, the IGBMC high-throughput sequencing platform, G. Duval, P. Eberlin, IGBMC purchase service for their help, and K. Bezstarosti for technical support. A.R.K. is the recipient of a fellowship from INSERM-Re´gion Alsace and the Association pour la Recherche sur le Cancer. This work was funded by grants from Agence Nationale de la Recherche (GenomATAC; ANR-09-BLAN-0266) and the European Union (EUTRACC, LSHG-CT-2007-037445).

CONFLICT OF INTEREST

The authors declare that they have no conflict ofi nterest.

REFERENCES

Jawdekar GW, Henry RW (2008) Transcriptional regulation of human small nuclear RNA genes. Biochim Biophys Acta 1779: 295–305

Krebs A, Frontini M, Tora L (2008) GPAT: retrieval of genomic annotation from large genomic position datasets. BMC Bioinformatics 9: 533

Lee MW, Chang AC, Sun DS, Hsu CY, Chang NC (2001) Restricted expression of LUZP in neural lineage cells: a study in embryonic stem cells. J Biomed Sci 8: 504–511

Liu X, Vorontchikhina M, Wang YL, Faiola F, Martinez E (2008) STAGA recruits Mediator to the MYC oncoprotein to stimulate transcription and cell proliferation. Mol Cell Biol 28: 108–121

Malik S, Roeder RG (2005) Dynamic regulation of pol II transcription by the mammalian Mediator complex. Trends Biochem Sci 30: 256–263

Mikkelsen TS et al (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448: 553–560

Nagy Z, Riss A, Fujiyama S, Krebs A, Orpinell M, Jansen P, Cohen A, Stunnenberg HG, Kato S, Tora L (2010) The metazoan ATAC and SAGA coactivator HAT complexes regulate different sets ofi nducible target genes. Cell Mol Life Sci 67: 611–628

Paoletti AC, Parmely TJ, Tomomori-Sato C, Sato S, Zhu D, Conaway RC, Conaway JW, Florens L, Washburn MP (2006) Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. Proc Natl Acad Sci USA 103: 18928–18933

Sun DS, Chang AC, Jenkins NA, Gilbert DJ, Copeland NG, Chang NC (1996) Identification, molecular characterization, and chromosomal localization of the cDNA encoding a novel leucine zipper motif-containing protein. Genomics 36: 54–62

Thomas MC, Chiang CM (2006) The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol 41: 105–178

Wang YL, Faiola F, Xu M, Pan S, Martinez E (2008) Human ATAC is a GCN5/PCAF-containing acetylase complex with a novel NC2-like histone fold module that interacts with the TATA-binding protein. J Biol Chem 283: 33808–33815

Zhang Y et al (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biol 9: R137

SUPPLEMENTARY INFORMATIONS

Supplementary Methods

Experimental methods:

Protein immuno-purification and MS analysis

Immunoprecipitations were carried as previously described (Nagy, Riss et al.). For MS analysis, 8 mg of mESC nuclear extract were subjected to tandem overnight IP followed by peptide competition elution (2 mg/mL). Retrieved purified complexes were concentrated on Vivaspin 500 PES columns (100MDa, VS0101) and loaded on acrylamide gels (Invitrogen NP0321BOX). Gel lanes were cut into 13 slices and each of them was washed and subjected to in-gel proteolysis with trypsin. Proteolytic peptides were then collected and analyzed by nanoflow LC-MS/MS on an LTQ-Orbitrap thermo mass spectrometer. The raw data from each of the slices were combined for each lane and used for a protein database search using Mascot searching software.

DNA removal from nuclear extracts:

DNase I treatment was performed by incubating the nuclear extracts with 0.1U/µL of DNase, in $CaCl_2$ 10mM, 30 minutes 37°C. Ethidium bromide treatment was performed as described in (Lai and Herr 1992).

siRNA transfection

siRNA transfection procedure for mESC was adapted from (Fazzio, Huff et al. 2008). Commercially available (Dharmacon) siRNA mixture against GCN5 (L-040665-01), MED1 (L-040964-01), ATAC2 (L-058367-00), scramble (D-001810-10-20) were transfected using Lipofectamin 2000 (Invitogen). Knock down was monired after 72h incubation.

Expression level monitoring

Total mRNA was prepared from mESC using Trizol Reagent (Invitrogen) and reverse transcribed using AMV reverse transcriptase kit (Roche) following manufacturer

recommendations. Relative RNA level was determined for precursor snRNA and ncRNA using specific set of primers (sequence upon request) and normalized to a house keeping gene (mARBP- NM_022402).

Bioinformatics analysis:

ChIP-seq data:

Illumina raw files were aligned against reference (mouse mm9) genome using the eland program allowing one mismatch (Quality Control data regarding sequencing can be found in Supplementary Table 1). Enrichment clusters were detected using MACS (Zhang, Liu et al. 2008). The retrieved peaks were filtered (max size 1000bp, min size 100bp) and repeat masked, to establish the final binding sites lists. Peaks were annotated using the GPAT web server (Krebs, Frontini et al. 2008) using ENSEMBL gene 57 database.

The binding site list of LUZP1 was used as a reference to collect densities from different tracks (GCN5, Pol II, mock) around the binding sites (−1000 to +1000 bp with respect to the centre of the peak). Density was calculated by counting the tags found in the window around the binding sites. Cut off value for presence/absence of a factor was estimated by comparing specific IP track densities with mock (no-antibody IP) density track.

A similar procedure was applied to collect densities around the transcription start site of the 861 snRNA genes present in the ENSEMBL gene 57 database.

Mass spectrometry data:

Peak lists were automatically created from raw data files using the Mascot Distiller software (version 2.1; MatrixScience). The Mascot search algorithm (version 2.2, MatrixScience) was used for searching against the NCBInr database (release NCBInr_20090222; taxonomy: *Mus musculus*). The peptide tolerance was typically set to 10 ppm and the fragment ion tolerance to 0.8 Da. A maximum number of 2 missed cleavages by trypsin were allowed and carbamidomethylated cysteine and oxidized methionine were set as fixed and variable modifications, respectively. The Mascot score cut-off value for a positive protein hit was set to 60. Individual peptide MS/MS spectra with Mascot scores below 40

were checked manually and either interpreted as valid identifications or discarded. Typical contaminants, also present in immunopurifications using beads coated with pre-immune serum or antibodies directed against irrelevant proteins were omitted from the table.

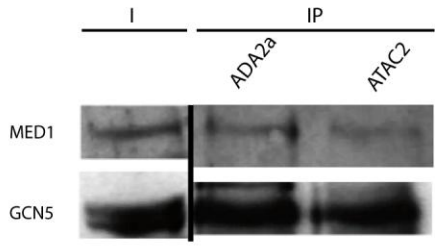
For the analysis of ATAC preparation purified with different salt concentrations, list of specifically identified proteins were established. In order to define protein groups behaving similarly, the proteins were clusterized according to their MASCOT score Rapid identification of proteins by peptide-mass fingerprinting in the primary identification(Pappin, Hojrup et al. 1993). Clustering was performed using Cluster 3.0 (http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/software.htm) using kmean clustering method. Visualisation as well as gene list extraction was achieved using Java treeview (Saldanha 2004).

Supplementary Material

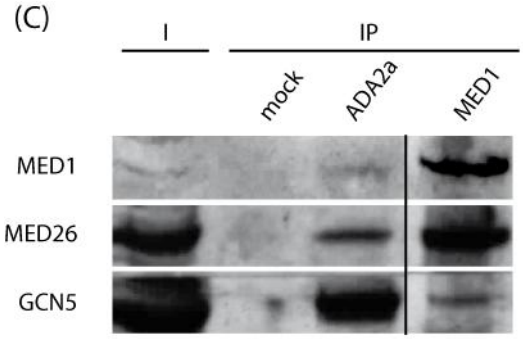
(A)

	Uniprot ID	GCN5-SAGA	
		Coverage (%)	Unique peptides
SAGA core	**TRRAP**	45,8	100
	GCNL2	29,9	25
	PCAF	-	-
	SPT20	41,7	20
	TAF6L	37,7	17
	TAF5L	37,7	16
	TAF9	61,4	12
	USP22	28,1	11
	ATXN7L2	19,5	10
	TAD3L	29,4	9
	ATX7L3	33,1	9
	TAD1L	25,4	8
	SPT3	37,7	8
	SGF29	25,6	7
	ATX7L1	16,6	7
	ATX7	10,6	6
	SPT7	16,9	5
	TAF12	18	3
	TAF10	20,6	2
	SPT7	14,1	5
	TAD2L	23,8	10

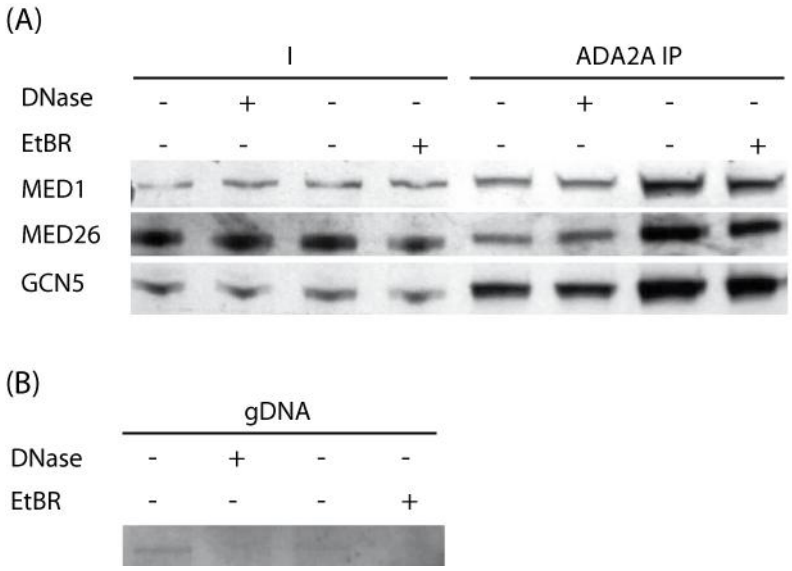
(B)



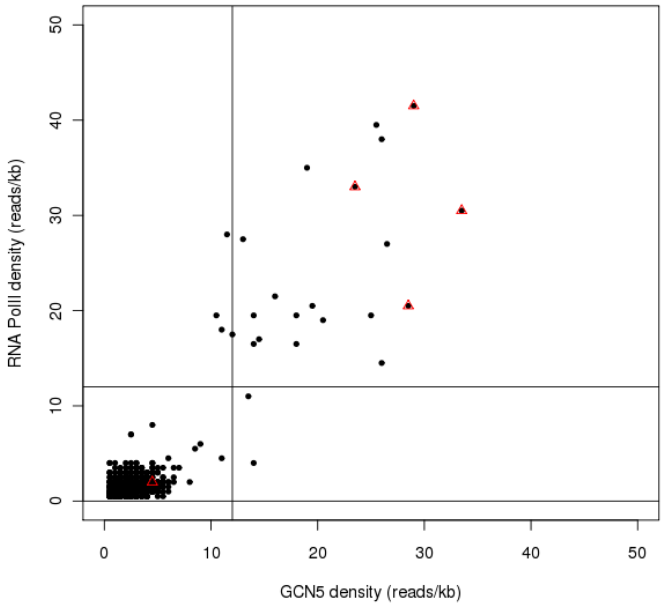
(C)



Supplementary Figure 1 online:
(A) Results of mass spec analysis of G-SAGA showing the percentage of coverage and the number of unique peptides found for each sub-unit. SAGA HAT sub-units are highlighted in gray. (B) Equivalent amounts of ATAC complexes purified by α-ADA2a IP and α-ATAC2 IP (normalized for their GCN5 content) were separated by SDS-PAGE and analysed by Western blot against ATAC-specific (GCN5) and MED-specific subunit (MED1). Input (I) and IP-ed fractions are shown. (C) α-ADA2a and α-MED1 co-immunoprecipitated proteins were separated by SDS-PAGE and analysed by Western blot against ATAC-specific (GCN5) and MED-specific subunit (MED1 and MED26). Input (I) and IP-ed fractions are shown.

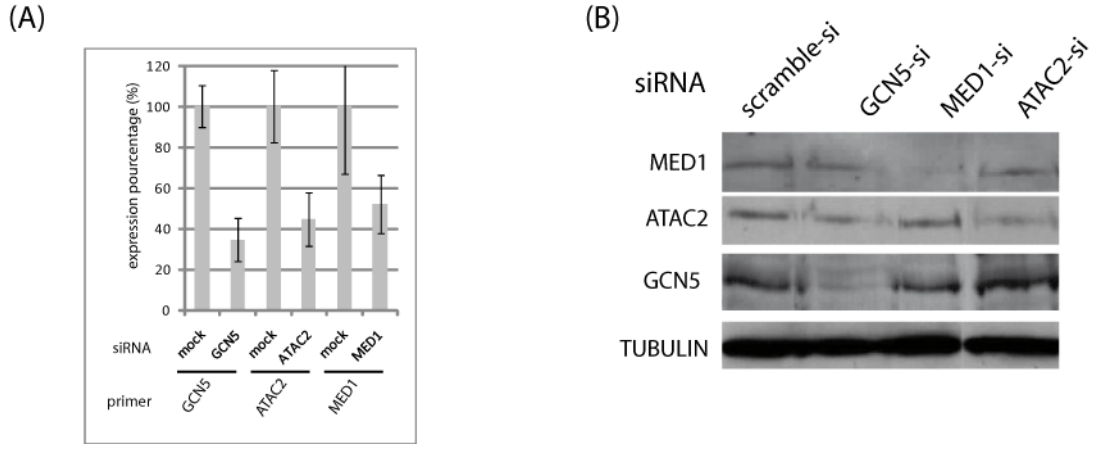


Supplementary Figure 2 online:
(A) MECO formation is not dependant on the presence of contaminating DNA in the Nuclear Extracts. ATAC complexes were IP-ed with an α-ADA2a antibody from nuclear extracts pre-treated with DNase I or Ethidium Bromide. The purified complexes were then analysed by Western blot with the indicated antibodies. (B) DNA removal from nuclear extracts. Ethidium bromide stained agarose gel showing depletion of gDNA after either DNase I or Ethidium Bromide treatment of the extracts. Input (I) and IP-ed fractions are shown.

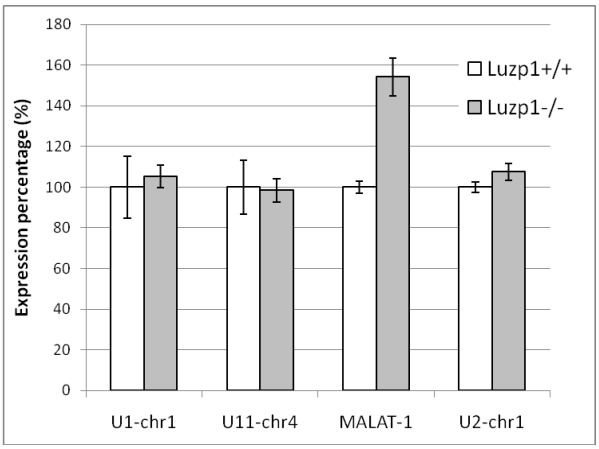


Supplementary Figure 3 online:
GCN5 systematically occupies active Pol II transcribed snRNA promoters. The tag density over a 2 kb region around the 5' of all snRNA genes (861 from EMSEMBL gene 57) was collected and plotted for GCN5 and Pol II ChIP-seq data. Same data collection was performed for LUZP1 data and sites presenting densities above a certain cut off were highlighted on the

plot (red triangle >15 read/kb). The plot is delimited by bars representing cut-off (15read/kb as a cut off value) used in further subdivision of the data.



Supplementary Figure 4 online:
Measurement of knockdown efficiency of siRNA treatment against MECO sub-units. (A) Total RNA extracts were prepared after 72 hours of treatment by the corresponding siRNA. The extracts were then reverse transcribed and relative transcript abundance was measured by qPCR using primers specific for the targeted mRNA. Results are presented relative to mRNA level in cells transfected with scramble siRNA. Mean and standard deviation over three biological replicates was calculated. (B) Whole cell extracts were prepared after 72 hours of treatment by the corresponding siRNA (horizontal legend). The extracts were then analysed by Western blot with the indicated antibodies (vertical legend).



Supplementary Figure 5 online:
Quantification of non coding gene expression changes in Luzp1 -/- ES cells. The expression levels of MECO target genes were quantified by Q-RT-PCR in wild type and Luzp1-/- ES cells. The expression levels of MECO target genes in *Luzp1*-/- cells (light grey bars), relative to *Luzp1* (+/+) cells (white bars, taken as 100%) are represented. Mean and standard deviation over three qPCR replicates was calculated.

Table S1 online: Overview of sequenced and mapped tags per data set.

ChIP-seq quality control data for the present study (grey lanes) are compared to recently published results (white lanes) with similar technology.

		total read number	input for alignment (mm9)	aligned tags (<1 mismatch)	unique tags	percentage aligned
present data	mock	1,75E+07	7,38E+06	7,31E+06	**2,73E+06**	42%
	GCN5	1,47E+07	8,21E+06	8,17E+06	**4,81E+06**	56%
	LUZP1	1,66E+07	6,83E+06	6,79E+06	**4,21E+06**	41%
(Welboren, van Driel et al. 2009)	ERα minus ligand	3,38E+06	1,44E+06	NA	NA	43%
	ERα E2	1,27E+07	8,31E+06	NA	NA	66%

Table S2 online: Full list of candidates for forming the molecular bridge between ATAC and MEDIATOR

Proteins found at all wash stringencies were sorted by mascot score, functionally classified (Highlighting. Green: ATAC ; orange: MEDIATOR ; purple RNA Pol II), known classical mass spectrometry contaminants were excluded (grey). A list of possible new candidates for forming ATAC-MED molecular bridge was isolated (bold) including LUZP1.

UNIPROT ID	MASCOT 250	MASCOT 500	MASCO 1000
Q3TUF7|YETS2_MOUSE	4992	4850	4781
Q8BTI8|SRRM2_MOUSE	3774	3178	2937
Q6KAQ7|ZZZ3_MOUSE	2851	2665	2476
Q3UHU5|K0802_MOUSE	2566	1706	2161
Q9JHD2|GCNL2_MOUSE	1648	1576	1804
Q9D0E1|HNRPM_MOUSE	2086	1501	1427
Q8CID0|CSR2B_MOUSE	1276	1402	1311
Q99LQ1|MBIP1_MOUSE	1230	1126	1028
Q8CHV6|TAD2A_MOUSE	1053	817	841
A2ABV5|MED14_MOUSE	1169	918	836
A2AGH6|MED12_MOUSE	914	435	701
P68033|ACTC_MOUSE	501	612	694
P68134|ACTS_MOUSE	501	612	694
P05784|K1C18_MOUSE	1034	704	632
Q924H2|MED15_MOUSE	583	540	630
Q8VDR9|DOCK6_MOUSE	1286	821	588
P63017|HSP7C_MOUSE	1118	780	586
P08775|RPB1_MOUSE	811	357	572
Q8R0L9|TAD3L_MOUSE	774	819	566
P61965|WDR5_MOUSE	744	601	556
Q9DA08|SGF29_MOUSE	659	629	451
P05213|TBA1B_MOUSE	610	545	436
P20029|GRP78_MOUSE	482	376	368
Q61656|DDX5_MOUSE	631	328	303
Q8R4U7|LUZP1_MOUSE	**387**	**311**	**301**
Q99PQ2|TRI11_MOUSE	**168**	**361**	**274**
Q9CWF2|TBB2B_MOUSE	515	467	264
P68372|TBB2C_MOUSE	453	337	262
Q5SWW4|MED13_MOUSE	342	123	247
Q8VCD5|MED17_MOUSE	472	326	233
Q8CFI7|RPB2_MOUSE	446	193	232
Q501J6|DDX17_MOUSE	468	185	231
Q99K74|MED24_MOUSE	333	443	230
Q05512|MARK2_MOUSE	203	0	186
Q8R361|RFIP5_MOUSE	303	268	180
Q6PGF3|MED16_MOUSE	275	201	177
Q9CQA5|MED4_MOUSE	225	300	175
P47753|CAZA1_MOUSE	387	132	172
Q9DB40|MED27_MOUSE	297	242	162

Q80YQ2|MED23_MOUSE	200	140	156
Q9Z2X1|HNRPF_MOUSE	373	205	150
Q921D4|MED6_MOUSE	213	185	146
Q9CXU1|MED31_MOUSE	191	114	143
Q8VIJ6|SFPQ_MOUSE	298	219	137
P47754|CAZA2_MOUSE	127	114	134
O54946|DNJB6_MOUSE	198	151	132
Q5SUF2|CROP_MOUSE	**226**	**116**	**127**
Q9CXY6|ILF2_MOUSE	**250**	**161**	**126**
P97760|RPB3_MOUSE	232	117	126

II. Genome wide mapping of ATAC complex reveals a tissue specific binding at enhancers (unpublished results).

Defining the loci bound by a specific transcriptional regulatory factor, (i) allows defining the gene networks that it regulates, (ii) allows the understanding on the type of loci regulated, (iii) and is a prerequisite to be able to mechanistically dissect its mode of action. In order to define ATAC binding sites, we performed ChIP-seq against ZZZ3, a specific sub-unit of the complex that is not present in SAGA. This experiment was performed in collaboration with a group involved in the large scale Encyclopedia Of DNA Elements (ENCODE), thus allowing the integration of ATAC data with multiple information on chromatin states and other factor binding.

Results

ZZZ3 ChIP-seq was performed in two human cell lines (GM12878-lymphoblastoid cells and HeLa cervical carcinoma cells) chosen by the ENCODE consortium to allow inter-cellular comparison of the results. The set of ATAC binding sites was defined by isolating reads enrichment clusters using MACS (Zhang, Liu et al. 2008). Using stringent search criteria, 372 ATAC binding sites could be identified with high confidence in GM12878 cells (GM).

Starting from this reference set of loci, we decided to define the type of genomic elements bound by ATAC using information on the chromatin state around the identified ATAC bound loci. First, we collected signal densities in ChIP-seq datasets for different chromatin marks known to be markers of promoters (H3K4me3) and enhancers (H3K4me1) (Heintzman, Hon et al. 2009). Additionally, we interrogated the transcription state using RNA Pol II data and the opening of the chromatin using DNAse Hypersensitiviy Sites (DHS) data. Second, we organized these loci by clustering to identify potentially existing categories. The heatmap of the clustered data (Figure 1A) made clearly appear three major groups of ATAC bound loci: (i) loci enriched in H3K4me1 and with low H3K4me3 (red squared) (ii) loci enriched in H3K4me3, Pol II and showing signal in DHS experiments (black squared) (iii) a third group of loci did not present any enrichment for any of the features tested (blue squared).While the group 2 loci have features corresponding to active promoters, group 1 loci have enhancer type features. Average profiles for the three categories have been calculated and comparatively plotted (Figure 1B), and they correspond to the previously described profiles of each category (Heintzman, Hon et al. 2009). We then analysed the genomic location of the loci by calculating the distance to the start site of the closest transcription unit in each category. The frequency plot of the calculated distances shows that while in category

2 most of the binding sites are located close to transcription start sites (TSS) (Figure 1C), in category 1 the binding sites are not showing any preference for TSS (Figure 1C). This observation reinforce our previous conclusion that category 1 would be enhancers loci while category 2 would represent promoters. Taken together, these results demonstrate that ATAC can bind active promoter elements as well as distal enhancer elements. This is the first direct demonstration that the GCN5-Containing Complexe (GCC) ATAC is recruited to enhancers. Thus our results demonstrate that p300 that was previously considered as a universal marker of enhancer is not the only HAT recruited to these functional elements.

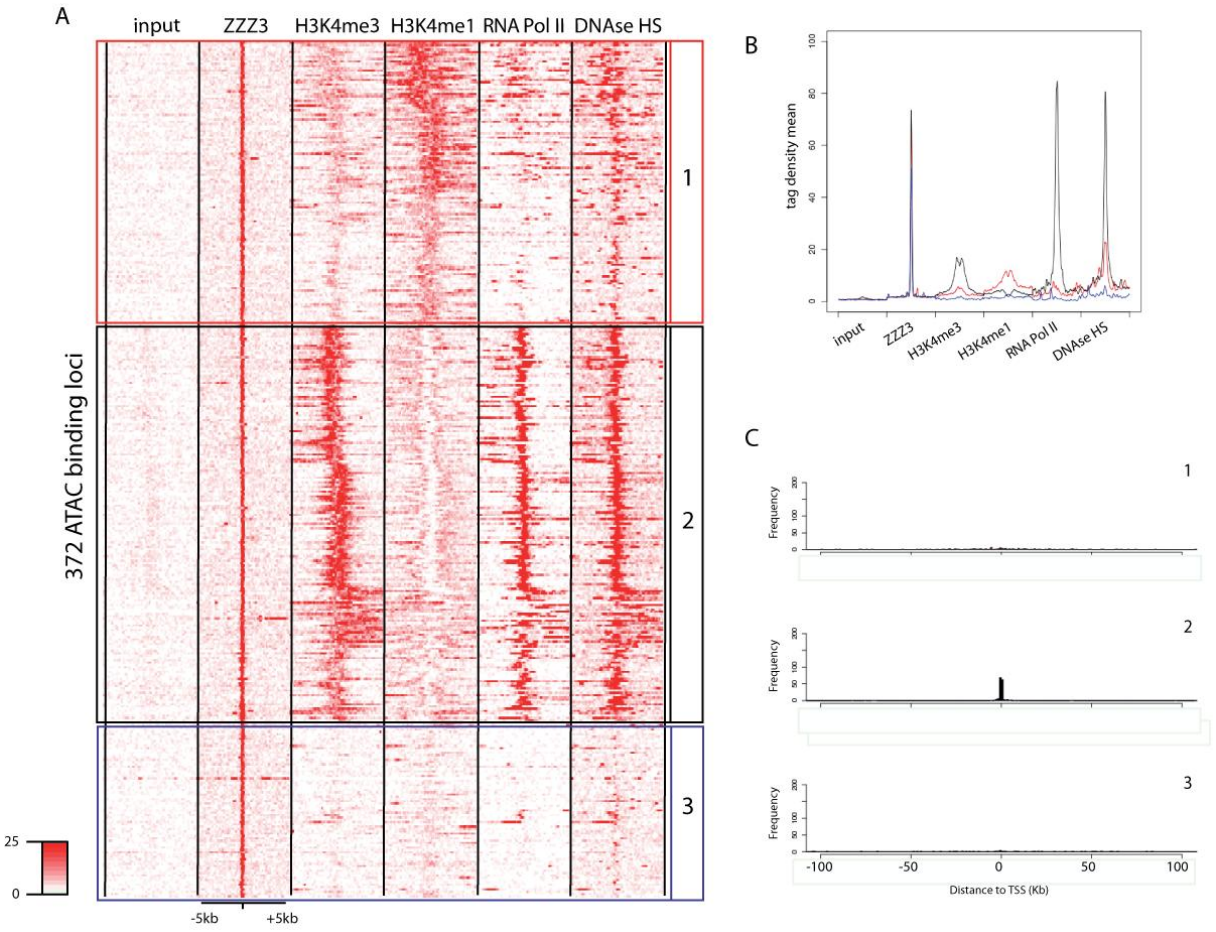


Figure 1: ATAC associates with both promoter and enhancers elements.

(A) Heatmap of the signal density observed on regions surrounding the 372 high confidence ZZZ3 (ATAC) binding sites (±5 kb) for different genomic features (as indicated). The density map was subjected to clustering in order to create groups of loci sharing the same genomic profile. Then loci were classified as (1-red) active enhancer when enriched for H3K4me1 and harbours reduced H3K4me3, (2-black) active promoters when enriched for H3K4me3, (3-blue) unassigned when none of the tested features were enriched. (B) Average density profiles of the three previously identified categories (1- red) active enhancers, (2-black) active promoters, (3-blue) unassigned. (C) Frequency plot representing the distance of the ATAC binding sites to the transcription start site (TSS, assumed to be 5' of the transcript)

of the closest gene in the genome in each category previously isolated. Binding sites in both the enhancer (1-red) and unassigned (3-blue) categories do not show any preferential location next to TSS while these in the promoter category (2-black) are preferentially found next to TSS.

Next, we comparatively examined the binding status of ATAC on the GM bound loci in the HeLa cells. To this end, we first used a similar approach as previously described and looked at ZZZ3 density in HeLa over the loci bound in GM (Figure 2A). We observed that while ZZZ3 binding is lost on the enhancers type loci (1-red squared), it seems invariant across cell lines on the promoter category (2-black squared). In order to confirm this observation we used a more quantitative approach by collecting enrichments in both the cell lines over a window around ZZZ3 binding sites in GM. Then for each loci category, we plotted enrichment in one cell line versus the other (Figure 2B). This plot quantitatively shows a correlation between cell lines on promoters (Pearson correlation coefficient=0.78), but not on enhancer (Pearson correlation coefficient=0.35). Taken together, these results demonstrate that ATAC is binding to both active promoters as well as enhancers. Moreover, while promoter binding is invariant across cell types, binding to enhancers appears to be cell type specific.

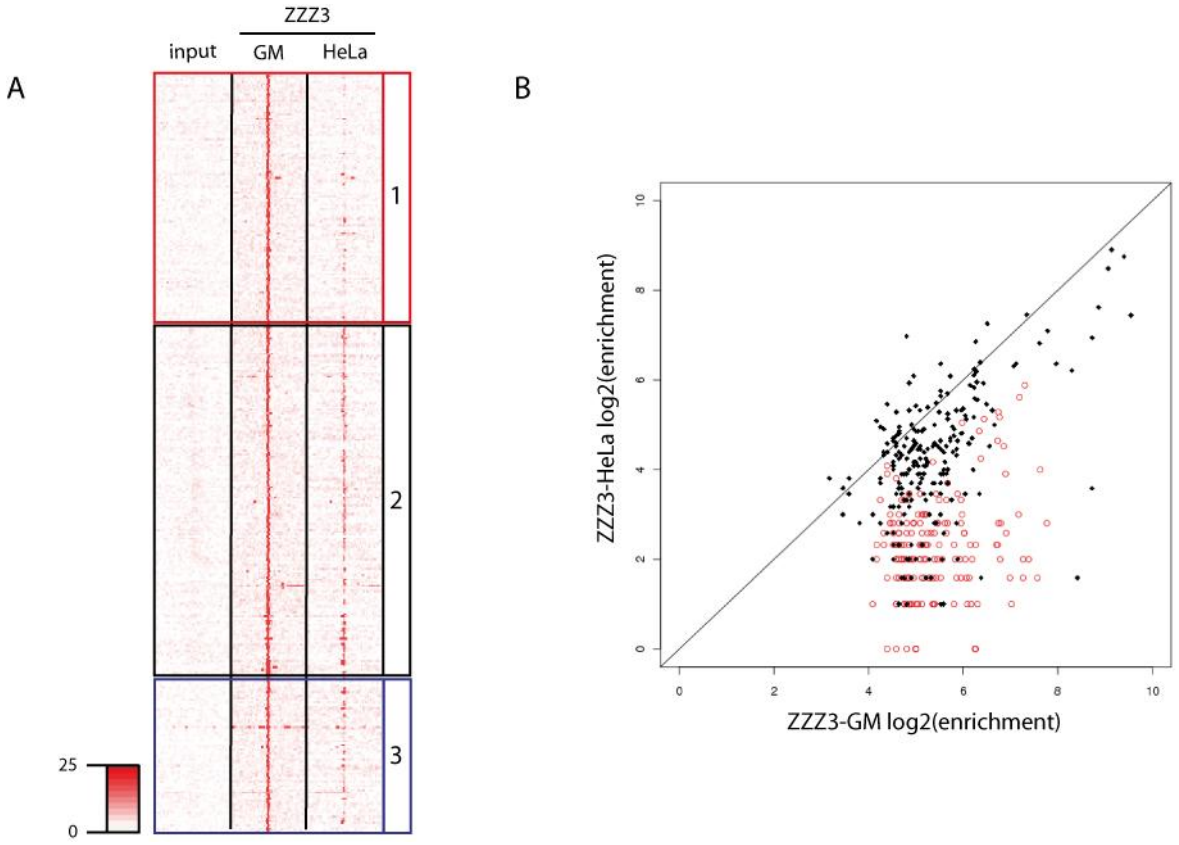


Figure 2: ATAC binding exerts high tissue specificity on enhancer but not promoters. Comparative heatmap of the ATAC (ZZZ3) ChIP-seq signal density in GM12878 and HeLa around the high confidence ZZZ3 binding sites previously identified. Similar loci organization was conserved in order to illustrate differences in cell type binding specificity between enhancer (red), promoter (black) and undefined (blue) type of loci. (B) Quantification of the signal density (log_2) for ZZZ3 in GM12878 (GM) versus HeLa cells over the isolated promoters (black) and enhancers (red).

Conclusions

These results show that p300 is not the only HAT recruited to the enhancer loci. However, comparison of the ATAC bound enhancers with p300 binding loci should be further carried out to determine if both HATs are recruited on the same enhancers or if different enhancers have different HAT dependencies. Interestingly, similar to what was observed for p300 (Heintzman, Hon et al. 2009), ATAC binding to enhancers appears to be cell type specific, suggesting functional similarities between the mode of action of ATAC and p300.

Material and methods

ChIP was performed as previously described (Nagy, Riss et al. 2010). For ChIP-seq analysis, 300µg of chromatin (DNA) was incubated with ZZZ3 antibodies (2616 (Nagy, Riss et al. 2010)). Retrieved purified DNA was sequenced using Illumina Genome Analyzer II following manufacturer recommendations. Illumina raw files were aligned against reference (human hg18) genome using the eland program allowing one mismatch. Enrichment clusters were detected using MACS (Zhang, Liu et al. 2008) using an input DNA as a negative control. The retrieved peaks were filtered (max size 1000bp, min size 100bp), to establish the final binding sites lists. Multiple comparisons, categories separation, heatmap and plots creation were performed using seqMINER functionalities (Ye T, Krebs A et al submitted (see detailed description in a latter section)) with the established ZZZ3 references binding sites list as an input. For the multiple comparisons, density data were collected from raw tracks from resource experiments of the ENCODE project (http://genome.ucsc.edu/ENCODE/downloads.html). Kmeans clustering was used to organize loci and separate categories.

III. Lessons from genome-wide studies: re-definition of HAT co-activator function (Anamika K, Krebs A et al, 2010, Epigenetics & Chromatin 2010, 3:18.)

Lessons from genome-wide studies: re-definition of HAT co-activator function

Krishanpal Anamika[1,2*], Arnaud R. Krebs[1*], Julie Thompson[2], Olivier Poch[2] and Làszlò Tora[1,3]

[1]Department of Functional Genomics, [2]Department of Structural Biology and Genomics, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR 7104, INSERM U 964, Université de Strasbourg, 1 Rue Laurent Fries, 67404 Illkirch Cedex, France

*Equal contribution

[3]Corresponding author. Tel: +33 388653444, Fax: +33 388653201, Email: laszlo@igbmc.fr

Running Title: Novel mode of action of HATs

Subject areas: Functional genomics, Genetics, Biochemistry

Abstract

Histone acetylation is one of the key regulatory mechanisms controlling transcriptional activity in eukaryotic cells. In higher eukaryotes, a number of nuclear histone acetyl-transferase (HAT) enzymes have been identified, most of which are part of a large multi-subunit complex. This diversity, combined with the large number of potentially acetylable lysines on histones suggested the existence of a specific regulatory mechanism based on the substrate specificity of HATs. Over the last decade, intensive characterisations of the HAT complexes have been carried out. However, the precise mode of action of HATs and particularly the functional differences amongst these complexes remain elusive. Here we review current insights into the functional role of HATs focusing on the specificity of their action. Studies based on biochemical as well as genetic approaches suggested that HATs exert a high degree of specificity in their acetylation spectra and in the cellular processes they regulate. A different picture recently emerged from genomic approaches that provided genome-wide maps of HATs recruitment. The careful analysis of these data suggests that all HAT complexes would be simultaneously recruited to a similar set of loci in the genome arguing for a low specificity in their function. In this review, we will discuss the significance of these contradictions and suggest a new model that integrates biochemical, genetic and genome-wide data to better describe the functional specificity of HAT complexes.

Introduction

Histone post-translational modifications have shown to be key regulators among transcription regulation mechanisms (Kingston, Bunker et al. 1996; Eberharter and Becker 2002). Histone acetylation is known to play an important role in the regulation of transcriptional activity in eukaryotic cells (Shahbazian and Grunstein 2007) by affecting higher-order folding of chromatin fibers, loosening of the contacts between the DNA and the nucleosomes, and/or histone-nonhistone protein interactions (Tse, Sera et al. 1998; Carruthers and Hansen 2000; Wolffe and Hansen 2001; Fischle, Wang et al. 2003; Shogren-Knaak, Ishii et al. 2006). Histone acetylation on various target lysines, is in general positively associated with gene expression. Thus, HATs are thought to increase the decompaction of chromatin, which in turn may increase the accessibility of factors that promote transcription (Krajewski and Becker 1998a; Akhtar and Becker 2000; Sterner and Berger 2000b; Shogren-Knaak, Ishii et al. 2006). In higher eukaryotes, two enzymatic families (GNAT and MYST), each containing a histone acetyl-transferase (HAT) enzyme, have been identified and often shown to be subunits of larger transcriptional co-activator complexes.

Over the last decade, two approaches were mainly used to better understand the specific mechanisms determining the functional specificity of HATs. First, *in vitro* acetylation assays were intensively conducted to investigate the substrate specificity of distinct HATs. These analyses showed that HATs exert a certain degree of specificity for particular lysine residues on different histone tails. Second, *in vivo* gene inactivation studies allowed testing the HAT specificity by observing phenotypical effects caused by ablation of a particular HAT. Interestingly, most of these studies argued for a high degree of specificity in the developmental or gene expression phenotypes. More recently, the recent appearance of new high throughput technologies (ChIP-seq) allowed the investigation of HATs recruitment and acetylation deposition at a genome wide scale (Wang, Zang et al. 2008; Wang, Zang et al.

2009). Contrary to previously discussed evidences, when analysed carefully these data suggest a low specificity in the recruitment and activity of HATs over the genome. In this manuscript, we will comparatively review the conclusions of the different approaches. Additionally we will discuss the significance of each lane of conclusions and try to reconcile these new genome-wide evidences with existing knowledge in a new model for the mode of action of HATs in transcriptional activation.

Biochemistry: each HAT has a specific histone acetylation spectra modulated by its macromolecular complex

The two aspects that were intensively investigated immediately after the discovery of HATs were (i) identification of the protein complexes in which the individual HAT enzyme is incorporated and (ii) determination of the acetylation spectra associated with each HAT in these complexes (both reviewed in Table 1 for mammalian cells). Several conclusions emerged from these studies. First, it was shown that most of the HAT enzymes are part of large multi-subunit complexes. For example, Tip60 is contained in the 18 subunits NuA4 complex (Sapountzi, Logan et al. 2006), GCN5 or PCAF are members of the 19 subunits SAGA complex and the 10 subunits ATAC complex. Second, several studies described that the association of the HAT enzymes with their partners in the corresponding complexes can modify the specificity of the HATs on histones (i.e. GCN5 was shown in *Drosophila* to have different targeting specificity in ATAC or SAGA (Guelman, Suganuma et al. 2006)). Third, each HAT has a defined spectrum of target lysines and these spectra are only partially overlapping between individual HAT complexes (see Table 1 for review in mammalian systems). Taken together, these observations suggest a high specificity of each HAT when incorporated in its corresponding macromolecular complex, suggesting that this integration is required for the full and specific activity *in vivo*. Moreover, these studies suggested that each

HAT complex should create a specific signature on its target loci since a distinguishable acetylation pattern is observed for each complex tested *in vitro*.

Genetics: HATs exert a high degree of functional specificity

Another aspect that was systematically addressed to better understand the function of HATs is the phenotypical analysis of the effect of distinct ablation of HATs in vivo (reviewed in Table 2). Genetic knock out (KO) studies of HATs during mouse development reveal a significant variability in the phenotypes of HAT ablations. For example, *Tip60*-/- embryos do not develop beyond blastocyste stage (die at embryonic day (ED) 3.5) (Cai, Jin et al. 2009), while *Gcn5*-/- embryos die around ED 10.5 (Xu, Edmondson et al. 2000). Moreover, the maintenance of embryonic stem cells (ESC) pluripotence can be used as a second phenotypical readout. Using RNAi knock down of Tip60, this enzyme was shown to be required for ESC pluripotency, while GCN5 or PCAF seems to be dispensable (Table 2) (Fazzio, Huff et al. 2008; Ding, Paszkowski-Rogacz et al. 2009). Together, these results suggest that each of the HAT has key roles at different stages of developmental processes that seem to be independent and distinct from other HATs.

A remarkable feature emerging from sequence alignment of vertebrate HATs is the presence of highly related paralogs (i.e. CBP/p300 or Gcn5/PCAF) whereas a single gene is present in other non-vertebrate species. This observation raised the question of the functional importance of closely related paralogs in vertebrates. Answers were partially obtained by crossing mice carrying individual homozygous or heterozygous KOs of the given paralogs and comparing double KO to that of the single KO. For example, while the single CBP or p300 heterozygous mutants exert no phenotype, the mice heterozygous for both HATs show similar phenotypes to the homozygous depletion of one or the other paralogs (Yao, Oh et al. 1998). This result is suggesting that there is a functional redundancy between these paralogs, but that the dose of expression of the two paralogs is crucial for the proper development of the animals.

Taken together, these studies suggest that amongst HATs a high degree of functional specificity exists, with the exception of closely related paralogs that rather reflect the importance of gene expression dosage than functional specificity.

Genome-wide mapping: HATs are co-occurring at high frequency creating hyper acetylated environment

In the past few years our understanding of genome regulation has tremendously progressed due to the introduction of chromatin immunoprecipitation (ChIP) combined with microarray (ChIP-chip) or with high throughput sequencing (ChIP-Seq). These experiments allow analysing the presence of a particular genomic feature at the genome- wide scale. Recently, several systematic ChIP-seq studies have been published, providing genome-wide mapping data for HATs and acetylation marks in resting human T cells (Wang, Zang et al. 2008; Wang, Zang et al. 2009).

This resource is an unique opportunity to analyse the specificity of HATs recruitment genome-wide. Then, we have extracted from these data sets the genome-wide distributions of five HATs (namely p300, CBP, MOF, PCAF, and Tip60) and 18 acetylation marks. We then systematically analysed the presence of each of the 18 acetylation marks at the genome-wide binding sites of the five distinct HATs. Interestingly, this analysis revealed that when comparing the genome-wide binding sites of a given HAT with that of the 18 different histone acetylation marks, no specific or unique acetylation pattern for any of the five HATs could be identified (Figure 1 A, B; shown for CBP). In each of these "HAT-Acetylation marks" comparisons, variations in acetylation levels between different categories of loci can be observed, but with the same patterns for all the residues tested (except for histone H2AK5ac, H3K14ac and H3K23ac for which the antibodies used in the ChIP may have been too weak for the analysis). These results suggest that either HATs are systematically co-occurring at

these sites or that the HATs have specific binding patterns, but no enzymatic specificity *in vivo*. To be able to distinguish between these two possibilities the binding specificity of a given HAT versus the four other HATs has been tested. Interestingly, it has been observed that each genomic locus bound by a given HAT is also bound by the four other HATs (Figure 2 A and B). It appears that the binding of each HAT highly correlate with the binding of all the other HATs tested (Figure 2A, Pearson correlation>0.6).

The detected co-occurrence of HATs and the consequent broad acetylation patterns observed, can suggest two different scenarios. First, since in ChIP not a single cell but cell populations are analysed, we cannot exclude that the recruitment HATs could occur in a stochastic manner. HATs may be interchangeable meaning that in a given cell one HAT is binding to a particular locus, but in another cell a distinct HAT is binding to the same locus. In this scenario, HATs may deposit specific histone acetylation marks, but the observed lack of histone acetylation mark specificity may come from the absence of mechanisms that would specifically recruit HAT to a particular loci. In this case, the acetylation specificity would not be an important feature for regulation, but only the acetylation *per se* would be requested for proper activation. In the second scenario, HATs would work collaboratively. Then, in all cells, all the studied HATs would be permanently recruited and released at each bound loci in a dynamic fashion. This scenario implies that many transcriptional co-activators are dynamically recruited all the time to a set of regulated loci to modify (acetylate) the given chromatin environment. Such a mechanism for co-activator action has already been suggested by Hager and Mistelli (Hager, McNally et al. 2009). Thus, gene regulation would mainly rely on the local abundance of the different co-activators, rather than on the recruitment of a specific co-activator and its corresponding specific pattern of histone modifications. In this case, there would be a collaborative effect of all the HATs on the deposition of histone acetylation marks at the regulated set of loci.

An integrated model depicting mode of action of HAT

Taken together, the evidences emerging from genomics, biochemical and genetic analyses of HAT action result in apparently contradictory conclusions. While both the biochemistry and the genetic data argue for a rather high degree of specificity in the functional spectrum of HATs (Figure 3A), the current genome-wide evidences suggest that HATs are in general redundantly recruited to the same set of loci (Figure 3B) suggesting low specificity in the action of HATs. To reconcile the different observations we will first confront each lane of evidence and then propose a new model that may explain the *in vivo* action of HAT complexes (Figure 3C).

The genome-wide models are not directly contradicting the biochemical evidences since the broad *in vivo* co-occurrence of HAT enzymes at many genomic loci would explain the observation of hyper acetylated loci *in vivo*. However, this model is challenged by the conclusions of the genetic studies. It is clear from the KO studies that HATs cannot compensate for each other. Moreover, if each process would require all the HATs, one would expect that all HAT KOs would result in the same phenotype since all HAT dependent processes should be affected. Thus a model assuming a complete overlap in HATs function can be excluded and new hypothesis have to be raised to integrate these new lanes of evidences.

The time is a variable that was not introduced in the previous formulation of hypotheses. The analysed ChIP-seq data were all produced at a *t* time that reflects a single stage of the transcription activation process. Since analysed genome-wide data were generated in resting somatic cells (CD4+ T cells) that are not challenged, one can assume that they exert only “routine” gene expression programs compared to the major transcriptional changes happening during differentiation of a tissue or an organism. We hypothesize that the genome-

wide data obtained in these cells reflect a “maintenance sage” in gene activation process (Figure 3C). In our novel model, we propose that HATs could have a dual mode of action by distinguishing “initiation” and a “maintenance” stage during the transcription activation process. The “initiation” would reflect the initial steps in the switch from an inactive to active transcription state while the “maintenance” would represent the stabilisation of an active transcription stage over the time. In the initiation stage, the transcriptional activation of a given gene would be highly dependent on the specific recruitment of a given HAT (Figure 3C left panel). Following this “initiation” stage, through a sequential process this initial activation would lead to the consequent recruitment or binding of multiple HATs, which may recognize the initially open and acetylated environment in a less specific manner. This hypothesis is in good agreement with the findings that many HAT-containing complexes contain subunits with bromodomains that are thought to bind to acetylated histone tails. The binding of several HATs to the initially acetylated locus would thus serve to maintain the acetylation level and by consequence the activation state of a given locus (Figure 3C right panel).

Our model would then explain why in dynamic developmental processes HATs show high functional specificity (genetic data) that may not be the case in resting differentiated somatic cells. During cell differentiation, the high complexity level of transcriptional regulatory networks is a prerequisite for orchestration of the time controlled initiation of the proper transcriptional programs. However, in resting somatic tissues, where the number of newly initiated transcription processes is likely to be lower than during cell commitment, this large panel of activators redundantly maintains activation sates. This could explain why high redundancy in recruitment of HATs is observed in resting cells, while their separate function is needed to achieve proper embryo development.

One can expect that in the future, generalization of the genome-wide mapping technology will allow new insights into the functional rule of HATs. For example, the time

course analysis of recruitment of HATs upon a dynamic activation of particular gene networks, like *in vitro* stem cells differentiation systems (Bibel, Richter et al. 2007), should allow to test the proposed model.

Conclusions

HAT containing complexes are key components of chromatin mediated transcriptional regulatory networks. Proper understanding and modelling of their mode of action and function within these networks is a prerequisite to accurate prediction of transcriptional systems behaviour. Recent technology breakthrough in the post-genomic area allowed new insights into HATs function but were often interpreted ignoring previously available biochemical and genetic data, leading to oversimplified models. Here we propose novel considerations that could reconcile the different lines of evidences in a unified model which describe the mode of action of HATs more reasonably. Our current model proposes that HATs are differently required depending on the stage of gene activation, with a high functional specificity in the early gene activation stage and non-specific functionality in later maintenance stage.

Figures:

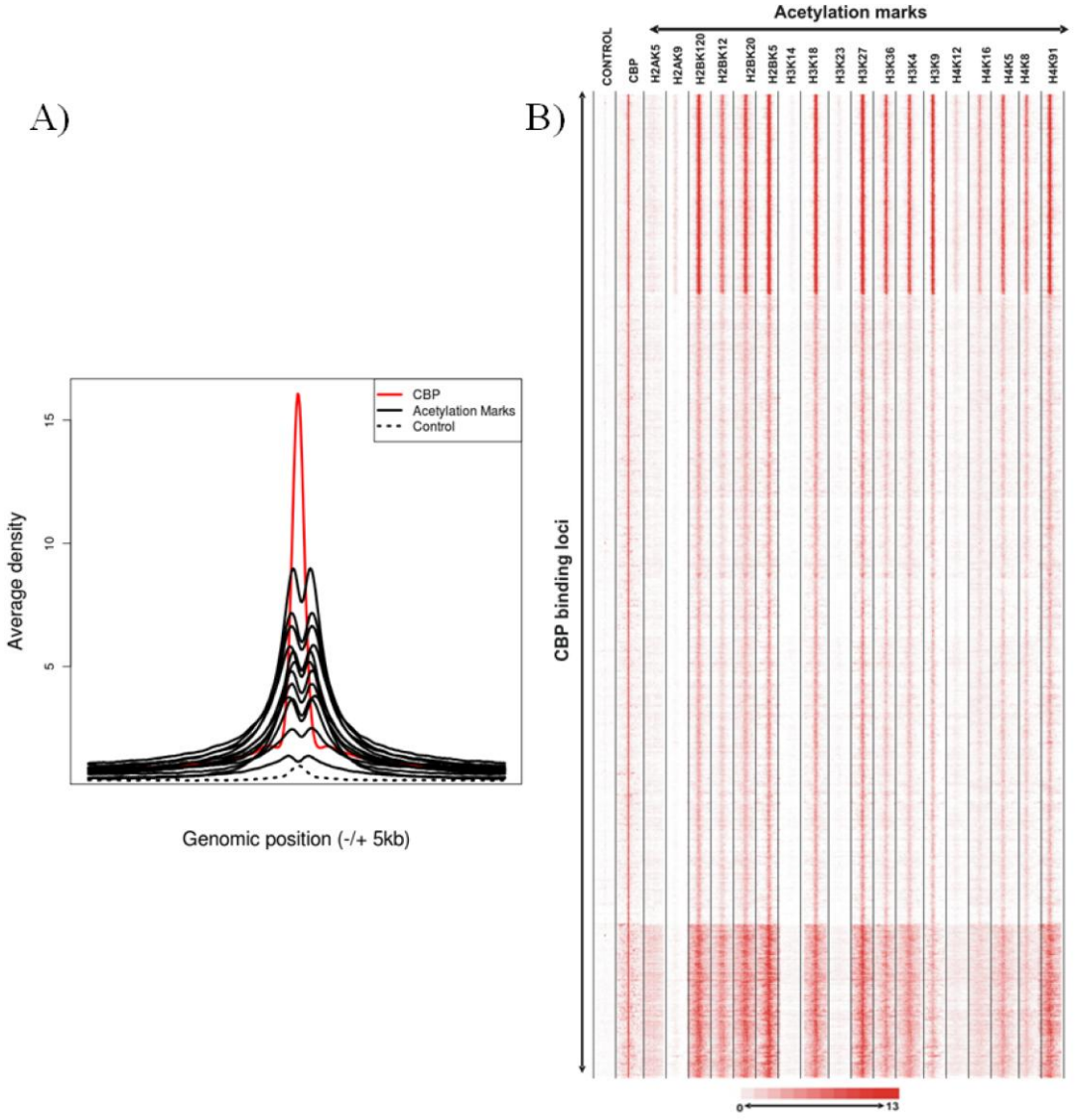


Figure 1: Genome-wide binding patterns of histone acetylations on binding sites of HATs.

Raw ChIP-seq data were extracted from (Wang, Zang et al. 2008; Wang, Zang et al. 2009). A) Co-occurrence of 15 acetylation marks on CBP binding sites: Average binding densities of 15 histone acetylation marks (marked in black), control [IgG] (marked in dashed black) and CBP (marked in red) surrounding -/+5kb region of a collection of 10360 CBP binding sites. From the raw datasets enrichment clusters representing CBP binding sites were determined. Around each CBP binding site four hundred 25 bp bins were created and densities were collected for each bin for the 15 acetylation tracks. The mean was calculated for each bin and used to represent average acetylation densities around the CBP binding sites. B) Co-occurrence of 18 acetylation marks on all the CBP binding sites: Binding densities of regions (-/+5kb) surrounding the 10360 binding sites of CBP. Densities are shown for control (IgG), CBP and 18 histone acetylation marks (see legend). In the heatmap each line represents a genomic location of a binding site with its surrounding -/+ 5kb region. CBP binding sites were used as reference to collect ChIP-seq densities over a 10kb (-/+ 5kb) window. This

matrix was subjected to k-means clustering. The heatmap representing the clustered density matrix is displayed.

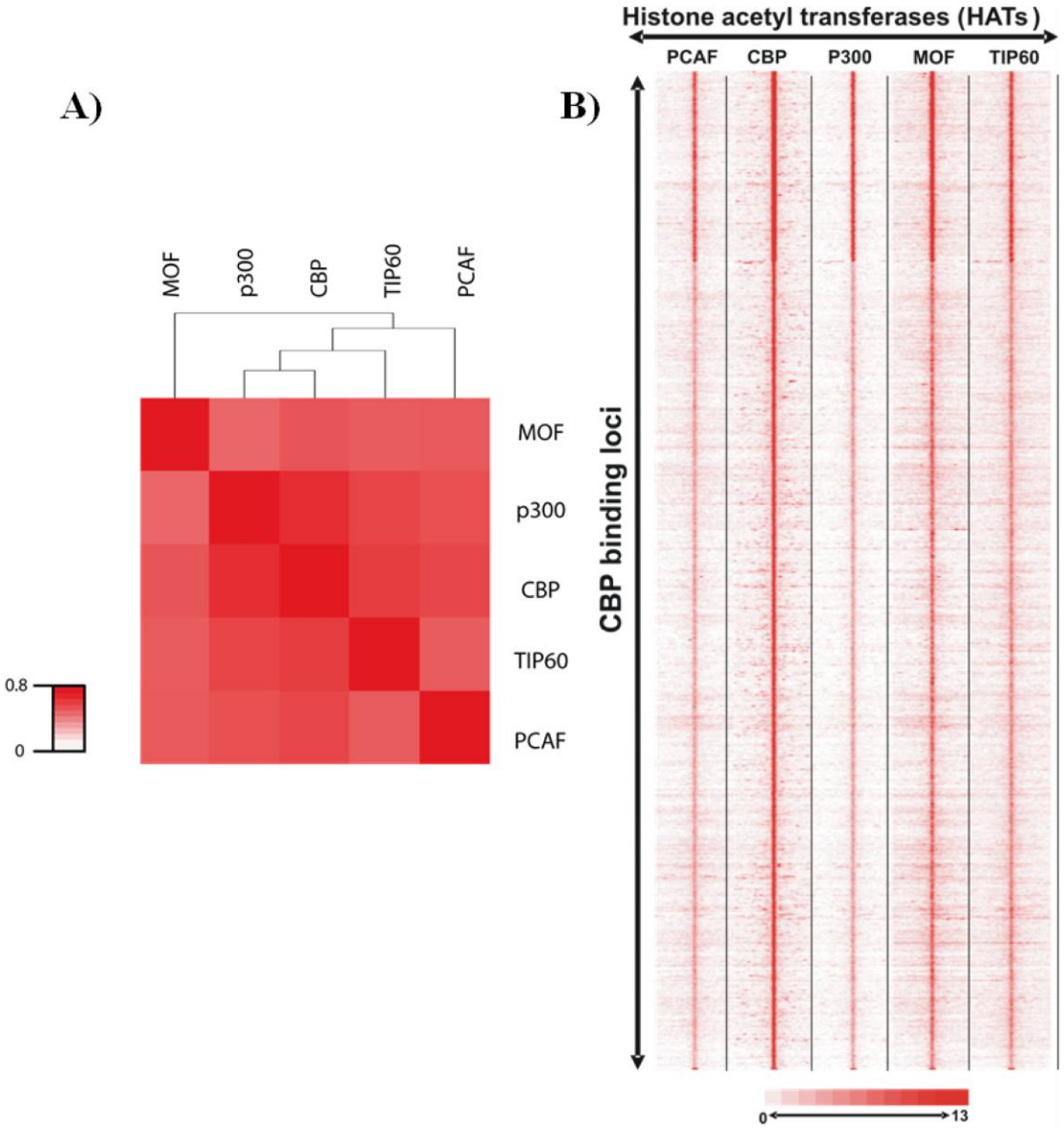


Figure 2: HATs are co-recruited at high frequency on their binding loci.
Raw ChIP-seq data were extracted from (Wang, Zang et al. 2008; Wang, Zang et al. 2009). A) Heatmap showing co-localization frequency of all the five HATs namely MOF, p300, CBP, TIP60 and PCAF. Colors in the heat map reflect the co-localization frequency (Pearson correlation coefficient) of each pair of HAT (Red means more frequently co-localized) over a 2kb region surrounding the complete set of HATs binding sites. HATs were clustered along both axes based on the similarity in their co-localization with other HATs. B) Co-occurrence of five HATs on CBP binding sites: Binding densities of PCAF, p300, MOF and TIP60 were clustered according to 10360 CBP binding sites. In the clustering each line represents a genomic location of a binding site with its surrounding -/+ 5kb region. HAT binding sites were used as reference to collect ChIP-seq densities over a 10kb (-/+ 5kb) window in each HAT density maps. This matrix was subjected to k-means clustering. The heatmap representing the clustered density matrix is displayed.

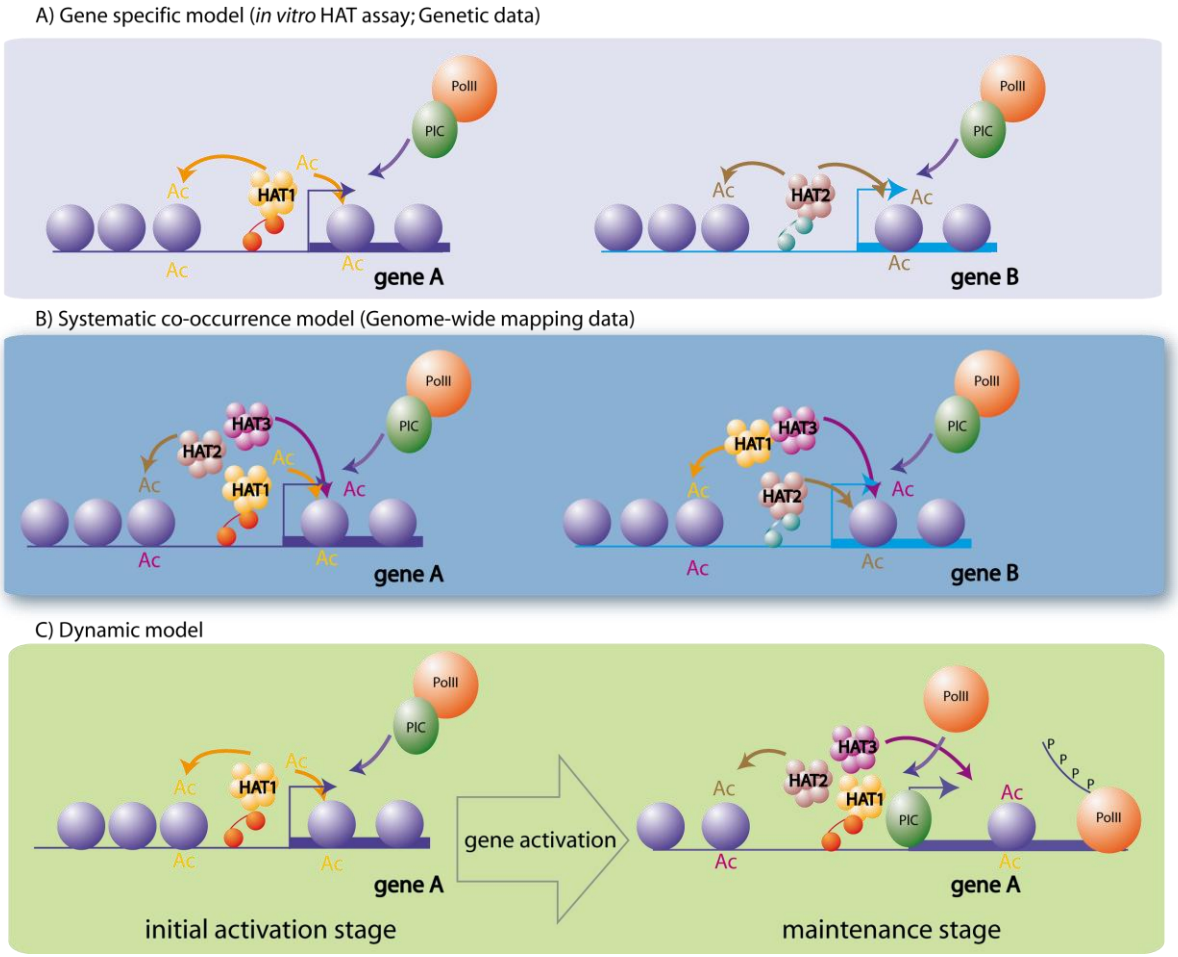


Figure 3: Schematic models describing the possible modes of action of HATs.
A) Gene specificity model. This model can be proposed based on the conclusions from biochemical and genetic studies. Each HAT complex is recruited by a particular DNA binding transcriptional activator to a defined set of genes, allowing their activation. Thus HATs exerts a high functional specificity. B) Systematic co-occurrence model. This model arises from conclusions of genome wide mapping studies. HATs would be co-recruited to all the transcriptionally active loci, creating a hyper-acetylated environment that would favor their activation. C) Dynamic model. Model proposed to reconcile the observations of the genome-wide mapping of HATs with previous biochemical and genetic observations. HATs would play a dual role in the gene activation process. In the first phase a specific HAT, recruited by a specific activator, allow the initiation of the activation process. Later, other HATs can bind the activated loci in a less specific manner thus maintaining a non-specific hyper acetylated environment.

		Histone Lysine	H3				H4				H2A		H2B				references
			9	14	18	23	5	8	12	16	5	9	5	12	15	20	
	molecular complex	HAT Familly															
Gcn5- KAT2A/ PCAF - KAT2B	SAGA	GNAT															(Brand, Yamamoto et al. 1999a; Nagy, Riss et al. 2009b),
	ATAC																(Wang, Faiola et al. 2008; Guelman, Kozuka et al. 2009; Nagy, Riss et al. 2009b)
	free																(Brand, Yamamoto et al. 1999a; Nagy, Riss et al. 2009b)
MOF - KAT8	MSL	MYST															(Cai, Jin et al. 2009)
	NSL																(Cai, Jin et al. 2009)
	free		ND	ND	ND	ND					ND	ND	ND	ND	ND	ND	(Cai, Jin et al. 2009)
TIP60 - KAT5	NuA4	MYST															(Altaf, Auger et al.; Allard, Utley et al. 1999; Ikura, Ogryzko et al. 2000) 2009
	free																(Yamamoto and Horikoshi 1997; Allard, Utley et al. 1999)
p300 - KAT3B	p300/CBP	p300/CBP	ND			ND				ND		ND				ND	(Schiltz, Mizzen et al. 1999)
CBP - KAT3A	p300/CBP	p300/CBP	ND			ND				ND		ND				ND	(Schiltz, Mizzen et al. 1999)

Table 1: Substrate specificity of HATs in vitro.
Review of the *in vitro* substrate specificity described in mammalian systems, for the studied five HAT enzymes, either as a free protein or within their respective macromolecular complexes. Dark grey boxes represent lysine residues highly acetylated; light grey boxes represent residues where weak acetylation activity has been observed and white boxes represent residues where no acetylation is detected. ND represents residues that have not been tested for acetylation.

			Phenotypic readout		
	molecular complex	**HAT family**	**mouse knock out**	**mESC pluripotency**	**references**
Gcn5-KAT2A	**SAGA/ATAC**	**GNAT**	ED 10.5	-	(Xu, Edmondson et al. 2000; Ding, Paszkowski-Rogacz et al. 2009)
PCAF - KAT2B			viable	-	(Xu, Edmondson et al. 2000; Fazzio, Huff et al. 2008)
MOF - KAT8	**MSL/NSL**	**MYST**	ED 3,5	+	(Fazzio, Huff et al. 2008; Thomas, Dixon et al. 2008)
TIP60 - KAT5	**NuA4**		ED 3,5	+	(Fazzio, Huff et al. 2008; Hu, Fisher et al. 2009)
CBP - KAT3A	**p300/CBP**	**p300/CBP**	ED 11.5	ND	(Yao, Oh et al. 1998)
p300 - KAT3B	**p300/CBP**		ED 11.5	-	(Yao, Oh et al. 1998; Fazzio, Huff et al. 2008)

Table 2: Phenotypes observed in gene disruption studies of HATs:
Phenotypes associated with HATs genetic knock out (KO) or RNAi targeting in mouse development and ES cells pluripotency respectively. For mouse KO, the day of embryonic death (ED) is presented as a phenotype readout parameter. For the effect on mouse embryonic stem cell (mESC) pluripotency, any observation (flattering of the cells, differentiation…) was simplified as a positive phenotype (+). ND represents HATs for which the parameter was not determined.

IV. GENOMIC POSITIONS ANNOTATION TOOL (GPAT)

(Krebs A *et al*, 2008)

BMC Bioinformatics

Bio Med Central

Software

Open Access

GPAT: Retrieval of genomic annotation from large genomic position datasets

Arnaud Krebs*, Mattia Frontini and Làszlò Tora*

Address: Department of Functional Genomics, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR 7104, INSERM U 596, Université Louis Pasteur de Strasbourg, BP 10142-67404 ILLKIRCH Cedex, CU de Strasbourg, France

Email: Arnaud Krebs* - krebs@igbmc.u-strasbg.fr; Mattia Frontini - frontini@igbmc.u-strasbg.fr; Làszlò Tora* - laszlo@igbmc.u-strasbg.fr

* Corresponding authors

Published: 15 December 2008

BMC Bioinformatics 2008, 9:533 doi:10.1186/1471-2105-9-533

Received: 16 September 2008
Accepted: 15 December 2008

This article is available from: http://www.biomedcentral.com/1471-2105/9/533

Abstract

Background: Recent genome wide transcription factor binding site or chromatin modification mapping analysis techniques, such as chromatin immunoprecipitation (ChIP) linked to DNA microarray analysis (ChIP on chip) or ChIP coupled to high throughput sequencing (ChIP-seq), generate tremendous amounts of genomic location data in the form of one-dimensional series of signals. After pre-analysis of these data (signal preclearing, relevant binding site detection), biologists need to search for the biological relevance of the detected genomic positions representing transcription regulation or chromatin modification events.

Results: To address this problem, we have developed a Genomic Position Annotation Tool (GPAT) with a simple web interface that allows the rapid and systematic labelling of thousands of genomic positions with several types of annotations. GPAT automatically extracts gene annotation information around the submitted positions from different public databases (Refseq or ENSEMBL). In addition, GPAT provides access to the expression status of the corresponding genes from either existing transcriptomic databases or from user-generated expression data sets. Furthermore, GPAT allows the localisation of the genomic coordinates relative to the chromosome bands and the well characterised ENCODE regions. We successfully used GPAT to analyse ChIP on chip data and to identify genes functionally regulated by the TATA binding protein (TBP).

Conclusion: GPAT provides a quick, convenient and flexible way to annotate large sets of genomic positions obtained after pre-analysis of ChIP-chip, ChIP-seq or other high throughput sequencing-based techniques. Through the different annotation data displayed, GPAT facilitates the interpretation of genome wide datasets for molecular biologists.

Background

One of the major issues in genomics is the genome wide mapping of transcription factor binding sites in order to study their function at the scale of the genome. The chromatin immunoprecipitation (ChIP) technique uses antibodies that are specific for a transcription factor or a chromatin modification, to isolate the DNA to which this factor or modified histone is bound in a cell at a given time. The recent appearance of several genome wide analysis techniques, where ChIP is either followed by DNA microarray analysis (ChIP on chip) or coupled to high throughput sequencing (ChIP-seq), made the genome wide mapping of DNA bound factors technically possible. However, these analyses generate tremendous amounts of

genomic location data in the form of one-dimensional series of signals.

Recently, efforts have been made to develop academic software to pre-analyse these datasets, (e.g. Mpeak [1]), in order to locate the signal peaks that correspond to functional elements, such as promoters, enhancers, repressors or insulators. However, once these datasets are cleaned and the significant signals are selected, biologists lack user-friendly tools to search for the biological relevance of the resulting binding site genomic positions.

Implementation

The web interface of GPAT is programmed using Python (v2.5.1) [2] running on an Apache [3] WWW server and forms an interactive layer between the user and the underlying processing applications. In order to increase the speed of data recovery, all the data are stored in a local PostgreSQL (v8.1.11) database [4]. The background processes are programmed in Python and take advantage of the PygreSQL [5] module to efficiently connect to the PostgreSQL database. The class diagram describing the GPAT object oriented python code is displayed in Additional file 1: GPAT class diagram.

Results

Comparison with existing tools

Existing on-line tools, such as the UCSC genome browser [6], allow the user to display so-called "custom tracks" and to browse locally defined genomic positions, thus facilitating the manual retrieval of biological information over hundreds of annotation tracks. However, the browser does not allow the batch processing of large numbers of positions. Complementary to this approach, studies based on genome wide analyses require a tool to systematically annotate data in a form suitable for further statistical analysis. To address this problem, several applications have been released recently.

Galaxy [7] is a framework giving access to popular sources of data, such as the UCSC Table Browser [6] or Biomart at ENSEMBL [8], using a variety of integrated tools. Although very powerful for certain applications, it remains too general for our purposes. In particular, the "Fetch the closest feature" module in the "Operate on genomic intervals" section that allows the annotation of genomic positions with gene identifiers has some limitations compared to a more specialized tool, such as GPAT (see below). As an example, the "Fetch the closest feature" module searches for the closest feature without distance limitations, and often matches very distant genes with questionable biological relevance. In addition, the output is limited to a plain text file corresponding to the concatenation of the input files with restricted information content (e.g. the distance to the TSS is not calculated and no hyperlinks to gene annotations are provided). Moreover, the direct cross-linking of the identified gene list with transcriptomic datasets is not directly possible.

CEAS [9] is a ChIP on chip analysis pipeline available via a web server, which includes a basic genomic position annotation function. One of the major limitations of CEAS is that the annotation search is completely automated and the user cannot adjust the default options. For example, the search window around the submitted position is fixed at 300 bp. This means that the software is not suitable for studies involving factors, which are known to bind enhancer regions located outside of the proximal promoter regions. Furthermore, since the full analysis is done at each dataset submission, the analysis time rapidly increases for large datasets (20 minutes for 1000 positions).

Cisgenome [10] is a powerful high throughput ChIP analysis package proposing a gene annotation function. However, the user has to perform a full installation process and must store the annotation files on a local computer (several gigabytes) in order to annotate his results. Furthermore, Cisgenome does not provide direct access to other annotations and experimental data (e.g. transcriptomic data), which is a necessary step in the extraction of biological meaning from high throughput data.

GPAT description

GPAT allows users to analyse large batches of genomic positions and to retrieve genomic annotations around these positions. Briefly, the user submits a flat file containing the genomic positions and then selects the annotation search options and the display options (Figure 1A). The annotated results can then be browsed or downloaded for subsequent analyses.

The input format is based on the GFF standard file format [11]. Mouse and human data produced by the two latest UCSC genome assembly releases (mm 8, mm 9, hg 17, hg 18) are currently supported.

For each submitted genomic position, GPAT calculates a mean value from the two boundary positions and compares this value to the gene annotation positions (Refseq release 31, ENSEMBL release 50). GPAT has three different gene annotation search modes corresponding to different biological questions (Figure 1B).

1. The "direct search" mode determines whether the submitted positions are detected inside a transcription unit. It was developed to allow searches for a binding site inside the transcription units (e.g. retrovirus insertion events).

2. In the "window search" mode, a user-defined window is calculated around the submitted positions. Then GPAT searches gene annotations located within this window. It

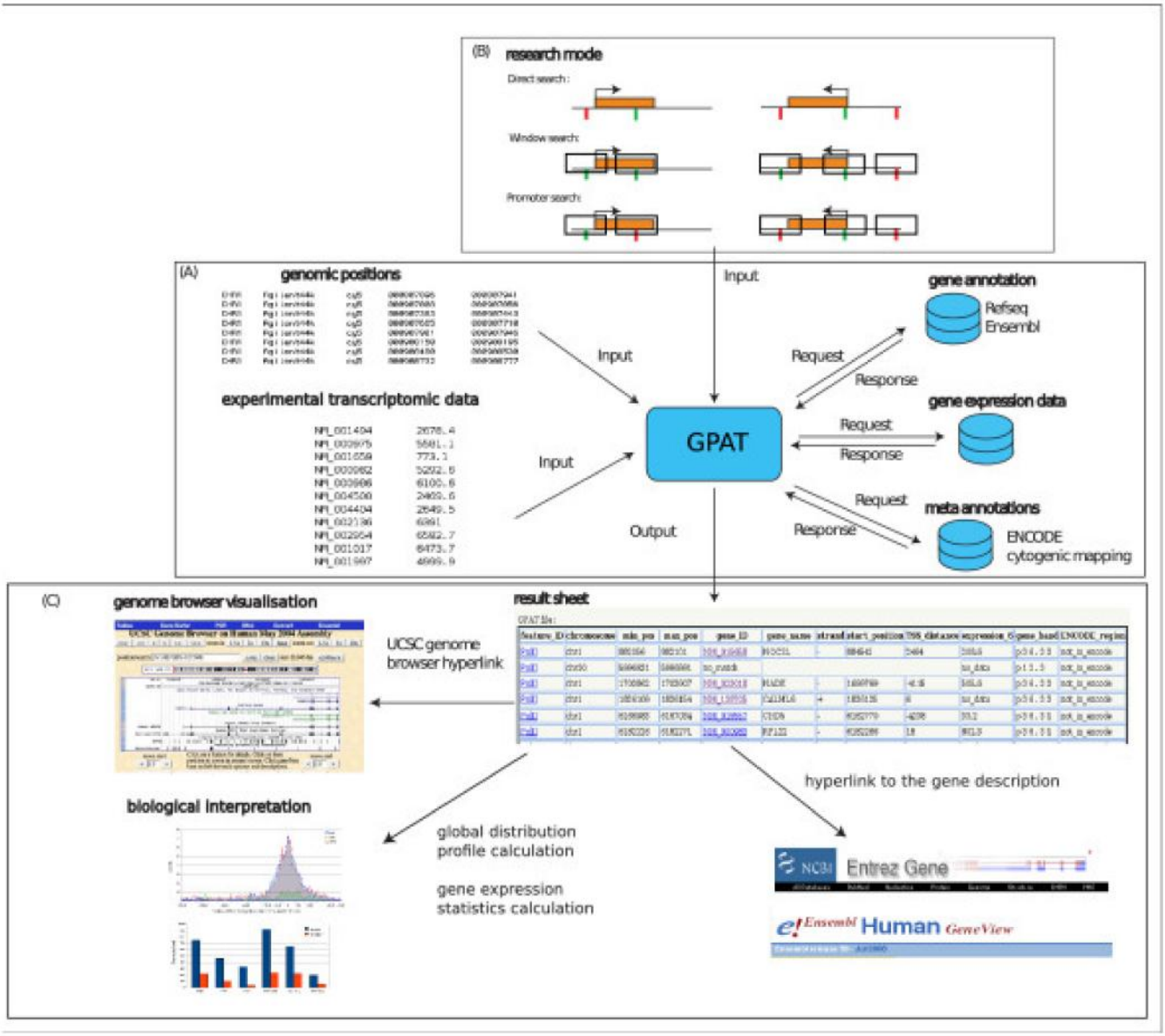


Figure 1
GPAT application flow chart : (A) Information flow of an annotation search in GPAT. (B) The three gene annotation search modes implemented in GPAT. The panel represents two transcription units oriented in opposite directions (orange boxes). The transcription start site (TSS) is symbolised by an arrow. User submitted positions are represented by vertical bars and the search window by open boxes. The colour of the vertical bar symbolizes the result of the GPAT search (green: annotation matched, red: not matched). The "direct search" mode searches the positions located inside a transcription unit. The "window search" mode allows the detection of transcription units located within a defined distance from the genomic positions. The "promoter search" mode allows the identification of transcription units having their TSS within a defined distance from the genomic positions. (C) Results table containing the annotated positions; links to UCSC genome browser and gene source information; global distribution profile of the matched genomic positions as compared to the TSSs of the corresponding genes and statistical values for the expression data of the corresponding genes (represented using a spreadsheet application).

was developed for datasets where no particular binding profile is expected or known (e.g. insulator elements or proteins of unknown function).

3. The "promoter search" mode uses the same window as the "window search" mode, but tests whether the transcription start site (TSS) of a transcription unit is found within this window. It was developed for datasets, where a binding in the neighbourhood of the TSS is expected (e.g. transcription factors).

Regardless of the search mode chosen by the user, GPAT provides a complete localisation report relative to the matched gene annotation (position of the TSS of the corresponding gene, distance of the detected location to the TSS). Furthermore, when multiple gene matches are found, the user can choose to retrieve either the closest annotation only or all the matched annotations.

Biologists are often interested in the expression status of the genes neighbouring the submitted genomic locations. Therefore, the GPAT software also gives access to the corresponding gene expression levels in several commonly used model cell lines as provided by the GNF Symatlas [12] and Stembase [13] (nine human and two mouse cell lines including mouse embryonic stem cells). Furthermore, users can upload their own transcriptomic data during the analysis process, thus allowing the retrieval of gene expression levels for datasets generated under various experimental conditions.

Several other genomic features, such as those extracted from the cytogenetic mapping [14] or the position relative to the ENCODE regions [15] can also be retrieved.

GPAT displays the results as a table containing the annotated positions hyperlinked to the data source (Figure 1C). A direct hyperlink to the UCSC genome browser [6] is provided to allow the user to browse other genomic features in the case of a successful match. In addition, the full set of results can be downloaded as a tabulated flat file and easily imported in any spreadsheet software for further analysis.

Finally, an additional analysis step allows the user to create a report file, containing summary information, in addition to the matched gene list. This option calculates the global distribution profile of the matched genomic positions as compared to the TSSs of the corresponding genes using parameters set by the user (see example in Figure 1C). Statistical values for the expression data of the corresponding genes are also calculated (Figure 1C).

Use of GPAT – Example:
The Tata Binding Protein (TBP) is a component of a number of complexes, including the TFIID complex involved in the RNA Polymerase II (Pol II) general transcription machinery. Surprisingly, it has been shown that Pol II transcription can occur in the absence of TBP [16,17]. However, the molecular mechanism leading to TBP-free transcription is still poorly understood. One of the strategies for studying this mechanism is to isolate target genes where Pol II transcription can be detected in the absence of TBP.

We tested GPAT using a promoter DNA tiling array dataset generated by hybridization of DNA (prepared from Hela cells), which was ChIPed using specific antibodies against Pol II, TBP and GST (as a negative control). For each experiment, three slides (Agilent G4483A – 013863, 013864 and 01387) representing approximately one-third of the most characterised promoters of the gene were hybridized (about 5600 genes covered from -8 kb upstream to +2.5 kb downstream with a ~350 bp resolution). After intra-array lowess normalization, peaks corresponding to the factor binding sites were detected using Agilent ChIP analytics (using the neighbourhood model).

A flat file containing the genomic positions generated by this pre-analysis was input to GPAT. The promoter search mode using a half window size of 5000 bp successfully retrieved the genes in the neighbourhood of the factor binding sites in less than a minute. This allowed us to build a list of genes potentially regulated by TBP and to obtain information about the presence of the Pol II at these sites (Figure 2A). Furthermore, using the distance to the 5' end of the matched gene transcripts provided in the GPAT output, the global distribution of the binding sites of TBP and Pol II relative to the 5' end could be computed (figure 2B). This shows, as expected, that the majority of the TBP and Pol II (but not GST) binding sites are located within +/- 1 kb of the TSS. This result is in agreement with observations made at the single gene scale and demonstrates the accuracy of GPAT results.

Finally, the output of GPAT allowed us to attribute an expression level for each of the differently regulated gene categories showing that the genes for which a co-occupancy of TBP and Pol II was detected have the strongest expression levels (Figure 2C). This correlation gives a high level of confidence in the list of binding sites established by our analysis. Furthermore, by combining these data, we obtained a list of expressed genes, for which Pol II, but not TBP is present (Figure 2A). These genes can be used as a model for further analysis of the mechanism of TBP-free transcription initiation. Finally, the group of genes for which TBP, but not Pol II is bound at the promoter shows low expression levels, implying that there are genes where TBP is bound, but Pol II, is not or is poorly recruited.

In conclusion, these results demonstrate the efficiency of GPAT for the extraction of biological meaning from large genomic position datasets.

Discussion

The current version of GPAT improves the bioinformatics analysis of ChIP on chip or ChIP-seq data. It considerably speeds up and facilitates the steps between data pre-analysis and the biological interpretation of the data. Notably, the amplitude in the stringency of the annotation search provided by the three different research modes should meet the requirements of most of the biological questions addressed. Furthermore, the full localisation report (including distance to the TSS) will help the biologist to easily understand the binding pattern of his studied factors. Finally, the connection of the binding sites and the gene expression data is an essential resource at different steps of the analysis. Firstly, it provides a supplementary filtering step to distinguish between relevant and background signals. Secondly, it adds another layer of complexity to the data interpretation, providing insight into the regulation processes taking place at particular genomic locations.

Currently, the investigation process of large scale ChIP data is complicated and involves multiple analysis steps. In order to make this technology available for laboratories without access to bioinformatics expertise, considerable efforts are needed to facilitate the data analysis (similar to the efforts dedicated to classical expression arrays). To

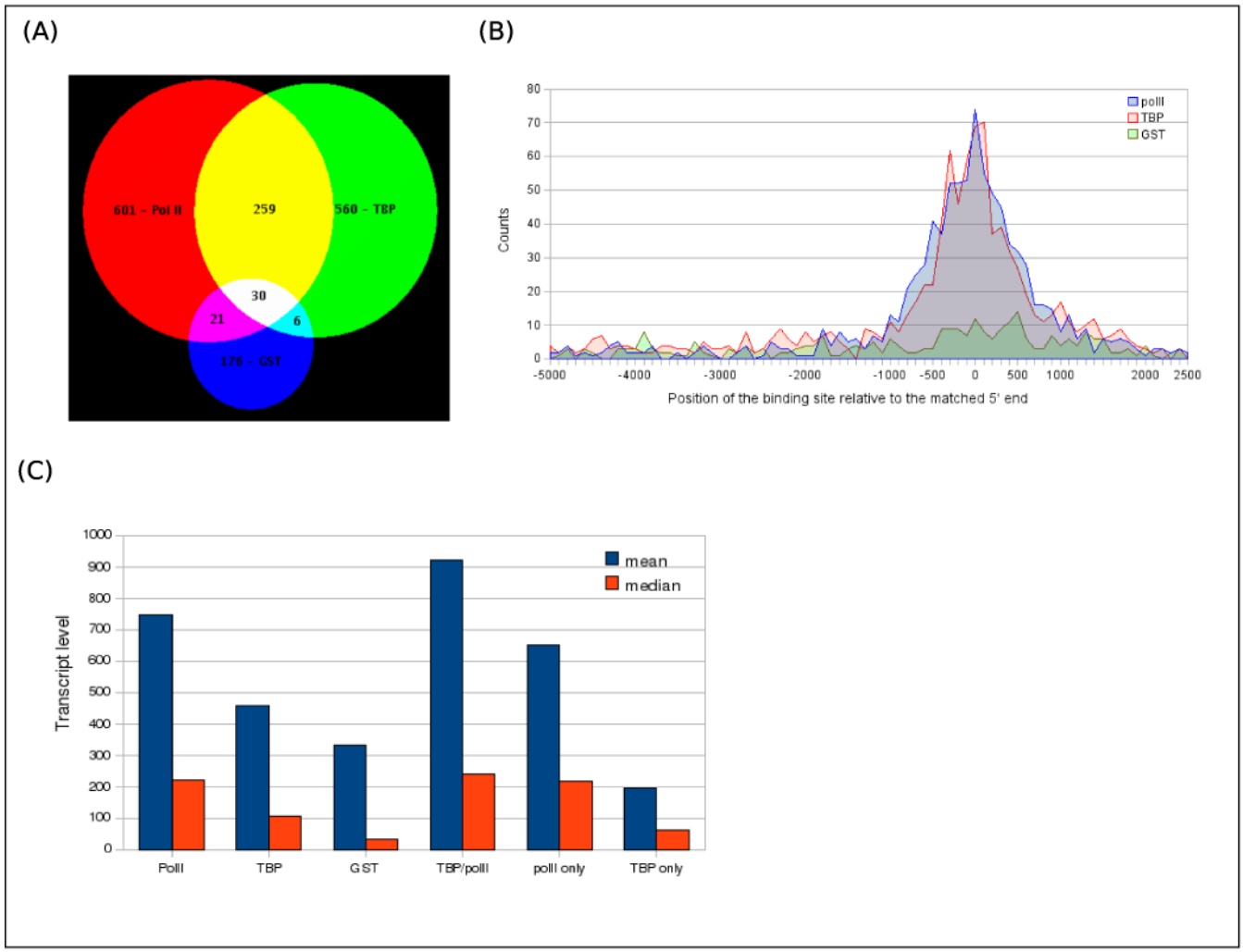


Figure 2

Example of exploitation of the GPAT results : (A) Venn diagram showing the genes with a single occupancy by Pol II (red) or TBP (green) respectively or a co-occupancy (yellow). (B) Distribution of Pol II (blue) and TBP (red) binding sites relative to the 5' end of the matched transcript. The distribution patterns of both Pol II and TBP, but not GST, cluster within +/- 1 kb around the 5' end of the matched transcripts. (C) Distribution of the expression level in each gene category. The highest expression level is observed for genes where both Pol II and TBP were detected at the promoter. Furthermore, genes bound only by Pol II, but not TBP show a high level of expression, suggesting the possibility of TBP independent genes.

address this issue, GPAT was designed to improve the steps following data pre-analysis (annotation, cross-linking with other experimental data). However, several improvements remain to be implemented. GPAT currently supports the mouse and human genomes, but adding other model organisms for which genome wide ChIP data are available (e.g. Drosophila) would make the impact of GPAT wider and could be easily implemented in the future.

Furthermore, GPAT provides expression data for several cell lines. One of the recent improvements of the ChIP method allows the use of tissues as a starting material. Since the GNF dataset was also generated for mouse tissues, these datasets could be easily integrated into the GPAT interface.

Conclusion

GPAT provides a quick, convenient and flexible way to annotate large sets of genomic positions obtained after pre-analysis of ChIP on chip, ChIP-seq or other high throughput sequencing based techniques. Thanks to the different annotation and experimental data provided (including the expression status of the identified genes), GPAT facilitates the interpretation of genome wide datasets. We hope that GPAT will be of great help to molecular biologists who wish to analyse large-scale genomic data.

Availability and requirements

Project home page http://bips.u-strasbg.fr/GPAT/Gpat_home.html

Source code

The GPAT source code can be freely downloaded from: http://bips.u-strasbg.fr/GPAT/Gpat_source_code.html

Operating systems

For use: Standard WWW browser

Programming language

Python, SQL, Javascript

Licence

GNU General Public License v3 (GPL 3)

Abbreviations

ChIP chip: Chromatin Immunoprecipitation followed by hybridation on DNA microarray; ChIP-seq: Chromatin Immunoprecipitation coupled to high throughput Sequencing; ChIP: Chromatin Immunoprecipitation; GPAT: Genomic Position Annotation Tool; RNA Pol II: RNA polymerase II; TBP: TATA Binding Protein; TSS: Transcription Start Site

Authors' contributions

AK conceived and implemented GPAT. MF generated the ChIP on chip data. AK did the bioinformatics analysis of the ChIP on chip data. AK and LT wrote the manuscript.

Additional material

Additional file 1

GPAT class diagram . Class diagram describing the python object oriented architecture of the GPAT software.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-9-533-S1.png]

Acknowledgements

This work was funded by the integrated European Transcriptome, Regulome & Cellular Commitment Consorsium (EUTRACC) project. AK is recipient of a fellowship from the INSERM and the Alsace region.

The authors thank the organisers and teachers of the "Informatics in biology" course of the Pasteur Institute (Paris, France) where GPAT was designed and implemented; M. Vanhoutteghaert and the IGBMC bioinformatics platform (particularly F. Plewniak) for technical support; C. Lethondal, C. Saveanu for critical reading of the manuscript; J. Thompson for advices and helpful comments on the manuscript.

References

1. Mpeak software [http://www.stat.ucla.edu/~zmdl/mpeak]
2. Python (v2.5.1) [http://www.python.org]
3. Apache [http://www.apache.org]
4. PostgreSQL [http://www.postgresql.org]
5. PygreSQL [http://www.druid.net/pygresql]
6. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D: The human genome browser at UCSC. *Genome Res* 2002, 12: 996-1006.
7. Galaxy [http://main.g2.bx.psu.edu/]
8. Biomart [http://www.biomart.org]
9. Ji X, Li W, Song J, Wei L, Liu XS: CEAS: cis-regulatory element annotation system. *Nucleic Acids Res* 2006, 34: W551-554.
10. Ji H, Jiang H, Ma W, Johnson DS, Myers RM, Wong WH: An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nat Biotechnol* 2008, 26: 1293-1300.
11. GFF Format [http://www.sanger.ac.uk/Software/formats/GFF/GFF_Spec.shtml]
12. Su AI, Wiltshire T, Batalov S, Lappe H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR, Hogenesch JB: A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA* 2004, 101: 6062-6067.
13. Perez-Iratxeta C, Palidwor G, Porter CJ, Sanchenia NA, Huska MR, Suomela BP, Muro EM, Krzyzanowski PM, Hughes E, Campbell PA, Rudnicki MA, Andrade MA: Study of stem cell function using microarray experiments. *FEBS Lett* 2005, 579: 1795-1801.
14. Furey TS, Haussler D: Integration of the cytogenetic map with the draft human genome sequence. *Hum Mol Genet* 2003, 12: 1037-1044.
15. ENCODE Project [http://www.genome.gov/10005107]
16. Martianov I, Viville S, Davidson I: RNA polymerase II transcription in murine cells lacking the TATA binding protein. *Science* 2002, 298: 1036-1039.
17. Muller F, Tora L: The multicoloured world of promoter recognition complexes. *Embo J* 2004, 23: 2-8.

V. seqMINER: An integrated ChIP-seq data interpretation platform (Ye T, Krebs A *et al* submitted)

seqMINER: An integrated ChIP-seq data interpretation platform (Ye T, Krebs A *et al* submitted)

Tao Ye*[1], Arnaud R. Krebs*[§1], Amin M. Choukrallah[1], Celine Keime[2], Frederic Plewniak[3], Irwin Davidson[1] and Laszlo Tora[§1]

Address: [1]Department of Functional Genomics, [2] Microarray and Sequencing Platform, [3] Bioinformatics Platform. Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR 7104, INSERM U 596, Université de Strasbourg, BP 10142-67404 ILLKIRCH Cedex, CU de Strasbourg, France.

[§]Corresponding authors

*These authors contributed equally to the work

Email addresses:

laszlo@igbmc.fr

krebs@igbmc.fr

ABSTRACT

In a single experiment, Chromatin Immunoprecipitation combined with high throughput sequencing (ChIP-seq) provides genomewide information about a given covalent histone modification or transcription factor occupancy. However, time efficient bioinformatics resources for extracting biological meaning out of these gigabyte-scale datasets are often a limiting factor for data interpretation by biologists. We created seqMINER, an integrated portable ChIP-seq data interpretation platform with optimized performances for efficient handling of multiple genomewide datasets. seqMINER allows comparison and integration of multiple ChIP-seq datasets and extraction of qualitative as well as quantitative information. seqMINER can handle the biological complexity of most experimental situations and proposes supervised methods to the user in data categorization according to the analysed features. In addition, through multiple graphical representations, seqMINER allows visualisation and modelling of general as well as specific patterns in a given dataset. Moreover, seqMINER proposes a module to quantitatively analyse correlations and differences between datasets. To demonstrate the efficiency of seqMINER, we have carried out a comprehensive analysis of genome wide chromatin modification data in mouse embryonic stem cells to understand the global epigenetic landscape and its change through cellular differentiation.

[seqMINER source code, documentation and wiki are available at http://bips.u-strasbg.fr/seqminer/]

INTRODUCTION

Chromatin immunoprecipitation (ChIP) allows the quantitative measurement of protein (i.e. transcription factor) occupancy or the presence of post-translational epigenetic histone modifications at defined genomic loci. Recent technological developments combining ChIP with direct high throughput sequencing of the immunoprecipated DNA fragments (ChIP-seq) allowed the analysis of transcription factor genomic occupancy or epigenetic chromatin marks at the genome wide (GW) scale in an relatively unbiased manner. Over the last years, numerous studies have used ChIP-seq as a central method to create GW binding maps for a particular genomic feature (Mikkelsen, Ku et al. 2007; Ku, Koche et al. 2008), installing ChIP-seq as the gold standard in the functional genomics toolbox.

Each ChIP-seq run generates data at the gigabyte scale that requires successive steps of bioinformatics treatment prior to biological interpretation (reviewed in (Johnson, Mortazavi et al. 2007)). In a standard analysis pipeline, two major steps can currently be distinguished. First, genomic locations presenting relevant enrichment in the ChIP-seq signal are identified and annotated with respect to known genomic sequence features (genes, transcripts, repeat elements etc). Second, by performing multiple rounds of analyses using various methods, the biological meaning of the dataset has to be extracted and often compared with other datasets. Many methods and softwares have been released over the past months in order to easily perform the initial analysis stage with good accuracy (reviewed in (Johnson, Mortazavi et al. 2007; Laajala, Raghav et al. 2009)). However, unlike the first step that distinguish relevant signal from noise to provide information on a given factor or chromatin mark, the second analysis stage requires the combined use of various methodologies to answer multiple complex biological questions. Thus, development of integrated complementary approaches is a prerequisite to make ChIP-seq analysis as easy and routine as possible.

Several initiatives devoted to particular biological questions have already contributed to enrich the ChIP-seq analysis toolbox. For example, numerous tools, integrated or not in larger analysis pipelines, have been proposed that allow: annotation of genomic features present in the neighbourhood of the relevant enrichment signals (Ji, Jiang et al. 2008; Krebs, Frontini et al. 2008; Shin, Liu et al. 2009), detection and *de novo* definition of consensus DNA motifs (Ji, Jiang et al. 2008), cross and compare information from distinct datasets (Blankenberg, Taylor et al. 2007), and comprehensive visualization of the obtained GW

results (Kent, Sugnet et al. 2002; Manske and Kwiatkowski 2009; Skinner, Uzilov et al. 2009). More recently, sophisticated statistical approaches were proposed to predict association between sets of genomic locations and numerous genomic features (Hon, Ren et al. 2008; Bock, Halachev et al. 2009). Nevertheless, due to the multiplicity of biological questions that may be asked by the ChIP-seq method, many analysis issues remain un-addressed.

In particular, the definition of the genomic rules governing the function of a particular factor is a complex, but central question. The most common method for addressing this question is to investigate the frequency of co-occurrence of the analysed factor/mark with other genomic features by (i) extracting relevant signals (peaks) in the two datasets to be compared and (ii) calculating the number of events that are close enough between the datasets to be considered as overlapping. This approach, that has been previously used (Blahnik, Dou et al.; Blankenberg, Taylor et al. 2007), presents several limitations. First, the analysis is biased towards the peak-searching step and the empirical cut-off values used to discriminate between relevant and background signals. The second drawback comes from comparisons of datasets containing either sharp (i.e. a transcription factor) or non-discrete binding sites (i.e. spread histone marks like H3K27me3), where peak detection is more difficult. Third, an increasing number of studies show that particular factors have more than a single function in the genome (i.e.(Bilodeau, Kagey et al. 2009)), challenging the direct interpretation of one to one comparisons. Finally, it becomes clear that definition of genomic regulatory elements cannot rely on a single feature, but need integration of multiple sources of information to be properly identified (Wang, Zang et al. 2008; Heintzman, Hon et al. 2009). Moreover, tools able to analyse multiple information sources are required to allow comprehensive qualitative and quantitative ChIP-seq analysis.

To overcome the above described limitations, we have developed seqMINER, an integrated user friendly platform that addresses central questions in the ChIP-seq analysis workflow. seqMINER was designed in order to make it as easy as possible for the biologist to carry out in depth interpretation of the analysed datasets to answer their biological question. The purpose of seqMINER is to allow qualitative and quantitative comparisons between a reference set of genomic positions and multiple ChIP-seq datasets. Different analysis modules have been implemented to allow users to search for full characteristics of a particular feature genomewide. Starting from a set of reference coordinates that can be a list of ChIP-seq

enrichment clusters (peaks) for a particular target (i.e. a transcription factor), seqMINER proposes two complementary methods to analyse the signal enrichment status in multiple other tracks: i) a qualitative method that computes a density array over a defined window around the reference coordinate; ii) a quantitative method that computes enrichment value over a defined widow around the reference coordinate. Following these steps, automated as well as manual reordering methods of the analysed loci are implemented to assist users in defining functional subgroups in their data. Finally graphical representations of the data are proposed through heatmaps and dotplots to illustrate particular, as well as general properties of the data. We show that, the different resources of seqMINER taken together allow a comprehensive analysis of genome-wide chromatin modification data to understand the global epigenetic landscape and its change through cellular differentiation.

RESULTS

Correlative integration of multiple datasets.

Most genomic studies aim to define the function of a particular regulatory factor by understanding globally how it affects other co-occurring events in a regulatory circuit (i.e. chromatin modifications, binding of other factors) and the consequences on outputs of the regulated system (i.e gene expression). Thus we designed seqMINER to allow the integration of multiple ChIP-seq datasets in a quick and user friendly manner.

seqMINER uses a list of BED formatted genomic coordinates (i.e. binding sites of a particular factor, set of genes, set of promoters etc.) as a reference for investigating information in other genomic datasets. Three stages can be distinguished in the analysis process (Figure 1). First, in the data collection module, seqMINER collects the read density from a reference dataset over a user-defined window around a set of coordinates and then calculates the read density in the same window in one or multiple other data sets. Second, in the clustering module, seqMINER, uses a clustering procedure (k-means) to organize the identified loci presenting similar read densities within the specified window. Third, in the visualization module, seqMINER allows at glance visualization of the entire output dataset through various graphical representations.

seqMINER proposes two related complementary tag (read) density based methods to analyse the signal enrichment status in multiple tracks (Figure 2). The first is a qualitative method that defines general patterns and functional sub-groups in the dataset: densities over a

window around the reference coordinates can be calculated at different resolutions in multiple tracks (Figure 2A). The created matrix can be organized by supervised k-means clustering to isolate groups of loci having similar features as developed by (Heintzman, Stuart et al. 2007; Heintzman, Hon et al. 2009). At this stage, clusters can optionally be reorganized manually according to the biological significance. Visualisation of the whole dataset is in this case achieved through heatmaps. This method allows easy visualisation of signal distribution over multiple loci and identifies general patterns over the dataset, which can be plotted as average profiles. Information on signal distribution, that can be an important biological feature, is conserved (i.e. broad, sharp enrichment peaks). However, visualisation of quantitative phenomenon is more complex. The second method allows extraction of quantitative information. Raw tag counts as well as normalised enrichment over a control track can be calculated (Figure 2B). The created matrix allows easy plotting of quantitative information and interpretation for one to one comparisons. Moreover, the method produces numerical data that are necessary to integrate sequencing data in larger mathematical models of a particular system.

All visual features as well as the list of loci can be exported for further analysis with other methodologies (i.e. gene annotation (Krebs, Frontini et al. 2008), ontology (Dennis, Sherman et al. 2003)). Moreover, output data can be used in new clustering rounds using different sets of data or analysis parameters. This possibility facilitates multiple iterative steps of analysis inherent to the genomic data analysis.

Performance optimisation

seqMINER takes advantage of the portability of the Java Virtual Machine (JVM) allowing an installation free multiplatform usage. seqMINER was designed to optimise all time limiting steps in the manipulation of the large raw sequencing datasets. For all the computational tasks, the implementation combines an optimal use of the Random Access Memory (RAM), avoiding repetitive access to the hard drive, and Simultaneous Multi Threading (SMT) to fully use available computational resources and lower the time of analysis. SMT permits the program to execute multiple independent threads to better utilize the resources provided by modern processor architectures. Particular attention was paid to the design of the genomic data storing objects to optimize space occupancy and efficiency of data

retrieval algorithms. seqMINER performances were assessed by training the analysis on three standard size ChIP-seq datasets (from 191-554MB) using as an input reference set different subsets of the ENSEMBL transcript database (33881 input coordinates) (Figure 3). All software trainings were performed on a PC running under windows with standard performances (CPU-3.0GHz core-duo; RAM – 4Go). For all the tested sets, the limiting step appears to be the loading of the dataset in the memory (Figure 3), which is usually performed only once and followed by multiple cycles of analysis. Nevertheless, the time required for a complete analysis by seqMINER for all tested datasets, regardless of the number of reference coordinates used, and including data loading is less than 30 seconds (Figure 3).

To bring the analysis time by seqMINER to acceptable levels we had to optimize the clustering process for the density arrays. The density array based method generates a large array of values (i.e. for a 10kb window at a 25bp resolution, an array of 400 values is created per dataset analysed) for each reference position. The use of existing implementation (de Hoon, Imoto et al. 2004) of clustering algorithms to organize these large datasets appeared to considerably slow down the analysis workflow (over an hour for the 8470 reference sites and two datasets). Thus, we decided to create a novel implementation of this algorithm using recent technology developments in programming resources. We implemented the k-means algorithm, taking advantage of SMT technology and using Java Machine Learning libraries (Abeel, de Peer et al. 2009) for the data structure implementation. This new algorithm dramatically improves the clustering speed compared to existing implementations (Figure 3) since less than 10 seconds are needed for clustering all the tested datasets, while more than an hour was needed for the smallest dataset tested with the former algorithm. Thus this new implementation used in seqMINER brings the efficiency of clustering routines to the requirements of ChIP-seq data analysis.

These combined efforts allow the analysis of multiple raw sequencing datasets in a few seconds, considerably speeding up the ChIP-seq analysis workflow and facilitating the testing of multiple biological assumptions in a time efficient manner.

Genome wide analysis of the chromatin landscape of genes in embryonic stem cells

Comparative analysis of epigenetic profiles in several cell types allows a global view and better understanding of chromatin dynamics and its role in gene regulation. In the last few years, numerous genomic studies focussed on embryonic stem cells (ESC) that can both self-renew indefinitely or differentiate in cell types that from the three primary germ layers. These interesting properties were broadly studied and were shown to be highly dependent on transcriptional and epigenetic regulatory networks (Bernstein, Mikkelsen et al. 2006; Boyer, Plath et al. 2006; Lee, Jenner et al. 2006). In order to train seqMINER and demonstrate its efficiency in addressing complex biological questions, we decided to (i) make a comprehensive description of the chromatin states at all annotated promoters of mouse ESCs, and (ii) quantitatively compare the active chromatin marking in ESCs with that observed in the differentiated brain tissue.

First, we used seqMINER to characterize the epigenetic profile of mouse genes in ESCs using three representative histone modifications known to mark active or inactive genes. The active histone marks H3K4me3 and H3K36me3, are enriched in the promoter regions and transcribed gene bodies, respectively. In contrast, the repressive mark, H3K27me3, is known to give a broader distribution over genes. As reference coordinates, we used transcription start sites (TSS; assumed to be at the 5' end of annotated genes) of the 33881 mouse genes referenced in ENSEMBL (v58) database. Tags densities from each dataset were collected in a window of 10 kb around the reference coordinates and the collected values were subjected to k-means clustering.

Out of this initial clustering, five groups of loci could be distinguished (Figure 4A) amongst which we can identify three clusters of active genes transcribed on the negative strand, on the positive strand and in both directions, as determined by the H3K36me3 mark (Figure 4A middle panel). Additionally, two clusters of inactive genes could be identified, one marked by both H3K4me3 and H3K27me3 marks (known as bivalent promoters) and one showing no significant enrichment of the three studied marks. Interestingly, while the bivalent loci are known to correspond to transiently repressed genes required for later differentiation of ESCs, the unmarked loci likely represent strongly repressed genes that could harbor constitutive heterochromatin marks. Thus, from heterogeneous starting datasets and based only on three chromatin features, we could identify consistent gene categories for which

average profiles can be automatically drawn (Figure 4B), illustrating general regulatory features of these loci.

Second, we aimed to test seqMINER in a study to compare quantitative changes in the histone H3K4me3 mark in ESCs and brain tissue. In order to compare these two datasets, we generated a list of reference coordinates enriched in H3K4me3 in ESCs. We used both the quantitative and the qualitative methods proposed in seqMINER to compare the signal in the two selected datasets over the reference loci. First, we used the qualitative method to organize groups in the dataset (Figure 5A). Two major clusters could be identified, the first cluster contains loci equally enriched in H3K4me3 mark in ES and brain cells, the second cluster contains loci with much higher H3K4me3 signal in ESCs relative to brain (cluster 2 in Figure 5A). This analysis suggests that cluster 2 comprises only ES specific genes. However, by performing an additional round of clustering on cluster 2, three sub-clusters were identified (2.1, 2.2 and 2.3), illustrating the importance of the possibility given by seqMINER to perform iterative rounds of analysis to organize the data with precision. Analysis of the sub-clusters after this second clustering step indicates that the seemingly ES specific loci can now be reclassified as weakly, moderately or strongly enriched in the brain (Figure 5B). The results now suggest for example that cluster 2.1 is also transcribed in the brain cells analyzed.

Second, in order to quantitatively determine the differences between ESC and brain, we collected enrichments over the input loci (H3K4me3 bound regions) using the quantitative method implemented in seqMINER (Figure 5B). Then using the visualization module of seqMINER, we plotted enrichment in ESC versus brain. Using this method, we again observe a heterogeneous pattern when using the total set of loci (Figure 5C) with a large proportion showing similar enrichments in ESC and brain tissue (Pearson=0.71), but also a subset of loci that are enriched in ESCs, but not in brain. Interestingly, when we collect the same data on subsets isolated with the qualitative method (subset 2 and 2.3) and overlay this information on the plot (Figure 5D, subset 2 in blue and 2.3 in green), we can confirm that these subsets correspond to the loci that change their status between the two conditions. Additional information is gained by this method since we observe that most of the loci that are differentially enriched in H3K4me3 between the two cell types display low enrichment in ESC. Consequently, loci that are highly enriched in the H3K4me3 mark are invariant in both cells types and probably correspond to housekeeping genes (Figure 5C).

These analyses showed how the quantitative and qualitative methods implemented in seqMINER can be combined to select populations of loci having similar features and quantitatively follow their behavior in different conditions. Importantly, we show how clusters can be used as reference coordinates for iterative rounds of analysis to detect potential sub-populations. Finally through some biological examples, we illustrate the possibilities given by seqMINER to create visual representations of the isolated genomic features.

DISCUSSION

Extracting the biological meaning from genomewide studies requires the use of various methodologies to answer complex biological questions. With seqMINER, we attempted to develop a standalone analysis platform that allows users to make biological interpretation of their high throughput sequencing data. We implemented two methods that allow high level correlative integration of multiple sequencing datasets to identify general as well as specific genomic characteristics that emerge from the analysed features. We provided a number of visualisation methods, in order to provide the possibility for rapid assessment of the datasets from different analysis perspectives (i.e. general patters, sub-groups, quantitative data comparisons). Thus, we provide a novel broad bioinformatics resource that should provide analysis solutions in many ChIP-seq based biological applications. Additionally, we significantly improved the efficiency of broadly used clustering algorithms by re-implementing them using recent technology developments (i.e. SMT). seqMINER should prove to be a valuable resource for a broad range of applications in the bioinformatics community.

By using Java Virtual Machine (JVM) for code interpretation, we aimed to avoid both operating system (OS) incompatibility and to limit user efforts for installation making seqMINER accessible to a broad range of users. However, on most OS, by default, only a small RAM memory space is attributed to the application, limiting the number of genomic features that can be analysed simultaneously. In order to overcome this limitation, the user has to manually set the desired RAM space dedicated for seqMINER analysis. Another implementation choice regarding memory usage was to load all the necessary data in the RAM prior to analysis to avoid repetitive hard disk access and speed up the computational tasks. This can be a limitation for seqMINER usage on local computers that are commonly equipped with 2-4 GB RAM and a 32bit operation system that limits the maximum RAM of

Java virtual machine at about 1.5GB. Thus, to perform multi-file comparison (~10< sequencing raw files), the use of seqMINER on a server is recommended.

A current limitation in ChIP-seq data inter-comparison comes from technical issues in the ChIP experiment itself. It appears to be relatively difficult to obtain quantitatively comparable data due to high variation in immunoprecipitation efficiency that is affected by a multitude of imponderable factors (antibody specificity and efficiency, natural abundance of the particular ChIPed feature etc). In its present version, seqMINER generally assumes that all datasets present a similar enrichment ranges. Particularly, for the clustering procedure, seqMINER equally weighs the density arrays of each data set. However, if a data set presents significantly higher enrichment compared to the others, it will likely over-contribute to the clustering and eventually lead to sub-optimal locus organisation in the final cluster.

A simple way to overcome this problem is to apply normalisation filters upstream of the seqMINER analysis to normalise the signal between datasets. Several methods have already been proposed for normalising high throughput sequencing datasets prior to comparison (i.e. (Rozowsky, Euskirchen et al. 2009; Taslim, Wu et al. 2009; Welboren, van Driel et al. 2009)). However, no real consensus has appeared in the field to date on whether methods that are suitable for a particular experimental design will be widely applicable. Thus, it is difficult to propose a robust and broadly accepted normalisation method to circumvent this problem. Another approach could be to interfere directly at the clustering stage. One can imagine developing a user supervised method allowing imposition of different weights to the clusters to correct for experimental variations between data sets. Methods for correcting experimental bias will likely be developed in the future that will facilitate seqMINER usage.

In the present version of seqMINER, we aimed to address numerous biological questions commonly raised in classical ChIP-seq analysis pipelines. As users may have specific questions they wish to ask and as the arrival of future technologies extends the scope of the questions that can be asked, we designed seqMINER in a flexible and open mode. We implemented seqMINER in an organized architecture (following Model-View-Controller guidelines) and making the source code available open source (GPL3). For example, support for new rich data formats (ie: SAM-BAM (Li, Handsaker et al. 2009)) or dedicated methods for integrating ChIP-seq with RNA-seq will be the next developments needed to improve seqMINER functionality.

METHODS

seqMINER algorithms

seqMINER uses two methods to quantify ChIP-seq signal depending on the type of analysis performed (Figure 2). For both methods, prior to analysis, all reads were extended from a user defined size (default 200bp). For the calculation of densities over a defined window, methods were derived from the one generally used to generate density files (i.e. (Zhang, Liu et al. 2008) except that tag extension was performed only on the direction of the tag (not in both strand orientations). Then a user defined number of bins are created around the reference coordinate and for each bin the maximal number of overlapping tags is computed (Figure 2).

To calculate a single quantitative value for a binding site, tag density was defined as the number of tags present or overlapping in user defined window (default 2kb) around the reference site (Figure 2). ChIP-seq enrichments (e) were defined as e = log2 ((foreground tags+q)/(background tags+q)) where q is an empirical constant in the range of 10 that normalises the enrichment value; foreground tags are the density value computed in the data track ; background tags, the density value in the control track. Use of the constant q reduces the influence of the signal variation in the noise measurement of the background sample on the ratio calculation.

Data source and treatment

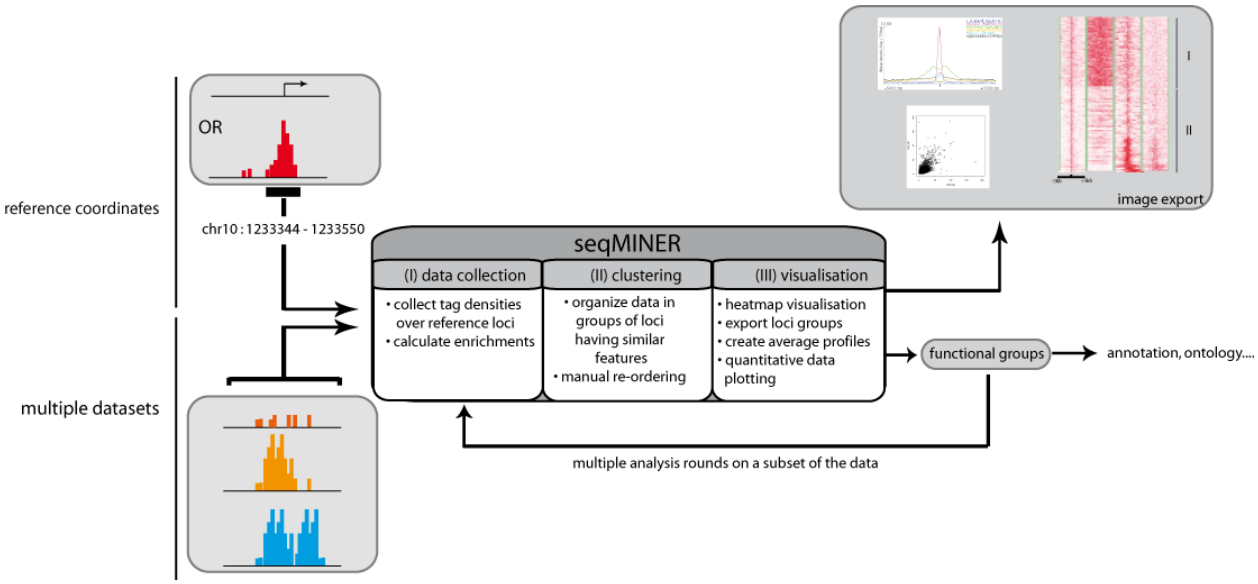
ChIP-seq datasets were downloaded from the public data bank Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/gds) under the accession number: GSM307618 (H3K4me3-ESC); GSM281696 (H3K4me3-brain); GSM307625 (input-ES); GSM307620 (H3K36me3-ES) and GSM307619 (H3K27me3-ES). Reference coordinates list was established using MACS (Wang, Zang et al. 2008).

Code repository

The source code, tutorial and wiki for seqMINER are available at http://bips.u-strasbg.fr/seqminer/ under General Public License (GPL3).

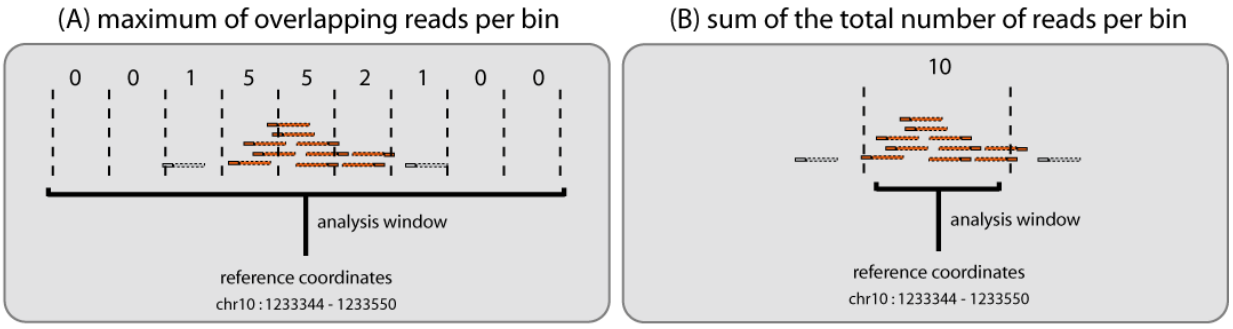
AKNOWLEGMENTS

We are grateful to Stadler M for technical suggestions, De Dieuleveult M, Legras S, Strub T for useful comments on the software. ARK is recipient of a fellowship from INSERM-Région Alsace and the Association pour la Recherche sur le Cancer (ARC). AMC is a recipient of a fellowship from the French Ministry of Education, Research and Technology (MNERT) and ARC. This work was funded by grants from ANR (GenomATAC; ANR-09-BLAN-0266), the EU (EUTRACC, LSHG-CT-2007-037445), the Institut National du Cancer, the Ligue National contre le cancer. ID is an 'équipe labélisée' of the Ligue National contre le cancer.



Ye T, et al Figure 1

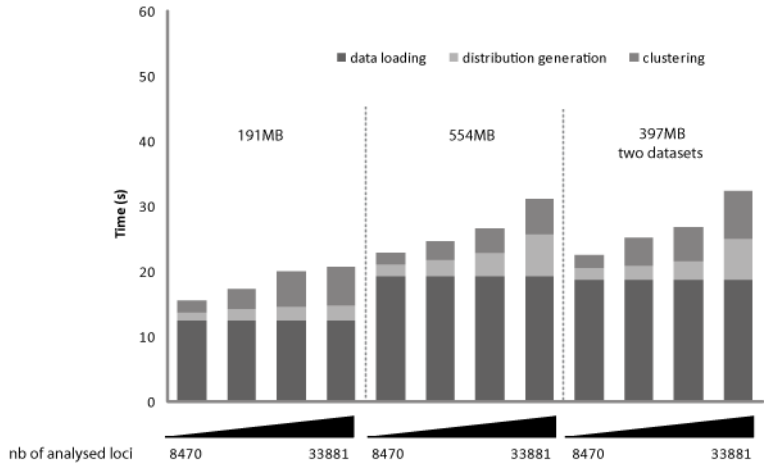
Figure 1. Schematic representation of the general workflow of seqMINER.
seqMINER takes as an input a single set of reference loci (i.e. gene promoters or binding sites) and multiple raw sequencing datasets. seqMINER can collect tag densities or calculate enrichment values around the set of reference coordinates. Using a combination of automated clustering and manual reordering methods, seqMINER helps the user to create functional groups within the reference set. Each specific sub-group in the dataset can be visualised as heatmaps, dotplots or average plots and the groups of loci can be exported for further analysis.



Ye T, et al Figure 2

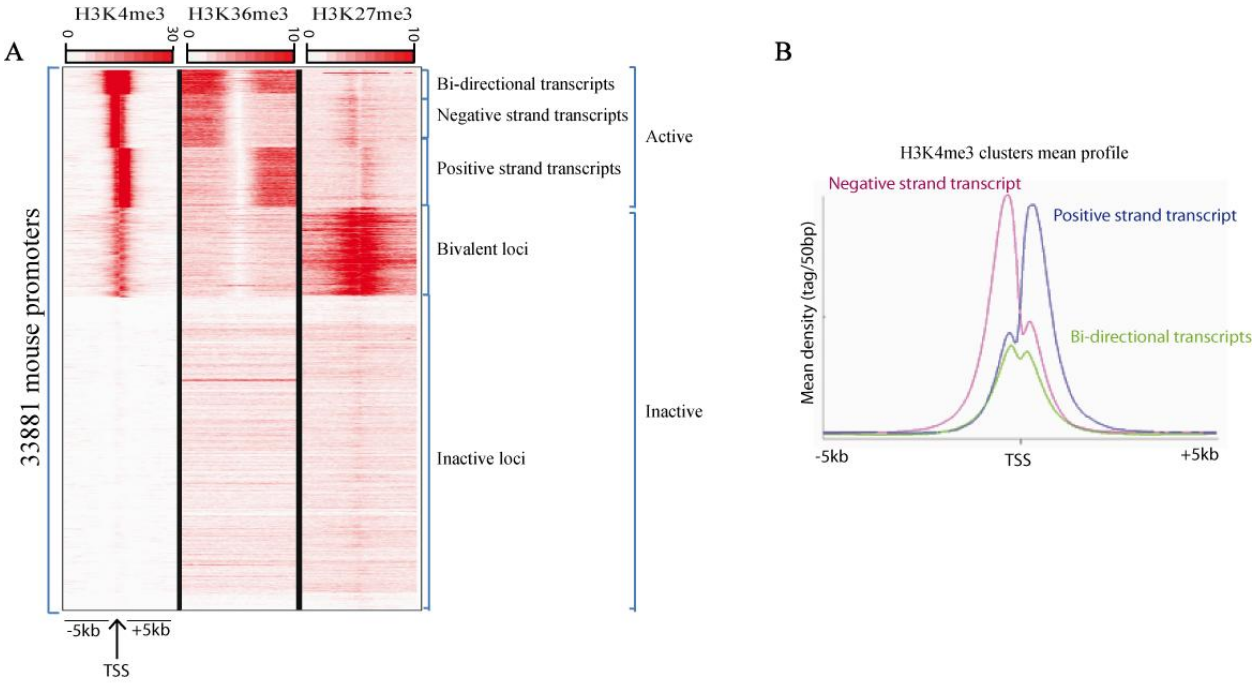
Figure 2. Schematic representation of data collection methods implemented in seqMINER.
In both methods, prior to quantification, reads are extended from a user defined value (default 200bp). (A) Maximal number of overlapping reads method: a user defined number of bins are created around the reference coordinate and for each bin the maximal number of overlapping

tags is computed. (B) Method summing the total number of tags: the number of tags presents or overlapping a user-defined window (default 2kb) around the reference site are counted.



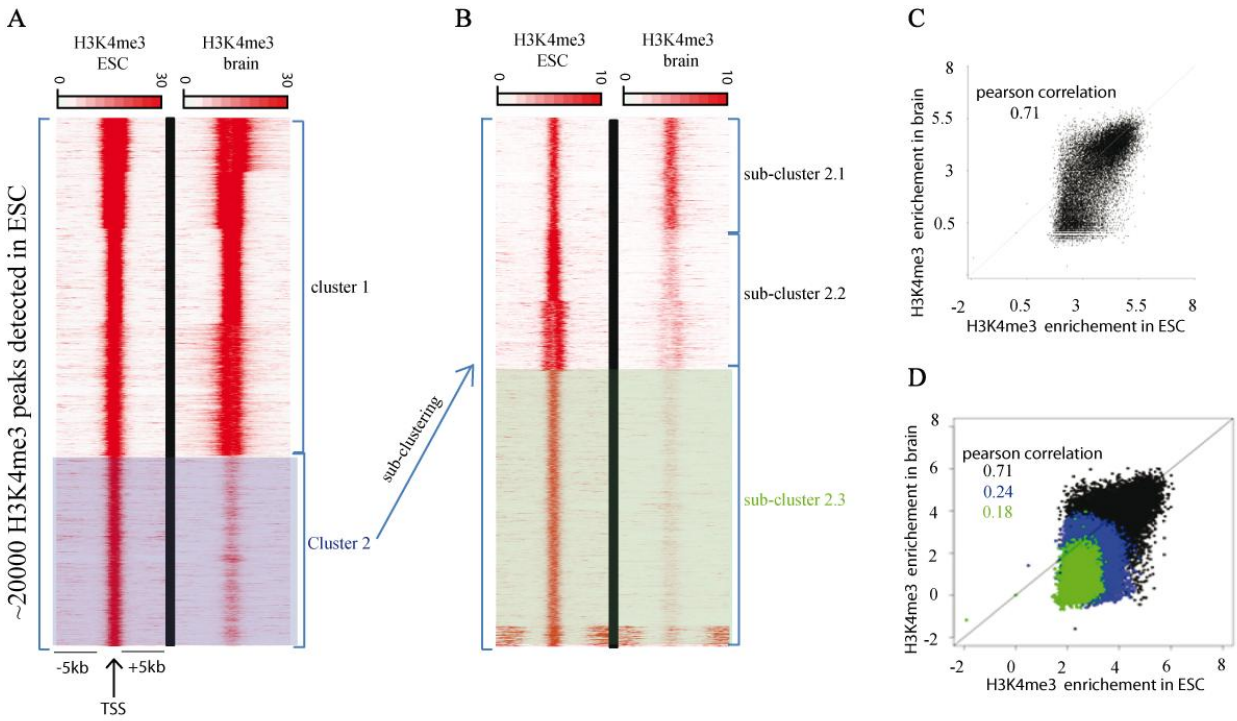
Ye T, et al Figure 3

Figure 3: Time to complete a typical analysis with seqMINER.
Time required for the different stages of the analysis (namely data loading, generation of distributions and clustering by seqMINER, using raw ChIP-seq datasets of various sizes (191, 554 and 397 MBs) and various number of reference coordinates (8470-33881). The analysis was performed on a PC running under windows with standard performances (CPU-3.0 GHz core-duo; RAM – 4Go).



Ye T, et al Figure 4

Figure 4: Characterisation of epigenetic profiles of mouse promoters in ESCs using seqMINER.
(A) Read densities of regions surrounding the whole set of TSS (assumed to be the 5'-end of the annotated transcript) of mouse genes from ENSEMBL (v58). TSSs were used as reference coordinates to collect data in publicly available H3K4me3 , H3K36me3 and H3K27me3 datasets. Tag densities from each ChIP-seq dataset were collected within a window of 10 kb around the reference coordinates, the collected data were subjected to k-means clustering. The five major clusters are indicated. (B) Using seqMINER, the average profile for selected clusters was automatically calculated and plotted. The H3K4me3 mean profile for transcripts actively transcribed on the negative strand (pink), positive (blue) strand and on both strands (green) was calculated and represented.



Ye T, et al Figure 5

Figure 5: Quantitative changes of H3K4me3 mark in mouse brain cells relative to mESCs
Tag densities of regions surrounding the H3K4me3 enriched loci in ESCs. Publicly available ChIP-seq datasets for H3K4me3 in ESCs and in brain cells were used in this comparative analysis. (A) H3K4me3 enriched loci in ESC were detected using MACS software, these loci were used as reference coordinates. Tag densities from H3K4me3-ESC and H3K4me3-brain datasets were collected within a window of 10 kb around the reference coordinates, and then the density files were subjected to k-means clustering. Two major clusters can be isolated: cluster 1 contains loci with significant and equal enrichment of H3K4me3 in both ESC and brain; cluster 2 contains loci with higher enrichment of H3K4me3 in ESC relative to brain tissue. (B) As a second step of analysis, the loci in cluster 2 were used as reference. The

densities around these loci were re-collected and a second round of clustering was performed. After the second round of clustering three sub-clusters can be isolated; sub cluster 2.1, 2.2 and 2.3 corresponding to loci weakly, moderately and strongly enriched in H3K4me3 mark in ESC relative to brain respectively. Note that the bottom of the sub cluster 2.3 that has H3K4me3 enrichment distant from the cluster center was not considered as a separate entity since we focused our analysis on the differential signal in the cluster center. (C) Quantification of the changes observed between the two conditions. Dotplot representing the H3K4me3 enrichment in ESC versus brain. Enrichments were calculated for H3K4me3-ESC and H3K4me3-brain datasets within a window of 2 kb around the complete set of reference coordinates (black dots), (D) and against previously isolated subsets of the reference coordinates (subset 2 in blue and subset 2.3 in green).

DISCUSSION

I. ATAC forms a stable meta structure with the Mediator complex (Krebs, Demmers et al.)

In this study, we have demonstrated the existence of a new meta coactivator complex (MECO) formed in cells by the stable association of the G-ATAC complex and the active PC2 form of the Mediator complex. We showed that the integrity of this complex depends partially on the presence of LUZP1 in MECO and that this complex is required for active transcription of a particular class of non-coding transcripts. Our findings support the novel idea that coactivator networks are functionally interconnected and under particular conditions, they can be modulated in stable meta structures to potentiate their actions at particular sites on the genome.

A. Formation of the molecular bridge

Interestingly, we could demonstrate that LUZP1 plays a role in the formation of the molecular bridge between the two complexes. However, it is clear that LUZP1 is not the only factor required for the formation of the bridge since upon LUZP1 depletion, only a partial decrease in the association is observed. Here I would like to expose and discuss the different hypotheses that could be raised on how this bridge could be formed besides the presence of LUZP1.

LUZP1 was described to contain two Leucine Zipper (LZ) motifs (Sun, Chang et al. 1996) that are usually involved in protein dimerization. Therefore, one could hypothesize that another LZ-containing partner could associate to the complex. We have tried to identify such a partner by systematically searching for LZ-containing LUZP1 associated proteins in the MS analysis by adapting existing bioinformatics tools to batch screening (Bornberg-Bauer, Rivals et al. 1998). Surprisingly, none of the identified proteins (including LUZP1) in different ATAC IPs were predicted to contain true leucine zipper motifs (LZ motif in a coiled coil structure) (data not shown). This analysis, would suggest, (1) that LUZP1 is probably not a real LZ-containing protein, and (2) by consequence that the presence of a LZ is not necessary a marker of the bridge forming factors.

The interaction with the Mediator is observed with GCN5 containing (G-ATAC), but not PCAF containing ATAC (P-ATAC) complexes (Krebs, Demmers et al.), Figure 1C). This suggests (1) that GCN5 has a specific role in the interaction, (2) that the sequences that diverge between the two paralogs are potentially relevant for this interaction. When looked carefully, the multiple alignments of GCN5/PCAF (Figure 1), reveals two regions that are highly divergent between the paralogs: the extreme (I) N-terminus; and a central region located between the "PCAF-homology domain" and (II) the catalytic region. It is then tempting to speculate that one of these two regions could have a role in the interaction with Mediator and therefore explaining why the interaction is only observed in the case of G-ATAC. However, the possibility that punctual difference in the PCAF/GCN5 sequences is contributing to the interaction cannot be excluded.

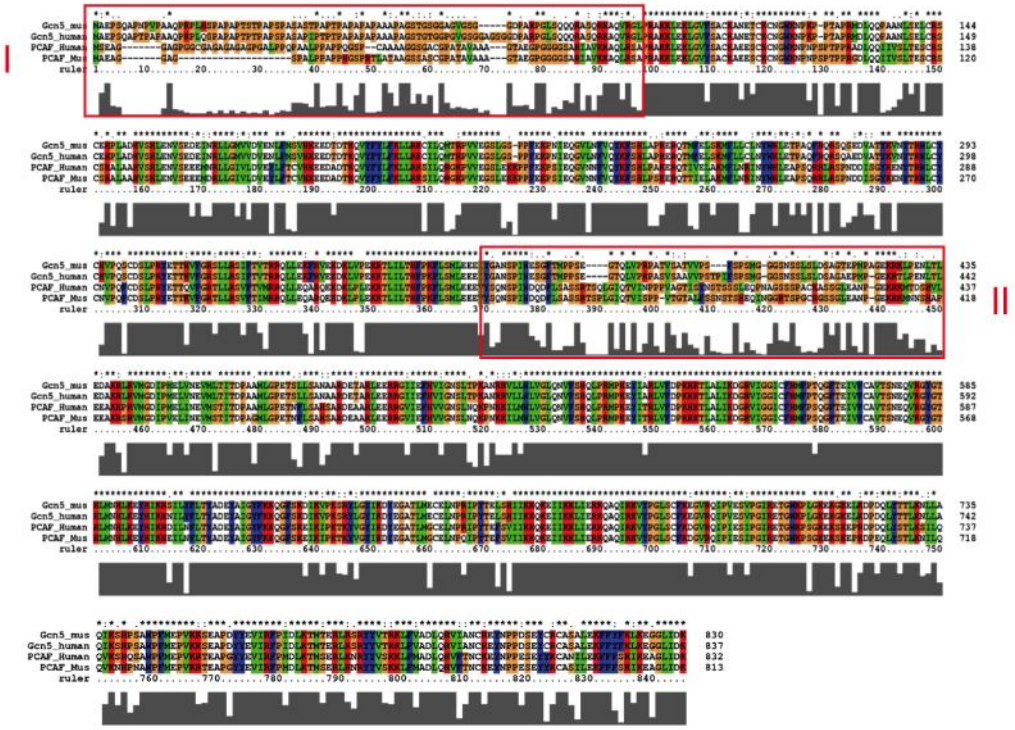


Figure 1: Multiple alignments of GCN5 and PCAF.
The human and mouse protein sequences were collected from REFSEQ for GCN5 and PCAF and aligned using the clustalW algorithm. (Larkin, Blackshields et al. 2007). The grey histograms represent the degree of conservation between the sequences. Red boxes highlight regions with poor conservation.

The observation that GCN5 may participate to the bridge is by itself striking since this protein is contained in both ATAC and SAGA complexes, but the interaction with Mediator is only observed in the case of ATAC. Thus additional to GCN5, some ATAC specific members should confer the specificity of the interaction with the Mediator. As discussed before, within

the complexes, GCN5 is contained in the ADA module that also participates in its HAT activity. One can then speculate that the ADA proteins that are thought to confer in part the catalytic specificity to the complexes (ADA2a in ATAC) would also play a role in the interaction with Mediator.

B. Implication on snRNA transcription regulation

snRNA are known to be mostly transcribed by Pol II (with the notable exception of the U6 snRNA), with a slightly adapted transcription initiation mechanism compared to mRNA genes (Hernandez 2001). The main specific feature is the requirement of a specific five subunit activation complex named SNAPc that binds snRNA promoter through a specific DNA element (PSE). Interestingly, it has been shown, contrary to mRNA genes, that a combination of GTFs and SNAPc were not sufficient to initiate Pol II transcription of U1 snRNA (Kuhlman, Cho et al. 1999). This suggests that additional co-factors are required for initiating transcription from snRNA promoters. It is then tempting to speculate that the newly identified ATAC-Mediator MECO complex could be involved in transcription of these genes. One possibility to test this hypothesis would be to perform the *in vitro* transcription experiments on snRNA template in the presence of the ATAC-Mediator (as described by (Kuhlman, Cho et al. 1999)). In any case, the recruitment of the MECO is a new layer of complexity in the regulation of the Pol II transcribed snRNA genes that expand our knowledge about current models snRNA transcription regulation.

C. Implication for co-activator mechanisms

In this study, we describe for the first time the association of two co-activator complexes in, a highly stable and cell specific manner. This led us to propose the novel concept that in particular conditions, co-activators (and eventually other transcription related complexes) can form stable meta co-activator structure to potentiate their action at particular places on the genome. This type of association is to be distinguished from low affinity transient associations that can be observed between co-activators (i.e. (Liu, Vorontchikhina et al. 2008)), most likely reflecting intermediates in the activation process, rather than functional inter-connections that could be expected for stable associated complexes. We anticipate that with the recent progress in proteomics, increasing sensitivity in the detection, more of these previously unexpected associations will be uncovered. In that case, this would imply that the

current models of sequential co-activators recruitment would evolve to a model of combinatorial co-activators action.

II. ATAC binds enhancer and promoter elements

In this study, we have established a genome wide catalogue of ATAC biding sites in various cell lines. Our results demonstrate that ATAC is binding to both active promoters and enhancers. Moreover, we observe that while the binding of ATAC on promoters is invariant across cell types, the binding to enhancers appears to be cell specific.

A. A diversity in enhancer regulation

Novel concepts on enhancer regulation have emerged from genome wide studies part of the larger ENCODE initiative (Heintzman, Stuart et al. 2007; Heintzman, Hon et al. 2009). As discussed in the Introduction, these studies have started to raise general concepts ruling the regulation of these genomic elements. Overall, the main conclusions being that enhancers were distinguishable based on the enrichment of a limited set of histone marks (H3K4me1 but not H3K4me3) together with the presence of the HAT p300/CBP. Our results extend the regulatory panel of a subset of enhancers, since we can clearly demonstrate the ATAC recruitment on these elements. This conclusion also raises a couple of interesting questions on the general regulatory rules of enhancers. First, we observe the presence of ATAC on ~400 loci while p300 was shown to bind ~35000 loci genome wide (Heintzman, Hon et al. 2009), suggesting that ATAC regulatory network would be much more specialized than the one governed by p300. Moreover, at this stage, we cannot know if this set of enhancer bound by ATAC is a subset of the previously identified p300 bound enhancers or if ATAC bound enhancers is a new category of enhancers. Interestingly, similarly to what was observed for p300 (Heintzman, Hon et al. 2009), we show that ATAC binding to enhancers appears to be cell type specific. This suggests functional similarities between the ATAC and p300 mode of action on enhancers. The comparison of the ATAC bound enhancers with p300 binding loci should be conducted to determine if both are recruited or if different enhancers have different HAT dependencies.

B. Enhancer/promoter communication

In our analysis, we showed that ATAC binds both enhancer and promoter elements. One missing step to reach a comprehensive description of ATAC regulatory network is to define the complete set of regulated genes. While the task is relatively straightforward in the case of binding of proximal elements, the attribution of a particular distal enhancer element to a gene is not easy, since they can be distant by several kilobases from the next gene. A first approximation assumes that genomes have evolved to reduce the distance between regulatory elements and consists to attribute to an enhancer the closest gene. Althrough this approach appears to work relatively well in a large number of cases (Visel, Blow et al. 2009), it was also shown not to be optimal in many other cases, creating a number of false positives in the identified gene lists. One of the main reason thought to impact distal/proximal element communication is the presence of high order chromatin structures mediated by insulator proteins (like CTCF) (Kim, Abdullaev et al. 2007). Insulator binding maps are known in several systems theoretically giving the possibility to integrate these signals in the prediction of distal/proximal element communication. However, the precise mechanism of insulators action is not fully understood yet, thus it is still difficult to predict how these signals influence this long range communications.

Besides the identification of the regulated genes, the understanding of the communication between enhancer and promoters that are kilobase away has been a subject of major debate in the field. Two major models are currently proposed to explain how this phenomenon could take place. Here I will present the dissection of one of these model in a recent study in yeast. Moreover, I will discuss the implication on the current knowledge on distal/proximal elements communication and how these conclusions could be extrapolated to mammalian cell (Krebs and Tora 2009).

(2005). PLoS Biol. *3*, e384. 10.1371/journal.pbio.0030384.

Jones, P.A., and Baylin, S.B. (2007). Cell *128*, 683–692.

Kraus, W.L. (2008). Curr. Opin. Cell Biol. *20*, 294–302.

Lin, J.C., Jeong, S., Liang, G., Takai, D., Fatemi, M., Tsai, Y.C., Egger, G., Gal-Yam, E.N., and Jones, P.A. (2007). Cancer Cell *12*, 432–444.

Tiwari, V.K., McGarvey, K.M., Licchesi, J.D., Ohm, J.E., Herman, J.G., Schubeler, D., and Baylin, S.B. (2008). PLoS Biol. *6*, 2911–2927.

Witcher, M., and Emerson, B.M. (2009). Mol. Cell *34*, 271–284.

Zilberman, D., Coleman-Derr, D., Ballinger, T., and Henikoff, S. (2008). Nature *456*, 125–129.

Zlatanova, J., and Caiafa, P. (2009). J. Cell Sci. *122*, 1275–1284.

Keys to Open Chromatin for Transcription Activation: FACT and Asf1

Arnaud Krebs[1] and László Tora[1,*]
[1]Functional Genomics Department, Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS UMR 7104, INSERM U 964, Université de Strasbourg, BP 10142-67404 ILLKIRCH Cedex, France
*Correspondence: laszlo@igbmc.u-strasbg.fr
DOI 10.1016/j.molcel.2009.05.004

In this issue of *Molecular Cell*, Takahata et al. (2009) dissect the mechanisms preceding transcription initiation, revealing a wave of nucleosome eviction, which spreads from distant *cis*-regulatory elements to promoters and which requires the activity of histone chaperones, FACT and Asf1.

In eukaryotes, the presence of chromatin represents a barrier for the transcription machinery to access the DNA that needs to be overcome to achieve transcription. Chromatin must be remodeled and post-translationally modified to allow productive transcription initiation and subsequent elongation of the pre-mRNA. Three families of functionally distinct protein complexes are involved in these processes: coactivator complexes, which harbor activities that covalently modify histones (e.g., histone acetyltranferases [HATs] or methyltransferases), and two other families involved in nucleosome displacement, the ATP-dependent remodeling complexes and histone chaperones. The histone remodeling and the modifying complexes are predominantly involved in preinitiation steps (Workman, 2006), whereas histone chaperones participate in nucleosome eviction and reassembly during transcription elongation (Park and Luger, 2008). Increasing evidence points to a role for histone chaperones in steps prior to transcription initiation (Eitoku et al., 2008), but their precise mechanisms of action in this process remained poorly understood.

In this issue, Takahata et al. (2009) dissect the mechanism of transcription activation of the *S. cerevisiae HO* gene and show that the coordinated action of two histone chaperones, Asf1 and FACT (Eitoku et al., 2008; Park and Luger, 2008; Reinberg and Sims, 2006), with that of several coactivators is required for the activation of this gene. They show that transcriptional coactivators (the Mediator complex, the SAGA HAT complex, and the ATP-dependent Swi/Snf remodeling complex) are initially recruited by the Swi5 transcription factor to a distant *cis*-regulatory element of the *HO* gene. Then, the complexes travel along a 2 kb region to reach the proximal promoter region. Using time course ChIP experiments, the authors show that this coactivator movement is concomitant with a wave of histone eviction along the region, suggesting interdependency between the two phenomena. Interestingly, this histone eviction process in the upstream promoter region requires the sequential action of Swi/Snf, FACT, and finally Asf1. Importantly, they demonstrate that the coactivator movement in this nontranscribed promoter region is dependent on the prior nucleosome removal by the histone chaperones.

Until now, the described action of FACT or Asf1 in transcription preinitiation steps was restricted to the region of the proximal promoter (Eitoku et al., 2008). Interestingly, Takahata et al. (2009) provide evidence for binding of these histone chaperones to *cis*-regulatory elements far upstream of this region. Additionally, they suggest that histone chaperones are required for the physical movement along large DNA regions of coactivator complexes in an RNA polymerase II-independent manner. Furthermore, this study emphasizes the complexity of transcription preinitiation processes, as no less than three different nucleosome displacement activities (FACT, Asf1, Swi/Snf) are necessary to promote coactivator movement to allow gene activation. It is interesting to note that complementary chaperone activities are used during this process, as FACT and Asf1 target H2A-H2B and H3-H4, respectively (Park and Luger, 2008).

The mechanism by which distant *cis*-regulatory elements and their binding

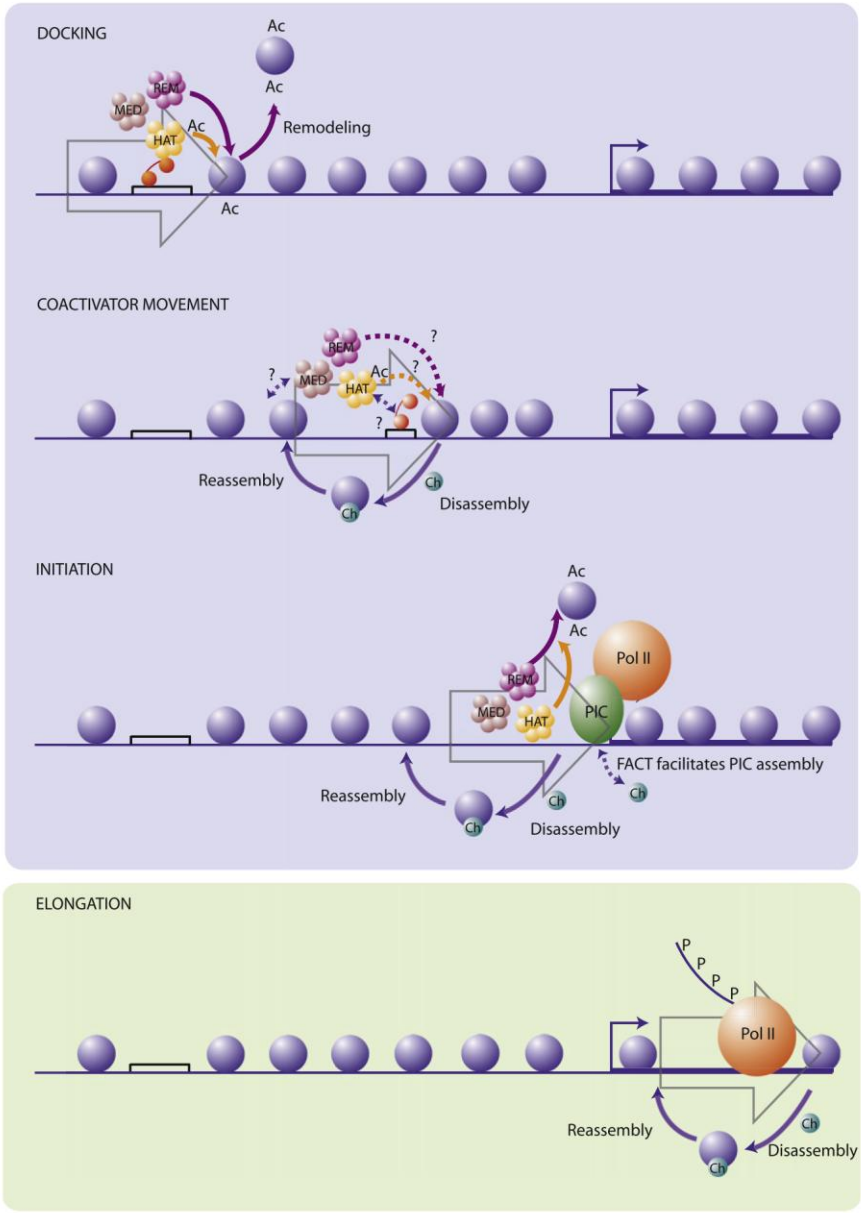


Figure 1. Involvement of Histone Chaperones in the Sequential Mechanisms Preceding Transcription Initiation

Docking: The coactivator complexes—Mediator ([MED] in brown), HAT (in light orange), and ATP-remodeling complexes ([REM] in purple)—are recruited by specific DNA-binding factors (orange) to a *cis*-regulatory element. Histones are evicted from the region by the combinatorial action of HAT and REM complexes (arrows). Coactivator Movement: The histone chaperones Asf1 and FACT ([Ch] in green) evict and reassemble nucleosomes to allow the tracking of the coactivators, possibly through interaction with other DNA-binding factors or histones (dotted arrows), towards the promoter. Initiation: The coactivators reach the proximal promoter region, where they enhance transcription initiation. Furthermore, FACT favors the formation of the preinitiation complex (PIC) through interactions with some of its subunits. Elongation: The histone chaperones evict and reassemble the nucleosomes allowing the progression of RNA polymerase II (Pol II).

factors communicate with proximal promoters and the preinitiation complex has been a long-standing question in the field of transcription. Several models are currently used to explain how activation or repression signals can be transmitted between elements that can be up to several kilobases away (reviewed in Palstra et al., 2008). In one of these models, known as the "looping model," the distant elements are brought into close spatial vicinity with the core promoter by formation of a loop in the DNA secondary structure. The existence of such a mechanism was demonstrated for the well-studied β-globin locus. The master regulator element (LCR; locus control region) of the β-globin locus is located over 50 kb from the proximal promoter. The LCR and the promoter were shown by several approaches, including 3C (chromosome conformation capture), to come into physical contact during locus activation (Vakoc et al., 2005), demonstrating a looping of the DNA separating the two elements. Another model, known as the "tracking model," proposes that coactivator complexes bound to the distant *cis*-regulatory elements track via small steps along the chromatin until they encounter the cognate promoter (Hatzis and Talianidis, 2002). Using the example of the *S. cerevisiae HO* gene, the study by Takahata et al. (2009) gives more precise insights on the mechanistic requirements for the communication of distant elements with the core promoter through a "tracking-type model" by underlining the importance of several histone chaperones (Figure 1).

As pointed out by the authors, the *HO* gene has a particularly long regulatory region when compared to other genes in the *S. cerevisiae* genome. However, in higher eukaryotic genomes distant *cis*-regulatory elements located several kilobases from the proximal promoter are a common feature (Heintzman et al., 2009). The similarity between the distant *cis*-regulatory element of the *HO* gene and some metazoan enhancers begs the question of whether the described coactivator movement mechanism could be extrapolated to higher eukaryotes. The identification of genome-wide patterns of enhancer distribution by Heintzman et al. (2009) shows that a large proportion of active human enhancers are located in a 10 kb window surrounding the transcription start sites (TSSs) of the reguated genes, but also that they can be distant up to 200 kb from the TSSs. Takahata et al. (2009) showed that no less than three nucleosome removal activities are necessary to allow the coactivator movement over a 2 kb region. This finding suggests

that the communication between very distantly located elements through the tracking model would be enzymatically very costly, and thus the nucleosome eviction wave facilitated by different chaperones will likely be limited to a set of relatively short-range interactions. The proposed mechanistic model (Figure 1) may apply only to enhancers located relatively close to the TSSs. For long-distance interactions, communication through the looping model seems more likely. The coexistence of at least two types of communication between distant *cis*-regulatory elements prior to transcription activation illustrates the complexity of transcription regulatory mechanisms. The use of high-throughput approaches, such as 5C (3C carbon copy) (Dostie et al., 2006), will likely help to classify *cis*-regulatory elements according to their regulatory mechanisms. Thus, future high-throughput approaches will help us to dissect mechanistically the possible coactivator movement mechanisms and their corresponding conservation in higher eukaryotes.

REFERENCES

Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C., et al. (2006). Genome Res. *16*, 1299–1309.

Eitoku, M., Sato, L., Senda, T., and Horikoshi, M. (2008). Cell. Mol. Life Sci. *65*, 414–444.

Hatzis, P., and Talianidis, I. (2002). Mol. Cell *10*, 1467–1477.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Nature *459*, 108–112.

Palstra, R.J., de Laat, W., and Grosveld, F. (2008). Adv. Genet. *61*, 107–142.

Park, Y.J., and Luger, K. (2008). Curr. Opin. Struct. Biol. *18*, 282–289.

Reinberg, D., and Sims, R.J., 3rd. (2006). J. Biol. Chem. *281*, 23297–23301.

Takahata, S., Yu, Y., and Stillman, D.J. (2009). Mol. Cell *34*, this issue, 405–415.

Vakoc, C.R., Letting, D.L., Gheldof, N., Sawado, T., Bender, M.A., Groudine, M., Weiss, M.J., Dekker, J., and Blobel, G.A. (2005). Mol. Cell *17*, 453–462.

Workman, J.L. (2006). Genes Dev. *20*, 2009–2017.

Let's Play Polo in the Field of Condensation

Tatsuya Hirano[1,*]
[1]RIKEN Advanced Science Institute, Chromosome Dynamics Laboratory, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan
*Correspondence: hiranot@riken.jp
DOI 10.1016/j.molcel.2009.05.002

In this issue of *Molecular Cell*, St-Pierre et al. (2009) present evidence that, in budding yeast, the polo kinase Cdc5 directly phosphorylates and activates condensin to promote anaphase-specific rDNA condensation.

Chromosome condensation is a fundamental process by which duplicated chromosomes are reorganized upon the onset of mitosis to render themselves competent for rapid and synchronous segregation in anaphase. This process not only makes long chromatin fibers compact and rigid, but also helps resolve entanglements between sister DNAs created during the preceding S phase. Any errors in chromosome condensation would cause defects in segregation, potentially resulting in aneuploidy or chromosome aberrations. Although a number of structural and regulatory factors participate in chromosome condensation, studies during the past decade have demonstrated that a class of multisubunit protein complexes, known as condensins, are among the central players in this process (reviewed in Hirano, 2005).

The canonical condensin complex (condensin I) comprises five subunits: a heterodimeric pair of SMC (structural maintenance of chromosomes) ATPase subunits and three non-SMC regulatory subunits. Higher eukaryotic cells possess a second complex (condensin II), which shares the same pair of SMC subunits and has a distinct, yet related, set of non-SMC subunits. How might this class of protein machineries induce a large-scale structural reorganization of chromosomes at a mechanistic level, and how might their actions be regulated precisely during the cell cycle? The first hints to these questions were provided more than a decade ago, immediately following the original identification of condensin I from *Xenopus* egg extracts. This work demonstrated that purified condensin I interacts directly with DNA in vitro and induces positive superhelical tension into DNA in an ATP hydrolysis-dependent manner (Kimura and Hirano, 1997). Further studies showed that this so-called supercoiling activity is mitosis specific and is activated by the master mitotic kinase Cdk1-cyclin B through direct phosphorylation of the condensin I regulatory subunits (Kimura et al., 1998). On the basis of these observations, it was proposed that the ATP-dependent supercoiling activity is one of the fundamental reactions that may drive mitotic chromosome condensation.

To what extent might this molecular mechanism be conserved during evolution? Might this seemingly simple picture be sufficient to explain the dynamic behavior and diverse functions of condensins that have accumulated in the literature since then? In this issue of *Molecular Cell*, St-Pierre et al. (2009) have now addressed these questions by using arguably the best-characterized eukaryotic organism, the budding yeast *Saccharomyces cerevisiae*. Yeast cells have

CONCLUSION

During my thesis, I have explored different aspects of the *in vivo* functionalities of GCN5 Containing Complexes (GCC). Thus using various methods, I could give new insights on the particular mode of action of these complexes as well as introducing new general concepts regarding the mechanisms ruling the action of co-activator complexes.

Through the study of the interactome of GCC in embryonic stem cells, I demonstrated that ATAC is able to work in concert with the Mediator, another co-activator complex. I showed that these complexes not only interact in a stable manner but also that this interaction mediates their synergic action on a specific set of loci. I demonstrated that this new entity is important for the regulation of snRNA genes expression. This observation led to the proposition of a couple of new concepts. First, this showed that coactivator complexes not only act in an independent but coordinated manner on their target loci, but also that they can form meta-coactivator (MECO) structures to potentiate their action at particular places on the genome. Second, our observations bring a new layer of complexity to the existing models of snRNA transcription regulation.

Through the genome wide mapping of the ATAC complex, I showed that this co-activator is recruited at two types of regulatory elements in the genome, promoter and enhancers. Interestingly, I could show that while the binding at promoters seems to be largely invariant across cell types, the binding at enhancer is highly variable from one cell to the other. These conclusions suggest that ATAC is a new regulator of distal regulatory elements. Moreover, it confirms the observation that transcription variability across cell types seems to be mostly regulated at the enhancer level.

Moreover, along the studies, I developed two integrated bioinformatics applications bringing innovative methods for analysing ChIP-seq data. The methods were implemented and spread out within the scientific community and are now broadly used.

In conclusion, my work contributed to a better understanding of several aspects of the biology of GCC in particular and their integration within the larger network of co-activators and other transcriptional regulators.

DETAIL OF CONTRIBUTIONS TO PUBLICATIONS

First author publications:

- Krishanpal Anamika*, Arnaud R. Krebs*, Julie Thompson, Olivier Poch and Làszlò Tora. Lessons from genome-wide studies: re-definition of HAT co-activator function (Epigenetics & Chromatin 2010, 3:18.).

  Participation to the scientific project design, bioinformatics tool implementation, literature review and paper writing.

  *Equal contribution

- **Krebs AR**, Demmers J, Karmodiya K, Chang NC, Chang AC, Tora L. ATAC and Mediator coactivators form a stable complex and regulate a set of non-coding RNA genes. EMBO Rep. 2010 Jul;11(7):541-7. 2010 May 28.

  Participation to the scientific project design, performed all wet experiments (except Mass Spectroemtry) and bioinformatics analysis, paper writing.

1. **Krebs A,** Tora L. Keys to open chromatin for transcription activation: FACT and Asf1. Mol Cell. 2009 May 14;34(4):397-9.

   Litterature review and paper writing (Preview).

First and corresponding author publications:

- Tao Ye*, **Arnaud R. Krebs***, Amin M. Choukrallah, Celine Keime, Frederic Plewniak, Irwin Davidson and Laszlo Tora. seqMINER: An integrated ChIP-seq data interpretation platform. (Submitted).

  Scientific design of the application, paper writing.

  *Equal contribution

2. **Krebs A,** Frontini M, Tora L. GPAT: retrieval of genomic annotation from large genomic position datasets. BMC Bioinformatics. 2008 Dec 15;9:533.

   Scientific design of the application, implementation of the software, paper writing.

Contributed publications:

3. Martianov I, Choukrallah MA, **Krebs A,** Ye T, Legras S, Rijkers E, Vanijcken W, Jost B, Sassone-Corsi P, Davidson I. Cell-specific occupancy of an extended repertoire of CREM and CREB binding loci in male germ cells. BMC Genomics. 2010 Sep 29;11(1):530.

   Bioinformatics support.

4. Orpinell M, Fournier M, Riss A, Nagy Z, **Krebs AR,** Frontini M, Tora L. The ATAC acetyl transferase complex controls mitotic progression by targeting non-histone substrates. EMBO J. 2010 Jul 21;29(14):2381-94. 2010 Jun 18.

   Production and characterisation of biological reagents (Antibodies).

5. Nagy Z, Riss A, Fujiyama S, **Krebs A,** Orpinell M, Jansen P, Cohen A, Stunnenberg HG, Kato S, Tora L. The metazoan ATAC and SAGA coactivator HAT complexes regulate different sets of inducible target genes. Cell Mol Life Sci.2010 Feb;67(4):611-28. Epub 2009 Nov 21.

   Production and characterisation of biological reagents (Antibodies).

BIBLIOGRAPHY

Abeel, T., Y. V. de Peer, et al. (2009). "Java-ML: A Machine Learning Library." Journal of Machine Learning Research **2009**(10): 931-934.

Ainbinder, E., M. Revach, et al. (2002). "Mechanism of rapid transcriptional induction of tumor necrosis factor alpha-responsive genes by NF-kappaB." Mol Cell Biol **22**(18): 6354-62.

Akhtar, A. and P. B. Becker (2000). "Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila." Mol Cell **5**(2): 367-75.

Akoulitchev, S., S. Chuikov, et al. (2000). "TFIIH is negatively regulated by cdk8-containing mediator complexes." Nature **407**(6800): 102-6.

Akoulitchev, S., T. P. Makela, et al. (1995). "Requirement for TFIIH kinase activity in transcription by RNA polymerase II." Nature **377**(6549): 557-60.

Albright, S. R. and R. Tjian (2000). "TAFs revisited: more data reveal new twists and confirm old ideas." Gene **242**(1-2): 1-13.

Allard, S., R. T. Utley, et al. (1999). "NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p." Embo J **18**(18): 5108-19.

Allison, L. A., M. Moyle, et al. (1985). "Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases." Cell **42**(2): 599-610.

Altaf, M., A. Auger, et al. "NuA4-dependent acetylation of nucleosomal histone H4 and H2A directly stimulates incorporation of H2A.Z by the SWR1 complex." J Biol Chem.

Angus-Hill, M. L., A. Schlichter, et al. (2001). "A Rsc3/Rsc30 zinc cluster dimer reveals novel roles for the chromatin remodeler RSC in gene expression and cell cycle control." Mol Cell **7**(4): 741-51.

Ares, M., Jr. and N. J. Proudfoot (2005). "The spanish connection: transcription and mRNA processing get even closer." Cell **120**(2): 163-6.

Armache, K. J., H. Kettenberger, et al. (2003). "Architecture of initiation-competent 12-subunit RNA polymerase II." Proc Natl Acad Sci U S A **100**(12): 6964-8.

Asturias, F. J., Y. W. Jiang, et al. (1999). "Conserved structures of mediator and RNA polymerase II holoenzyme." Science **283**: 985-987.

Auble, D. T., K. E. Hansen, et al. (1994). "Mot1, a global repressor of RNA polymerase II transcription, inhibits TBP binding to DNA by an ATP-dependent mechanism." Genes Dev **8**(16): 1920-34.

Bajic, V. B., S. L. Tan, et al. (2006). "Mice and men: their promoter properties." PLoS Genet **2**(4): e54.

Balasubramanian, R., M. G. Pray-Grant, et al. (2002). "Role of the Ada2 and Ada3 transcriptional coactivators in histone acetylation." J Biol Chem **277**(10): 7989-95.

Barberis, A., C. W. Muller, et al. (1993). "Delineation of two functional regions of transcription factor TFIIB." Proc.Natl.Acad.Sci.U.S.A. **90**: 5628-5632.

Barrera, L. O., Z. Li, et al. (2008). "Genome-wide mapping and analysis of active promoters in mouse embryonic stem cells and adult organs." Genome Res **18**(1): 46-59.

Barski, A., S. Cuddapah, et al. (2007). "High-resolution profiling of histone methylations in the human genome." Cell **129**(4): 823-37.

Bartolomei, M. S., N. F. Halden, et al. (1988). "Genetic analysis of the repetitive carboxyl-terminal domain of the largest subunit of mouse RNA polymerase II." Mol Cell Biol **8**(1): 330-9.

Bellorini, M., J. C. Dantonel, et al. (1996). "The major histocompatibility complex class II Ea promoter requires TFIID binding to an initiator sequence." Mol.Cell Biol. **16**: 503-512.

Bernstein, B. E., C. L. Liu, et al. (2004). "Global nucleosome occupancy in yeast." Genome Biol **5**(9): R62.

Bernstein, B. E., T. S. Mikkelsen, et al. (2006). "A bivalent chromatin structure marks key developmental genes in embryonic stem cells." Cell **125**(2): 315-26.

Bibel, M., J. Richter, et al. (2007). "Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells." Nat Protoc **2**(5): 1034-43.

Bilodeau, S., M. H. Kagey, et al. (2009). "SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state." Genes Dev **23**(21): 2484-9.

Bird, D. M. and D. L. Riddle (1989). "Molecular cloning and sequencing of ama-1, the gene encoding the largest subunit of Caenorhabditis elegans RNA polymerase II." Mol Cell Biol **9**: 4119-4130.

Birney, E., J. A. Stamatoyannopoulos, et al. (2007). "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project." Nature **447**(7146): 799-816.

Blackwood, E. M. and J. T. Kadonaga (1998). "Going the distance: a current view of enhancer action." Science **281**(5373): 60-3.

Blahnik, K. R., L. Dou, et al. "Sole-Search: an integrated analysis program for peak detection and functional annotation using ChIP-seq data." Nucleic Acids Res **38**(3): e13.

Blankenberg, D., J. Taylor, et al. (2007). "A framework for collaborative analysis of ENCODE data: making large-scale analyses biologist-friendly." Genome Res **17**(6): 960-4.

Bock, C., K. Halachev, et al. (2009). "EpiGRAPH: user-friendly software for statistical analysis and prediction of (epi)genomic data." Genome Biol **10**(2): R14.

Boeger, H., J. Griesenbeck, et al. (2003). "Nucleosomes unfold completely at a transcriptionally active promoter." Mol Cell **11**(6): 1587-98.

Bornberg-Bauer, E., E. Rivals, et al. (1998). "Computational approaches to identify leucine zippers." Nucleic Acids Res **26**(11): 2740-6.

Boyer, L. A., K. Plath, et al. (2006). "Polycomb complexes repress developmental regulators in murine embryonic stem cells." Nature **441**(7091): 349-53.

Brand, M., C. Leurent, et al. (1999). "Three-dimensional structures of the TAFII-containing complexes TFIID and TFTC." Science **286**(5447): 2151-3.

Brand, M., K. Yamamoto, et al. (1999b). "Identification of TATA-binding protein-free TAFII-containing complex subunits suggests a role in nucleosome acetylation and signal transduction." J.Biol Chem. **274**: 18285-18289.

Breathnach, R. and P. Chambon (1981). "Organization and expression of eucaryotic split genes coding for proteins." Annu.Rev.Biochem. **50**: 349-383.

Brownell, J. E. and C. D. Allis (1995). "An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in Tetrahymena macronuclei." Proc Natl Acad Sci U S A **92**(14): 6364-8.

Buratowski, S. (1994). "The basics of basal transcription by RNA polymerase II." Cell **77**: 1-3.

Buratowski, S. (2009). "Progression through the RNA polymerase II CTD cycle." Mol Cell **36**(4): 541-6.

Buratowski, S., S. Hahn, et al. (1989). "Five intermediate complexes in transcription initiation by RNA polymerase II." Cell **56**: 549-561.

Burke, T. W. and J. T. Kadonaga (1997). "The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila." Genes Dev. **11**: 3020-3031.

Burley, S. K. and R. G. Roeder (1996). "Biochemistry and structural biology of transcription factor IID (TFIID)." Annu.Rev.Biochem. **65**: 769-799.

Bushnell, D. A. and R. D. Kornberg (2003). "Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription." Proc Natl Acad Sci U S A **100**(12): 6969-73.

Bushnell, D. A., K. D. Westover, et al. (2004). "Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms." Science **303**(5660): 983-8.

Butler, J. E. and J. T. Kadonaga (2001). "Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs." Genes Dev **15**(19): 2515-9.

Cai, Y., J. Jin, et al. (2009). "Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the male-specific lethal (MSL) complex." J Biol Chem.

Calhoun, V. C., A. Stathopoulos, et al. (2002). "Promoter-proximal tethering elements regulate enhancer-promoter specificity in the Drosophila Antennapedia complex." Proc Natl Acad Sci U S A **99**(14): 9243-7.

Carninci, P., T. Kasukawa, et al. (2005). "The transcriptional landscape of the mammalian genome." Science **309**(5740): 1559-63.

Carninci, P., A. Sandelin, et al. (2006). "Genome-wide analysis of mammalian promoter architecture and evolution." Nat Genet **38**(6): 626-35.

Carrozza, M. J., R. T. Utley, et al. (2003). "The diverse functions of histone acetyltransferase complexes." Trends Genet **19**(6): 321-9.

Carruthers, L. M. and J. C. Hansen (2000). "The core histone N termini function independently of linker histones during chromatin condensation." J Biol Chem **275**(47): 37285-90.

Chalkley, G. E. and C. P. Verrijzer (1999). "DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250- TAF(II)150 complex recognizes the initiator." Embo J **18**(17): 4835-45.

Chandy, M., J. L. Gutierrez, et al. (2006). "SWI/SNF displaces SAGA-acetylated nucleosomes." Eukaryot Cell **5**(10): 1738-47.

Chang, C., C. F. Kostrub, et al. (1993). "RAP30/74 (transcription factor IIF) is required for promoter escape by RNA polymerase II." J Biol Chem **268**(27): 20482-9.

Chang, W. H. and R. D. Kornberg (2000). "Electron crystal structure of the transcription factor and DNA repair complex, core TFIIH." Cell **102**(5): 609-13.

Clements, A., J. R. Rojas, et al. (1999). "Crystal structure of the histone acetyltransferase domain of the human PCAF transcriptional regulator bound to coenzyme A." EMBO J. **18**: 3521-3532.

Cler, E., G. Papai, et al. (2009). "Recent advances in understanding the structure and function of general transcription factor TFIID." Cell Mol Life Sci **66**(13): 2123-34.

Conaway, R. C. and J. W. Conaway (1990). "Transcription initiated by RNA polymerase II and purified transcription factors from liver. Transcription factors alpha, beta gamma, and delta promote formation of intermediates in assembly of the functional preinitiation complex." J.Biol.Chem. **265**: 7559-7563.

Conaway, R. C., S. Sato, et al. (2005). "The mammalian Mediator complex and its role in transcriptional regulation." Trends Biochem Sci **30**(5): 250-5.

Cooper, S. J., N. D. Trinklein, et al. (2006). "Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome." Genome Res **16**(1): 1-10.

Corden, J. L., D. L. Cadena, et al. (1985). "A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II." Proc Natl Acad Sci U S A **82**(23): 7934-8.

Cortes, P., O. Flores, et al. (1992). "Factors involved in specific transcription by mammalian RNA polymerase II: purification and analysis of transcription factor IIA and identification of transcription factor IIJ." Mol.Cell Biol. **12**: 413-421.

Cosma, M. P., T. Tanaka, et al. (1999). "Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter." Cell **97**: 299-311.

Cote, J., J. Quinn, et al. (1994). "Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex." Science **265**(5168): 53-60.

Coulombe, B. and Z. F. Burton (1999). "DNA bending and wrapping around RNA polymerase: a "revolutionary" model describing transcriptional mechanisms." Microbiol Mol Biol Rev **63**(2): 457-78.

Damelin, M., I. Simon, et al. (2002). "The genome-wide localization of Rsc9, a component of the RSC chromatin-remodeling complex, changes in response to stress." Mol Cell **9**(3): 563-73.

de Hoon, M. J., S. Imoto, et al. (2004). "Open source clustering software." Bioinformatics **20**(9): 1453-4.

De Jong, J. and R. G. Roeder (1993). "A single cDNA, hTFIIA/alpha, encodes both the p35 and p19 subunits of human TFIIA." Genes.Dev. **7**: 2220-2234.

de Laat, W. L., N. G. Jaspers, et al. (1999). "Molecular mechanism of nucleotide excision repair." Genes Dev **13**(7): 768-85.

Dennis, G., Jr., B. T. Sherman, et al. (2003). "DAVID: Database for Annotation, Visualization, and Integrated Discovery." Genome Biol **4**(5): P3.

Dilworth, F. J., K. J. Seaver, et al. (2004). "In vitro transcription system delineates the distinct roles of the coactivators pCAF and p300 during MyoD/E47-dependent transactivation." Proc Natl Acad Sci U S A **101**(32): 11593-8.

Ding, L., M. Paszkowski-Rogacz, et al. (2009). "A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity." Cell Stem Cell **4**(5): 403-15.

Dynlacht, B. D., T. Hoey, et al. (1991). "Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation." Cell **66**: 563-576.

Eberharter, A. and P. B. Becker (2002). "Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics." EMBO Rep **3**(3): 224-9.

Emanuel, P. A. and D. S. Gilmour (1993). "Transcription factor TFIID recognizes DNA sequences downstream of the TATA element in the Hsp70 heat shock gene." Proc.Natl.Acad.Sci.U.S.A. **90**: 8449-8453.

Evans, E., J. G. Moggs, et al. (1997). "Mechanism of open complex and dual incision formation by human nucleotide excision repair factors." Embo J **16**(21): 6559-73.

Evans, R., J. A. Fairley, et al. (2001). "Activator-mediated disruption of sequence-specific DNA contacts by the general transcription factor TFIIB." Genes Dev **15**(22): 2945-9.

Fazzio, T. G., J. T. Huff, et al. (2008). "An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity." Cell **134**(1): 162-74.

Fazzio, T. G., C. Kooperberg, et al. (2001). "Widespread collaboration of Isw2 and Sin3-Rpd3 chromatin remodeling complexes in transcriptional repression." Mol Cell Biol **21**(19): 6450-60.

Feaver, W. J., N. L. Henry, et al. (1994). "Yeast TFIIE. Cloning, expression, and homology to vertebrate proteins." J.Biol.Chem. **269**: 27549-27553.

Feaver, W. J., N. L. Henry, et al. (1997). "Genes for Tfb2, Tfb3, and Tfb4 subunits of yeast transcription/repair factor IIH. Homology to human cyclin-dependent kinase activating kinase and IIH subunits." J.Biol Chem. **272**: 19319-19327.

Feaver, W. J., J. Q. Svejstrup, et al. (1994). "Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIH/TFIIK." Cell **79**: 1103-1109.

Fischer, L., M. Gerard, et al. (1992). "Cloning of the 62-kilodalton component of basic transcription factor BTF2." Science **257**(5075): 1392-5.

Fischle, W., Y. Wang, et al. (2003). "Histone and chromatin cross-talk." Curr Opin Cell Biol **15**(2): 172-83.

Fondell, J. D., H. Ge, et al. (1996). "Ligand induction of a transcriptionally active thyroid hormone receptor coactivator complex." Proc.Natl.Acad.Sci.U.S.A. **93**: 8329-8333.

Francis, N. J., R. E. Kingston, et al. (2004). "Chromatin compaction by a polycomb group protein complex." Science **306**(5701): 1574-7.

Gal-Yam, E. N., S. Jeong, et al. (2006). "Constitutive nucleosome depletion and ordered factor assembly at the GRP78 promoter revealed by single molecule footprinting." PLoS Genet **2**(9): e160.

Gamper, A. M., J. Kim, et al. (2009). "The STAGA subunit ADA2b is an important regulator of human GCN5 catalysis." Mol Cell Biol **29**(1): 266-80.

Gangloff, Y., C. Romier, et al. (2001). "The histone fold is a key structural motif of transcription factor TFIID." Trends Biochem Sci **26**(4): 250-7.

Gangloff, Y. G., S. Werten, et al. (2000). "The human TFIID components TAF(II)135 and TAF(II)20 and the yeast SAGA components ADA1 and TAF(II)68 heterodimerize to form histone-like pairs." Mol Cell Biol **20**(1): 340-51.

Geiger, J. H., S. Hahn, et al. (1996). "Crystal structure of the yeast TFIIA/TBP/DNA complex." Science **272**: 830-836.

Gerard, M., L. Fischer, et al. (1991). "Purification and interaction properties of the human RNA polymerase B(II) general transcription factor BTF2." J.Biol.Chem. **266**: 20940-20945.

Gershenzon, N. I. and I. P. Ioshikhes (2005). "Synergy of human Pol II core promoter elements revealed by statistical sequence analysis." Bioinformatics **21**(8): 1295-300.

Geyer, P. K. and V. G. Corces (1992). "DNA position-specific repression of transcription by a Drosophila zinc finger protein." Genes Dev **6**(10): 1865-73.

Goodrich, J. A. and R. Tjian (1994). "TBP-TAF complexes: selectivity factors for eukaryotic transcription." Curr.Opin.Cell Biol. **6**: 403-409.

Grant, P. A., D. Schieltz, et al. (1998). "A subset of TAF(II)s are integral components of the SAGA complex required for nucleosome acetylation and transcriptional stimulation." Cell **94**: 45-53.

Guelman, S., K. Kozuka, et al. (2009). "The double-histone-acetyltransferase complex ATAC is essential for mammalian development." Mol Cell Biol **29**(5): 1176-88.

Guelman, S., T. Suganuma, et al. (2006). "Host cell factor and an uncharacterized SANT domain protein are stable components of ATAC, a novel dAda2A/dGcn5-containing histone acetyltransferase complex in Drosophila." Mol Cell Biol **26**(3): 871-82.

Guenther, M. G., S. S. Levine, et al. (2007). "A chromatin landmark and transcription initiation at most promoters in human cells." Cell **130**(1): 77-88.

Guermah, M., Y. Tao, et al. (2001). "Positive and negative TAF(II) functions that suggest a dynamic TFIID structure and elicit synergy with traps in activator-induced transcription." Mol Cell Biol **21**(20): 6882-94.

Guillemette, B., A. R. Bataille, et al. (2005). "Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning." PLoS Biol **3**(12): e384.

Ha, I., W. S. Lane, et al. (1991). "Cloning of a human gene encoding the general transcription initiation factor IIB." Nature **352**: 689-695.

Ha, I., S. Roberts, et al. (1993). "Multiple functional domains of human transcription factor IIB: distinct interactions with two general transcription factors and RNA polymerase II." Genes.Dev. **7**: 1021-1032.

Hager, G. L., J. G. McNally, et al. (2009). "Transcription dynamics." Mol Cell **35**(6): 741-53.

Hardy, S., M. Brand, et al. (2002). "TATA-binding protein-free TAF-containing complex (TFTC) and p300 are both required for efficient transcriptional activation." J Biol Chem **277**(36): 32875-82.

Heintzman, N. D., G. C. Hon, et al. (2009). "Histone modifications at human enhancers reflect global cell-type-specific gene expression." Nature **459**(7243): 108-12.

Heintzman, N. D., R. K. Stuart, et al. (2007). "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome." Nat Genet **39**(3): 311-8.

Hekmatpanah, D. S. and R. A. Young (1991). "Mutations in a conserved region of RNA polymerase II influence the accuracy of mRNA start site selection." Mol Cell Biol **11**(11): 5781-91.

Henry, N. L., A. M. Campbell, et al. (1994). "TFIIF-TAF-RNA polymerase II connection." Genes.Dev. **8**: 2868-2878.

Hernandez, N. (1993). "TBP, a universal eukaryotic transcription factor?" Genes.Dev. **7**: 1291-1308.

Hernandez, N. (2001). "Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription." J Biol Chem **276**(29): 26733-6.

Hirose, Y. and J. L. Manley (2000). "RNA polymerase II and the integration of nuclear events." Genes Dev **14**(12): 1415-29.

Hoffmann, A., C. M. Chiang, et al. (1996). "A histone octamer-like structure within TFIID." Nature **380**(6572): 356-9.

Hoffmann, A., T. Oelgeschlager, et al. (1997). "Considerations of transcriptional control mechanisms: do TFIID-core promoter complexes recapitulate nucleosome-like functions?" Proc.Natl.Acad.Sci.U.S.A. **94**: 8928-8935.

Holstege, F. C., P. C. van der Vliet, et al. (1996). "Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH." EMBO J. **15**: 1666-1677.

Hon, G., B. Ren, et al. (2008). "ChromaSig: a probabilistic approach to finding common chromatin signatures in the human genome." PLoS Comput Biol **4**(10): e1000201.

Horikoshi, M., T. Yamamoto, et al. (1990). "Analysis of structure-function relationships of yeast TATA box binding factor TFIID." Cell **61**: 1171-1178.

Hu, Y., J. B. Fisher, et al. (2009). "Homozygous disruption of the Tip60 gene causes early embryonic lethality." Dev Dyn **238**(11): 2912-21.

Ikura, T., V. V. Ogryzko, et al. (2000). "Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis." Cell **102**(4): 463-73.

Imbalzano, A. N., H. Kwon, et al. (1994). "Facilitated binding of TATA-binding protein to nucleosomal DNA." Nature **370**(6489): 481-5.

Ingvarsdottir, K., N. J. Krogan, et al. (2005). "H2B ubiquitin protease Ubp8 and Sgf11 constitute a discrete functional module within the Saccharomyces cerevisiae SAGA complex." Mol Cell Biol **25**(3): 1162-72.

Inostroza, J. A., F. H. Mermelstein, et al. (1992). "Dr1, a TATA-binding protein-associated phosphoprotein and inhibitor of class II gene transcription." Cell **70**: 477-489.

Ito, M., C. X. Yuan, et al. (2000). "Involvement of the TRAP220 component of the TRAP/SMCC coactivator complex in embryonic development and thyroid hormone action." Mol Cell **5**(4): 683-93.

Ji, H., H. Jiang, et al. (2008). "An integrated software system for analyzing ChIP-chip and ChIP-seq data." Nat Biotechnol **26**(11): 1293-300.

Johnson, D. S., A. Mortazavi, et al. (2007). "Genome-wide mapping of in vivo protein-DNA interactions." Science **316**(5830): 1497-502.

Johnsson, A., M. Durand-Dubief, et al. (2009). "HAT-HDAC interplay modulates global histone H3K14 acetylation in gene-coding regions during stress." EMBO Rep **10**(9): 1009-14.

Kadonaga, J. T. (2002). "The DPE, a core promoter element for transcription by RNA polymerase II." Exp Mol Med **34**(4): 259-64.

Kaufmann, J. and S. T. Smale (1994). "Direct recognition of initiator elements by a component of the transcription factor IID complex." Genes Dev **8**(7): 821-9.

Kellum, R. and P. Schedl (1991). "A position-effect assay for boundaries of higher order chromosomal domains." Cell **64**(5): 941-50.

Kenneth, N. S., B. A. Ramsbottom, et al. (2007). "TRRAP and GCN5 are used by c-Myc to activate RNA polymerase III transcription." Proc Natl Acad Sci U S A **104**(38): 14917-22.

Kent, W. J., C. W. Sugnet, et al. (2002). "The human genome browser at UCSC." Genome Res **12**(6): 996-1006.

Kephart, D. D., M. P. Price, et al. (1993). "Cloning of a Drosophila cDNA with sequence similarity to human transcription factor RAP74." Nucleic Acids Res. **21**: 13|.

Kettenberger, H., K. J. Armache, et al. (2003). "Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage." Cell **114**(3): 347-57.

Killeen, M., B. Coulombe, et al. (1992). "Recombinant TBP, transcription factor IIB, and RAP30 are sufficient for promoter recognition by mammalian RNA polymerase II." J.Biol.Chem. **267**: 9463-9466.

Kim, T. H., Z. K. Abdullaev, et al. (2007). "Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome." Cell **128**(6): 1231-45.

Kim, T. H., L. O. Barrera, et al. (2005a). "Direct isolation and identification of promoters in the human genome." Genome Res **15**(6): 830-9.

Kim, T. H., L. O. Barrera, et al. (2005b). "A high-resolution map of active promoters in the human genome." Nature **436**(7052): 876-80.

Kim, T. K., R. H. Ebright, et al. (2000). "Mechanism of ATP-dependent promoter melting by transcription factor IIH." Science **288**(5470): 1418-22.

Kind, J., J. M. Vaquerizas, et al. (2008). "Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in Drosophila." Cell **133**(5): 813-28.

Kingston, R. E., C. A. Bunker, et al. (1996). "Repression and activation by multiprotein complexes that alter chromatin structure." Genes Dev **10**(8): 905-20.

Kitajima, S., T. Chibazakura, et al. (1994). "Regulation of the human general transcription initiation factor TFIIF by phosphorylation." J Biol Chem **269**(47): 29970-7.

Kohler, A., E. Zimmerman, et al. "Structural basis for assembly and activation of the heterotetrameric SAGA histone H2B deubiquitinase module." Cell **141**(4): 606-17.

Kohler, A., E. Zimmerman, et al. (2010). "Structural basis for assembly and activation of the heterotetrameric SAGA histone H2B deubiquitinase module." Cell **141**(4): 606-17.

Komashko, V. M., L. G. Acevedo, et al. (2008). "Using ChIP-chip technology to reveal common principles of transcriptional repression in normal and cancer cells." Genome Res **18**(4): 521-32.

Krajewski, W. A. and P. B. Becker (1998b). "Reconstitution of hyperacetylated, DNase I-sensitive chromatin characterized by high conformational flexibility of nucleosomal DNA." Proc.Natl.Acad.Sci.U.S.A. **95**: 1540-1545.

Krebs, A., M. Frontini, et al. (2008). "GPAT: retrieval of genomic annotation from large genomic position datasets." BMC Bioinformatics **9**: 533.

Krebs, A. and L. Tora (2009). "Keys to open chromatin for transcription activation: FACT and Asf1." Mol Cell **34**(4): 397-9.

Krebs, A. R., J. Demmers, et al. "ATAC and Mediator coactivators form a stable complex and regulate a set of non-coding RNA genes." EMBO Rep.

Kretzschmar, M., G. Stelzer, et al. (1994). "RNA polymerase II cofactor PC2 facilitates activation of transcription by GAL4-AH in vitro." Mol Cell Biol **14**(6): 3927-37.

Ku, M., R. P. Koche, et al. (2008). "Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains." PLoS Genet **4**(10): e1000242.

Kuhlman, T. C., H. Cho, et al. (1999). "The general transcription factors IIA, IIB, IIF, and IIE are required for RNA polymerase II transcription from the human U1 small nuclear RNA promoter." Mol Cell Biol **19**(3): 2130-41.

Kutach, A. K. and J. T. Kadonaga (2000). "The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters." Mol Cell Biol **20**(13): 4754-64.

Laajala, T. D., S. Raghav, et al. (2009). "A practical comparison of methods for detecting transcription factor binding sites in ChIP-seq experiments." BMC Genomics **10**: 618.

Lagrange, T., A. N. Kapanidis, et al. (1998). "New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB." Genes Dev. **12**: 34-44.

Lai, J. S. and W. Herr (1992). "Ethidium bromide provides a simple tool for identifying genuine DNA-independent protein associations." Proc Natl Acad Sci U S A **89**(15): 6958-62.

Larkin, M. A., G. Blackshields, et al. (2007). "Clustal W and Clustal X version 2.0." Bioinformatics **23**(21): 2947-8.

Lee, D. H., N. Gershenzon, et al. (2005). "Functional characterization of core promoter elements: the downstream core element is recognized by TAF1." Mol Cell Biol **25**(21): 9674-86.

Lee, K. K., L. Florens, et al. (2005). "The deubiquitylation activity of Ubp8 is dependent upon Sgf11 and its association with the SAGA complex." Mol Cell Biol **25**(3): 1173-82.

Lee, T. I., R. G. Jenner, et al. (2006). "Control of developmental regulators by Polycomb in human embryonic stem cells." Cell **125**(2): 301-13.

Leuther, K. K., D. A. Bushnell, et al. (1996). "Two-dimensional crystallography of TFIIB- and IIE-RNA polymerase II complexes: implications for start site selection and initiation complex formation." Cell **85**(5): 773-9.

Lewis, B. A., T. K. Kim, et al. (2000). "A downstream element in the human beta-globin promoter: evidence of extended sequence-specific transcription factor IID contacts." Proc Natl Acad Sci U S A **97**(13): 7172-7.

Li, B., M. Carey, et al. (2007). "The role of chromatin during transcription." Cell **128**(4): 707-19.

Li, H., B. Handsaker, et al. (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics **25**(16): 2078-9.

Lim, C. Y., B. Santoso, et al. (2004). "The MTE, a new core promoter element for transcription by RNA polymerase II." Genes Dev **18**(13): 1606-17.

Lin, Y. Y., J. Y. Lu, et al. (2009). "Protein acetylation microarray reveals that NuA4 controls key metabolic target regulating gluconeogenesis." Cell **136**(6): 1073-84.

Littlefield, O., Y. Korkhin, et al. (1999). "The structural basis for the oriented assembly of a TBP/TFB/promoter complex." Proc Natl Acad Sci U S A **96**(24): 13668-73.

Liu, X., J. Tesfai, et al. (2003). "c-Myc transformation domain recruits the human STAGA complex and requires TRRAP and GCN5 acetylase activity for transcription activation." J Biol Chem **278**(22): 20405-12.

Liu, X., M. Vorontchikhina, et al. (2008). "STAGA recruits Mediator to the MYC oncoprotein to stimulate transcription and cell proliferation." Mol Cell Biol **28**(1): 108-21.

Lo, W. S., L. Duggan, et al. (2001). "Snf1--a histone kinase that works in concert with the histone acetyltransferase Gcn5 to regulate transcription." Science **293**(5532): 1142-6.

Lu, H., L. Zawel, et al. (1992). "Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II [see comments]." Nature **358**: 641-645.

Ma, B. and N. Hernandez (2001). "A map of protein-protein contacts within the small nuclear RNA-activating protein complex SNAPc." J Biol Chem **276**(7): 5027-35.

Makela, T. P., J. D. Parvin, et al. (1995). "A kinase-deficient transcription factor TFIIH is functional in basal and activated transcription." Proc Natl Acad Sci U S A **92**(11): 5174-8.

Malik, S. and R. G. Roeder (2005). "Dynamic regulation of pol II transcription by the mammalian Mediator complex." Trends Biochem Sci **30**(5): 256-63.

Manske, H. M. and D. P. Kwiatkowski (2009). "LookSeq: a browser-based viewer for deep sequencing data." Genome Res **19**(11): 2125-32.

Martinez, E., T. K. Kundu, et al. (1998). "A human SPT3-TAFII31-GCN5-L acetylase complex distinct from transcription factor IID." J.Biol Chem. **273**: 23781-23785.

Matsui, T., J. Segall, et al. (1980). "Multiple factors required for accurate initiation of transcription by purified RNA polymerase II." J.Biol Chem. **255**: 1|92-1|96.

Mavrich, T. N., I. P. Ioshikhes, et al. (2008). "A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome." Genome Res **18**(7): 1073-83.

Maxon, M. E., J. A. Goodrich, et al. (1994). "Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance." Genes Dev **8**(5): 515-24.

McMahon, S. B., H. A. Van Buskirk, et al. (1998). "The novel ATM-related protein TRRAP is an essential cofactor for the c-Myc and E2F oncoproteins." Cell **94**(3): 363-74.

McMahon, S. B., M. A. Wood, et al. (2000). "The essential cofactor TRRAP recruits the histone acetyltransferase hGCN5 to c-Myc." Mol Cell Biol **20**(2): 556-62.

Merino, A., K. R. Madden, et al. (1993). "DNA topoisomerase I is involved in both repression and activation of transcription." Nature **365**(6443): 227-32.

Mikkelsen, T. S., M. Ku, et al. (2007). "Genome-wide maps of chromatin state in pluripotent and lineage-committed cells." Nature **448**(7153): 553-60.

Mittal, V., B. Ma, et al. (1999). "SNAP(c): a core promoter factor with a built-in DNA-binding damper that is deactivated by the Oct-1 POU domain." Genes Dev **13**(14): 1807-21.

Mohn, F., M. Weber, et al. (2008). "Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors." Mol Cell **30**(6): 755-66.

Morales, V., T. Straub, et al. (2004). "Functional integration of the histone acetyltransferase MOF into the dosage compensation complex." EMBO J **23**(11): 2258-68.

Naar, A. M., S. Ryu, et al. (1998). "Cofactor requirements for transcriptional activation by Sp1." Cold.Spring.Harb.Symp.Quant.Biol **63**: 189-199.

Nagy, Z., A. Riss, et al. (2010). "The metazoan ATAC and SAGA coactivator HAT complexes regulate different sets of inducible target genes." Cell Mol Life Sci **67**(4): 611-28.

Nagy, Z., A. Riss, et al. (2009c). "The human SPT20-containing SAGA complex plays a direct role in the regulation of endoplasmic reticulum stress-induced genes." Mol Cell Biol **29**(6): 1649-60.

Nagy, Z. and L. Tora (2007). "Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation." Oncogene **26**(37): 5341-57.

Nakajima, N., M. Horikoshi, et al. (1988). "Factors involved in specific transcription by mammalian RNA polymerase II: purification, genetic specificity, and TATA box-promoter interactions of TFIID." Mol Cell Biol **8**: 4028-4040.

Nechaev, S. and K. Adelman (2008). "Promoter-proximal Pol II: when stalling speeds things up." Cell Cycle **7**(11): 1539-44.

Nikolov, D. B., H. Chen, et al. (1995). "Crystal structure of a TFIIB-TBP-TATA-element ternary complex." Nature **377**: 119-128.

Nikolov, D. B., S. H. Hu, et al. (1992). "Crystal structure of TFIID TATA-box binding protein." Nature **360**(6399): 40-6.

Nobrega, M. A., I. Ovcharenko, et al. (2003). "Scanning human gene deserts for long-range enhancers." Science **302**(5644): 413.

Nonet, M., D. Sweetser, et al. (1987). "Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II." Cell **50**(6): 909-15.

O'Shea-Greenfield, A. and S. T. Smale (1992). "Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription." J.Biol.Chem. **267**: 6450.

Oelgeschlager, T., C. M. Chiang, et al. (1996). "Topology and reorganization of a human TFIID-promoter complex." Nature **382**: 735-738.

Ogryzko, V. V., T. Kotani, et al. (1998). "Histone-like TAFs within the PCAF histone acetylase complex " Cell **94**: 35-44.

Ohkuma, Y. and R. G. Roeder (1994). "Regulation of TFIIH ATPase and kinase activities by TFIIE during active initiation complex formation." Nature **368**(6467): 160-3.

Ohkuma, Y., H. Sumimoto, et al. (1990). "Factors involved in specific transcription by mammalian RNA polymerase II: purification and characterization of general transcription factor TFIIE." Proc.Natl.Acad.Sci.U.S.A. **87**: 9163-9167.

Ohler, U., G. C. Liao, et al. (2002). "Computational analysis of core promoters in the Drosophila genome." Genome Biol **3**(12): RESEARCH0087-7.

Ohler, U. and D. A. Wassarman "Promoting developmental transcription." Development **137**(1): 15-26.

Okamoto, T., S. Yamamoto, et al. (1998). "Analysis of the role of TFIIE in transcriptional regulation through structure-function studies of the TFIIEbeta subunit [In Process Citation]." J.Biol Chem. **273**: 19866-19876.

Orphanides, G., T. Lagrange, et al. (1996). "The general transcription factors of RNA polymerase II." Genes Dev. **10**: 2657-2683.

Orpinell, M., M. Fournier, et al. (2010). "The ATAC acetyl transferase complex controls mitotic progression by targeting non-histone substrates." Embo J.

Ouzounis, C. and C. Sander (1992). "TFIIB, an evolutionary link between the transcription machineries of archaebacteria and eukaryotes." Cell **71**(2): 189-90.

Ozer, J., P. A. Moore, et al. (1994). "Molecular cloning of the small (gamma) subunit of human TFIIA reveals functions critical for activated transcription." Genes Dev **8**(19): 2324-35.

Ozsolak, F., J. S. Song, et al. (2007). "High-throughput mapping of the chromatin structure of human promoters." Nat Biotechnol **25**(2): 244-8.

Paoletti, A. C., T. J. Parmely, et al. (2006). "Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors." Proc Natl Acad Sci U S A **103**(50): 18928-33.

Pappin, D. J., P. Hojrup, et al. (1993). "Rapid identification of proteins by peptide-mass fingerprinting." Curr Biol **3**(6): 327-32.

Parker, C. S. and J. Topol (1984). "A Drosophila RNA polymerase II transcription factor contains a promoter-region-specific DNA-binding activity." Cell **36**(2): 357-69.

Peterson, M. G., N. Tanese, et al. (1990). "Functional domains and upstream activation properties of cloned human TATA binding protein [published erratum appears in Science 1990 Aug 24; 249(4971):844]." Science **248**: 1625-1630.

Phatnani, H. P. and A. L. Greenleaf (2006). "Phosphorylation and functions of the RNA polymerase II CTD." Genes Dev **20**(21): 2922-36.

Pinto, I., D. E. Ware, et al. (1992). "The yeast SUA7 gene encodes a homolog of human transcription factor TFIIB and is required for normal start site selection in vivo." Cell **68**(5): 977-88.

Powell, D. W., C. M. Weaver, et al. (2004). "Cluster analysis of mass spectrometry data reveals a novel component of SAGA." Mol Cell Biol **24**(16): 7249-59.

Pray-Grant, M. G., J. A. Daniel, et al. (2005). "Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation." Nature **433**(7024): 434-8.

Proudfoot, N. J., A. Furger, et al. (2002). "Integrating mRNA processing with transcription." Cell **108**(4): 501-12.

Purrello, M., C. Di Pietro, et al. (1994). "Localization of the human genes encoding the two subunits of general transcription factor TFIIE." Genomics **23**(1): 253-5.

Qin, Y., L. K. Kong, et al. (2004). "Long-range activation of Sox9 in Odd Sex (Ods) mice." Hum Mol Genet **13**(12): 1213-8.

Qureshi, S. A. and S. P. Jackson (1998). "Sequence-specific DNA binding by the S. shibatae TFIIB homolog, TFB, and its effect on promoter strength." Mol Cell **1**(3): 389-400.

Rahl, P. B., C. Y. Lin, et al. (2010). "c-Myc regulates transcriptional pause release." Cell **141**(3): 432-45.

Ranish, J. A., W. S. Lane, et al. (1992). "Isolation of two genes that encode subunits of the yeast transcription factor IIA." Science **255**(5048): 1127-9.

Reed, R. (2003). "Coupling transcription, splicing and mRNA export." Curr Opin Cell Biol **15**(3): 326-31.

Reinke, H. and W. Horz (2003). "Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter." Mol Cell **11**(6): 1599-607.

Ringrose, L., H. Ehret, et al. (2004). "Distinct contributions of histone H3 lysine 9 and 27 methylation to locus-specific stability of polycomb complexes." Mol Cell **16**(4): 641-53.

Robert, F., D. K. Pokholok, et al. (2004). "Global position and recruitment of HATs and HDACs in the yeast genome." Mol Cell **16**(2): 199-209.

Roberts, S. M. and F. Winston (1996). "SPT20/ADA5 encodes a novel protein functionally related to the TATA- binding protein and important for transcription in Saccharomyces cerevisiae." Mol Cell Biol **16**: 3206-3213.

Roeder, R. G. (1996). "The role of general initiation factors in transcription by RNA polymerase II." Trends.Biochem.Sci. **21**: 327-335.

Roeder, R. G. (2003). "Lasker Basic Medical Research Award. The eukaryotic transcriptional machinery: complexities and mechanisms unforeseen." Nat Med **9**(10): 1239-44.

Rojas, J. R., R. C. Trievel, et al. (1999). "Structure of Tetrahymena GCN5 bound to coenzyme A and a histone H3 peptide [In Process Citation]." Nature **401**: 93-98.

Rosaleny, L. E., A. B. Ruiz-Garcia, et al. (2007). "The Sas3p and Gcn5p histone acetyltransferases are recruited to similar genes." Genome Biol **8**(6): R119.

Roth, S. Y., J. M. Denu, et al. (2001). "Histone acetyltransferases." Annu Rev Biochem **70**: 81-120.

Rozowsky, J., G. Euskirchen, et al. (2009). "PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls." Nat Biotechnol **27**(1): 66-75.

Saldanha, A. J. (2004). "Java Treeview--extensible visualization of microarray data." Bioinformatics **20**(17): 3246-8.

Samara, N. L., A. B. Datta, et al. "Structural insights into the assembly and function of the SAGA deubiquitinating module." Science **328**(5981): 1025-9.

Samara, N. L., A. B. Datta, et al. (2010). "Structural insights into the assembly and function of the SAGA deubiquitinating module." Science **328**(5981): 1025-9.

Sandelin, A., P. Carninci, et al. (2007). "Mammalian RNA polymerase II core promoters: insights from genome-wide studies." Nat Rev Genet **8**(6): 424-36.

Sanders, S. L., J. Jennings, et al. (2002). "Proteomics of the eukaryotic transcription machinery: identification of proteins associated with components of yeast TFIID by multidimensional mass spectrometry." Mol Cell Biol **22**(13): 4723-38.

Santos-Rosa, H., R. Schneider, et al. (2003). "Methylation of histone H3 K4 mediates association of the Isw1p ATPase with chromatin." Mol Cell **12**(5): 1325-32.

Sapountzi, V., I. R. Logan, et al. (2006). "Cellular functions of TIP60." Int J Biochem Cell Biol **38**(9): 1496-509.

Sartorelli, V., P. L. Puri, et al. (1999). "Acetylation of MyoD directed by PCAF is necessary for the execution of the muscle program." Mol Cell **4**(5): 725-34.

Saurin, A. J., Z. Shao, et al. (2001). "A Drosophila Polycomb group complex includes Zeste and dTAFII proteins." Nature **412**(6847): 655-60.

Sawadogo, M. and R. G. Roeder (1985). "Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region." Cell **43**(1): 165-75.

Schaeffer, L., V. Moncollin, et al. (1994). "The ERCC2/DNA repair protein is associated with the class II BTF2/TFIIH transcription factor." Embo J **13**(10): 2388-92.

Schaeffer, L., R. Roy, et al. (1993). "DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor [see comments]." Science **260**: 58-63.

Schiltz, R. L., C. A. Mizzen, et al. (1999). "Overlapping but distinct patterns of histone acetylation by the human coactivators p300 and PCAF within nucleosomal substrates." J Biol Chem **274**(3): 1189-92.

Schones, D. E., K. Cui, et al. (2008). "Dynamic regulation of nucleosome positioning in the human genome." Cell **132**(5): 887-98.

Schultz, P., S. Fribourg, et al. (2000). "Molecular structure of human TFIIH." Cell **102**(5): 599-607.

Schwabish, M. A. and K. Struhl (2007). "The Swi/Snf complex is important for histone eviction during transcriptional activation and RNA polymerase II elongation in vivo." Mol Cell Biol **27**(20): 6987-95.

Selleck, W., R. Howley, et al. (2001). "A histone fold TAF octamer within the yeast TFIID transcriptional coactivator." Nat Struct Biol **8**(8): 695-700.

Shahbazian, M. D. and M. Grunstein (2007). "Functions of site-specific histone acetylation and deacetylation." Annu Rev Biochem **76**: 75-100.

Sharp, P. A. (1992). "TATA-binding protein is a classless factor." Cell **68**: 8|-21.

Shen, X., G. Mizuguchi, et al. (2000). "A chromatin remodelling complex involved in transcription and DNA processing." Nature **406**(6795): 541-4.

Shin, H., T. Liu, et al. (2009). "CEAS: cis-regulatory element annotation system." Bioinformatics **25**(19): 2605-6.

Shivaswamy, S., A. Bhinge, et al. (2008). "Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation." PLoS Biol **6**(3): e65.

Shogren-Knaak, M., H. Ishii, et al. (2006). "Histone H4-K16 acetylation controls chromatin structure and protein interactions." Science **311**(5762): 844-7.

Shogren-Knaak, M. and C. L. Peterson (2006). "Switching on chromatin: mechanistic role of histone H4-K16 acetylation." Cell Cycle **5**(13): 1361-5.

Sims, R. J., 3rd, S. S. Mandal, et al. (2004). "Recent highlights of RNA-polymerase-II-mediated transcription." Curr Opin Cell Biol **16**(3): 263-71.

Singer, V. L., C. R. Wobbe, et al. (1990). "A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation." Genes.Dev. **4**: 636-645.

Skinner, M. E., A. V. Uzilov, et al. (2009). "JBrowse: a next-generation genome browser." Genome Res **19**(9): 1630-8.

Smale, S. T. and D. Baltimore (1989). "The "initiator" as a transcription control element." Cell **57**: 103-113.

Smale, S. T. and J. T. Kadonaga (2003). "The RNA polymerase II core promoter." Annu Rev Biochem **72**: 449-79.

Sopta, M., Z. F. Burton, et al. (1989). "Structure and associated DNA-helicase activity of a general transcription initiation factor that binds to RNA polymerase II." Nature **341**: 410-414.

Sterner, D. E. and S. L. Berger (2000a). "Acetylation of histones and transcription-related factors." Microbiol Mol Biol Rev **64**(2): 435-59.

Suganuma, T., J. L. Gutierrez, et al. (2008). "ATAC is a double histone acetyltransferase complex that stimulates nucleosome sliding." Nat Struct Mol Biol **15**(4): 364-72.

Sun, D. S., A. C. Chang, et al. (1996). "Identification, molecular characterization, and chromosomal localization of the cDNA encoding a novel leucine zipper motif-containing protein." Genomics **36**(1): 54-62.

Sun, F. L. and S. C. Elgin (1999). "Putting boundaries on silence." Cell **99**(5): 459-62.

Svejstrup, J. Q., P. Vichi, et al. (1996). "The multiple roles of transcription/repair factor TFIIH." Trends Biochem Sci **21**(9): 346-50.

Syntichaki, P., I. Topalidou, et al. (2000). "The Gcn5 bromodomain co-ordinates nucleosome remodelling." Nature **404**(6776): 414-7.

Szutorisz, H., N. Dillon, et al. (2005). "The role of enhancers as centres for general transcription factor recruitment." Trends Biochem Sci **30**(11): 593-9.

Takahata, S., Y. Yu, et al. (2009). "FACT and Asf1 regulate nucleosome dynamics and coactivator binding at the HO promoter." Mol Cell **34**(4): 405-15.

Tan, S., Y. Hunziker, et al. (1996). "Crystal structure of a yeast TFIIA/TBP/DNA complex." Nature **381**: 127-151.

Taslim, C., J. Wu, et al. (2009). "Comparative study on ChIP-seq data: normalization and binding pattern characterization." Bioinformatics **25**(18): 2334-40.

Thomas, T., M. P. Dixon, et al. (2008). "Mof (MYST1 or KAT8) is essential for progression of embryonic development past the blastocyst stage and required for normal chromatin architecture." Mol Cell Biol **28**(16): 5093-105.

Tora, L. (2002). "A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription." Genes Dev **16**(6): 673-5.

Tran, H. G., D. J. Steger, et al. (2000). "The chromo domain protein chd1p from budding yeast is an ATP-dependent chromatin-modifying factor." EMBO J **19**(10): 2323-31.

Tsai, F. T. and P. B. Sigler (2000). "Structural basis of preinitiation complex assembly on human pol II promoters." Embo J **19**(1): 25-36.

Tse, C., T. Sera, et al. (1998). "Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III." Mol Cell Biol **18**(8): 4629-38.

Tsukiyama, T. (2002). "The in vivo functions of ATP-dependent chromatin-remodelling factors." Nat Rev Mol Cell Biol **3**(6): 422-9.

Valenzuela, L. and R. T. Kamakaka (2006). "Chromatin insulators." Annu Rev Genet **40**: 107-38.

Venters, B. J. and B. F. Pugh (2009a). "A canonical promoter organization of the transcription machinery and its regulators in the Saccharomyces genome." Genome Res **19**(3): 360-71.

Venters, B. J. and B. F. Pugh (2009b). "How eukaryotic genes are transcribed." Crit Rev Biochem Mol Biol **44**(2-3): 117-41.

Vermeulen, W., A. J. van Vuuren, et al. (1994). "Three unusual repair deficiencies associated with transcription factor BTF2(TFIIH): evidence for the existence of a transcription syndrome." Cold Spring Harb Symp Quant Biol **59**: 317-29.

Verrijzer, C. P., J. L. Chen, et al. (1995). "Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II." Cell **81**: 1115-1125.

Visel, A., M. J. Blow, et al. (2009). "ChIP-seq accurately predicts tissue-specific activity of enhancers." Nature **457**(7231): 854-8.

Wang, J. C. and M. W. Van Dyke (1993). "Initiator sequences direct downstream promoter binding by human transcription factor IID." Biochim Biophys Acta **1216**(1): 73-80.

Wang, X., S. K. Hansen, et al. (1997). "Drosophila TFIIE: purification, cloning, and functional reconstitution." Proc Natl Acad Sci U S A **94**(2): 433-8.

Wang, Y. L., F. Faiola, et al. (2008). "Human ATAC Is a GCN5/PCAF-containing acetylase complex with a novel NC2-like histone fold module that interacts with the TATA-binding protein." J Biol Chem **283**(49): 33808-15.

Wang, Z., S. Buratowski, et al. (1995). "The yeast TFB1 and SSL1 genes, which encode subunits of transcription factor IIH, are required for nucleotide excision repair and RNA polymerase II transcription." Mol Cell Biol **15**(4): 2288-93.

Wang, Z., C. Zang, et al. (2009). "Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes." Cell **138**(5): 1019-31.

Wang, Z., C. Zang, et al. (2008). "Combinatorial patterns of histone acetylations and methylations in the human genome." Nat Genet **40**(7): 897-903.

Welboren, W. J., M. A. van Driel, et al. (2009). "ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands." Embo J **28**(10): 1418-28.

Wieczorek, E., M. Brand, et al. (1998). "Function of TAF(II)-containing complex without TBP in transcription by RNA polymerase II." Nature **393**: 187-191.

Winston, F. and M. Carlson (1992). "Yeast SNF/SWI transcriptional activators and the SPT/SIN chromatin connection." Trends Genet **8**(11): 387-91.

Winston, F. and P. Sudarsanam (1998). "The SAGA of Spt proteins and transcriptional analysis in yeast: past, present, and future." Cold Spring Harb Symp Quant Biol **63**: 553-61.

Wolffe, A. P. and J. C. Hansen (2001). "Nuclear visions: functional flexibility from structural instability." Cell **104**(5): 631-4.

Wong, J. M. and E. Bateman (1994). "TBP-DNA interactions in the minor groove discriminate between A:T and T:A base pairs." Nucleic Acids Res **22**(10): 1890-6.

Wu, P. Y., C. Ruhlmann, et al. (2004). "Molecular architecture of the S. cerevisiae SAGA complex." Mol Cell **15**(2): 199-208.

Xie, X., T. Kokubo, et al. (1996). "Structural similarity between TAFs and the heterotetrameric core of the histone octamer." Nature **380**(6572): 316-22.

Xu, W., D. G. Edmondson, et al. (2000). "Loss of Gcn5l2 leads to increased apoptosis and mesodermal defects during mouse development." Nat Genet **26**(2): 229-32.

Yamamoto, T. and M. Horikoshi (1997). "Novel substrate specificity of the histone acetyltransferase activity of HIV-1-Tat interactive protein Tip60." J Biol Chem **272**(49): 30595-8.

Yamashita, S., K. Wada, et al. (1992). "Isolation and characterization of a cDNA encoding Drosophila transcription factor TFIIB." Proc.Natl.Acad.Sci.U.S.A. **89**: 2839-2843.

Yanagisawa, J., H. Kitagawa, et al. (2002). "Nuclear receptor function requires a TFTC-type histone acetyl transferase complex." Mol Cell **9**(3): 553-62.

Yao, T. P., S. P. Oh, et al. (1998). "Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300." Cell **93**(3): 361-72.

Yokomori, K., A. Admon, et al. (1993). "Drosophila TFIIA-L is processed into two subunits that are associated with the TBP/TAF complex." Genes.Dev. **7**: 2235-2245.

Young, R. A. (1991). "RNA polymerase II." Annu Rev Biochem **60**: 689-715.

Yuan, C. X., M. Ito, et al. (1998). "The TRAP220 component of a thyroid hormone receptor-associated protein (TRAP) coactivator complex interacts directly with nuclear receptors in a ligand-dependent fashion." Proc.Natl.Acad.Sci.U.S.A. **95**: 7939-7944.

Yuan, G. C., Y. J. Liu, et al. (2005). "Genome-scale identification of nucleosome positions in S. cerevisiae." Science **309**(5734): 626-30.

Yudkovsky, N., C. Logie, et al. (1999). "Recruitment of the SWI/SNF chromatin remodeling complex by transcriptional activators." Genes Dev **13**(18): 2369-74.

Zawel, L. and D. Reinberg (1993). "Initiation of transcription by RNA polymerase II: a multi-step process." Prog.Nucleic Acid.Res.Mol.Biol. **44**: 67-108.

Zehring, W. A., J. M. Lee, et al. (1988). "The C-terminal repeat domain of RNA polymerase II largest subunit is essential in vivo but is not required for accurate transcription initiation in vitro." Proc Natl Acad Sci U S A **85**(11): 3698-702.

Zeitlinger, J., A. Stark, et al. (2007). "RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo." Nat Genet **39**(12): 1512-6.

Zhang, H., D. N. Roberts, et al. (2005). "Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss." Cell **123**(2): 219-31.

Zhang, X. Y., M. Varthi, et al. (2008). "The putative cancer stem cell marker USP22 is a subunit of the human SAGA complex required for activated transcription and cell-cycle progression." Mol Cell **29**(1): 102-11.

Zhang, Y., T. Liu, et al. (2008). "Model-based analysis of ChIP-Seq (MACS)." Genome Biol **9**(9): R137.

Zhao, Y., G. Lang, et al. (2008). "A TFTC/STAGA module mediates histone H2A and H2B deubiquitination, coactivates nuclear receptors, and counteracts heterochromatin silencing." Mol Cell **29**(1): 92-101.

Zhou, Q., T. G. Boyer, et al. (1993). "Factors (TAFs) required for activated transcription interact with TATA box-binding protein conserved core domain." Genes.Dev. **7**: 180-187.

Zhu, W., Q. Zeng, et al. (1996). "The N-terminal domain of TFIIB from Pyrococcus furiosus forms a zinc ribbon." Nat Struct Biol **3**(2): 122-4.

APPENDIX

Contributed Studies

1. Orpinell M, Fournier M, Riss A, Nagy Z, **Krebs AR,** Frontini M, Tora L. The ATAC acetyl transferase complex controls mitotic progression by targeting non-histone substrates. EMBO J. 2010 Jul 21;29(14):2381-94. 2010 Jun 18.

2. Nagy Z, Riss A, Fujiyama S, **Krebs A,** Orpinell M, Jansen P, Cohen A, Stunnenberg HG, Kato S, Tora L. The metazoan ATAC and SAGA coactivator HAT complexes regulate different sets of inducible target genes. Cell Mol Life Sci.2010 Feb;67(4):611-28. Epub 2009 Nov 21.

The EMBO Journal (2010) 29, 2381–2394 | © 2010 European Molecular Biology Organization | All Rights Reserved 0261-4189/10
www.embojournal.org

THE EMBO JOURNAL

The ATAC acetyl transferase complex controls mitotic progression by targeting non-histone substrates

Meritxell Orpinell, Marjorie Fournier, Anne Riss, Zita Nagy, Arnaud R Krebs, Mattia Frontini[1] and Làszlò Tora*

Department of Functional Genomics, Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS UMR 7104, INSERM U964, Université de Strasbourg, Illkirch Cedex, France

All DNA-related processes rely on the degree of chromatin compaction. The highest level of chromatin condensation accompanies transition to mitosis, central for cell cycle progression. Covalent modifications of histones, mainly deacetylation, have been implicated in this transition, which also involves transcriptional repression. Here, we show that the Gcn5-containing histone acetyl transferase complex, *A*da *T*wo *A* containing (ATAC), controls mitotic progression through the regulation of the activity of non-histone targets. RNAi for the ATAC subunits Ada2a/Ada3 results in delayed M/G1 transition and pronounced cell division defects such as centrosome multiplication, defective spindle and midbody formation, generation of binucleated cells and hyperacetylation of histone H4K16 and α-tubulin. We show that ATAC localizes to the mitotic spindle and controls cell cycle progression through direct acetylation of Cyclin A/Cdk2. Our data describes a new pathway in which the ATAC complex controls Cyclin A/Cdk2 mitotic function: ATAC/Gcn5-mediated acetylation targets Cyclin A for degradation, which in turn regulates the SIRT2 deacetylase activity. Thus, we have uncovered an essential function for ATAC in regulating Cyclin A activity and consequent mitotic progression.

The EMBO Journal (2010) **29,** 2381–2394. doi:10.1038/emboj.2010.125; Published online 18 June 2010
Subject Categories: chromatin & transcription; cell cycle
Keywords: Ada; chromatin; histone; spindle; α-tubulin

Introduction

Eukaryotic cells must regulate accurately the packaging and unfolding of their chromatin throughout the cell cycle to ensure precise transcription and timely replication of their genetic material. The structural features of chromatin are controlled partially by post-translational modifications occurring on histones, among which acetylation has a major function (Kouzarides, 2007). Histone acetylation levels are defined by the co-ordinated but opposite action of histone acetyl transferase (HAT) and deacetylase (HDAC) enzymes, which regulate essential cellular processes, such as DNA replication, transcription and/or cell division. During mitotic chromosome condensation, HDAC activity is favoured on the histones resulting in a predominantly deacetylated state (Valls *et al*, 2005). HDAC enzymes also regulate mitotic progression by targeting non-histone substrates that drive chromosome separation (Dryden *et al*, 2003; Ishii *et al*, 2008). The high level of histone deacetylation during mitosis suggested that the activity of HAT complexes is downregulated during this process, and thus their potential contribution to cell division remained largely unexplored.

Gcn5, the founding member of a GNAT protein family, is a subunit of several transcriptional coactivator complexes (Brownell and Allis, 1996; Lee and Workman, 2007). In addition to the function of Gcn5 in transcription regulation, its potential involvement in cell cycle regulation has been recently described (Vernarecci *et al*, 2008; Paolinelli *et al*, 2009). In metazoans, at least two Gcn5 containing HAT complexes exist: *S*pt–*A*da–*Gcn5 a*cetyltransferase (SAGA) and *A*da *T*wo *A* containing (ATAC) (Lee and Workman, 2007; Nagy and Tora, 2007; Suganuma *et al*, 2008; Wang *et al*, 2008; Guelman *et al*, 2009; Nagy *et al*, 2010). These two complexes share a number of components, but differ in molecular size, subunit composition and substrate specificity (Martinez, 2002; Ciurciu *et al*, 2008; Suganuma *et al*, 2008; Nagy *et al*, 2010). Gcn5 and two adaptor proteins, Ada2b/Ada3 in SAGA or Ada2a/Ada3 in ATAC, form the catalytic core of the complexes, respectively (Suganuma *et al*, 2008; Wang *et al*, 2008; Gamper *et al*, 2009; Nagy *et al*, 2010). In addition, the mammalian ATAC complex harbours several *bona fide* subunits with distinct properties, such as a second putative HAT enzyme (Atac2), other subunits involved in transcription regulation (NC2β), nucleosome remodelling (Wdr5, Sgf29), cell growth (Yeats2) and potential DNA binding (Zzz3) (Wang *et al*, 2008; Guelman *et al*, 2009; Nagy *et al*, 2010). Recently, it has been shown that the presence of Gcn5-HAT, or its vertebrate paralogue, Pcaf, is mutually exclusive in mammalian ATAC complexes (Nagy *et al*, 2010).

Drosophila ATAC possesses different substrate specificity than dSAGA, as it mainly acetylates histone H4 (Ciurciu *et al*, 2006; Guelman *et al*, 2006; Suganuma *et al*, 2008). The H4-specific activity was suggested to result from the presence of the second HAT, Atac2, in the complex (Suganuma *et al*, 2008). However, when testing the HAT activity of different human ATAC preparations on free histones and nucleosomes, it acetylated histone H3 and H4, with histone H3 being the preferential target (Wang *et al*, 2008; Guelman *et al*, 2009; Nagy *et al*, 2010). As in human, both SAGA and ATAC complexes have same specificity towards histone H3 and

*Corresponding author. Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS UMR 7104, INSERM U964, Université de Strasbourg, 1 Rue Laurent Fries, BP 10142, Illkirch Cedex 67404, France.
Tel.: +33 388 65 34 44, Fax: +33 388 65 32 01;
E-mail: laszlo@igbmc.u-strasbg.fr
[1]Present address: Clinical Science Center, Hammersmith Hospital Campus, Du Cane Road, London W12 0NN, UK

Received: 3 November 2009; accepted: 14 May 2010; published online: 18 June 2010

H4, acetylation of different non-histones targets could give functional specificity for each complex. However, at present the function of the metazoan ATAC complex is not clear, and the physiological targets of this complex await further analysis.

Here, we identify a function for the mammalian ATAC complex in orchestrating mitotic progression. We provide evidence that the specific depletion of the Ada core of ATAC leads to severe mitotic abnormalities including centrosome multiplication, defective midbody formation and completion of cytokinesis, appearance of binucleated cells, H4K16 and α-tubulin hyperacetylation, and impaired mitotic localization and activity of the SIRT2 deacetylase. We report that the presence of the ATAC complex is essential during mitosis to inhibit Cyclin A/Cdk2 activity by favouring Cyclin A degradation through acetylation. As the Cyclin A/Cdk2 kinase is essential for correct centrosome formation and inhibits SIRT2 function, our data positions the ATAC complex as an important regulator of mitosis, and thus uncovers an essential function for the ATAC acetyl transferase (AT) complex in cell division.

Results

Identification of the ATAC complex at the mitotic spindle

As Gcn5 has been implicated in cell cycle regulation (see Introduction), we aimed to investigate which of the two mammalian Gcn5-containing complexes, SAGA or ATAC, was involved in this function *in vivo*. Thus, we tested whether ablation of Spt20 (SAGA specific) and Ada2a (ATAC specific) by RNAi would affect normal cell division rates (Figure 1A–C). The indicated subunits were knocked down by either transfecting a mixture of four specific siRNAs against the respective mRNAs (Spt20 or Ada2a) into mouse NIH3T3 cells, or by transfecting different small hairpin (sh) DNA constructs targeting Spt20 or Ada2a into human HeLa or 293T cells. First, the efficiency of the depletion was verified and the specific effects were compared with a mixture of non-targeting siRNAs (Mock) (Supplementary Figure 1). As Spt20 or Ada2a knockdown (KD) was efficient and specific in the different cellular systems used, we tested whether the ablation of either SAGA or ATAC function would influence cell division. To this end, we scored the number of cells undergoing mitosis using time-lapse microscopy. The initial number of cells on the image field was considered the 'Total' cell number, and the cell cycle of each cell was followed for 30 h. For each cell, we determined whether it was (1) dividing properly or (2) displaying cell division defects (such as asymmetrical, delayed or failed division) and (3) multinucleation. Depletion of Ada2a led to reduced number of cells undergoing proper division ($<50\%$), as these cells show several defects to complete mitosis. On the contrary, depletion of Spt20 had no significant effect (Figure 1A; Supplementary Figure 2). Consistent with this, depletion of Ada2a, but not that of Spt20, lead to increased mitotic abnormalities (delayed, asymmetric or incomplete cell divisions) and concomitantly, increased population of bi- or multinucleated cells (Figure 1B and C; Supplementary Figure 2). These data suggest that the ATAC complex is required for proper cell cycle progression in mammalian cells.

To define the cellular events in which ATAC is involved during cell cycle and to compare it with SAGA, we investigated the localization of these two complexes along different stages of the cell cycle in mouse fibroblasts using immunofluorescence labelling. Interestingly, ATAC specific subunits (Ada2a and Yeats2) and subunits of ATAC that are also present in SAGA, localized to the mitotic spindle (Gcn5 and Ada3) (Figure 1D, panels g–j; Supplementary Figure 3B). In contrast, SAGA-specific subunits, such as Spt20 or Usp22, were excluded from the chromatin and the mitotic spindle during mitosis (Figure 1D, panels k and l; Supplementary Figure 3A). These observations suggest that the whole ATAC complex localizes to the mitotic spindle. Note that in interphasic cells all the antibodies used gave a nuclear staining for the tested factors (Figure 1D, panels a–f).

To confirm that the observed specific localization corresponds to the ATAC complex, we compared the composition of ATAC in asynchronized and G2/M synchronized cells. Cells were either non-treated or synchronized with nocodazole and cell extracts prepared. From both cell extracts, ATAC complexes were immunopurified using three different antibodies against ATAC-specific subunits and the ATAC composition was then verified by western blot (WB) analysis (Figure 1E). The fact that no differences were detected between the compositions of the immunoprecipitated complexes prepared from non-synchronized (A) or mitotic cells (M) suggested that the ATAC complex does not dissociate during mitosis. This is in good agreement with the immunofluorescence experiments (Figure 1D; Supplementary Figure 3B). Consistent with these observations, ATAC subunits, such as Ada2a and Ada3 (hereafter Ada2a/3), co-localized during all mitotic stages (Figure 2A, panel g; Figure 2C, panels g–i). In addition, the KD of either Ada2a or Ada3 impaired the localization of Ada3 and Ada2a, respectively (Figure 2A, panels b–f). Finally, KD of Ada2a or Ada3 resulted in the dissociation of the Gcn5-HAT subunit from the complex (Figure 2B, lanes 4 and 8, for KD efficiency see Supplementary Figure 1A–C). Altogether, our data indicate that the ATAC complex requires its integrity to localize to the mitotic spindle.

To get a more precise description of the localization of ATAC subunits to the mitotic spindle, we compared the localization of Ada2a/3 to that of the microtubule network (exemplified by α-tubulin) and members of the chromosome passenger complex (CPC) (exemplified by Aurora B). These two markers were selected because of their important functions during cell division: the microtubule network, because it provides the pulling force for chromosome segregation (Dumont and Mitchison, 2009), and the CPC, because it is controlling several mitotic features ranging from chromosome-microtubule attachment to cytokinesis (Ruchaud *et al*, 2007). Interestingly, from early mitotic stages to anaphase, the ATAC subunits Ada2a and Ada3 strongly co-localized with the microtubule network, as exemplified by the α-tubulin staining (Figure 2C, panels p–r), but not with CPC components, such as Aurora B (Figure 2C, panels y, z, aa). These data suggest that during mitosis the ATAC complex is mainly associated with the microtubule network. The co-localization of Ada2a/3 with α-tubulin is restricted to mitosis, as during interphase, all ATAC subunits analysed were predominantly nuclear (Figure 1D, panels a–d), whereas the microtubule network is cytoplasmic.

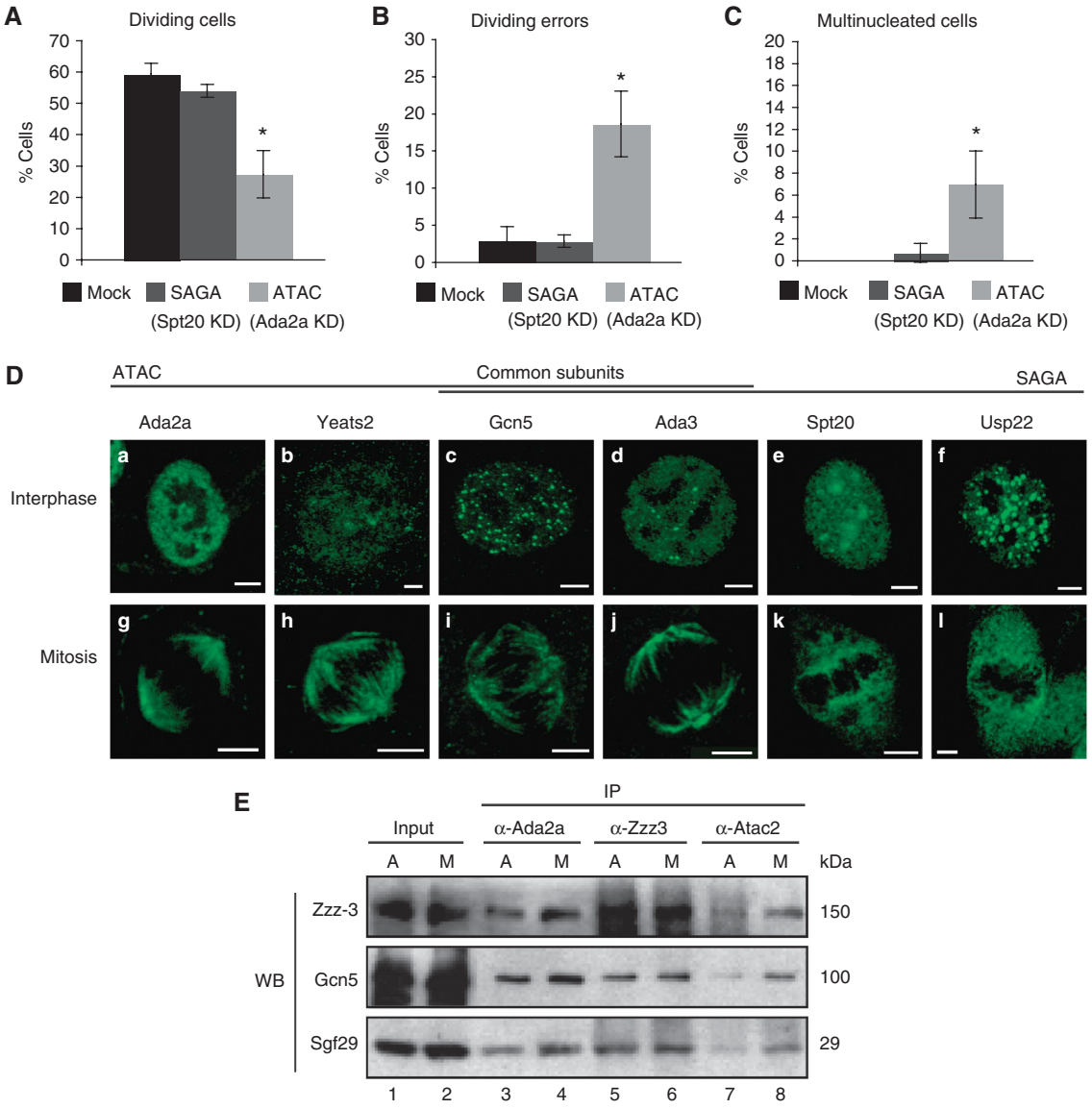


Figure 1 Identification of ATAC as a potential mitotic regulator. (**A**–**C**) Cell division analysis of NIH-3T3 cells transfected with control siRNA (Mock) or siRNA to achieve either SAGA (Spt20)- or ATAC (Ada2a)-specific subunit knockdowns (KDs). siRNA-transfected cells in prophase were randomly chosen and video imaged for 30 h. See Supplementary Figure 2 for representative images from the time-lapse analysis. The initial number of cells on the image field was considered the 'Total' cell number, and the life cycle of each cell was followed. For each cell, we determined whether it was (**A**) undergoing normal cell division, (**B**) displaying division errors (delayed, asymmetric or uncompleted cell divisions), or (**C**) multinucleation. The number of cells in (**A**), (**B**) or (**C**) is expressed as percentage of total cells analysed (>100 cells/condition, mean and s.d. from four independent experiments, $*P<0.05$). (**D**) Interphasic and mitotic localization of endogenous ATAC or SAGA subunits in mouse NIH-3T3 cells. Cells were stained with anti-ATAC (Ada2a, Yeats2)-, anti-ATAC/SAGA (Gcn5, Ada3)- or anti-SAGA (Usp22, Spt20)-specific antibodies, as indicated. Scale bars: 4 μm. (**E**) Composition of the ATAC complex in asynchronous (A) or mitotic (M) cell cultures. Cell extracts were prepared and immunoprecipitation (IP) of the endogenous ATAC complex was carried out using antibodies against the indicated subunits from the two cell populations. A total of 10% of the input extracts (Input) and the immune pellets were analysed by western blot (WB) using antibodies as indicated.

Depletion of the Ada core of ATAC leads to mitotic abnormalities

The specific positioning of ATAC subunits to the mitotic spindle prompted us to further investigate the function of ATAC at different mitotic stages. We explored how cells progressed through the cell cycle after G2/M arrest on Ada2a or Ada3 RNAi. Consistent with the above results, cells depleted for Ada2a or Ada3 showed a delayed M/G1 transition after nocodazole synchronization (Figure 3A). In agreement with this prolonged mitotic phase after Ada2a or Ada3 depletion, we also observed an increase in the mitotic index of these cells, which was measured on the basis of cells positive for mitotic-specific histone marks such as H3S10P and H4K20me1 (Figure 3B) (Wei *et al*, 1998; Oda *et al*, 2009). Moreover, visualization of centrosomes by γ-tubulin and chromatin by Hoechst staining in the Ada2a/3-depleted cells showed a drastic induction of centrosome multiplication, with a four-fold increase in cells containing more than two centrosomes (Figure 3C). Furthermore, imaging of α-tubulin or chromatin in these cells revealed that the ablation of Ada2a and Ada3 lead to the appearance of aberrant midbodies, which are thicker than those in the

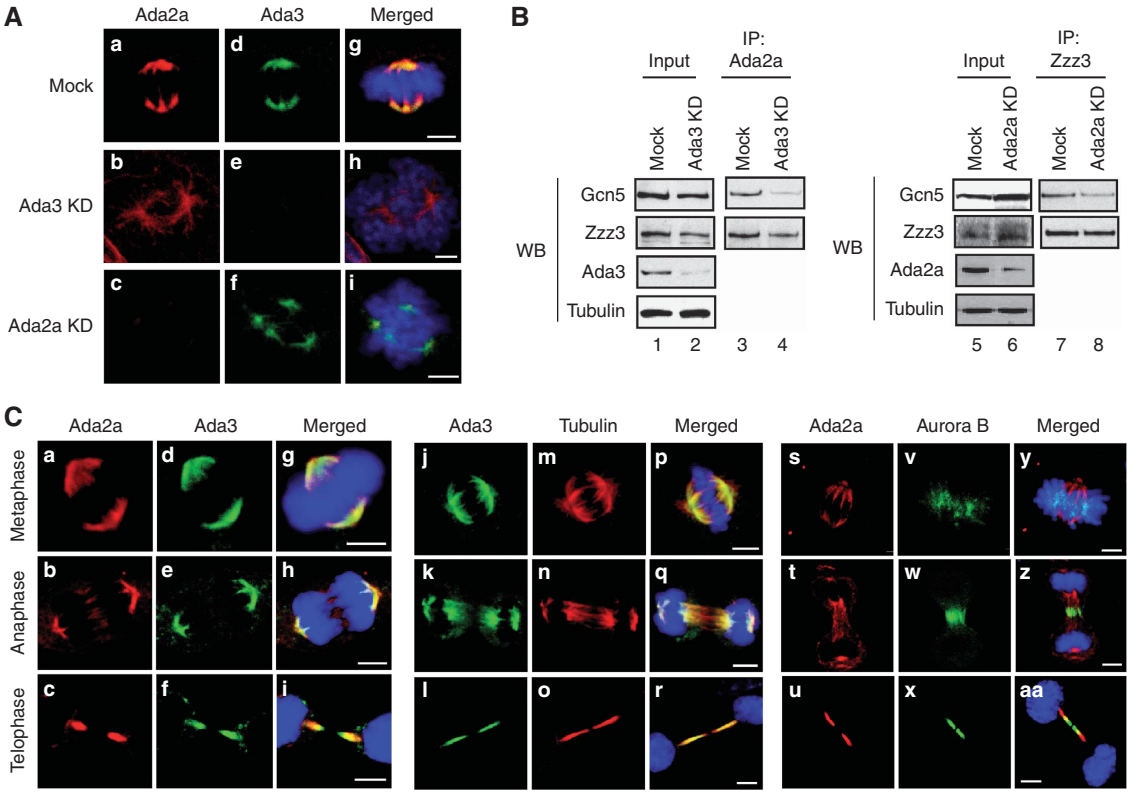


Figure 2 The ATAC complex stays associated during mitosis and colocalizes with the microtubule network. (**A**) Knock down (KD) of endogenous Ada2a (panels c, f, i) and Ada3 (panels b, e, h) protein expression after RNAi was verified by immunofluorescence and compared with cells transfected with a non-targeting control siRNA (Mock, panels a, d, g). The antibody used to label the cells is indicated at the top of the panels. Images are representative of $n>3$ independent experiments. Scale bars: 4 μm. (**B**) Composition of the ATAC complex on Ada3 and Ada2a knockdown. Cells were synchronized in G2/M phase with nocodazole and immunoprecipitation (IP) of the endogenous ATAC complex was carried out using an anti-Ada2a or anti-Zzz3 antibody, as marked. A total of 10% of the input extracts (Input) and the immune pellets were analysed by western blot (WB) using the antibodies indicated on the left of the panels. (**C**) The co-localization of Ada2a with Ada3 (panels a–i), Ada3 with α-tubulin (panels j–r) and Ada2a with AuroraB (panels s–aa) was tested by immunofluorescence in metaphase, anaphase and telophase cells (marked on the left). The merged images show the specific localization of the tested factors together with DNA (blue). Images are representative of at least three independent experiments. Scale bars: 4 μm.

control cells and twisted in shape, as well as to shortened distances between the dividing nuclei (Figure 3D). All these data show that cells lacking Ada2a or Ada3 are unable to proceed properly along cytokinesis and suggest that the mammalian ATAC complex is required for correct cell division, by affecting both initial and late stages of mitosis.

Depletion of the Ada core of ATAC leads to H4K16 and tubulin hyperacetylation

In metazoan organisms, the function of the ATAC complex has been associated with acetylation of histones (see Introduction), particularly H4K16. As the H4K16Ac mark is known to prevent chromatin compaction (Shogren-Knaak *et al*, 2006), the need of a functional ATAC complex during mitosis seems paradoxical with the notion that histone hypoacetylation is a prerequisite for mitotic chromosome formation (Kruhlak *et al*, 2001; Cimini *et al*, 2003; Valls *et al*, 2005) (Supplementary Figure 4). We thus examined whether global levels of acetylation on specific histone residues were affected in mitotic cells upon RNAi for Ada2a/3. Unexpectedly, knocking down the Ada2a or Ada3 components of the ATAC complex resulted in a specific increase of H4K16 acetylation (Figure 4A, panels d–l; Figure 4C), whereas H3K9, H3K14, H4K5 and H4K12 acetylation levels remained unchanged (Supplementary Figure 5). The hyperacetylation of H4K16 suggests a defective chromosome condensation following Ada2a/3 KD. As we identified ATAC subunits at the mitotic spindle together with α-tubulin, we also explored the state of α-tubulin acetylation under the same RNAi conditions. A similar hyperacetylation was observed on α-tubulin, whose mitotic function is also critically regulated by its acetylation (Zilberman *et al*, 2009 and refs therein). Thus, RNAi of either Ada2a or Ada3 induced hyperacetylation of both H4K16 and α-tubulin, which would consequently affect chromatin architecture and the tubulin network (Figure 4A and B, panels d–l; Figure 4C). Note, however, that the KD of the SAGA-specific subunit Spt20 did not result in hyperacetylation of α-tubulin (Supplementary Figure 6).

Earlier observations suggested that the Gcn5-HAT activity in the ATAC complex may be positively regulated by the Ada2a and Ada3 subunits that are directly interacting with Gcn5 (Balasubramanian *et al*, 2002; Gamper *et al*, 2009). Consistent with this, we observed that Gcn5 dissociates from the ATAC complex after Ada2a/3 KD (Figure 2B). Therefore, the increased acetylation of H4K16 and α-tubulin on Ada2a or Ada3 depletion may not be the result of an increased AT activity of ATAC. Rather, ATAC might affect other substrates

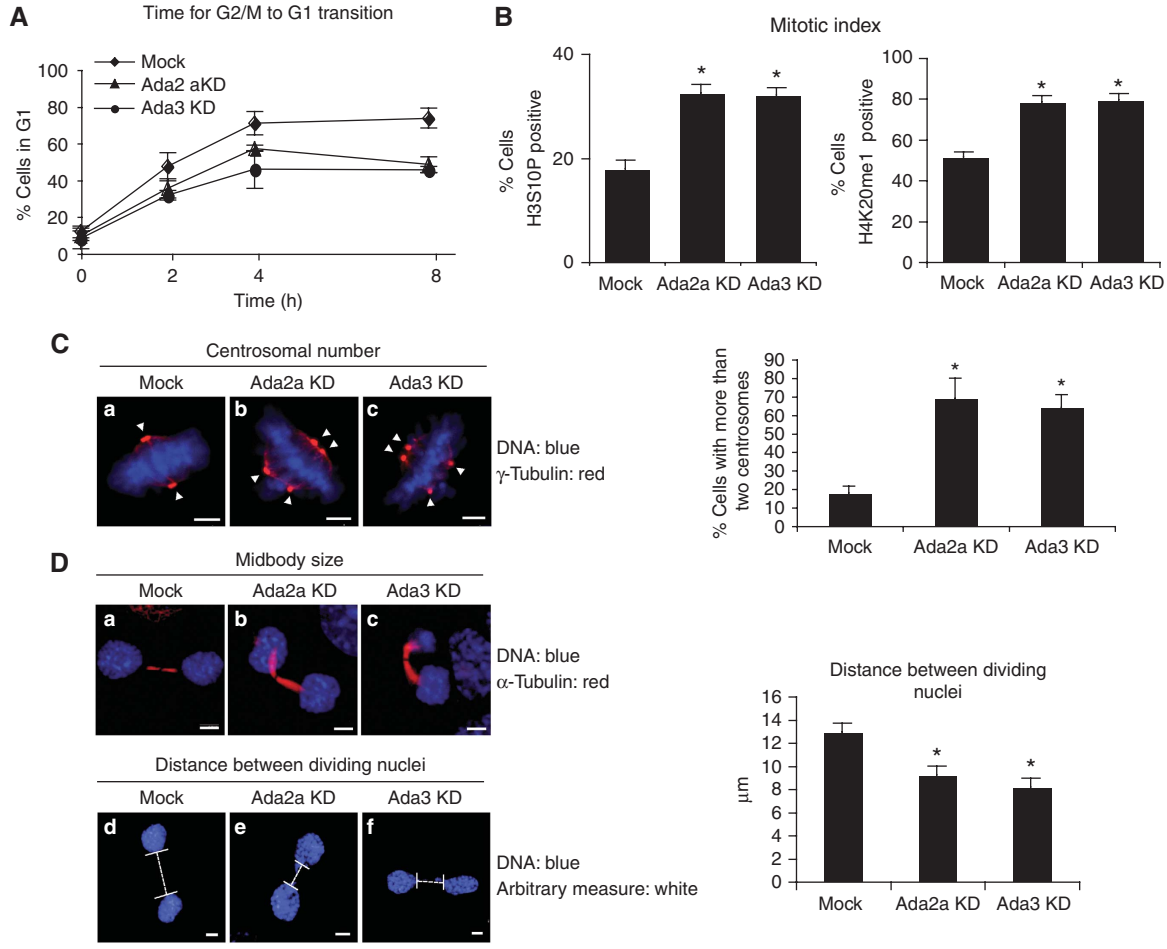
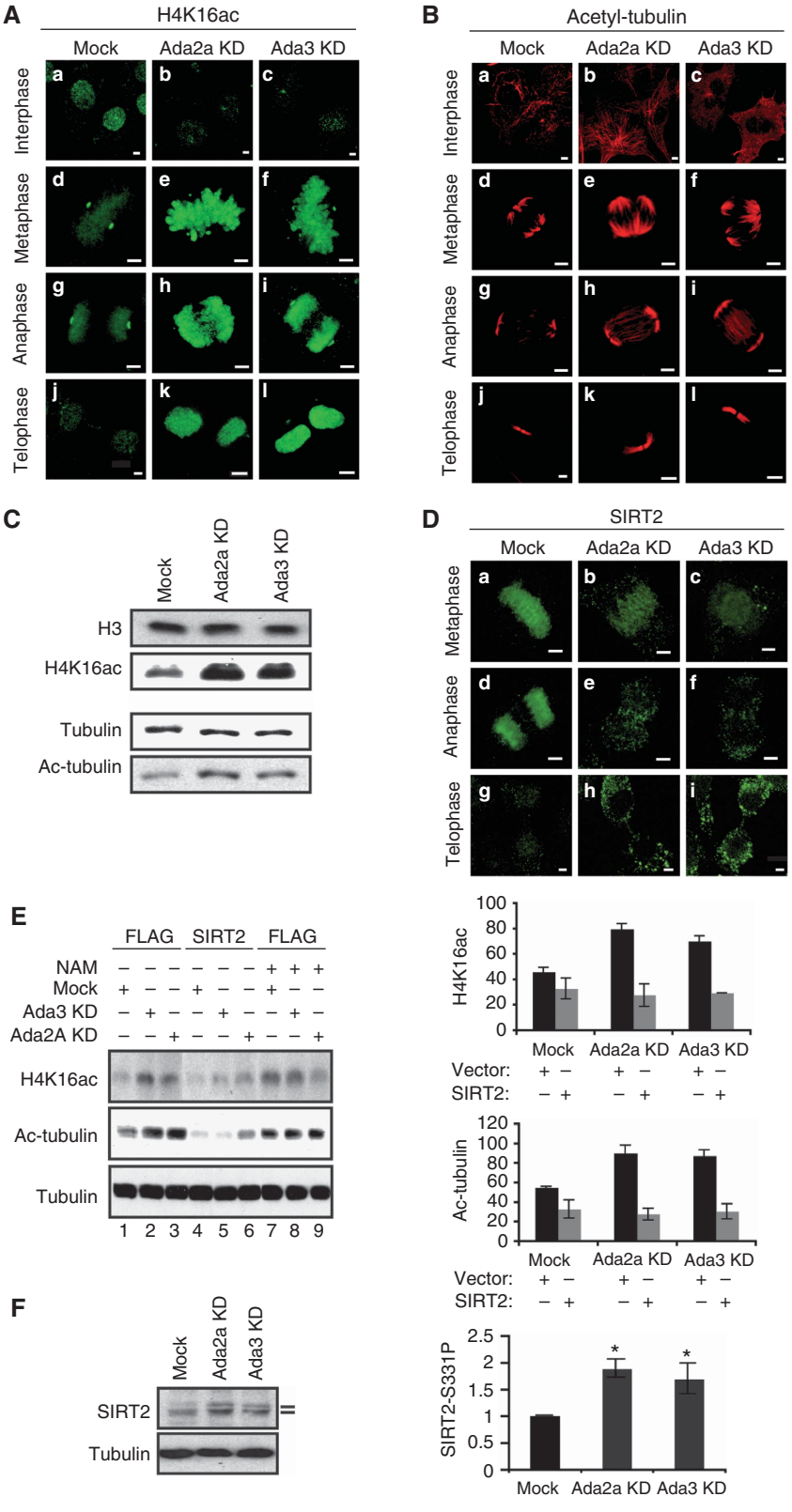


Figure 3 ATAC-depleted cells display delayed mitosis and defects in cytokinesis. NIH-3T3 cells were transfected with non-targeting control siRNA (Mock) or siRNAs against Ada2a or Ada3. After 72–96 h of transfection, cells were further analysed. (**A**) The different siRNA-treated cells were treated overnight with nocodazole to arrest them in G2/M and then released. After release, cells were harvested at different time points (as indicated) and the G1 population was monitored by FACS analysis. The results are expressed as percentage of total cells analysed (mean and s.d. from four independent experiments). (**B**) To analyse the mitotic index of siRNA-transfected cell populations, cells were stained with an anti-H3S10P and an anti-H4K20me1 antibody, and the DNA was visualized by Hoechst staining. The number of cells positive for the two marks were determined by immunofluorescence for each condition (>400 cells/condition, mean and s.d. from $n>4$ independent experiments, $^*P<0.05$). (**C**) To visualize centrosomes in siRNA-transfected mitotic cells, the cells were stained with an anti-γ-tubulin antibody and the DNA was visualized by Hoechst staining. Centrosome numbers were determined for each condition (>200 cells/condition, mean and s.d. from $n>4$ independent experiments, $^*P<0.05$). (**D**) To visualize the microtubule network in siRNA-transfected mitotic cells, cells were co-stained with an anti-α-tubulin antibody and Hoechst (DNA). A line (dashed-white line), placed on the basis of the position of the midbodies (α-tubulin staining), was used for measuring the distance between the daughter cells (>100 cells/condition, mean and s.d. from $n>3$ independent experiments, $^*P<0.05$).

that consequently lead to changes in the acetylation status of H4K16 and α-tubulin. Interestingly, the mammalian HDAC SIRT2 mediates both mitotic H4K16 and α-tubulin deacetylation (North *et al*, 2003; Vaquero *et al*, 2006, 2007), suggesting that it may be a potential target of the ATAC complex. SIRT2 belongs to the class III HDAC enzymes (sirtuins), which require nicotinamide (NAM) adenine dinucleotide [NAD(+)] for catalysis. SIRT2 resides in the cytoplasm during interphase, but at the onset of mitosis relocates to the nucleus, in which it deacetylates its substrates thereby ensuring mitotic progression (Vaquero *et al*, 2006, 2007). Thus, next we analysed the SIRT2 localization in cells in which ATAC subunits have been knocked down. In Ada2a/3-depleted cells, SIRT2 did not localize properly to the chromatin during mitosis (Figure 4D) that can explain the observed increase in the levels of acetylation of H4K16 and α-tubulin. To confirm our hypothesis, we tested whether overexpression of SIRT2 could rescue the deacetylation of H4K16 and α-tubulin in Ada2a/3-depleted cells. Indeed, overexpression of wild-type SIRT2 rescued both H4K16 and α-tubulin deacetylation in Ada2a/3-depleted cells (Figure 4E, compare lanes 1–3 with 4–6). In contrast, overexpressing the catalytically inactive SIRT2-H150Y mutant did not restore the H4K16 and α-tubulin deacetylation (Supplementary Figure 7) (Pandithage *et al*, 2008), further suggesting that the HDAC activity of SIRT2 is needed for this function. Consistent with this, inhibition of endogenous SIRT2 activity in cells with NAM (Bitterman *et al*, 2002) mimicked the effects of Ada2a/3 RNAi on H4K16 and α-tubulin acetylations (Figure 4E, compare lane 1 with lanes 2, 3 and 7, and see quantification on the right of the panel). Finally, *in vitro* SIRT2 tubulin deacetylase (TDAC) assays using overexpressed SIRT2 protein in Ada2a or Ada3 KD cell backgrounds further showed that the deacetylase activity of SIRT2 is impaired when the ATAC complex is lacking the Ada core (Supplementary Figure 8A). These observations together indicate that the

increased H4K16 and α-tubulin acetylation caused by KDs of Ada2a/3 is due to reduced/mislocalized SIRT2 activity and not because of increased activity of ATAC. To prove our hypothesis that SIRT2 is the major deacetylase involved in the observed phenotype, we aimed to exclude the potential contribution from the principal α-TDAC in the cells, HDAC6 (Hubbert *et al*, 2002; Zhang *et al*, 2003). To this end, we first tested the *in vitro* HDAC6 enzymatic activity in TDAC assays by using overexpressed HDAC6 protein in Ada2a or Ada3 KD cell backgrounds. This experiment showed that HDAC6 is fully active in the absence of the Ada core of the ATAC complex (Supplementary Figure 8B). Next, we



inhibited the endogenous activity of HDAC6 with TSA (Minoru *et al*, 1995; Furumai *et al*, 2001) and compared the levels of α-tubulin acetylation to those observed on either NAM treatment or Ada2a/3 depletion. TSA treatments led to a more than 10-fold increase on α-tubulin acetylation levels, clearly larger than the 2.5-fold increase observed on either NAM treatment or Ada2a/3 depletion (Supplementary Figure 8C; Figure 4E, compare lanes 1–7). This result is consistent with the notion that HDAC6 is indeed the main α-TDAC acting in the cells. Furthermore, as earlier tested for SIRT2, we examined whether overexpression of HDAC6 could rescue the deacetylation of α-tubulin in Ada2a/3-depleted cells. Contrary to the SIRT2 overexpression experiment (Figure 4E), exogenous HDAC6 expression did not restore α-tubulin acetylation levels in Ada2a/3 backgrounds (compare Supplementary Figure 8D, lanes 2–6 with Figure 4E, lanes 2–6).

Overall, these results show that the hyperacetylation of H4K16 and α-tubulin observed on depletion of Ada2a or Ada3 is due to inefficient SIRT2 deacetylase activity. Furthermore, our data indicate that the ATAC complex modulates SIRT2 activity during mitosis, placing these two opposite enzymatic activities as parts of the same regulatory pathway.

ATAC regulates the phosphorylation state of SIRT2

SIRT2 activity can be regulated through post-translational modifications on several residues, among which S331 phosphorylation has been shown to partially block its HDAC activity (North and Verdin, 2007; Pandithage *et al*, 2008). Interestingly, after Ada2a/3 KD, a slower SIRT2 migrating form appeared on WB, suggesting that ATAC could regulate SIRT2 phosphorylation (Figure 4F, upper panel). The use of SIRT2-S331P-specific antibodies (Pandithage *et al*, 2008) confirmed that SIRT2 was phosphorylated on S331 in cells in which either Ada2a or Ada3 was depleted (Figure 4F, right panel). Expression of an SIRT2 mutant in which the S331 residue is mutated to alanine (S331A) was enough to prevent the hyperacetylation effects of Ada2a/3 KD on α-tubulin (Supplementary Figure 7). In contrast, overexpression of a SIRT2-S331E phospho-mimicking mutant phenocopied the effect of the Ada2a/3 KD, suggesting that Ada2a and Ada3 would indirectly regulate SIRT2 activity through changes in its phosphorylation status (Supplementary Figure 7).

ATAC acetylates Cyclin A

The inhibitory phosphorylation of SIRT2 on S331 is catalysed by the Cyclin A/Cdk2 complex (Pandithage *et al*, 2008). As ATAC had no detectable kinase activity on SIRT2 (data not shown), we next explored whether ATAC might indirectly affect SIRT2 phosphorylation, and thus its activity, by regulating Cyclin A/Cdk2 function through its acetylation. On the basis of the specific localization of Gcn5 to the spindle during mitosis (Figure 1D, panel i), we first checked whether this enzyme could acetylate Cyclin A/Cdk2. Using *in vitro* acetylation assays, we could detect that recombinant Gcn5 acetylates recombinant Cyclin A alone or in the context of the Cyclin A/Cdk2 complex (Figure 5A, lanes 11 and 12). However, as the ATAC complex contains other acetyltransferase activities such as Atac2 or Pcaf (see Introduction), we also tested these enzymes as potential Cyclin A acetyltransferases, along with Gcn5 and a catalytically inactive form of Gcn5 (rGCN5-ΔHAT) (Bu *et al*, 2007). As earlier described, the GCN5-ΔHAT mutant does not acetylate histone H3 and H4 peptides *in vitro* (Supplementary Figure 9A). We observed that Cyclin A is acetylated by Gcn5 and Pcaf, in agreement with a recent report (Mateo *et al*, 2009) (Figure 5B, lanes 8 and 9). On the contrary, neither the catalytically inactive Gcn5 (rGCN5-ΔHAT) nor Atac2 acetylated Cyclin A *in vitro* (Figure 5B, lanes 7 and 10), which further indicates that only the paralogue HATs, Gcn5 and Pcaf can acetylate Cyclin A. Next, we tested whether these enzymes could also mediate Cyclin A acetylation when incorporated in their respective endogenous complexes (ATAC or SAGA). To this end, we obtained highly pure ATAC or SAGA complexes from HeLa cells (Figure 5C) and used them to carry out *in vitro* acetylation assays as before. This experiment shows that only the ATAC complex mediates Cyclin A acetylation (Figure 5D, compare lane 10 with 11 and 12). Therefore, this result indicates that Cyclin A is a target for the acetyltransferase activity of ATAC and that the activity of the Cyclin A/Cdk2 kinase complex may be controlled by acetylation in the cells. To test whether Gcn5 can acetylate Cyclin A in a cellular context, cells were transfected with expression vectors for either Flag-Cyclin A together with Flag-Gcn5 wild type (Gcn5wt), or with a Flag-Gcn5 HAT-defective mutant (GCN5-ΔHAT). After transfection, an anti-Flag immunoprecipitation (IP) was carried out and the acetylation status of Cyclin A was analysed by using an antibody against Ac-Lys. We detected acetylated Cyclin A only in the wt Gcn5-expressing cells indicating that Cyclin A is a target of the Gcn5-HAT activity (Figure 5E).

Identification of Cyclin A residues targets for Gcn5 acetylation

In a recent study, lysines K54, K68, K95 and K112 of Cyclin A were identified as potential *in vitro* target sites for Pcaf

Figure 4 ATAC regulates the mitotic function of the SIRT2 HDAC. (**A–C**) Ada2a/3 depletion causes H4K16 and α-tubulin hyperacetylation. NIH-3T3 cells transfected with control (mock), Ada2a or Ada3 siRNA were visualized by immunofluorescence along mitosis using either (**A**) anti-H4K16ac or (**B**) anti-Acetyl-α-tubulin antibodies. (**C**) Western blot (WB) analysis of mitotic whole cell extracts from Ada2a/3 siRNA-transfected cells (indicated on the top) using the antibodies as indicated on the left. (**D**) Ada2a/3 KD disturbs the mitotic positioning of SIRT2. NIH-3T3 cells were transfected with the indicated siRNAs and mitotic cells were visualized using the anti-SIRT2 antibody in immunofluorescence. (**E**) SIRT2 overexpression restores H4K16 and α-tubulin deacetylation, whereas the inhibition of SIRT2 mimics the KD of either Ada2a or Ada3. The 293T cells were co-transfected with an empty vector (FLAG) or an expression vector for SIRT2-FLAG (SIRT2), and DNA constructs expressing shRNAs against Ada2a (Ada2a KD), Ada3 (Ada3 KD) or a scramble shRNA (Mock) as indicated on the top of the panel. Transfected cells were either untreated or treated O/N with 5 mM nicotinamide (NAM). After the indicated transfections and treatments, WCE was prepared from the cells and analysed by western blot (WB) with the indicated antibodies (left panel). Panels on the right: the results of the quantification of $n>10$ independent experiments by densitometry ($^*P<0.05$). (**F**) Ada2a or Ada3 knockdown increases the phosphorylation of SIRT2 on S331. NIH-3T3 cells were treated with control (mock), Ada2a or Ada3 siRNAs for 48 h. Cell extracts were prepared after transfection and analysed by WB using an anti-SIRT2 antibody (left panel). Endogenous SIRT2 was immunoprecipitated from control and Ada2a or Ada3 KD 293T cells and analysed by an antibody that recognizes SIRT2 in general or specifically recognizing SIRT2 phosphorylated on S331P. WBs obtained with the phospho-specific antibody were quantified by densitometry. Indicated fold changes are mean from $n=4$ independent experiments, $^*P<0.05$.

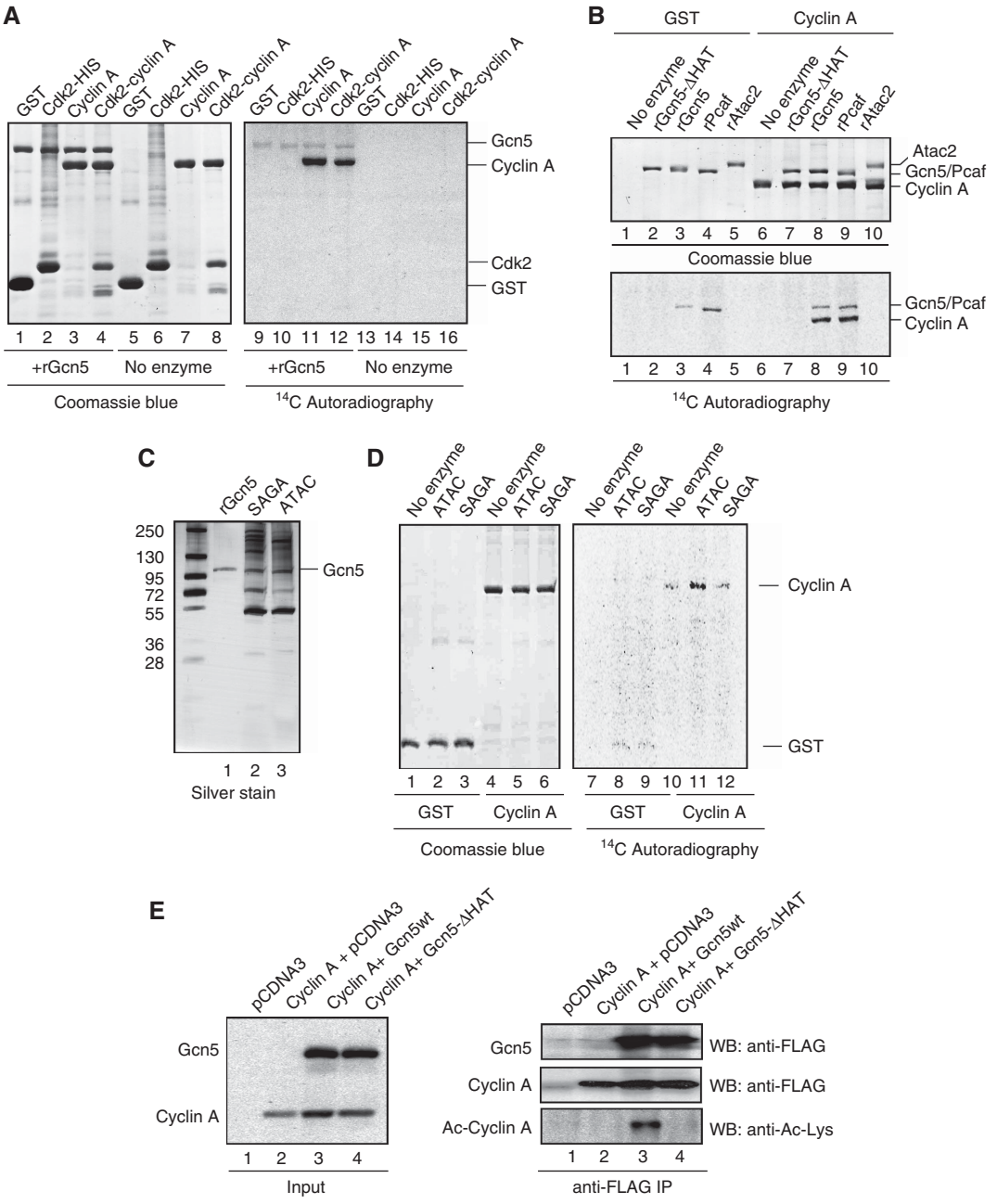


Figure 5 The ATAC complex acetylates Cyclin A/Cdk2 through Gcn5 or Pcaf. (**A**) Gcn5 acetylates the Cyclin A/Cdk2 complex *in vitro*. *In vitro* acetylation assays have been carried out using purified recombinant (r) Cdk2 and Cyclin A as substrates, and rGcn5 as enzyme. Reactions were separated by SDS–PAGE, visualized by coomassie blue staining (left panel) and labelled proteins were visualized by autoradiography (right panel). (**B**) Gcn5 and Pcaf, but not Gcn5-ΔHAT or Atac2, efficiently acetylate Cyclin A *in vitro*. *In vitro* acetylation assays have been carried out using purified recombinant (r) Cyclin A as substrate, and rGcn5, rGcn5-ΔHAT, rPcaf or rAtac2 as enzyme. Reactions were separated by SDS–PAGE, visualized by coomassie blue staining (upper panel) and labelled proteins were visualized by autoradiography (lower panel). (**C**, **D**) The ATAC complex, but not SAGA, acetylates Cyclin A *in vitro*. *In vitro* acetylation assays have been carried out using purified recombinant (r) Cyclin A as substrate, and the purified hATAC or hSAGA complexes as enzymes. (**C**) Shows a silver staining for the respective ATAC and SAGA purifications, with equivalent Gcn5 content. (**D**) Shows the acetylation experiment performed with the purified ATAC and SAGA complexes: reactions were separated by SDS–PAGE, visualized by coomassie blue staining (left panel) and labelled proteins were visualized by autoradiography (right panel). (**E**) Gcn5 acetylates Cyclin A *in vivo*. 293T cells were transfected with an empty vector (pCDNA3) (lane 1), or co-transfected with FLAG-Cyclin A and pCDNA3 (lane 2); FLAG-Cyclin A and FLAG-Gcn5 wild type (Gcn5wt) (lane 3), or Cyclin A and FLAG-Gcn5 mutated in the HAT domain (Gcn5-ΔHAT) (lane 4). WCEs from the transfected cells was prepared and analysed by western blot (WB) with the indicated antibodies (left panel). FLAG-tagged proteins were immunoprecipitated from WCEs, eluted with FLAG peptide and analysed with an antibody recognizing acetyl-lysines, Cyclin A and Gcn5 (as indicated).

acetylation (Mateo *et al*, 2009). These four lysine residues are located on the N-terminal domain of Cyclin A, in the so-called canonical degradation (D)-box (positions 46–63) or the extended D-box (65–82) (Tin Su, 2001 and references therein). Interestingly, these domains have been implicated in regulating the stability of the protein (Wolthuis *et al*, 2008) (Figure 6A). Moreover, it has been reported that the replacement of K54 and K68 of Cyclin A by arginines stabilizes

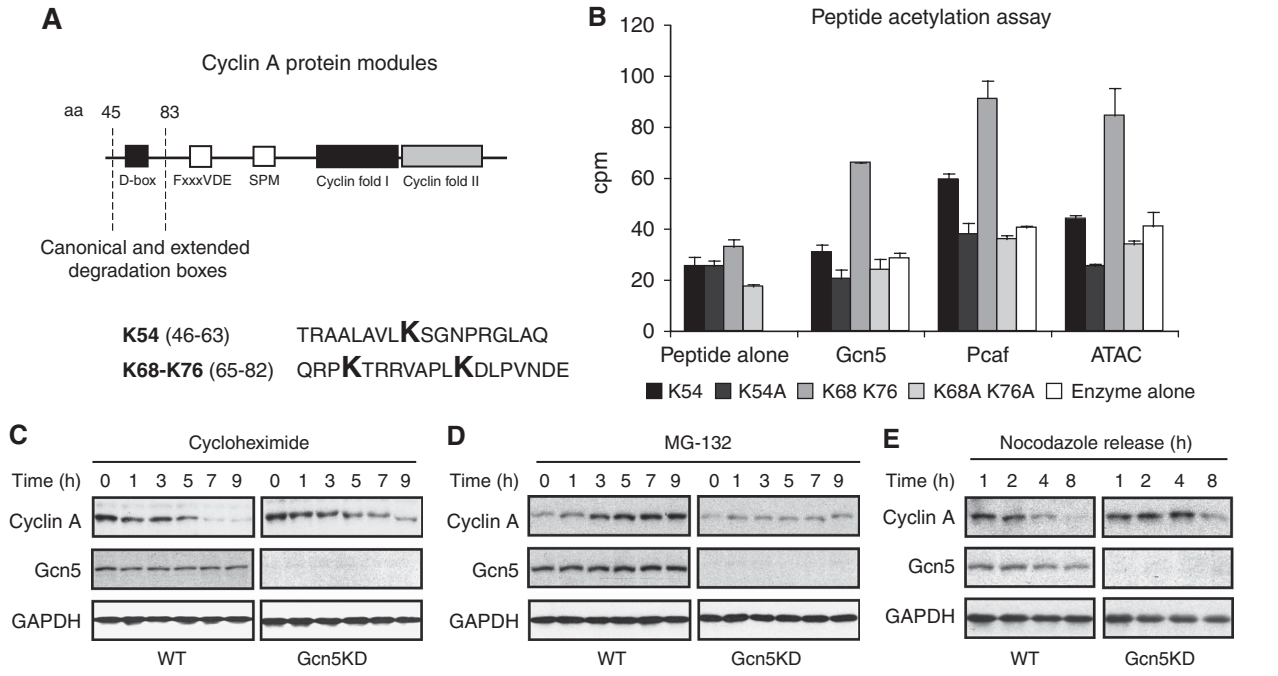


Figure 6 The ATAC complex regulates Cyclin A stability through acetylation. (**A**) Upper panel: schematic representation of the Cyclin A protein and its characteristic domains. The amino-acid (aa) positions of the canonical (D-box) and the extended degradation boxes are indicated. Lower panel: summary of the peptides synthetized to be used as substrate in the peptide acetylation assays shown in (**B**). (**B**) Peptide acetylation assays were performed using rGcn5, rPcaf or the ATAC complex as enzymes, and the peptides containing potential lysine target sites [wt (K) or mutated to Alanine (A)] as indicated. (**C**, **D**) Gcn5 acetylation promotes Cyclin A degradation *in vivo* through the proteasome. Inducible HeLa shRNA (Luc-KD or Gcn5-KD) cell lines maintained in the presence of doxicyclin were left untreated or exposed to cycloheximide (**C**) or MG-132 (**D**) for different time points (as indicated). Cells were collected, WCEs prepared and analysed by WB with the indicated antibodies. (**E**) Gcn5 acetylation promotes Cyclin A degradation during mitosis. HeLa shRNA (Luc-KD or Gcn5-KD) cell lines maintained with doxicyclin, blocked at the G2 phase of the cell cycle with Nocodazole for 18 h and then released from this block for several hours as indicated. Cells were collected, WCEs prepared and analysed by WB with the indicated antibodies.

Cyclin A in an ubiquitin-independent manner (Fung *et al*, 2005). As our mass spectrometric analysis suggested that Gcn5 may acetylate Cyclin A at positions K54 and K68 positions (data not shown) and as these lysines are located in the D-box of Cyclin A, we generated two peptides covering these positions, and their corresponding mutant versions in which the lysines (K) were substituted with alanines (A) (Figure 6A). Using the wild-type or mutated peptides in *in vitro* peptide acetylation assays, we observed that either the recombinant rGcn5 or rPcaf enzymes or the ATAC complex acetylated the Cyclin A peptide (from amino-acid 65–82) containing K68 and K76, which corresponds to the extended D-box (Geley *et al*, 2001) (Figure 6B). In contrast, the enzymatic activities were much less efficient on the Cyclin A peptide (from amino-acid 46–63) containing K54, present within the canonical D-box (Figure 6B). Note that the corresponding mutant peptides showed only background acetylation levels. Importantly, an earlier study showed that the single deletion of the canonical D-box (positions 45–58) does not affect Cyclin A protein turnover. However, the deletion of both the canonical and extended D-box (position 47–83) leads to stable Cyclin A protein levels (Geley *et al*, 2001). Our present results together with earlier observations suggest that acetylation of Cyclin A on the extended D-box may serve as a regulatory mechanism for Cyclin A protein stability along the cell cycle, and concomitantly, regulate Cyclin A mitotic function through its degradation.

Gcn5 regulates Cyclin A stability through acetylation

The function of Cyclin A/Cdk2 is restricted to early mitotic stages as Cdk2 is inactivated and Cyclin A degraded when cells enter in prometaphase (den Elzen and Pines, 2001). Thus, we tested whether the acetylation of Cyclin A by Gcn5 would affect the stability of Cyclin A in the cells. To this end, we generated an inducible HeLa cell line in which Gcn5 expression could be knocked down by the doxicyclin inducible expression of a specific shRNA (http://tronolab.epfl.ch/). As a control, we generated a cell line with an inducible shRNA against Luciferase (Luc-KD). After 48 h of induction by doxicyclin treatment, Luc-KD or Gcn5-KD cells were treated with cycloheximide to block *de novo* protein synthesis and the Cyclin A levels at different time points were measured by WB analysis (Figure 6C). In wild-type cells, degradation of Cyclin A was completed after 5–7 h of cycloheximide treatment. In contrast, in Gcn5-KD cells, this degradation was clearly delayed, as after 9 h of cycloheximide treatment Cyclin A was still detectable. This result is in good agreement with our hypothesis that Gcn5 activity could regulate Cyclin A protein stability. To further confirm this observation, we performed the same type of experiment, but on cells transfected with expression vectors for Gcn5wt or for Gcn5-ΔHAT. After transfection, cells were treated with cycloheximide and Cyclin A levels measured by WB analysis (Supplementary Figure 9B). In non-transfected cells, degradation of Cyclin A started at 6 h of cycloheximide treatment, whereas in Gcn5-expressing cells, this degradation started

earlier (around 3 h of treatment), further confirming that Gcn5-mediated acetylation can trigger Cyclin A degradation. In contrast, in cells expressing the Gcn5-ΔHAT mutant, Cyclin A degradation started much later (around 9 h of treatment) further underlining the importance of the acetyltransferase activity of Gcn5 in Cyclin A degradation (Supplementary Figure 9B).

On the basis of these results and on the identification of the potential lysines as target of Gcn5 acetyltransferase activity within the extended D-box of Cyclin A, we next tested whether Gcn5-mediated Cyclin A degradation was proteasome dependent. For this aim, after 48 h of doxicyclin treatment, wild-type or Gcn5-KD cells were treated with MG-132 to block proteasome-mediated protein degradation, and Cyclin A levels were then measured by WB analysis at different time points of MG-132 treatment (Figure 6D). In wild-type cells, Cyclin A levels steadily increased over time, whereas in Gcn5-KD cells, Cyclin A levels remained unchanged. This result suggests that in the absence of Gcn5, Cyclin A is not targeted to proteasome degradation. This observation is in agreement with our working hypothesis that Gcn5-mediated acetylation on Cyclin A triggers Cyclin A degradation. Finally, we evaluated whether the mitotic degradation of Cyclin A was affected in the absence of Gcn5. Thus, we blocked both wt and Gcn5-KD cells at the G2 phase of the cell cycle by nocodazole treatment for 18 h, and then monitored Cyclin A protein levels by WB at different time points after releasing the block (Figure 6E). Consistent with our above results, we observed a delayed Cyclin A degradation on G2/M synchronization and release in Gcn5-KD cells, when compared with the wild-type cells (Figure 6E). These results together show that Gcn5 AT activity has a function in the regulation of the mitotic degradation of Cyclin A and thus the activity of the Cyclin A/cdk2 kinase. Our data show that Cyclin A is degraded in a Gcn5-dependent manner at the beginning of mitosis to ensure timely completion of cell division (Tin Su, 2001 and references therein).

Ada core of the ATAC complex regulates Cyclin A stability

Next, we tested whether depletion of the Ada2a/3 subunits of the ATAC complex could lead to changes in Cyclin A levels in the cells. To this end, we examined Cyclin A levels under Ada2a/3 KD conditions by immunofluorescence, either in late G2 or metaphase. In agreement with our above observations, in Ada2a/3-depleted cells, Cyclin A levels remained stable during late G2, suggesting that Cyclin A degradation at early mitotic stages did not occur (Figure 7A, compare panel a with c and e, or panel g with i and k). Next, we analysed endogenous Cyclin A levels by WB analysis in asynchronous, G1 or G2/M arrested cell populations. Asynchronous cultures showed increased Cyclin A levels after Ada2a/3 depletion (Figure 7B). After the release of control cells arrested in G1 or G2/M, Cyclin A levels steadily increased from G1 to G2 (Figure 7C, lanes 1–3), and were degraded on G2 to M transition (lanes 4 and 5 in Figure 7C). In contrast, Ada2a/3-depleted (KD) cells displayed higher and stable Cyclin A levels along G1–G2 (Figure 7C, lanes 1–3), and were unable to trigger Cyclin A degradation on mitosis (Figure 7C, lanes 4 and 5).

To test whether the increased Cyclin A protein levels on Ada2a/3 KD could be also due to transcriptional effects, we evaluated the impact of Ada2a KD on the transcription of the *Cyclin A* gene in cell populations synchronized at the G2 phase of the cell cycle (Figure 7D, left panels). After synchronization, control or Ada2a KD cells were collected and mRNA expression for *Cyclin A* or *Cyclin B* was measured. This experiment clearly showed that *Cyclin A* or *B* transcripts did not change on Ada2a KD (Figure 7D, right panel), further confirming our above results that the ATAC KD induced Cyclin A stabilization effects are taking place at the protein level. The lack of Cyclin A degradation on G2/M release in Gcn5-, Ada2a- and Ada3-depleted cells strongly suggests that ATAC-mediated acetylation contributes to Cyclin A degradation in wild-type conditions.

In summary, here we describe a new pathway in which the ATAC complex controls Cyclin A/Cdk2 and, indirectly, SIRT2 activity. The ATAC/Gcn5-mediated acetylation of Cyclin A targets it for degradation, which is indispensable for obtaining the non-phosphorylated form of SIRT2. This SIRT2 is consequently fully active and able to deacetylate its mitotic targets, H4K16 and tubulin.

Discussion

Although the function of HAT complexes in regulating chromatin structure and transcription activation is widely studied (Lee and Workman, 2007; Nagy and Tora, 2007), less is known about their non-histone substrates and the function they fulfil through acetylating other proteins than histones. In the present report, we describe that the Gcn5-containing ATAC complex localizes to the mitotic spindle in which it has an essential function in orchestrating the progression through mitosis. This is a specific feature of the ATAC complex that contrasts with the other Gcn5-containing complex, SAGA, showing a different localization and not being involved in mitosis. We show that cells display a number of mitotic abnormalities upon depletion of the Ada core of ATAC. Our observations suggest that in the absence of the Ada core, the Gcn5 AT, probably as a free protein or associated with a partial ATAC complex, acetylates inefficiently its mitotic substrates, and in turn globally impacts on mitotic progression.

Indeed, analysis of the phenotype arising on Ada2a/3 depletion reveals a number of defects in crucial stages of cell division. Altogether, Ada2a/3-depleted cells proceed with major difficulties through mitosis, by either dividing at very slow rates or asymmetrically. In the most severe cases, these cells fail to complete their division and generate multinucleated cells or cells that die soon after mitosis.

Depletion of Ada2a/3 critically affects both early and late mitotic states. Ada2a/3 KD leads to a four-fold increase in the number of cells possessing super-numeral centrosomes compared with control cells. Deregulated centrosome amplification has a major impact on cell division, as centrosome duplication is a prerequisite for proper bipolar spindle formation, correct chromosome segregation and symmetrical cell division (Heald *et al*, 1997). Importantly, centrosome duplication requires, among other activities, the function of the Cdk2 kinase in association with Cyclin A (Meraldi *et al*, 1999). In addition, Cyclin A/Cdk2 has been described to be essential to coordinate centrosomal and mitotic events (De Boer *et al*, 2008). Therefore, any modification of normal Cyclin A/Cdk2 activity would influence not only proper

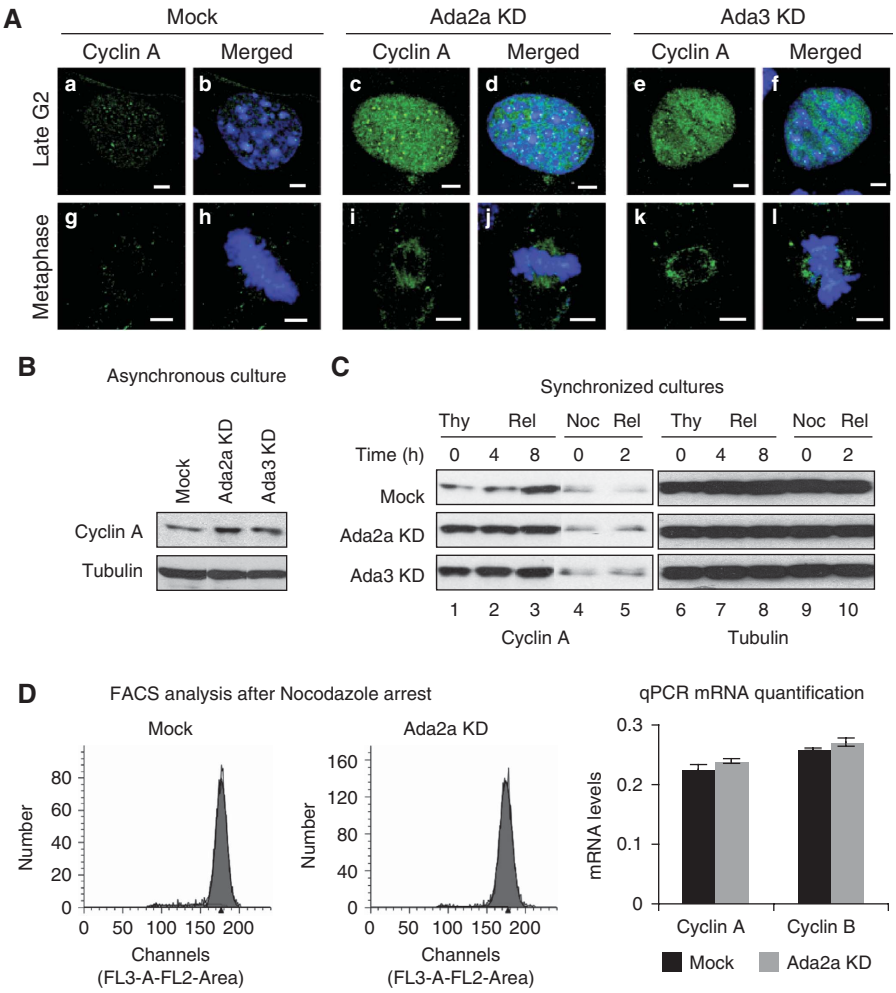


Figure 7 The Ada core of the ATAC complex regulates Cyclin A stability. (**A**) Mock or Ada2a/3 siRNA-transfected cells were visualized by immunofluorescence in late G2 and metaphase using an anti-Cyclin A antibody. Images are representative of three independent experiments. (**B**, **C**) WB analysis of Cyclin A levels in cell extracts prepared from wild-type or Ada2a and Ada3 KD cells. Asynchronous (**B**), G1 (thymidine: Thy) or G2/M (nocodazole: Noc) arrested cells (**C**) were released and Cyclin A protein level monitored using the indicated antibodies. (**D**) *Cyclin A* transcription is not affected after depletion of an ATAC subunit (Ada2a). Mock or Ada2a siRNA-transfected cells were blocked at the G2/M phase of the cell cycle by Nocodazole treatment for 18 h (as verified by the FACS analyses shown on the left). In the two left panels the x-axis represents the DNA content and the y-axis the cell numbers ($\times$ 1000). Cells were collected and their respective RNA extracted. This RNA was used for quantitative PCR analysis of the *Cyclin A* and *Cyclin B* transcripts for each condition (panel on the right). Mean and s.d. were calculated from three independent experiments.

centrosome duplication rates, but also centrosomal activities and mitotic events. Our results uncover the involvement of the ATAC complex, through Gcn5-mediated acetylation, in the regulation of Cyclin A stability and as a consequence, Cdk2 activity. Importantly, the depletion of Ada2a and Ada3 deregulates the enzymatic activity of ATAC, as the Ada core is essential to sustain Gcn5-HAT activity (Balasubramanian *et al*, 2002). Furthermore, we show that the depletion of the Ada2 or Ada3 subunits results in the disassembly of Gcn5 from the ATAC complex. Overall, the deregulated ATAC activity because of the Ada2a/3 KD would directly affect Cyclin A acetylation, stability and function, thus correlating with the observed centrosome abnormalities. Consistent with the aberrant centrosome multiplication, Ada2a/3-depleted cells manifest defects also in spindle and midbody formation, which in turn correlates with the observed difficulties to proceed through cytokinesis and complete cell division. Altogether, these abnormalities result in a time delay of mitosis and retarded entry into the G1 phase of the cell cycle.

In addition, Ada2a and Ada3 KD lead to changes in chromatin post-translational modifications, such an increase in H4K16 acetylation levels and a reduction of H3S10 phosphorylation levels (Ciurciu *et al*, 2008; Nagy *et al*, 2010 and data not shown). These two features oppose the normal mitotic scenario characterized by high H3S10P and low H4K16Ac levels required for normal chromatin condensation. Furthermore, we also show abnormal hyperacetylation of the microtubule network in ATAC-depleted cells. Microtubules act as pulling forces for chromosome segregation and their dynamics relies on hypoacetylated states, as α-tubulin acetylation marks static microtubule-based structures (Westermann and Weber, 2003; Hammond *et al*, 2008). Therefore, the observed hyperacetylation of α-tubulin would impair proper chromosome segregation and is consistent with the phenotype of binucleated cells that we describe. Thus, the absence of the Ada core of ATAC leads to inefficient chromosome compaction (because of H4K16 hyperacetylation) and increased microtubule stability (because of α-tubulin hyperacetylation), which together will impair mitotic progression.

It seemed striking that the depletion of an AT complex could lead to the hyperacetylation of two proteins whose deacetylated state is a prerequisite for mitotic progression: α-tubulin and H4K16. We solved this apparent paradox by showing that the effects observed in the absence of the Ada core of ATAC derive from impaired SIRT2 activity on these substrates. Consistent with this, in the absence of Ada2a and Ada3, SIRT2 overexpression is enough to restore normal H4K16 and α-tubulin deacetylation. Moreover, we also show that the Cyclin A/Cdk2 complex constitutes the link between the activities of ATAC and SIRT2, and that ATAC-mediated acetylation of Cyclin A determines the fate of the Cyclin A/Cdk2 complex by priming Cyclin A for degradation. These observations are in good agreement with recent findings describing that Cyclin A degradation could be triggered by acetylation (Mateo *et al*, 2009). Furthermore, in this report, we show that both Gcn5 and the ATAC complex mediate lysine acetylation in the extended D-box of Cyclin A and thus regulate Cyclin A degradation along the cell cycle. Consequently, any dysfunction of the ATAC complex will have an impact on Cyclin A activity. At the onset of mitosis, Cyclin A has to be degraded to ensure faithful mitotic progression. Indeed, an earlier report shows that after deletion of its extended D-box, Cyclin A becomes non-degradable, which leads to mitotic delay, as cells show anaphase arrest and difficulties to complete cytokinesis (Geley *et al*, 2001). Our study shows a situation that mimics the phenotype arising after deletion of the Cyclin A extended D-box, as depletion of Ada2a or Ada3 also leads to Cyclin A stabilization and difficulties for mitotic completion. Overall, our data illustrate a pathway in which, during mitosis, ATAC inhibits Cyclin A/Cdk2 function by promoting Cyclin A degradation through Gcn5-mediated acetylation. This consequently renders SIRT2 non-phosphorylated and fully active, inducing deacetylation of H4K16 and α-tubulin, which allows chromatin compaction and segregation. In the absence of the Ada2a/3 core of ATAC, Gcn5 activity is not correctly targeted and thus, Cyclin A levels remain high, as Cyclin A fails to be efficiently degraded. This in turn would lead to amplified Cyclin A/Cdk2 kinase activity, which then would result in abnormal centrosome duplication and defective bipolar spindle formation, SIRT2 phophorylation and inactivation. All these effects together would then lead to defective chromatin compaction, microtubule dynamics, and thus, mitotic failure.

Altogether, our data highlight an essential implication of a mammalian AT complex in mitosis, by regulating the activity of crucial non-histone substrates. We have uncovered a novel function for a HAT complex (ATAC) that earlier has been mainly implicated in transcription regulation. Acetylation of non-histone substrates is now being generally accepted as a regulatory mechanism of protein activity. As we describe in this study for Cyclin A, protein acetylation seems to be a more general mechanism to control protein turnover and, therefore, the function of important cellular protein machines is controlled by regulating their cellular amounts (Caron *et al*, 2005; Sadoul *et al*, 2008). The involvement of the ATAC complex in targeting cell cycle kinase complex(es) reveals a novel regulatory function for this complex in controlling cell cycle progression. Moreover, our new results challenge the classical concept that HAT activities must be replaced by HDACs during mitosis, and indicate that ATs must remain active and act coordinately with HDACs to regulate cell division.

Materials and methods

Cell culture, reagents, treatments and transfections

Cell culture conditions, specific treatments and transfection reagents are detailed in the Supplementary data.

Generation of inducible HeLa shRNA cell lines

During a first infection, HeLa cells (ATCC) were transduced with a lentivirus encoding the tetraCycline repressor DNA-binding domain fused to a KRAB domain (pLV-tTRKRAB-red). Five days after infection, a cell population expressing similar levels of the dsRED marker was sorted using an FACS Diva. These cells were hereafter grown in DMEM supplemented with 10% Tet-free FBS. In a second round of infections, these cells were transduced with lentiviruses harbouring shRNA cloned in pLVTH either against Gcn5 (target sequences: CGTGCTGTCACCTCGAATGA) or GL2 luciferase as control (target sequences: CCTTACGCTGAGTACTTCGA) at a MOI of 20. All cell lines generated were checked for the presence of the EGFP marker after 2 days of doxicylin induction (1 μg/ml) by FACS. Lentivirus production and titre evaluation were carried out according to Professor D Trono's laboratory protocols (http://tronolab.epfl.ch/) pLV-tTRKRAB-red, pLVTH and the packaging systems were kindly provided by Professor D Trono and are described in Wiznerowicz and Trono (2003).

Antibodies

Antibodies used in this study are detailed in the Supplementary data.

Protein overexpression in baculoviruses

Expression and purification of human Cyclin A in complex with Cdk2 was performed as earlier described (Sarcevic *et al*, 1997). The baculovirus expressing rGcn5 was described in Demeny *et al* (2007). The rGcn5-ΔHAT construct is essentially the same containing two point mutations (E575A and D615A) generated by site-directed mutagenesis. The rATAC2 expressing baculovirus contains the amplified human ATAC2 cDNA obtained from the IMAGE clone IRAUp969G0838D. Primers used for the amplification were 5′-TCCTCGAGCTGTATTCGCATCAGCGCC-3′ and 5′-AAGAATTCGATG GATAGTAGCATCCACCTGAG-3′. The amplicon was inserted into the EcoRI XhoI sites of the HA tag containing pCDNA 3.1 vector and further cloned into pVL1393 (BD Biosciences) to generate recombinant viruses. The rPCAF was kindly provided by N Rochel.

Preparation of cell extracts and IP

Details concerning preparation of total or nuclear cell extracts from mammalian cells or insect cells, and immunoprecipation are described in the Supplementary data.

HAT assays

Acetylation assays: proteins, purified either by GST pull down or by His-tag purification, were incubated in the presence of recombinant Gcn5, Gcn5-ΔHAT, Pcaf or Atac2 (purified from baculovirus-infected insect cells by anti-FLAG or anti-HA IP followed by elution with FLAG or HA peptides) or the human ATAC or SAGA complexes (purified from HeLa nuclear extracts) and ^{14}C-Acetyl-CoA. The reaction mixture (25 μl) containing 5X HAT buffer (250 mM Tris [pH 7.9], 50% glycerol, 0.5 mM EDTA, 250 mM KCl, 100 mM sodium butyrate, 5 mM and the protease inhibitor C-Complete (Roche) was incubated for 1 h at 30°C. The reaction was stopped by adding Laemmli buffer with 10 mM DTT and boiled for 5–10 min. Proteins from the reactions were separated on a 13% SDS–PAGE and analysed by coomassie brillant blue staining and then by radiography.

Peptide acetylation assays were performed as described in Nagy *et al* (2010).

TDAC assays were performed as described in North *et al* (2003).

Immunofluorescence

Details concerning indirect immunofluorescence, microscopy, and image analysis are described in the Supplementary data.

FACS analysis

Cells were trypsinized, washed with PBS, and fixed with ice-cold 70% ethanol O/N at 4°C. DNA was stained using a solution with 50 μg/ml propidium iodide (Sigma) and 1 mg/ml RNase A in PBS. The cells were analysed on FACScalibur (BD Biosciences) using CellQuest and ModFit data analysis software.

Supplementary data
Supplementary data are available at *The EMBO Journal* Online (http://www.embojournal.org).

Acknowledgements

We thank V Band, B Lüscher, B Sarcevic, X Zhao for reagents, BJ North for reagents and advice; D Devys, ME Torres-Padilla and C Cantó for critically reading the paper and the IGBMC facilities for technical support. ZN and MO were supported by a fellowship from the European Community grant and a fellowship from the Fondation pour la Recherche Médicale, and AK and AR by a fellowship of the Alsace Region. This work was supported by funds from CNRS, and European Community (HPRN-CT 00504228, STREP LSHG-CT-2004-20 502950 and EUTRACC LSHG-CT-2007-037445), INCA (2008-UBICAN) and AICR (09-0258) grants.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Balasubramanian R, Pray-Grant MG, Selleck W, Grant PA, Tan S (2002) Role of the Ada2 and Ada3 transcriptional coactivators in histone acetylation. *J Biol Chem* **277:** 7989–7995

Bitterman KJ, Anderson RM, Cohen HY, Latorre-Esteves M, Sinclair DA (2002) Inhibition of silencing and accelerated aging by nicotinamide, a putative negative regulator of yeast Sir2 and human SIRT1. *J Biol Chem* **277:** 45099–45107

Brownell JE, Allis CD (1996) Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation. *Curr Opin Genet Dev* **6:** 176–184

Bu P, Evrard YA, Lozano G, Dent SY (2007) Loss of Gcn5 acetyltransferase activity leads to neural tube closure defects and exencephaly in mouse embryos. *Mol Cell Biol* **27:** 3405–3416

Caron C, Boyault C, Khochbin S (2005) Regulatory cross-talk between lysine acetylation and ubiquitination: role in the control of protein stability. *BioEssays* **27:** 408–415

Cimini D, Mattiuzzo M, Torosantucci L, Degrassi F (2003) Histone hyperacetylation in mitosis prevents sister chromatid separation and produces chromosome segregation defects. *Mol Biol Cell* **14:** 3821–3833

Ciurciu A, Komonyi O, Boros IM (2008) Loss of ATAC-specific acetylation of histone H4 at Lys12 reduces binding of JIL-1 to chromatin and phosphorylation of histone H3 at Ser10. *J Cell Sci* **121:** 3366–3372

Ciurciu A, Komonyi O, Pankotai T, Boros IM (2006) The Drosophila histone acetyltransferase Gcn5 and transcriptional adaptor Ada2a are involved in nucleosomal histone H4 acetylation. *Mol Cell Biol* **26:** 9413–9423

De Boer L, Oakes V, Beamish H, Giles N, Stevens F, Somodevilla-Torres M, DeSouza C, Gabrielli B (2008) Cyclin A//cdk2 coordinates centrosomal and nuclear mitotic events. *Oncogene* **27:** 4261–4268

Demeny MA, Soutoglou E, Nagy Z, Scheer E, Janoshazi A, Richardot M, Argentini M, Kessler P, Tora L (2007) Identification of a small TAF complex and its role in the assembly of TAF-containing complexes. *PLoS One* **2:** e316

den Elzen N, Pines J (2001) Cyclin A is destroyed in prometaphase and can delay chromosome alignment and anaphase. *J Cell Biol* **153:** 121–136

Dryden SC, Nahhas FA, Nowak JE, Goustin A-S, Tainsky MA (2003) Role for human SIRT2 NAD-dependent deacetylase activity in control of mitotic exit in the cell cycle. *Mol Cell Biol* **23:** 3173–3185

Dumont S, Mitchison TJ (2009) Force and length in the mitotic spindle. *Curr Biol* **19:** R749–R761

Fung TK, Yam CH, Poon RY (2005) The N-terminal regulatory domain of cyclin A contains redundant ubiquitination targeting sequences and acceptor sites. *Cell Cycle* **4:** 1411–1420

Furumai R, Komatsu Y, Nishino N, Khochbin S, Yoshida M, Horinouchi S (2001) Potent histone deacetylase inhibitors built from trichostatin A and cyclic tetrapeptide antibiotics including trapoxin. *Proc Natl Acad Sci USA* **98:** 87–92

Gamper AM, Kim J, Roeder RG (2009) The STAGA subunit ADA2b is an important regulator of human GCN5 catalysis. *Mol Cell Biol* **29:** 266–280

Geley S, Kramer E, Gieffers C, Gannon J, Peters J-M, Hunt T (2001) Anaphase-promoting complex/cyclosome-dependent proteolysis of human cyclin A starts at the beginning of mitosis and is not subject to the spindle assembly checkpoint. *J Cell Biol* **153:** 137–148

Guelman S, Kozuka K, Mao Y, Pham V, Solloway MJ, Wang J, Wu J, Lill JR, Zha J (2009) The double-histone-acetyltransferase complex ATAC is essential for mammalian development. *Mol Cell Biol* **29:** 1176–1188

Guelman S, Suganuma T, Florens L, Swanson SK, Kiesecker CL, Kusch T, Anderson S, Yates III JR, Washburn MP, Abmayr SM, Workman JL (2006) Host cell factor and an uncharacterized SANT domain protein are stable components of ATAC, a novel dAda2A/dGcn5-containing histone acetyltransferase complex in Drosophila. *Mol Cell Biol* **26:** 871–882

Hammond JW, Cai D, Verhey KJ (2008) Tubulin modifications and their cellular functions. *Curr Opin Cell Biol* **20:** 71–76

Heald R, Tournebize R, Habermann A, Karsenti E, Hyman A (1997) Spindle assembly in Xenopus egg extracts: respective roles of centrosomes and microtubule self-organization. *J Cell Biol* **138:** 615–628

Hubbert C, Guardiola A, Shao R, Kawaguchi Y, Ito A, Nixon A, Yoshida M, Wang X-F, Yao T-P (2002) HDAC6 is a microtubule-associated deacetylase. *Nature* **417:** 455–458

Ishii S, Kurasawa Y, Wong J, Yu-Lee LY (2008) Histone deacetylase 3 localizes to the mitotic spindle and is required for kinetochore-microtubule attachment. *Proc Natl Acad Sci USA* **105:** 4179–4184

Kouzarides T (2007) Chromatin modifications and their function. *Cell* **128:** 693–705

Kruhlak MJ, Hendzel MJ, Fischle W, Bertos NR, Hameed S, Yang X-J, Verdin E, Bazett-Jones DP (2001) Regulation of global acetylation in mitosis through loss of histone acetyltransferases and deacetylases from chromatin. *J Biol Chem* **276:** 38307–38319

Lee KK, Workman JL (2007) Histone acetyltransferase complexes: one size doesn't fit all. *Nat Rev Mol Cell Biol* **8:** 284–295

Martinez E (2002) Multi-protein complexes in eukaryotic gene transcription. *Plant Mol Biol* **50:** 925–947

Mateo F, Vidal-Laliena M, Canela N, Busino L, Martinez-Balbas MA, Pagano M, Agell N, Bachs O (2009) Degradation of cyclin A is regulated by acetylation. *Oncogene* **28:** 2654–2666

Meraldi P, Lukas J, Fry AM, Bartek J, Nigg EA (1999) Centrosome duplication in mammalian somatic cells requires E2F and Cdk2-Cyclin A. *Nat Cell Biol* **1:** 88–93

Minoru Y, Sueharu H, Teruhiko B (1995) Trichostatin A and trapoxin: novel chemical probes for the role of histone acetylation in chromatin structure and function. *BioEssays* **17:** 423–430

Nagy Z, Riss A, Fujiyama S, Krebs A, Orpinell M, Jansen P, Cohen A, Stunnenberg H, Kato S, Tora L (2010) The metazoan ATAC and SAGA coactivator HAT complexes regulate different sets of inducible target genes. *Cell Mol Life Sci* **67:** 611–628

Nagy Z, Tora L (2007) Distinct GCN5//PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation. *Oncogene* **26:** 5341–5357

North BJ, Marshall BL, Borra MT, Denu JM, Verdin E (2003) The human Sir2 ortholog, SIRT2, is an NAD+-dependent tubulin deacetylase. *Mol Cell* **11:** 437–444

North BJ, Verdin E (2007) Mitotic regulation of SIRT2 by Cyclin-dependent kinase 1-dependent phosphorylation. *J Biol Chem* **282:** 19546–19555

Oda H, Okamoto I, Murphy N, Chu J, Price SM, Shen MM, Torres-Padilla ME, Heard E, Reinberg D (2009) Monomethylation of histone H4-lysine 20 is involved in chromosome structure and stability and is essential for mouse development. *Mol Cell Biol* **29:** 2278–2295

Pandithage R, Lilischkis R, Harting K, Wolf A, Jedamzik B, Luscher-Firzlaff J, Vervoorts J, Lasonder E, Kremmer E, Knoll B, Luscher B (2008) The regulation of SIRT2 function by cyclin-dependent kinases affects cell motility. *J Cell Biol* **180:** 915–929

Paolinelli R, Mendoza-Maldonado R, Cereseto A, Giacca M (2009) Acetylation by GCN5 regulates CDC6 phosphorylation in the S phase of the cell cycle. *Nat Struct Mol Biol* **16:** 412–420

Ruchaud S, Carmena M, Earnshaw WC (2007) Chromosomal passengers: conducting cell division. *Nat Rev Mol Cell Biol* **8:** 798–812

Sadoul K, Boyault C, Pabion M, Khochbin S (2008) Regulation of protein turnover by acetyltransferases and deacetylases. *Biochimie* **90:** 306–312

Sarcevic B, Lilischkis R, Sutherland RL (1997) Differential phosphorylation of T-47D human breast cancer cell substrates by D1-, D3-, E-, and A-type cyclin-CDK complexes. *J Biol Chem* **272:** 33327–33337

Shogren-Knaak M, Ishii H, Sun J-M, Pazin MJ, Davie JR, Peterson CL (2006) Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311:** 844–847

Suganuma T, Gutierrez JL, Li B, Florens L, Swanson SK, Washburn MP, Abmayr SM, Workman JL (2008) ATAC is a double histone acetyltransferase complex that stimulates nucleosome sliding. *Nat Struct Mol Biol* **15:** 364–372

Tin Su T (2001) Cell cycle: how, when and why cells get rid of cyclin A. *Curr Biol* **11:** R467–R469

Valls E, Sanchez-Molina S, Martinez-Balbas MA (2005) Role of histone modifications in marking and activating genes through mitosis. *J Biol Chem* **280:** 42592–42600

Vaquero A, Scher MB, Lee DH, Sutton A, Cheng HL, Alt FW, Serrano L, Sternglanz R, Reinberg D (2006) SirT2 is a histone deacetylase with preference for histone H4 Lys 16 during mitosis. *Genes Dev* **20:** 1256–1261

Vaquero A, Sternglanz R, Reinberg D (2007) NAD+-dependent deacetylation of H4 lysine 16 by class III HDACs. *Oncogene* **26:** 5505–5520

Vernarecci S, Ornaghi P, Bagu A, Cundari E, Ballario P, Filetici P (2008) Gcn5p plays an important role in centromere kinetochore function in budding yeast. *Mol Cell Biol* **28:** 988–996

Wang Y-L, Faiola F, Xu M, Pan S, Martinez E (2008) Human ATAC is a GCN5/PCAF-containing acetylase complex with a novel NC2-like histone fold module that interacts with the TATA-binding protein. *J Biol Chem* **283:** 33808–33815

Wei Y, Mizzen CA, Cook RG, Gorovsky MA, Allis CD (1998) Phosphorylation of histone H3 at serine 10 is correlated with chromosome condensation during mitosis and meiosis in tetrahymena. *Proc Natl Acad Sci USA* **95:** 7480–7484

Westermann S, Weber K (2003) Post-translational modifications regulate microtubule function. *Nat Rev Mol Cell Biol* **4:** 938–948

Wiznerowicz M, Trono D (2003) Conditional suppression of cellular genes: lentivirus vector-mediated drug-inducible RNA interference. *J Virol* **77:** 8957–8961

Wolthuis R, Clay-Farrace L, van Zon W, Yekezare M, Koop L, Ogink J, Medema R, Pines J (2008) Cdc20 and Cks direct the spindle checkpoint-independent destruction of Cyclin A. *Mol Cell* **30:** 290–302

Zhang Y, Li N, Caron C, Matthias G, Hess D, Khochbin S, Matthias P (2003) HDAC-6 interacts with and deacetylates tubulin and microtubules *in vivo*. *EMBO J* **22:** 1168–1179

Zilberman Y, Ballestrem C, Carramusa L, Mazitschek R, Khochbin S, Bershadsky A (2009) Regulation of microtubule dynamics by inhibition of the tubulin deacetylase HDAC6. *J Cell Sci* **122:** 3531–3541

Cell. Mol. Life Sci. (2010) 67:611–628
DOI 10.1007/s00018-009-0199-8

Cellular and Molecular Life Sciences

RESEARCH ARTICLE

The metazoan ATAC and SAGA coactivator HAT complexes regulate different sets of inducible target genes

Zita Nagy · Anne Riss · Sally Fujiyama · Arnaud Krebs · Meritxell Orpinell · Pascal Jansen · Adrian Cohen · Henk G. Stunnenberg · Shigeaki Kato · Làszlò Tora

Received: 21 July 2009 / Revised: 23 October 2009 / Accepted: 30 October 2009 / Published online: 21 November 2009

Abstract Histone acetyl transferases (HATs) play a crucial role in eukaryotes by regulating chromatin architecture and locus-specific transcription. The GCN5 HAT was identified as a subunit of the SAGA (Spt-Ada-Gcn5-Acetyltransferase) multiprotein complex. Vertebrate cells express a second HAT, PCAF, that is 73% identical to GCN5. Here, we report the characterization of the mammalian ATAC (Ada-Two-A-Containing) complexes containing either GCN5 or PCAF in a mutually exclusive manner. In vitro ATAC complexes acetylate lysine 14 of histone H3. Moreover, ATAC- or SAGA-specific knockdown experiments suggest that both ATAC and SAGA are involved in the acetylation of histone H3K9 and K14 residues. Despite their catalytic similarities, SAGA and ATAC execute their coactivator functions on distinct sets of inducible target genes. Interestingly, ATAC strongly influences the global phosphorylation level of histone H3S10, suggesting that in mammalian cells a cross-talk exists linking ATAC function to H3S10 phosphorylation.

Keywords ADA2a · GCN5 · ATAC2 · Histone acetyltransferase · Immediate early gene · PCAF · H3S10P

Z. Nagy and A. Riss contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00018-009-0199-8) contains supplementary material, which is available to authorized users.

Z. Nagy · A. Riss · A. Krebs · M. Orpinell · L. Tora (✉)
Department of Functional Genomics,
Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR 7104, INSERM U 964,
Université de Strasbourg, BP 10142, 67404 Illkirch Cedex,
CU de Strasbourg, France
e-mail: laszlo@igbmc.u-strasbg.fr; laszlo@titus.u-strasbg.fr

S. Fujiyama · S. Kato
Institute of Molecular and Cellular Biosciences,
The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku,
Tokyo 113-0032, Japan

P. Jansen · A. Cohen · H. G. Stunnenberg
Department of Molecular Biology, NCMLS 274,
Radboud University Nijmegen, PO Box 9101,
6500 HB Nijmegen, The Netherlands

Introduction

Post-translational modifications of histones are known to play fundamental roles on the biology of the cell [1]. One of these modifications, the acetylation of lysine residues, has immediate effects on gene regulation, affects the folding of chromatin fibers, and also plays a role in the interaction between histones and other proteins [2–5]. Histone acetylation in general is a mark of active transcription, although the different lysine residues seem to be important for distinct biological processes. For example, it has been reported that histone H3 K9/K14 acetylation defines distinct chromatin regions permissive for gene expression [3, 6]. Furthermore, histone H4K5Ac and H4K12Ac have been suggested to function as deposition marks for newly synthesized histones [7], while acetylation of H4K16 leads to global opening of the chromatin due to changes in the physico-chemical properties of the chromatin fiber [8]. The histone acetyl transferases that add the acetyl group to lysines are classified into several HAT enzyme families and have distinct substrate specificities [9, 10]. They function as co-activators/adapters of transcription, but can also acetylate non-histone substrates, thereby regulating their activity or stability [4, 11]. GCN5 is the founding member of the GNAT family of HATs.

While most of the metazoan genomes code for one GCN5 type factor, vertebrates have a second gene encoding PCAF (p300/CBP associated factor), which is highly homologous to GCN5 [4, 11, 12]. GCN5 (or PCAF) is a subunit of the SAGA (Sgf-Ada-Gcn5 containing acetyltransferase) complex that is conserved from yeast to human cells [4, 13, 14]. Eukaryotic SAGA complexes harbor 19 subunits, which include TRRAP (or yTra1), ENY2 (or ySus1), USP22 (or yUbp8), and subunits belonging to the ADA, SPT, TAF, and SGF group of proteins [13]. Recently, a second GCN5-containing complex has been described from metazoans, called ATAC (ADA-Two-A containing) [15–18]. The ATAC complex shares a common core with SAGA, composed of GCN5-ADA3-SGF29 and either ADA2a (in ATAC) or ADA2b (in SAGA). Besides these subunits, the human ATAC complexes contain a second putative HAT, called hATAC2 (or hCSRP2BP), and five other subunits; hYEATS2, hZZZ3, hMBIP, hWDR5, and hNC2β in human cells or their *Drosophila* orthologues in dATAC. At present, the biological function of the ATAC complex is not well understood.

In vitro, yeast and human GCN5 acetylates mainly histone H3K14. However, when incorporated into the hSAGA complex, although GCN5 still shows a preference for histone H3K14, it also acetylates H3K9 and H3K18 [19, 20]. The *Drosophila* ATAC possesses different substrate specificity than dSAGA, since it mainly acetylates histone H4 [16, 17, 21]. The H4 specific activity was suggested to result from the presence of the second HAT, ATAC2, in the complex [17]. When testing the HAT activity of different hATAC preparations on free histones, the Flag-MBIP immunoprecipitated (IPed) hATAC acetylated histone H3 and H4, while the Flag-YEATS2 IPed hATAC acetylated only histone H3 [15, 18]. Thus, at present, the function of the distinct metazoan ATAC complexes is not clear, and the physiological targets of the two known subunits with enzymatic activities (ATAC2 and GCN5) await further analysis. Moreover, it is not yet well understood why metazoan cells have two different GCN5-containing HAT complexes and whether these complexes carry out redundant or specific functions.

Here, we report the identification of the endogenous human ATAC complexes. We determined the composition and the histone acetylation specificity of the hATAC complex in vitro. Importantly, we propose that the biological relevance of having two GCN5-containing HAT complexes, hSAGA and hATAC, manifests mainly in vivo. We show in both *Drosophila* salivary glands and human cells that the SAGA and ATAC HAT complexes respond to different stimuli and thus play a transcriptional coactivator function at distinct sets of target genes. We also provide evidence for ATAC playing a role in the regulation of global histone H3 Ser10 phosphorylation suggesting a cross-talk between histone acetylation and phosphorylation on histone H3 tails.

Materials and methods

Antibodies

The anti-hADA2a monoclonal antibody (2AD2A1) was raised against the peptide MDRLGSFSNDPSDKPP(C) and its specificity tested in immunofluorescent staining (electronic supplementary material, ESM, Fig. 1). To raise polyclonal antibodies for the ATAC subunits, the following pepides were synthetized and coupled to ovalbumine: for the anti-ADA2a antibody (#2619), the MDRLGSF SNDPSDKPP(C) peptide; for the anti-hZZZ3 antibody (#2616), the (C)GNNNGRTTDLKQQSTRESW peptide; for the anti-mADA3 antibody (#2678), the LEGKTGHG PGPGPGRPKSKN(C) peptide; for the anti-ATAC2 antibody (#2734), the IRSHLHRSDPHWTPEPD(C) peptide, for the anti-hYEATS2 antibody (#2783), the LSQHND FLSDKDNNSNM(C) peptide; for the anti-hMBIP antibody (#2786), the TRPEGIPGSGHKPNSMLR(C) peptide; for the anti-mGCN5 antibody (#2676), the MAEPSQAP TPAPAAQPRPL(C) peptide; and for the anti-mPCAF (#2760) antibody, the MAEAGGAGJPALPPAPPHG(C) peptide. The sera were affinity purified by using the Sulfolink Coupling Gel (Pierce) following the manufacturer's recommendations.

Antibodies against the following proteins have been described earlier: GNC5: 2GC2C11 [22], TAF10: 23TA1H8 [23], TRRAP: 2TRR2D5 [24], TBP: 3G3 [25], USP22: #2391 [26], SPT20 (p38IP): #4112 [27] and #2487 [13], WDR5: [28], HCF-1: N18 [29], NC2α and NC2β [30].

Antibodies used for detection of histone modifications on western blot were: H3 core: ab1791 Abcam; H3K9Ac: #06-942 Millipore; H3S10P: #05-598 Millipore; H3K14Ac: #07-353 Millipore; H4K5Ac: ab51997 Abcam; H4K12Ac: ab1761 Abcam; H4K16Ac: #07-329 Millipore.

Antibodies used for ChIP were: H3 core: ab1791 Abcam; H3K9/K14Ac: #06-599 Millipore.

Drosophila antibodies used for polytene chromosome staining

Primary antibodies were rabbit anti-FLAG (F-7425; Sigma–Aldrich) at 1:50, monoclonal antibody against RNA Polymerase II phosphorylated on Ser5 (H14; COVANCE) at 1:50 and rabbit anti-Ada2b [31] at 1:25 dilution. Secondary antibodies were Texas Red Goat Anti-Mouse IgM, μ chain specific (Jackson ImmunoResearch), and Alexa

Fluor® 488 Donkey anti-rabbit IgG (Molecular Probes) at 1:1,000 dilution.

Baculovirus infections and protein overexpression

Overexpression and protein preparation was done as described in [32]. For the HA-ATAC2 construct, the cDNA clone IRAUp969G0838D was purchased from ImaGENES, the coding sequence amplified with the following primers: 5′ AAGAATTCGATGGATAGTAGCATCCACCTGAG 3′; 5′ TCCTCGAGCTGTATTCGCATCAGCGCC 3′, and then cloned to the *Eco*RI *Xho*I sites of pCDNA3 vector modified to contain an N-terminal HA tag. The tagged cDNA was further cloned to pVL1393 baculovirus transfer vector to generate recombinant virus. The hGCN5 expressing construct was described in [33].

Immunoprecipitations were carried out as described earlier [33].

Cell growth condition, stress treatments

HeLa cells were grown in Dulbecco's modified Medium supplemented with 1 g/l glucose, 5% FCS, phenol red, and gentamycine. Before TPA treatment, cells were serum straved O/N (0.5% FCS) and then treated with 50 ng/ml TPA during 1 h. Control treatment was carried out with DMSO.

Chromatin immunoprecipitation (ChIP) was carried out as described in [13].

RNA purification, reverse transcription and qPCR

Total RNA was purified using Trizol reagent (Invitrogen), reverse transcribed by MMLV reverse transcriptase using random hexamers, and analyzed by the quantitative PCR (Q-PCR) machine Roche LightCycler480 with Syber Green (Quiagen) Master mix. All the detected values represented in the manuscript have been normalized to *CyclophilinB* and represent biologically independent replicates.

siRNA

Negative control (ref. number D-001810-10), anti hZZZ3 siRNA (L-013939-01), anti hATAC2 (L-008481-00), anti hADA2a siRNA (ref number L-017516-00), and anti hSPT20 siRNA (ref number L-013820-00) was purchased from Dharmacon and transfected using Lipofectamine 2000 and OptiMEM serum-free medium following the manufacturer's recommendations.

Nuclear extract preparation was described in [33].

Purification of histone H3/H4 dimers, octamers, mono- and polynucleosomes was performed as described in [34].

Fly stock and generation of transgenic flies

Fly maintenance and crosses were performed as described previously [26]. The expression constructs of FLAG-tagged D12 and CG10238 in pUAST vector were sent to bestGene (CA, USA) *Drosophila* embryo injection services for generation of transgenic flies. Dpp-Gal4 (to obtain UAS-FLAG-CG10238/dpp-Gal4 flies) and dpp.blk1-Gal4 (to obtain UAS-FLAG-D12/+;dpp.blk1-Gal4/+ flies) stocks were obtained from the Bloomington Drosophila Stock centre.

Preparation and treatment of polytene chromosomes

Drosophila salivary glands were dissected from third instar larvae in 40% acetic acid and fixed for 5 min in 3.7% formaldehyde, 1% Triton-X 100, 1/2× PBS, and followed by 2 min in 3.7% formaldehyde, 50% acetic acid on poly-L lysine-treated slide glass. For TPA treatment, salivary glands were dissected in 0.7% NaCl, 0.1% NP-40 solution, incubated with 4 nM TPA in Schneider's Drosophila Medium (1×) (Invitrogene), 10% FBS for 30 min at 25°C, and then fixed. After fixation, polytene chromosomes were squashed by squeezing the slide and siliconized cover slip and were frozen in liquid nitrogen. The cover slip was removed and dehydrated for 10 min in ice-cold 50% acetone, 50% methanol solution. The slide was washed in PBS for 10 min and, for blocking, incubated in 5% skim milk, PBST (PBS with 0.1% Tween 20) for 1 h at RT. The squashed polytene chromosomes were covered overnight at 4°C with the diluted primary antibody in a humid chamber. The slides were washed three times for 5 min in PBST and incubated with the secondary antibody for 2 h at RT. Slides were washed three times for 5 min with PBST and mounted with VECTASHIELD Mounting Medium with DAPI (Vector Laboratories, Burlingame). The specimens were photographed in confocal microscopy LSM510 (Carl Zeiss MicroImaging).

Acetylation assays

Peptide acetylations were performed as described in [13, 35]. Briefly, 1.2 µg of peptide (corresponding to the N-terminal tail of H3 at positions 6–20 or the N-terminal tail of H4 from positions 1–19) was added to the immunopurified protein sample together with 1 µl of H^3 Acetyl CoenzymeA (Amersham; 50 µCi/ml) in the reaction buffer (50 mM TrisHCl pH 8, 20 mM KCl, 5 mM DTT, 4 mM EDTA) and incubated at 30°C for 1 h. Samples were spotted on Whatmann P81 nitrocellulose filters, washed 3 times for 10 min in ice-cold 50 mM $NaHCO_3$ pH 9 buffer, and dried. Filters were then dropped into 5 ml of Ready-Safe liquid scintillation cocktail (Beckman Coulter) and

radioactivity was quantified by an LS6000SC Beckman counter.

Histone acetylation assay: histones were incubated with ^{14}C-Acetyl-CoA and ATAC complex (purified from HeLa nuclear extracts by anti-ADA2a IP followed by elution with peptide) or SAGA complex (purified from HeLa nuclear extracts by anti-USP22 IP followed by peptide) in 1×HAT buffer (50 mM Tris pH 7.9, 10% glycerol, 0,1 mM EDTA, 50 mM KCl, 20 mM Sodium Butyrate, 1 mM DTT and protease inhibitors). The reactions were incubated during 1 h at 30°C, stopped by adding Laemmli buffer with 100 mM DTT and boiled for 10 min. Proteins were then loaded on a 13% SDS–PAGE and analyzed by coomassie brillant blue staining. The gel was then incubated for 20 min in "Amplify" fluorographic reagent (GE Healthcare) and dried. Blank phosphor screen (Fuji) was placed overnight on the gel and the radioactive signal was analysed with phosphorimager scanner Typhoon 8600.

Primers used for the RT qPCR analysis

EGR-1 RT L: ACCTGACCGCAGAGTCTTTTCC;
EGR-1 RT R: CAGGGAAAAGCGGCCAGTATAG;
FRA-1 RT L: CAGGAACCGGAGGAAGGAACT;
FRA-1 RT R: TGCTTCTGCAGCTCCTCAATCT;
c-FOS RT L: GGGGCAAGGTGGAACAGTTATC;
c-FOS RT R: TAGTTGGTCTGTCTCCGCTTGG;
CyclophilinB RT L: CTTCCCCGATGAGAACTTCAA ACT;
CyclophilinB RT R: CACCTCCATGCCCTCTAGAAC TTT;
GAPDH RT L: ACAGTCCATGCCATCACTGCC;
GAPDH RT R: GCCTGCTTCACCACCTTCTTG.

Primers used in the ChIP quantification reactions

EGR-1 P L: CTAGGGTGCAGGATGGAGGTG;
EGR-1 P R: TATGGGAAGCAGGAAGCCCTAA;
FRA-1 P L: GTTCCCCGAAGTCTCGGAACAT;
FRA-1 P R: GTGGTTCAGCCCGAGAACTTTT;
c-FOS P L: TTGAGCCCGTGACGTTTACACT;
c-FOS P R: TTCTCAGATGCTCGCTGCAGAT;
non-coding L:TGGAACTTCTGGAAGACACTGGAA;
non-coding R: TACACCACTCAAGGGAAACTGGAA.

Results

Composition of the endogenous hATAC complex

To isolate endogenous human ATAC complexes we carried out immunopurification (IP) using a monoclonal antibody raised against ADA2a (ESM, Fig. 1) that is the defining subunit of the complex (ADA-Two-A-Containing) [15–18]. The IP was carried out on a HeLa cell nuclear extract without overexpressing any of the putative ATAC subunits and the proteins that coprecipitated with hADA2a were resolved by SDS–PAGE and visualized by silver nitrate staining (ESM, Fig. 2a). The identification of the components of this endogenous complex by mass spectrometry indicated mostly the presence of the same polypeptides as the ones reported by Guelman and colleagues [15] (ESM, Fig. 2a, and see Table 1). To further verify the presence of all the reported subunits in our endogenous complex, we raised specific antibodies against most of the known subunits of hATAC. In addition, to carry out a comparative examination of the two human GCN5-containing complexes, we have also purified endogenous hSAGA complex by using an antibody raised against the recently identified hSPT20 subunit [13]. Consequent comparison of the two complexes by western blotting analysis revealed the similarities and differences between hSAGA and hATAC. Common subunits present in both complexes are: hGCN5, hADA3 and hSGF29 (Fig. 1a). Human ATAC specific components are: hYEATS2, a large protein containing a so-called YEATS domain with no known function; hZZZ3, a potential transcription factor with zinc finger and SANT domains; hATAC2 containing a putative acetyl transferase domain; hADA2a, a known adaptor protein affecting the activity of hGCN5 [36]; hMBIP that was shown to interact with the MAPK upstream kinase [37]; and hWDR5 also known as a subunit of the MLL complex [38] (Fig. 1a lane 2). Importantly, SAGA specific subunits such as hTRRAP, hUSP22 and hTAF10 did not copurify with hADA2a (Fig. 1a lane 2). As TBP and HCF1 were reported to interact with the anti-Flag-YEATS2 purified hATAC complex [18] (Table 1), we tested the presence of these factors in our endogenous ATAC complex preparations. In good agreement with our mass spectrometry data, neither TBP nor HCF1 could be detected in the complex by western blot analysis (data not shown).

To further confirm our results on the composition of the endogenous hATAC complex, we carried out reciprocal IPs with antibodies raised against the putative ATAC subunits, such as hGCN5, hADA3, hATAC2, hZZZ3, hMBIP, and hYEATS2. Western blot analyses of samples eluted by excess of the antigene peptide show that besides the common subunits (hGCN5, hADA3, and hSGF29), the ATAC specific proteins (hZZZ3, hYEATS2, hATAC2, hMBIP, and hWDR5) copurify with the precipitated proteins in all the cases (Fig. 1b). On the other hand, antibodies against common subunits of SAGA and ATAC (anti-hGCN5 or anti-hADA3) also pull down hSAGA specific subunits, such as hTRRAP, hUSP22, and hTAF10 (Fig. 1b lanes 2, 3). Based on our observations, we conclude that hADA2a, hGCN5, hADA3, hSGF29, hZZZ3,

Table 1 Subunit composition of the known *Drosophila* and human ATAC complexes

Drosophila ATAC subunits [16, 17]	Human ATAC purified by anti-flagYEATS2 IP [18]	Human ATAC purified by anti-flagMBIP IP [15]	Human ATAC purified by anti-hADA2a IP (present work)
dGCN5	hGCN5	hGCN5	hGCN5
	hPCAF		hPCAF
dADA2a	hADA2a	hADA2a	hADA2a
dADA3	hADA3	hADA3	hADA3
dATAC1 (CG9200)	ZZZ3	ZZZ3	ZZZ3
dATAC2 (CG10414)	hATAC2 (hCSRP2BP)	hATAC2 (hCSRP2BP)	hATAC2 (hCSRP2BP)
CG30390	hSGF29	hSGF29	hSGF29
dATAC3 (CG32343)			
HCF	HCF1		
WDS	WDR5	WDR5	WDR5
D12	YEATS2	YEATS2	YEATS2
NC2β	NC2β	NC2β	NC2β
CG10238	MBIP	MBIP	MBIP
	UBAP2L		
CHRAC14	POLE3		
	POLE4		
	MAP3K7		
	TBP		

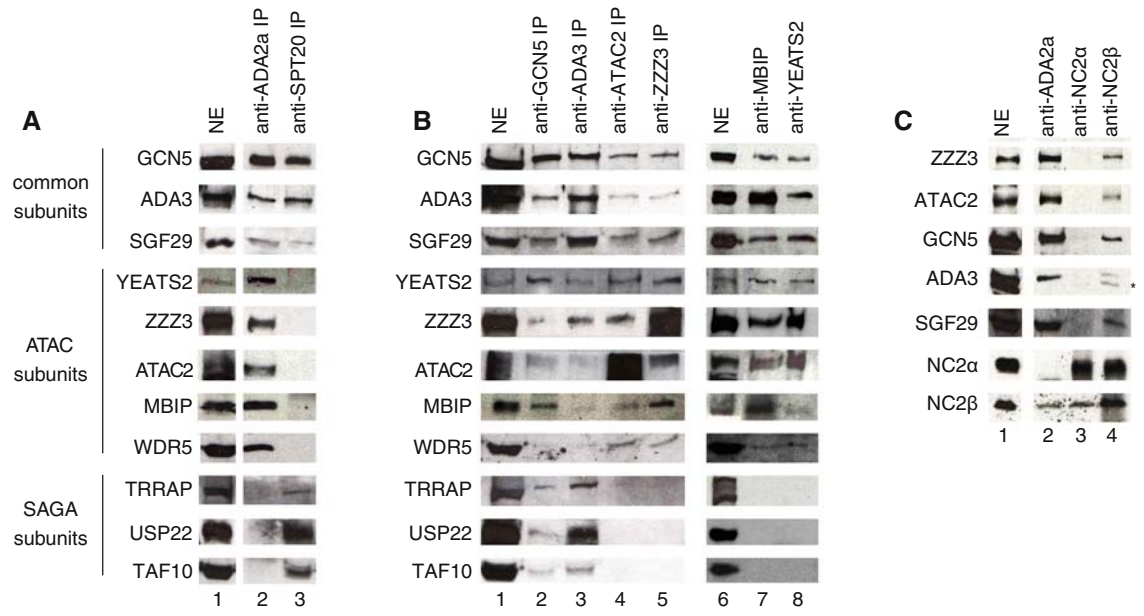


Fig. 1 Subunit composition of the endogenous human ATAC complex. **a** Comparison of an anti-ADA2a and an anti-hSPT20 IP reveals the differences in composition of hATAC and hSAGA, respectively. The two complexes were purified from HeLa nuclear extract (NE) by means of antibodies developed against hADA2a (*lane 2*) or hSPT20 (*lane 3*) and the coprecipitated proteins were detected by western blotting. **b** Different IPs demonstrate the existing interactions between hSAGA and hATAC subunits. The common, the ATAC-specific and the SAGA-specific, subunits are marked on the *left side of the figure*. **c** NC2β, but not NC2α is a component of the hATAC complex, as it copurifies with other subunits of the complex both in anti-ADA2a (*lane 2*) and in anti-NC2β (*lane 4*) IPs. *Asterisk* the heavy chain of the antibody

hATAC2, hYEATS2, hMBIP, and hWDR5 are all bona fide subunits of the endogenous ATAC complex in human cells.

As we have also obtained one peptide by mass spectrometry corresponding to the NC2β (also called Dr1; [39–41]) in one of our ATAC purifications, and as NC2β has been reported to be a component of the Flag-purified ATAC complexes [15, 18], we have investigated the interaction between hADA2a and the two subunits of the NC2 complex, NC2α and NC2β [30, 42]. Our results clearly show that only NC2β is associated with the endogenous hATAC complex since an antibody against NC2β coprecipitated the tested hATAC subunits and, vice versa, the anti-ADA2a IP precipitated NC2β (Fig. 1c lanes 2, 4). In contrast, the anti-NC2α antibody coprecipitated only NC2β, but no ATAC subunits (Fig. 1c lane 3). Thus, NC2β is also a bona fide subunit of the endogenous hATAC complex.

In summary, our results provide evidence for the existence of an endogenous ATAC complex containing at least ten subunits (see Table 1).

In vitro hATAC and SAGA acetylate histone H3 K14

After the purification of the endogenous ATAC complex from human cells, we aimed to test its substrate specificity and compare it to that of hSAGA. To this end, we performed in vitro acetylation assays on wild-type and mutated N-terminal histone tail peptides. In each mutant peptide, we changed one acetylable lysine (K) residue to arginine (R) that mimics the non-acetylated form of the amino acid. First, we analyzed the substrate specificity of the two GCN5-containing complexes, ATAC and SAGA, in peptide acetylation assay. Both complexes were purified from HeLa cells by specific IPs and normalized for the content of the common subunits (ESM, Fig. 3a). Surprisingly, in this in vitro peptide acetylation test, we observed no difference between hSAGA and hATAC, as both complexes showed a preference for acetylating histone H3K14 (Fig. 2a, b).

Next, we analyzed the activity of the two ATAC-associated HAT enzymes alone. Both recombinant enzymes (flag-GCN5 and HA-ATAC2) were purified from baculovirus infected SF9 insect cells by immunoprecipitation and consecutive peptide elution (ESM, Fig. 3b). While GCN5 showed specificity toward H3K14 in these reactions (Fig. 2c), ATAC2 had no detectable activity even when ten times more enzyme was added to the reactions (Fig. 2d). Thus, our data suggest that the second putative acetyltransferase subunit of the ATAC complex is inactive in vitro on the tested histone tail peptides.

To exclude the possibility that the use of non-physiological substrates (i.e., short peptides) changes the specificity of the enzymes in the complexes, we tested the HAT activity of ATAC and SAGA on purified full-length histone H3–H4 dimers and histone octamers containing all the four core histones. In these acetyltransferase assays, both SAGA and ATAC complexes acetylated mainly histone H3 (Fig. 2e). Similar results were obtained when we tested the acetyltransferase activity of the two complexes on mono- and polynucleosomes (Fig. 2f). Note, however, that polynucleosomes seemed to be better substrates than mononucleosomes. Our results are in good agreement with those of the Martinez group [18], and suggest that, in spite of the evolutionary conservation of the protein sequences from *Drosophila* to human, the human complex has no or very weak H4 specificity in vitro. Thus, the functional difference between the human SAGA and ATAC complexes seem not to be related to their histone substrate specificity.

ATAC complexes containing either GCN5 or PCAF exist in mouse fibroblasts

The fact that the ATAC complex contains two HATs (GCN5 and ATAC2) raised the possibility that the two enzymes may have overlapping or redundant activities within the complex. To be able to measure the activity of ATAC2 alone in the context of the intact ATAC complex, we decided to purify ATAC complexes from *Gcn5*$^{hat/hat}$ mouse embryo fibroblasts (MEFs), in which the GCN5 catalytic activity was inactivated by double E568A and D608A mutations in the HAT domain [43]. We prepared nuclear extracts from wild-type (+/+) and *Gcn5*$^{hat/hat}$ MEFs and carried out IPs using the above described anti-ADA2a antibody (Fig. 3a, elution 1). Surprisingly, in nuclear extracts prepared from *Gcn5*$^{hat/hat}$ mutant MEFs, PCAF expression was significantly up-regulated (2- to 3-fold), indicating that, when GCN5 is inactivated, cells compensate the loss of its activity with that of its paralogue (see Introduction and Fig. 3b lane 2). Note that similar results were described in a different cellular system [44]. Moreover, our western blot analysis indicated that ATAC, purified by a simple anti-ADA2a IP, contained both PCAF and GCN5 (Fig. 3b lanes 3, 4). The question thus rose whether these complexes can be separated into fractions containing exclusively GCN5 or PCAF. Alternatively, a given ATAC complex might contain both GCN5 and PCAF. To decide, the ADA2a-containing complexes were re-IPed with an anti-GCN5 antibody (see Fig. 3a). When the GCN5-free supernatant of this second anti-GCN5 IP was compared to the GCN5-containing elution fraction by western blot (elution 2 on Fig. 3a), it became clear that we have separated PCAF-containing ATAC complexes (hereafter called P-ATAC) from GCN5-containing complexes (hereafter called G-ATAC)

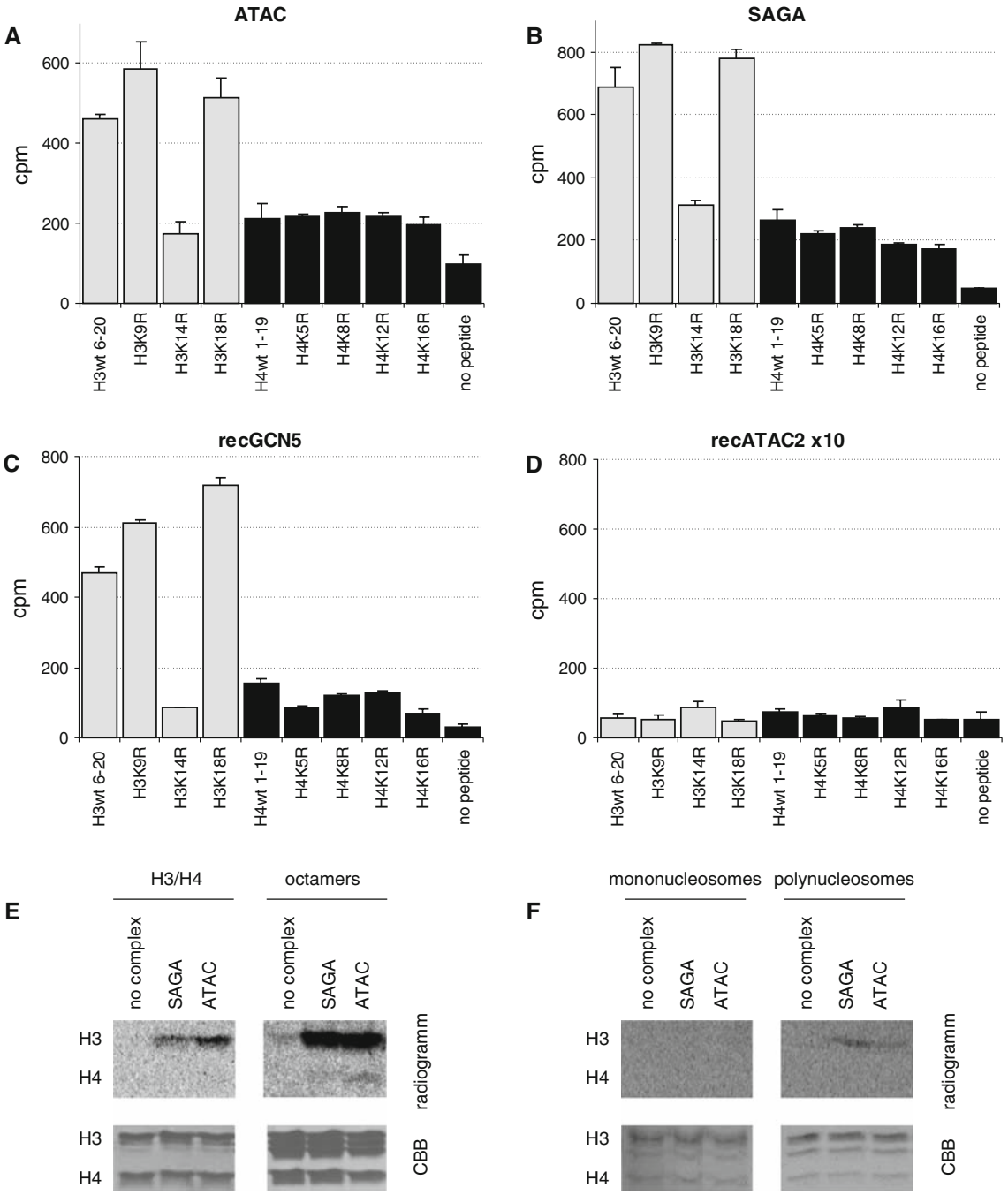


Fig. 2 Histone acetyltransferase activity of the human GCN5 containing complexes and their HAT subunits. **a–d** Acetylation activity of endogenous purified hATAC (**a**), hSAGA (**b**), recombinant (rec) GCN5 (**c**), and recombinant ATAC2 (**d**) on histone tail peptides was measured by liquid scintillography. Histone H3 peptides are shown in *gray*, histone H4 peptides and the reaction without peptide are in *black*. Note that, due to the described weak HAT activity of recombinant hATAC2 [15], in (**d**), ten times more recombinant protein was used than in the reactions shown in (**c**). **e** Acetylation activity of endogenous complexes on H3–H4 dimers and histone octamers. **f** Acetylation activity of hATAC and hSAGA complexes on mono- and polynucleosomes. In **e** and **f**, *upper panels* show the autoradiography and *lower panels* the corresponding coomassie stained histones

(Fig. 3b, compare lanes 5–6 with 7–8). These results clearly demonstrate that vertebrate cells contain both G-ATAC and P-ATAC complexes and that the presence of GCN5 or PCAF in these complexes is mutually exclusive.

The in vitro HAT activities of the different ATAC complexes

To measure the acetyltransferase activity of ATAC2 in ATAC, we have compared the HAT activities of the above

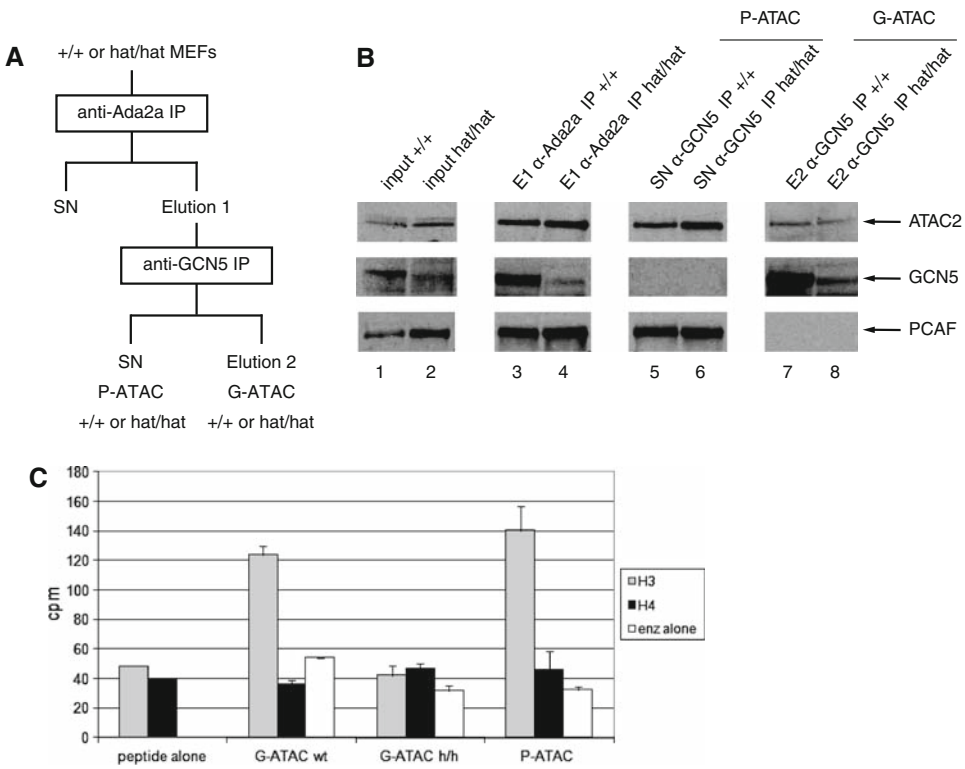


Fig. 3 Composition of different ATAC complexes from MEFs and their HAT activity on histone tail peptides. **a** Purification scheme of the separation of GCN5- (G-ATAC) and PCAF-containing (P-ATAC) ATAC complexes from wild-type (+/+) and mutant (*hat*/*hat*) MEFs. **b** Western blot analysis of the different fractions obtained during the purification shown in (**a**) using the indicated antibodies. **c** Acetylation of histone H3 and H4 tail peptides using the indicated purified complexes was measured. The amount of the purified complexes was normalized to their ATAC2 content

purified GCN5- or GCN5^{hat}-containing ATAC complexes on histone tail peptides as previously (see Fig. 2). Surprisingly, in this in vitro peptide acetylation test using either the H3 (aa 6–20) or the H4 peptide (aa 1–19), we observed no activity of the G-ATAC^{hat} complex, while wild-type G-ATAC acetylated the H3 peptide as before (Fig. 3c). This result, together with the lack of activity obtained with recATAC2 on histone tails (Fig. 2d), suggests that the mammalian ATAC2 is inactive in vitro in conditions that are appropriate for the GCN5 HAT activity. Moreover, in this in vitro test, the HAT specificity of P-ATAC was similar to that of G-ATAC (Fig. 3c). Thus, our observations suggest that (1) in vertebrates at least two different ATAC complexes exist with very similar composition and in vitro substrate specificity, (2) the mouse or human ATAC2 may have non-histone substrates, and (3) that in vitro no significant differences could be determined between the HAT activities of SAGA and ATAC.

ATAC- or SAGA-specific knock-downs lead to a drop in global histone H3 acetylation on K9 and K14, but do not affect histone H4 acetylation levels

To further analyze the differences between the two human GCN5-containing complexes in vivo, we tested how the global level of post translational modifications of histones are affected in cells where either an ATAC-specific or a SAGA-specific subunit was knocked down by using siRNAs. HeLa cells were transfected with control non-targeting siRNA (NC), anti-SPT20 siRNA (SAGA-specific) or anti-ADA2a (ATAC-specific) siRNA (Fig. 4a). At 48 h post-transfection, cell extracts were prepared and analysed by western blotting. The knock-down of SPT20 or ADA2a was efficient since we obtained a clear reduction in the respective protein levels (Fig. 4a). Next, the amount of histone content in each extract was normalized using an antibody recognizing the core domain of histone H3 (Fig. 4b upper panel). Surprisingly, compared to the control, only the H3K9Ac and the H3K14Ac marks were reduced and this reduction was of similar extent in both siADA2a and siSPT20 treated cells (Fig. 4b, c). On the other hand, the acetylation marks located on histone H4, such as K5Ac, K12Ac, and K16Ac, were unaffected in our siRNA-transfected cells. These experiments show that, similarly to the in vitro experiments, on global histone acetylation levels no significant differences could be determined between the HAT activities of SAGA and ATAC. Also, in HeLa cells, the knock-down of ATAC- and SAGA-specific subunits do not influence globally the histone H4 acetylation.

ATAC-specific knock-down leads to a drop in global histone H3 phosphoryalation on serine 10

As the *Drosophila* ATAC complex was shown to influence histone H3S10 phosphorylation [45], we have also tested whether ATAC- or SAGA specific knock-downs affect

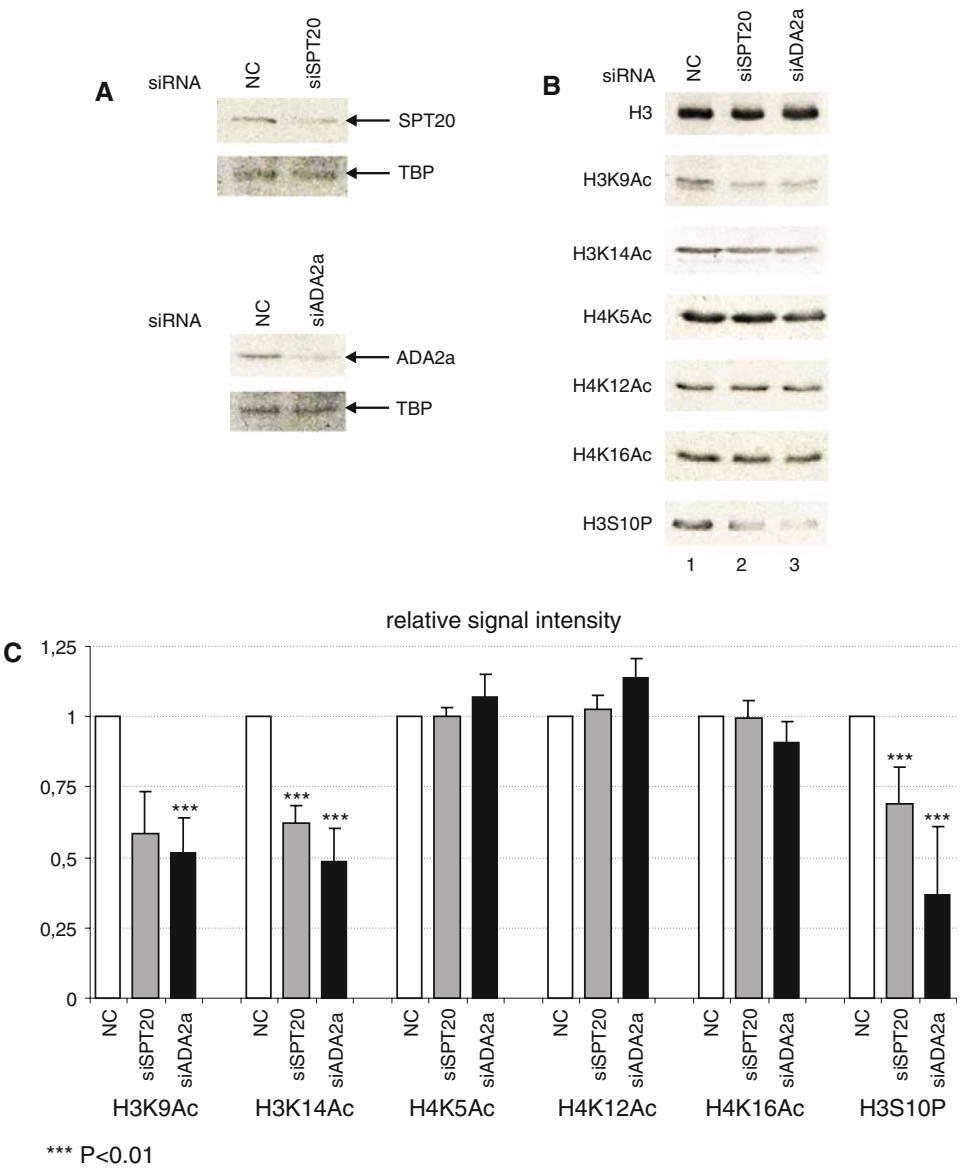


Fig. 4 Global changes of post translational histone modifications in cells deficient for ATAC or SAGA. **a** Transfection of non-targeting control siRNA (NC) or siRNAs directed against SPT20 or ADA2a leads to the specific knock-down of the targeted subunit as compared in western blot. TBP served as a loading conrol. **b** Changes in the histone acetylation and phosphorylation marks in siRNA-transfected cells. The global levels of the given modifications were analyzed by western blot. Blotting with an antibody recognizing the core domain of histone H3 served as a loading control. **c** Histone H3 post-translational modifications are perturbed in siRNA-transfected HeLa cells. Quantification of signal intensities of four independent knock-down experiments shows global changes in H3 modifications, while H4 acetylation remains unaffected. The H3S10P signal is strikingly low in siADA2a-transfected cells

H3S10 phosphorylation in human cells. Interestingly, we found that in HeLa cells with decreased ADA2a levels the H3S10 phosphorylation mark was reduced to 36% of the control, while the effect of SAGA knock-down was much milder (70%) (Fig. 4b, c). Importantly, this result defines an evolutionarily conserved cross-talk linking ATAC HAT function to H3S10 phophorylation.

dATAC and dSAGA regulate different inducible genes in *Drosophila* salivary glands in vivo

To explore the in vivo functional differences between ATAC and SAGA complexes, first we examined the localization of the complexes in salivary glands on polytene chromosomes of *Drosophila*. To be able to visualize the recruitment of dATAC specific subunits and compare the function of dATAC to dSAGA, transgenic flies were generated in which two dATAC-specific subunits, D12 (homologue of hYEATS2) and CG10238 (homologue of hMBIP, see Table 1), were Flag tagged. The localization of ADA2b, a SAGA specific subunit, was followed by an anti-dADA2b antibody labeling. For detecting the polymerase, we used an antibody raised against the serine 5 phosphorylation of the CTD of the large subunit of Pol II (Pol II Ser5P), which is a marker for Pol II incorporating in a functional preinitiation complex (PIC). Interestingly, under non-stimulated conditions, all the bands stained by antibodies against Flag-D12, Flag-CG10238, or ADA2b localized to euchromatic segments giving weak DAPI signal (Fig. 5a, c, e). Although ATAC and SAGA were suggested to function in histone modification and transcriptional regulation, our in vivo results in these non-

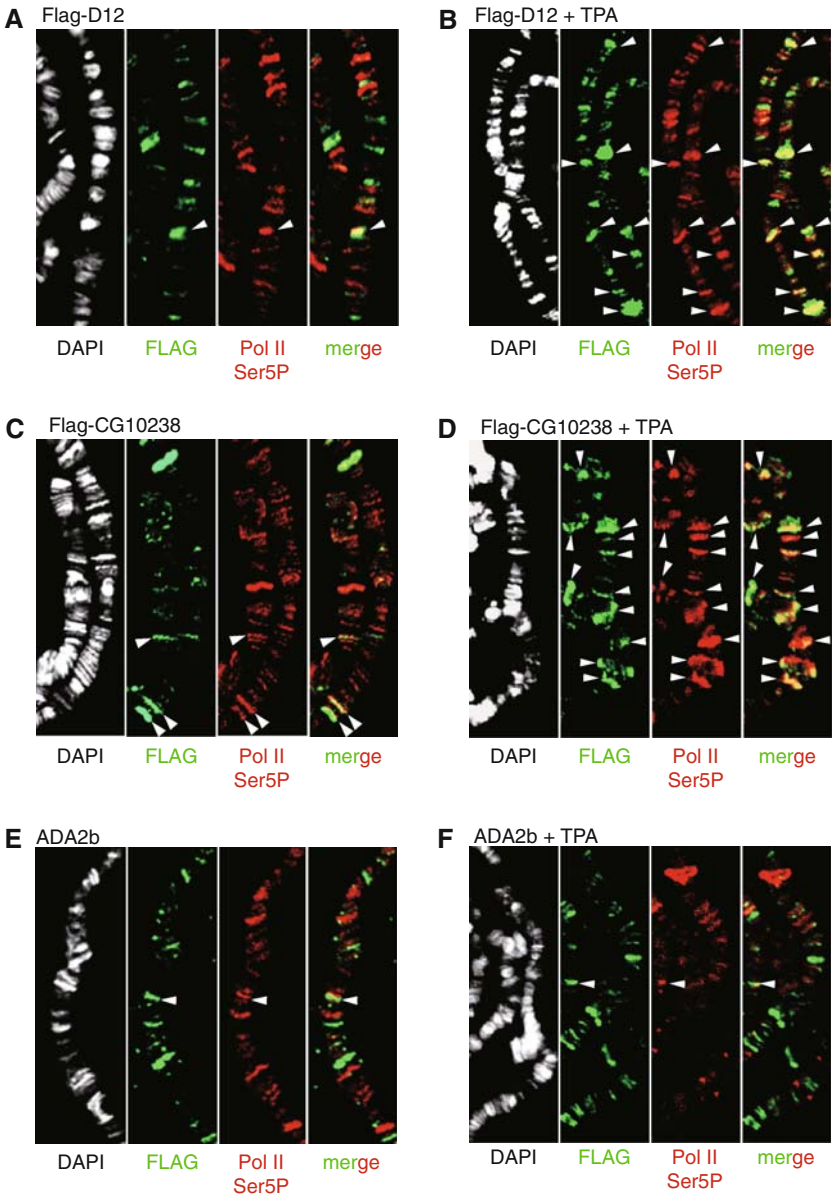


Fig. 5 dATAC, but not dSAGA, is recruited to TPA-induced transcription sites on *Drosophila* polytene chromosomes. Polytene chromosome co-staining of non-treated (**a**, **c**, **e**) and TPA treated (**b**, **d**, **f**) samples is shown. The DAPI staining (*white*), the anti-flag (**a**–**d**) or the anti-ADA2b staining (**e**, **f**) (*green*), and the anti-RNA Pol II Ser5P staining (*red*) is shown in each *panel* together with the merged picture. A dramatic increase in RNA Pol II and flag-D12 (homologue of hYEATS2) or flag-CG10238 (homologue of hMBIP) colocalization occurs after TPA treatment (compare **a** to **b** and **c** to **d**) marked with *arrowheads*. In contrast, no increase in the colocalization was detected for ADA2b and RNA Pol II (compare **e** to **f**)

stimulated conditions showed only rare colocalization of dATAC or dSAGA with RNA Pol II. To establish the role of ATAC in transcription activation, we induced genes by a TPA (12-O-tetradecanoylphorbol-13-acetate)-treatment and tested the recruitment of the dATAC complex to the transcriptionally active loci. Following the TPA-treatment, we observed an increased recruitment of dATAC (visualized by D12 and CG10238) to Pol II positive chromosome regions (see white arrows in Fig. 5b, d). In striking contrast to dATAC subunits, the dSAGA-specific subunit ADA2b showed no recruitment to the active bands following TPA induction (see Fig. 5f). These in vivo results suggest for the first time that SAGA and ATAC regulate different set of genes depending on the cellular stress received.

hATAC is recruited to the promoter of immediate early genes in human cells

The above-described in vivo results obtained on *Drosophila* salivary gland polytene chromosomes hinted at regulatory mechanisms in which the function of SAGA and ATAC do not overlap. In addition, our recent results showed that human SAGA is not involved in the transcription activation of immediate early (IE) genes [13]. Thus, we tested whether hATAC would participate in the regulation of IE genes following stimulation. To this end, we analyzed the recruitment of the hATAC complex to IE gene promoters after TPA treatment by chromatin immunoprecipitation (ChIP) in HeLa cells. First, we examined the mRNA level

of three IE genes (*c-FOS*, *FRA-1*, and *EGR-1*) after 1 h of TPA treatment. As a result, we obtained a 3- to 20-fold stimulation of these mRNA species compared to the control, while the expression of GAPDH mRNA remained unchanged (Fig. 6a). We also carried out a control DMSO treatment, where we obtained no significant effect on the expression of IE genes (not shown). Then, we prepared chromatin from both non-treated and TPA-induced cells and carried out ChIPs by using antibodies raised against RNA Pol II, hATAC subunits (hADA3, hZZZ3), and a hSAGA-specific subunit (hSPT20). The results showed that, together with RNA Pol II, subunits of the hATAC complex (hZZZ3 and hADA3) got recruited to the three IE promoters after TPA treatment in human cells (Fig. 6b–d). At the same time, no increase in the occupation of these loci was observed for hSPT20, which is a hSAGA specific subunit (Fig. 6e). The amount of the IPed control region (a genomic region not harboring any Pol II transcription unit) remained unchanged following the TPA treatment (Fig. 6b–e) and was close to levels obtained with negative control anti-GST ChIP (not shown). Note that TPA did not induce the expression of the tested ATAC or SAGA subunits in the cells (ESM Fig. 4). These results together with those obtained with *Drosophila* salivary gland stainings show that ATAC, but not SAGA, is recruited to TPA-induced gene promoters. Thus, our observations indicate for the first time a differential recruitment of the two HAT complexes, ATAC and SAGA, to stress-regulated genes in mammalian cells.

Knock-down of ATAC subunits leads to defects in the TPA-induced gene expression

Next, we analyzed whether the knock-down of ATAC subunits influences the regulation of the above tested TPA-regulated genes. After transfection of HeLa cells with siRNA against hZZZ3 or hATAC2, we observed a ~50% decrease in the corresponding mRNA levels (Fig. 7a) and a ~ 75% decrease in the protein levels of these two ATAC subunits, respectively (Fig. 7b), when compared to the control siRNA-transfected cells. To analyze the effect of the knock-down of a SAGA specific subunit, we carried out anti-SPT20 RNAi experiments in parallel. Following TPA-treatment in ZZZ3- or ATAC2-siRNA-transfected cells, the transcriptional activation of the tested IE genes was significantly reduced to ~50% of the negative control situation (Fig. 7c–e). At the same time, knock-down of SPT20 had no significant effect on the up-regulation of IE genes. This is in good agreement with our previous observations showing that SAGA is not required for IE gene induction [13]. The siRNA transfections had no effect on the GAPDH mRNA level (Fig. 7f). Altogether, these results confirm that ATAC, but not SAGA, is recruited to the promoters and is required for the induction of the studied TPA-induced genes.

ATAC is indispensable for the correct histone H3 acetylation status of IE gene promoters both under non-induced and activated conditions

As the above results indicated that the knock-down of ATAC subunits leads to defects in the induction of IE gene expression and that the global levels of both H3K9Ac and H3K14Ac were decreased in cells deficient for either their SAGA or ATAC function (Figs. 4 and 7), we tested whether the knock-down of ATAC- and SAGA-specific subunit(s) influences the acetylation at specific gene loci. Thus, we carried ChIP experiments to test the H3 acetylation at positions K9 and K14 on histone H3 at the promoters of IE genes that we found to be regulated only by ATAC (see Figs. 6 and 7). HeLa cells were transfected either with non-targeting siRNA (NC), or siRNA against ADA2a or SPT20. At 48 h after transfection, the cells were serum starved overnight and then half of them treated with TPA for an hour as above. Chromatin was prepared and subjected to two different IPs. The anti-H3 core IP served as a control, while the IP using the anti-H3K9/K14Ac antibody highlighted the changes of this active chromatin mark at the tested genomic regions. All values obtained were normalized with those obtained at a non-coding region, where no transcriptional regulation takes place, thus the level of the histone marks remains unchanged. No significant changes were obtained in H3 levels at the tested IE gene promoters following TPA treatment (Fig. 8a). As expected, in the control siRNA-treated cells, the histone H3 acetylation level increased considerably (3- to 6-fold) at all the three IE promoters tested (see NC in Fig. 8b). Similarly, an increase in H3K9 and K14 acetylation was also observed in siSPT20-transfected cells; however, the increase was less pronounced (2-fold; Fig. 8b). Surprisingly, in siADA2a-transfected cells, the H3K9 and K14 acetylation pattern was completely deregulated before activation. In the non-treated cells, the H3K9 and H3K14 positions became highly acetylated at the promoters of these three, normally silent, genes. The fact that the presence of the positive H3K9/K14Ac histone mark at the promoter of IE genes does not correlate with the transcriptional status of the genes (Fig. 8, see also Figs. 6a and 7c–e) suggests that the H3K9 and H3K14 acetylation marks alone are not sufficient for recruiting the Pol II transcription machinery (see also Discussion). Additionally, in the siADA2a-transfected cells, following TPA induction, the H3K9/K14 acetylation dropped at the three tested promoters, when compared to the non-treated cells (Fig. 8b). This deregulated acetylation balance at the IE promoters may thus be responsible for the impaired

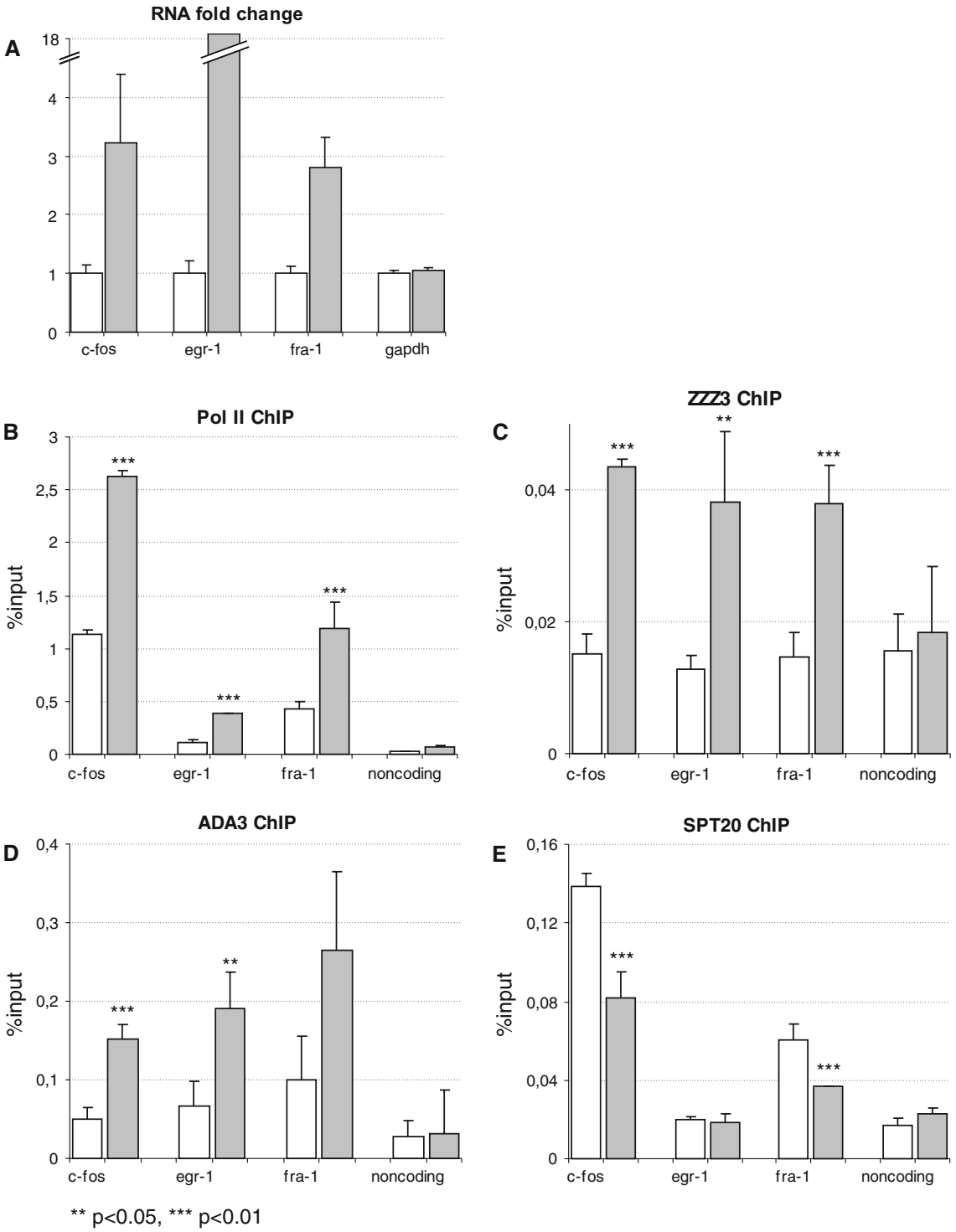


Fig. 6 hATAC subunits get recruited, together with RNA Pol II, to the promoter of immediate early genes after TPA treatment. **a** Expression level of IE genes *c-FOS*, *EGR-1* and *FRA-1*. Fold change of mRNA levels normalized by *cyclophylinB* are shown on the graph after 1 h of TPA treatment. *White bars* represent the nontreated sample, *gray bars* show the results obtained after TPA treatment. **b–e** ChIP results obtained with antibodies against RNA Pol II (**b**), ZZZ3 (**c**), ADA3 (**d**), and SPT20 (**e**) are shown on three IE gene promoters and the control non-coding region. Non-treated values (% input) are represented as *white bars*, the treated samples are shown by *gray bars*. Similar results were obtained in two biological replicates. Results obtained in a representative experiment are shown with SD values calculated from qPCR triplicates for each time point

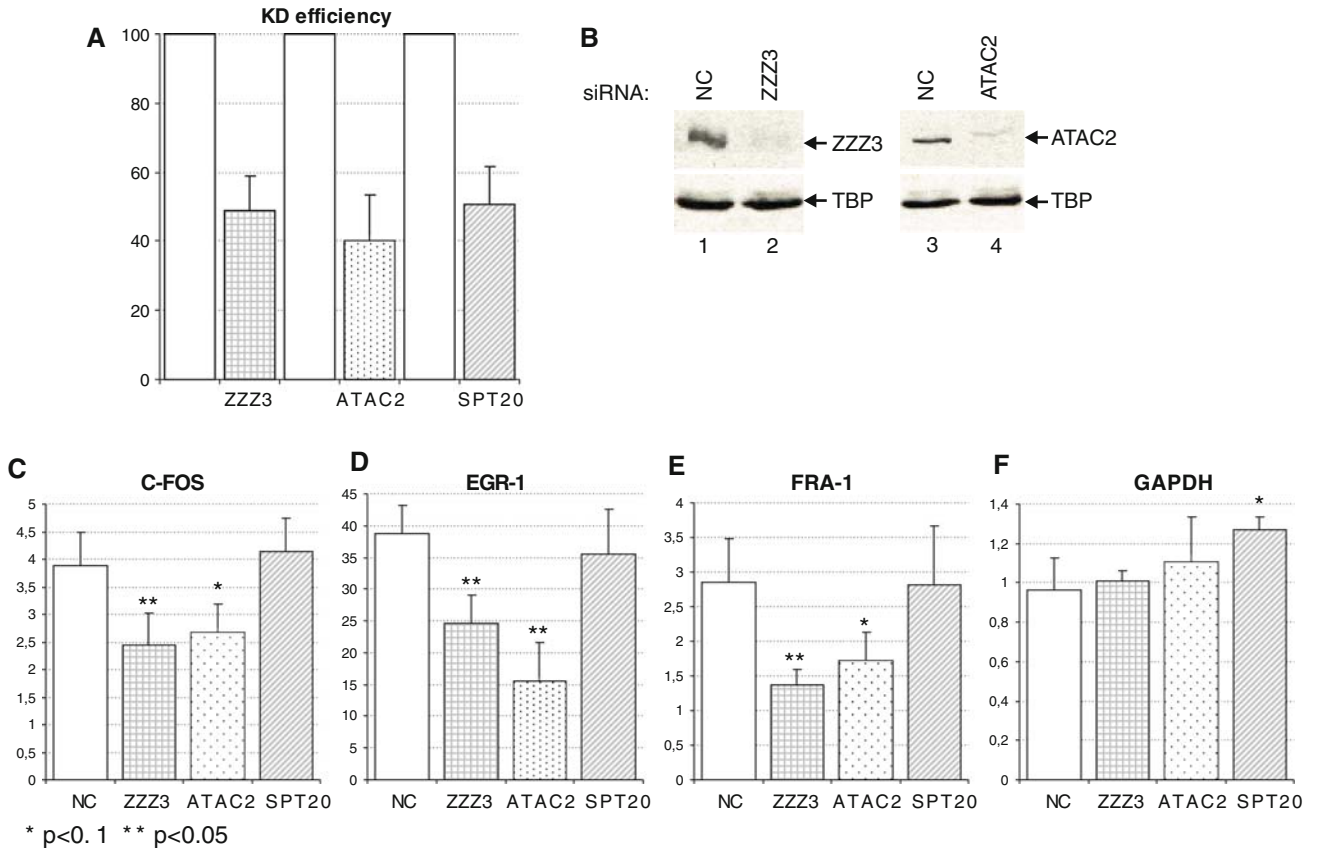


Fig. 7 hATAC but not hSAGA is indispensable for correct up regulation of IE genes after TPA treatment. **a** Knock-down efficiency of the three siRNAs (against ZZZ3, ATAC2 and SPT20) was analyzed by RT qPCR. The amount of residual mRNA is shown (*bars with different patterns*) compared to the negative control (*NC*) siRNA-transfected cells (*white bars*). **b** Knock-down efficiency of the siZZZ3 (*lane 2*) and siATAC2 (*lane 4*) was analyzed by western blotting and compared to negative control (*NC*) siRNA-transfected cells (*lanes 1, 3*). As a loading control, the same blot was developed with an anti-TBP antibody. **c–f.** Induction of IE gene expression after TPA treatment in cells transfected with different siRNAs (indicated on the *bottom of the graphs*). Results of IE genes or *GAPDH* mRNA quantification are shown as fold change over the non-treated samples and represent three independent experiments. The *different patterns of the bars* refer to **a**

activation of the IE genes in ATAC knock-down cells (see Fig. 7c–e). In all, our observations suggest that ATAC is indispensable for the transcriptional regulation of IE genes both in non-induced and in activated situation, while SAGA is not required.

Discussion

Metazoan ATAC complexes are conserved through evolution

Although GCN5 was the first enzyme identified to link histone modification and regulation of gene expression [46], it was only during recent years that data shed light on the existence of several GCN5-containing complexes in vivo in metazoans. The first indication came from the discovery that in metazoans and in plants the yeast Ada2 protein, an adaptor having effect on the activity of GCN5 activity [36, 47], has two orthologues: ADA2a and ADA2b [31, 48, 49]. These two orthologues in plants and *Drosophila* have distinct biological functions [50, 51]. Both ADA2a and ADA2b are essential in *Drosophila* [51], and the two proteins associate with dGCN5 in the context of two different complexes. While SAGA is a well-studied transcriptional co-activator complex [4, 52], the second complex, ATAC, was long over-looked in the different studies. The first signs of its presence in metazoans emerged during the determination of size of ADA2a- and ADA2b-containing complexes [16, 31]. The genetic analysis of mutant flies also suggested that two different assemblies are at play on the genome and that the substrate specificities of SAGA and ATAC differ in vivo [21, 51]. Further analysis of the dATAC complex established that a second HAT enzyme, dATAC2, is also part of the complex that seems to acetylate histone H4 at position H4K16 during *Drosophila* embryogenesis [17]. This observation, however, contradicts the results obtained on the polytene

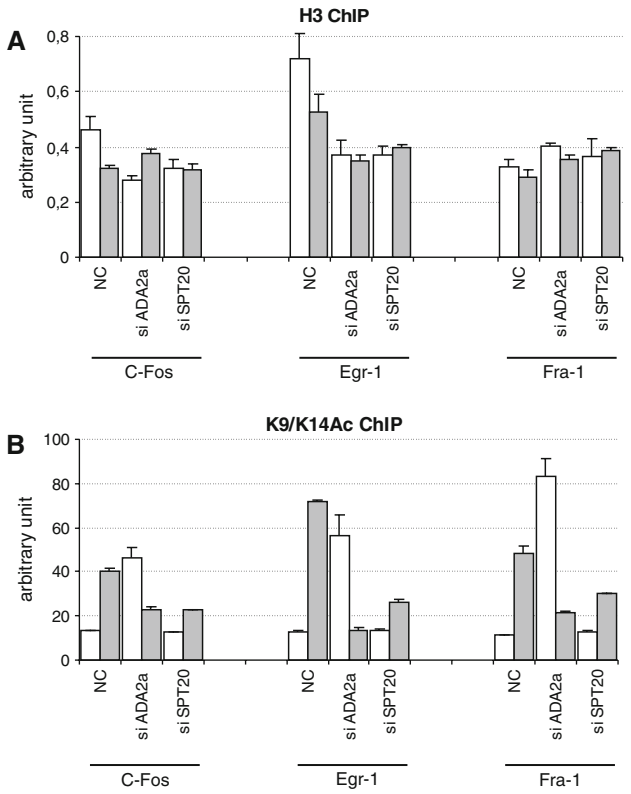


Fig. 8 Histone H3 acetylation is strongly affected in ATAC deficient cells at the promoter of IE genes. **a** ChIP results of siRNA-transfected cells obtained at the promoter of three IE genes (see Fig. 6a for indications) are represented before and after induction with TPA treatment (*white* and *gray bars*, respectively). The chromatin was immunoprecipitated with an antibody recognizing the core domain of H3. Results were normalized with those obtained at a non-transcribed region of the genome. The siRNA used for knocking down specific subunits of ATAC or SAGA is *marked on the axis*. **b** ChIP results obtained with an antibody specifically recognizing the H3K9/K14Ac signal shows perturbation at the promoter of TPA induced genes in siADA2a-transfected cells compared to the control (*NC*). In siSPT20-transfected cells, the response to stimulation at the acetylation level seems to be only slightly affected. Results are shown with SD values calculated from qPCR triplicates for each time point

chromosomes prepared from ADA2a mutant flies, which show a clear decrease in acetylated H4K5 and H4K12 levels, while the H4K16 hyperacetylation on the male X chromosome is not altered [21, 53]. The different results of the two systems still await a precise analysis and explanation.

Considering the high conservation of the known subunits of both dSAGA and dATAC through evolution (reviewed in [4, 54]), the identification of an ATAC-like complex in vertebrates was expected. Early publications provided evidence for the interaction between hADA2a and hGCN5 or its mammalian-specific homologue PCAF [14, 55], but results clearly showing that both ATAC and SAGA exist in human cells were lacking until the last year [15, 18]. Our present study validates the results obtained by overexpressing one of the ATAC subunits and using it for consequent immunopurification [15, 18]. At the same time, some of the identified components of the hATAC purified by an anti-flag IP seem to be missing in our endogenous system (see Table 1). Wang and colleagues reported the presence of additional proteins in the ATAC purification associating with the Flag-tagged YEATS2 (i.e., UBAP2L, MAP3K7, POLE4, and TBP) [18]. These subunits or their *Drosophila* orthologues have not been identified in our endogenous complex or in the other reported ATAC complexes ([15, 17] and our study) (Table 1). Interestingly, dHCF1 and its human homologue were described to be components of the dATAC- and the Flag-tagged hYEATS2-containing hATAC complexes (Table 1); however, in our endogenous ATAC preparations and in that purified by Guelman and colleagues [15], no hHCF1 was identified either by mass spectrometry or by western blot analysis.

A strikingly high level of conservation exist between the human and *Drosophila* ATAC complexes in composition (see Table 1), though some differences remain. So far no human homologue of dATAC4 (GABPβ2) has been found in human ATAC preparations. Moreover, the *Drosophila* and the human ATAC complexes seem to vary both in their substrate specificity and their overall size. While the dATAC complex is clearly smaller than dSAGA [17, 31, 48], analysis of the endogenous or the Flag-purified human ATAC highlights a surprisingly large and variable size of the human complexes (from 2 MDa to about 600 kDa) ([15], and our unpublished results). These observations predict the future identification of new subunits that might be human specific, as the sum of the masses of the already identified subunits is only about 800 kDa. Alternatively, hATAC complexes may have heterogeneous stoichiometry of certain subunits or simply heterogeneous shapes possibly due to binding of ATAC to nucleic acids or to other substrates. In contrast, the human SAGA complex seems to have a well-conserved size (of about 2 MDa) and structure.

Our results together with those published recently [15, 18] point to a difference between the human and the *Drosophila* ATAC complexes. While in the case of the fly ATAC complex a strong histone H4 specific activity was documented, in the case of hATAC this activity is hardly detectable in vitro and in vivo. The in vitro observed differences between the substrate specificity of the *Drosophila* and the human ATAC complexes could be explained by eventual non-optimal conditions used in the reactions; however, the in vivo differences are more difficult to reconcile. In this respect, it is worth noting that, while the *Drosophila* ATAC2 enzyme was shown to acetylate H4K16 [17], its human homologue shows no measurable activity on histone tail peptides as substrates in vitro, neither when isolated nor in the context of the ATAC complex (see Figs. 2d and 3c). This discrepancy may be explained

by the fact that the *Drosophila* ATAC2 protein contains a canonical PHD domain on its N-terminal end, while the vertebrate ATAC2 proteins lack key residues in this domain and contain only a putative Zn finger (ESM, Fig. 5). Thus, it is possible that, while dATAC2 is able to bind to histone tails via its PHD domain, as described for the PHD domain of TAF3 [56, 57], the vertebrate proteins are unable to do so. Consequently, the human ATAC complex may bind to other histone marks than the *Drosophila* complex, for example, via Tudor, WD40, or SANT domains present in its different subunits, and thus could acetylate different histone tail residues than the *Drosophila* complex. At the same time mice lacking ATAC2 die early in development and possess decreased global histone acetylation levels. However, since mATAC2 seems also to be essential for the integrity of the mammalian ATAC complex [15], future experiments should decide whether the drop in histone acetylation levels in *Atac2* knock-out mice is due to the lack of ATAC2 as a HAT enzyme per se, or rather to the lower level of the entire ATAC complex. The mouse ATAC2 ablation in *Atac2* knock-down or knock-out systems leads to a decrease in global H3K9, H4K5, H4K12 and H4K16 acetylation levels [15]. The decrease of H3K9Ac is in agreement with our results obtained following ADA2a knock-down in human cells; however, we did not detect any change in the different histone H4 acetylations. Thus, understanding the exact biological function, the precise acetyltrasferase specificity of the mammalian ATAC complexes (G-ATAC and P-ATAC), and the role of the second potential acetyltransferase, ATAC2, in these complexes still awaits further analysis.

ATAC and SAGA regulate different set of stress inducible target genes

Yeast SAGA was suggested to play a role in stress-regulated genes [58] and act as a locus-specific coactivator complex that binds close to the nuclosome-free region formed upstream of the +1 nucleosomes on expressed genes in the yeast genome [59]. Similarly, *Drosophila* SAGA subunits (TRRAP and GCN5) were detected at the inducible *hsp70* gene promoters following heat shock [60]. Human SAGA was shown to play a direct role in the up-regulation of p53-dependent genes following UV-C irradiation [36] and also in the regulation of endoplasmatic reticulum stress-induced genes [13]. In contrast, hSAGA did not seem to be involved in the regulation of IE genes induced by Na-arsenite stress [13], suggesting that hSAGA is not required as a promoter-specific coactivator at every stress-regulated gene promoter.

A stimulation-specific coactivator role of ATAC can be drawn from our new results, which seems to be conserved between *Drosophila* and human. ATAC gets recruited to TPA-induced transcription puffs on the polytene chromosomes of *Drosophila*, while these sites are deprived from dSAGA as no ADA2b, a SAGA specific subunit, was detected at the TPA-induced puffs. The same scenario stands for human cells, where hATAC subunits accumulate at the promoter of activated IE genes together with RNA Pol II after TPA induction. In agreement, the induction of the tested human IE genes was seriously compromised when cells harbored decreased level of different ATAC subunits (Fig. 7). Also, when a core ATAC subunit was knocked down by siRNA, the H3K9 and K14 acetylation marks decreased at the promoters following TPA stimulation (Fig. 8). Surprisingly, under "non-activated" conditions, the same acetylation marks were about 3- to 7-fold higher at the tested promoters than in the control cells. This increased basal histone H3 acetlyation level might be the indirect result of the perturbation we observed in the global H3S10P mark (Fig. 4). It is possible that the decreased H3S10 phosphorylation makes the chromatin at the IE promoters more permissive for the recruitment of another HAT complex than ATAC, which normally would not act at these sites in the cell. Thus, our results shed light on a dual function of ATAC at the IE gene promoters. On one hand, ATAC is indispensable for the induced transcription of these genes after stress. On the other hand, the complex is also required, probably indirectly, for the maintenance of the low level of H3K9/K14Ac marks at the same promoters in basal conditions.

Our results show that, in contrast to ATAC, TPA-inducible promoters lack hSAGA, providing evidence for the different recruitment pattern of the two complexes on the genome. In good agreement with this differential coactivator recruitment model, the group of R. Roeder has shown that to UV-stress-regulated gene promoters are occupied only by SAGA, but not ATAC [36]. These observations together demonstrate that two distinct types of GCN5- (or PCAF)-containing HAT complexes with potentially different coactivator activities exist in the cells to regulate different subset of induced genes.

One distinction between IE genes and other inducible genes is that the histones at IE genes get both acetylated and phosphorylated upon induction [61]. In this respect, it is worth noting that dATAC was shown to play a role in histone H3S10 phosphorylation [45]. In ADA2a mutant flies, decreased histone acetylation led to consequent decrease in H3S10 phosphorylation by the JIL kinase [45]. Our novel observations show that the mammalian ATAC complex is also required for global H3S10 phosphorylation establishing a conserved link between ATAC function and H3S10 phophorylation during evolution. Human cells with knock-down levels of ADA2a (hATAC subunit) also possess a strongly decreased global H3S10P level. Thus,

ATAC is acting both at global and locus-specific levels on the genome. Our results suggest that in ADA2a knockdown condition the loss of the ATAC HAT complex results in unbalanced H3 tail acetylation at the IE gene promoters, that in combination with the global decrease of the H3S10 phophorylation makes the activation of immediate early genes deficient. Thus, our data demonstrate that ATAC plays a crucial role in the transcriptional regulation of IE genes.

During evolution, the complexity increased not only at the level of gene number and genome size but also at regulatory circuits. Our present understanding on the composition and functioning of GCN5-containing HAT complexes is a nice example for this increase. In the unicellular organism *Saccharomyces cerevisiae*, Gcn5 is the component of two complexes (ySAGA and yADA) of which the catalytic core remains exactly the same, formed by the Ada2-Ada3-Gcn5 triad [62]. In *Drosophila*, we find two genes encoding ADA2 paralogues, *ADA2a* and *ADA2b*, and, as a consequence, two different complexes, dATAC and dSAGA, have evolved [17, 21, 31]. Furthermore, in dATAC, a second potential acetyltransferase enzyme was identified that brings another activity to the complex [17]. The complexity reaches its maximum in mammalian cells, where on the top of the two ADA2 proteins, two GCN5 homologues are also present (GCN5 and PCAF). Our data suggest, together with that of Gamper and colleagues [36], that both GCN5 and PCAF form SAGA- and ATAC-type complexes and that even all the four possibilities may coexist in one cell (this study, and not shown). At the same time, our results also provide evidence that the functional differences of these complexes materialize mainly in vivo. The numerous subunits surrounding the enzymes (GCN5 or PCAF) in such complexes can on the one hand affect their activity, as shown for hADA2b in the context of hSAGA [36], while on the other hand, these subunits possibly function as interaction surfaces for different transcription activators playing roles in distinct signaling pathways during development or stress response. In vitro dissection of the interactions within the complexes might provide data that will help understanding the biological role of each subunit. However, further in vivo genetic studies are indispensable for the comprehension of function in the cellular context. In the present study, we identified the components of the endogenous hATAC complex, and we also provide evidence that SAGA and ATAC complexes do not regulate the same subset of inducible genes that could be the starting point for future more detailed genome-wide studies.

Acknowledgments We are grateful to W. Herr, B. Malecova and T. Oelgeschläger for reagents, to R. and J. Conaway for sharing unpublished results, to R. Schneider, K. Kamieniarz, J. Bonnet and G. Lang for suggestions and help with the HAT assays, to the IGBMC core facilities, and to D. Devys, B. Malecova and T. Pankotai for critically reading the manuscript. Z.N. was supported by a fellowship from the European Community grant (HPRN-CT-00504228) and by a fellowship from the Fondation pour la Recherche Médicale (FRM). A.R. and A.K. were supported by fellowships of the Alsace Region. This work was supported by funds from CNRS, INSERM, Université de Strasbourg, the FRM, and European Community (HPRN-CT 00504228, STREP LSHG-CT-2004-502950 and EUTRACC LSHG-CT-2007-037445) and INCA (2008-UBICAN) grants.

References

1. Kouzarides T (2007) Chromatin modifications and their function. Cell 128:693–705
2. Hayes JJ, Hansen JC (2001) Nucleosomes and the chromatin fiber. Curr Opin Genet Dev 11:124–129
3. Li B, Carey M, Workman JL (2007) The role of chromatin during transcription. Cell 128:707–719
4. Nagy Z, Tora L (2007) Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation. Oncogene 26:5341–5357
5. Shahbazian MD, Grunstein M (2007) Functions of site-specific histone acetylation and deacetylation. Annu Rev Biochem 76:75–100
6. Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Lee TI, Bell GW, Walker K, Rolfe PA, Herbolsheimer E, Zeitlinger J, Lewitter F, Gifford DK, Young RA (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. Cell 122:517–527
7. Sobel RE, Cook RG, Perry CA, Annunziato AT, Allis CD (1995) Conservation of deposition-related acetylation sites in newly synthesized histones H3 and H4. Proc Natl Acad Sci USA 92:1237–1241
8. Shogren-Knaak M, Ishii H, Sun JM, Pazin MJ, Davie JR, Peterson CL (2006) Histone H4–K16 acetylation controls chromatin structure and protein interactions. Science 311:844–847
9. Carrozza MJ, Utley RT, Workman JL, Cote J (2003) The diverse functions of histone acetyltransferase complexes. Trends Genet 19:321–329
10. Roth SY, Denu JM, Allis CD (2001) Histone acetyltransferases. Annu Rev Biochem 70:81–120
11. Sterner DE, Berger SL (2000) Acetylation of histones and transcription-related factors. Microbiol Mol Biol Rev 64:435–459
12. Yang XJ (2004) The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases. Nucleic Acids Res 32:959–976
13. Nagy Z, Riss A, Romier C, le Guezennec X, Dongre AR, Orpinell M, Han J, Stunnenberg H, Tora L (2009) The human SPT20-containing SAGA complex plays a direct role in the regulation of endoplasmic reticulum stress-induced genes. Mol Cell Biol 29:1649–1660
14. Ogryzko VV, Kotani T, Zhang X, Schiltz RL, Howard T, Yang XJ, Howard BH, Qin J, Nakatani Y (1998) Histone-like TAFs within the PCAF histone acetylase complex. Cell 94:35–44
15. Guelman S, Kozuka K, Mao Y, Pham V, Solloway MJ, Wang J, Wu J, Lill JR, Zha J (2009) The double-histone-acetyltransferase complex ATAC is essential for mammalian development. Mol Cell Biol 29:1176–1188
16. Guelman S, Suganuma T, Florens L, Swanson SK, Kiesecker CL, Kusch T, Anderson S, Yates JR 3rd, Washburn MP, Abmayr SM, Workman JL (2006) Host cell factor and an uncharacterized SANT domain protein are stable components of ATAC, a novel

dAda2A/dGcn5-containing histone acetyltransferase complex in *Drosophila*. Mol Cell Biol 26:871–882
17. Suganuma T, Gutierrez JL, Li B, Florens L, Swanson SK, Washburn MP, Abmayr SM, Workman JL (2008) ATAC is a double histone acetyltransferase complex that stimulates nucleosome sliding. Nat Struct Mol Biol 15:364–372
18. Wang YL, Faiola F, Xu M, Pan S, Martinez E (2008) Human ATAC Is a GCN5/PCAF-containing acetylase complex with a novel NC2-like histone fold module that interacts with the TATA-binding protein. J Biol Chem 283:33808–33815
19. Brand M, Yamamoto K, Staub A, Tora L (1999) Identification of TATA-binding protein-free TAFII-containing complex subunits suggests a role in nucleosome acetylation and signal transduction. J Biol Chem 274:18285–18289
20. Grant PA, Eberharter A, John S, Cook RG, Turner BM, Workman JL (1999) Expanded lysine acetylation specificity of Gcn5 in native complexes. J Biol Chem 274:5895–5900
21. Ciurciu A, Komonyi O, Pankotai T, Boros IM (2006) The *Drosophila* histone acetyltransferase Gcn5 and transcriptional adaptor Ada2a are involved in nucleosomal histone H4 acetylation. Mol Cell Biol 26:9413–9423
22. Brand M, Moggs JG, Oulad-Abdelghani M, Lejeune F, Dilworth FJ, Stevenin J, Almouzni G, Tora L (2001) UV-damaged DNA-binding protein in the TFTC complex links DNA damage recognition to nucleosome acetylation. EMBO J 20:3187–3196
23. Wieczorek E, Brand M, Jacq X, Tora L (1998) Function of TAF(II)-containing complex without TBP in transcription by RNA polymerase II. Nature 393:187–191
24. Helmlinger D, Hardy S, Sasorith S, Klein F, Robert F, Weber C, Miguet L, Potier N, Van-Dorsselaer A, Wurtz JM, Mandel JL, Tora L, Devys D (2004) Ataxin-7 is a subunit of GCN5 histone acetyltransferase-containing complexes. Hum Mol Genet 13:1257–1265
25. Brou C, Chaudhary S, Davidson I, Lutz Y, Wu J, Egly JM, Tora L, Chambon P (1993) Distinct TFIID complexes mediate the effect of different transcriptional activators. EMBO J 12:489–499
26. Zhao Y, Lang G, Ito S, Bonnet J, Metzger E, Sawatsubashi S, Suzuki E, Le Guezennec X, Stunnenberg HG, Krasnov A, Georgieva SG, Schule R, Takeyama K, Kato S, Tora L, Devys D (2008) A TFTC/STAGA module mediates histone H2A and H2B deubiquitination, coactivates nuclear receptors, and counteracts heterochromatin silencing. Mol Cell 29:92–101
27. Zohn IE, Li Y, Skolnik EY, Anderson KV, Han J, Niswander L (2006) p38 and a p38-interacting protein are critical for downregulation of E-cadherin during mouse gastrulation. Cell 125:957–969
28. Wysocka J, Myers MP, Laherty CD, Eisenman RN, Herr W (2003) Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3–K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1. Genes Dev 17:896–911
29. Julien E, Herr W (2003) Proteolytic processing is necessary to separate and ensure proper cell growth and cytokinesis functions of HCF-1. EMBO J 22:2360–2369
30. Malecova B, Gross P, Boyer-Guittaut M, Yavuz S, Oelgeschlager T (2007) The initiator core promoter element antagonizes repression of TATA-directed transcription by negative cofactor NC2. J Biol Chem 282:24767–24776
31. Muratoglu S, Georgieva S, Papai G, Scheer E, Enunlu I, Komonyi O, Cserpan I, Lebedeva L, Nabirochkina E, Udvardy A, Tora L, Boros I (2003) Two different *Drosophila* ADA2 homologues are present in distinct GCN5 histone acetyltransferase-containing complexes. Mol Cell Biol 23:306–321
32. Leurent C, Sanders SL, Demeny MA, Garbett KA, Ruhlmann C, Weil PA, Tora L, Schultz P (2004) Mapping key functional sites within yeast TFIID. EMBO J 23:719–727
33. Demeny MA, Soutoglou E, Nagy Z, Scheer E, Janoshazi A, Richardot M, Argentini M, Kessler P, Tora L (2007) Identification of a small TAF complex and its role in the assembly of TAF-containing complexes. PLoS ONE 2:e316
34. Schnitzler GR (2001) Isolation of histones and nucleosome cores from mammalian cells. Curr Protoc Mol Biol Ch 21, Unit 21.5
35. Kouskouti A, Scheer E, Staub A, Tora L, Talianidis I (2004) Gene-specific modulation of TAF10 function by SET9-mediated methylation. Mol Cell 14:175–182
36. Gamper AM, Roeder RG (2008) Multivalent binding of p53 to the STAGA complex mediates coactivator recruitment after UV damage. Mol Cell Biol 28:2517–2527
37. Fukuyama K, Yoshida M, Yamashita A, Deyama T, Baba M, Suzuki A, Mohri H, Ikezawa Z, Nakajima H, Hirai S, Ohno S (2000) MAPK upstream kinase (MUK)-binding inhibitory protein, a negative regulator of MUK/dual leucine zipper-bearing kinase/leucine zipper protein kinase. J Biol Chem 275:21247–21254
38. Wysocka J, Swigut T, Milne TA, Dou Y, Zhang X, Burlingame AL, Roeder RG, Brivanlou AH, Allis CD (2005) WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. Cell 121:859–872
39. Inostroza JA, Mermelstein FH, Ha I, Lane WS, Reinberg D (1992) Dr1, a TATA-binding protein-associated phosphoprotein and inhibitor of class II gene transcription. Cell 70:477–489
40. Meisterernst M, Roeder RG (1991) Family of proteins that interact with TFIID and regulate promoter activity. Cell 67:557–567
41. Thomas MC, Chiang CM (2006) The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol 41:105–178
42. Kamada K, Shu F, Chen H, Malik S, Stelzer G, Roeder RG, Meisterernst M, Burley SK (2001) Crystal structure of negative cofactor 2 recognizing the TBP-DNA transcription complex. Cell 106:71–81
43. Bu P, Evrard YA, Lozano G, Dent SY (2007) Loss of Gcn5 acetyltransferase activity leads to neural tube closure defects and exencephaly in mouse embryos. Mol Cell Biol 27:3405–3416
44. Kikuchi H, Takami Y, Nakayama T (2005) GCN5: a supervisor in all-inclusive control of vertebrate cell cycle progression through transcription regulation of various cell cycle-related genes. Gene 347:83–97
45. Ciurciu A, Komonyi O, Boros IM (2008) Loss of ATAC-specific acetylation of histone H4 at Lys12 reduces binding of JIL-1 to chromatin and phosphorylation of histone H3 at Ser10. J Cell Sci 121:3366–3372
46. Brownell JE, Zhou J, Ranalli T, Kobayashi R, Edmondson DG, Roth SY, Allis CD (1996) Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell 84:843–851
47. Balasubramanian R, Pray-Grant MG, Selleck W, Grant PA, Tan S (2002) Role of the Ada2 and Ada3 transcriptional coactivators in histone acetylation. J Biol Chem 277:7989–7995
48. Kusch T, Guelman S, Abmayr SM, Workman JL (2003) Two Drosophila Ada2 homologues function in different multiprotein complexes. Mol Cell Biol 23:3305–3319
49. Stockinger EJ, Mao Y, Regier MK, Triezenberg SJ, Thomashow MF (2001) Transcriptional adaptor and histone acetyltransferase proteins in Arabidopsis and their interactions with CBF1, a transcriptional activator involved in cold-regulated gene expression. Nucleic Acids Res 29:1524–1533
50. Hark AT, Vlachonasios KE, Pavangadkar KA, Rao S, Gordon H, Adamakis ID, Kaldis A, Thomashow MF, Triezenberg SJ (2009) Two Arabidopsis orthologs of the transcriptional coactivator ADA2 have distinct biological functions. Biochim Biophys Acta 1789:117–124

51. Pankotai T, Komonyi O, Bodai L, Ujfaludi Z, Muratoglu S, Ciurciu A, Tora L, Szabad J, Boros I (2005) The homologous *Drosophila* transcriptional adaptors ADA2a and ADA2b are both required for normal development but have different functions. Mol Cell Biol 25:8215–8227
52. Baker SP, Grant PA (2007) The SAGA continues: expanding the cellular role of a transcriptional co-activator complex. Oncogene 26:5329–5340
53. Carre C, Ciurciu A, Komonyi O, Jacquier C, Fagegaltier D, Pidoux J, Tricoire H, Tora L, Boros IM, Antoniewski C (2008) The *Drosophila* NURF remodelling and the ATAC histone acetylase complexes functionally interact and are required for global chromosome organization. EMBO Rep 9:187–192
54. Lee KK, Workman JL (2007) Histone acetyltransferase complexes: one size doesn't fit all. Nat Rev Mol Cell Biol 8:284–295
55. Forsberg EC, Lam LT, Yang XJ, Nakatani Y, Bresnick EH (1997) Human histone acetyltransferase GCN5 exists in a stable macromolecular complex lacking the adapter ADA2. Biochemistry 36:15918–15924
56. van Ingen H, van Schaik FM, Wienk H, Ballering J, Rehmann H, Dechesne AC, Kruijzer JA, Liskamp RM, Timmers HT, Boelens R (2008) Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3. Structure 16:1245–1256
57. Vermeulen M, Mulder KW, Denissov S, Pijnappel WW, van Schaik FM, Varier RA, Baltissen MP, Stunnenberg HG, Mann M, Timmers HT (2007) Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell 131:58–69
58. Huisinga KL, Pugh BF (2004) A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in *Saccharomyces* cerevisiae. Mol Cell 13:573–585
59. Venters BJ, Pugh BF (2009) A canonical promoter organization of the transcription machinery and its regulators in the *Saccharomyces* genome. Genome Res 19:360–371
60. Lebedeva LA, Nabirochkina EN, Kurshakova MM, Robert F, Krasnov AN, Evgen'ev MB, Kadonaga JT, Georgieva SG, Tora L (2005) Occupancy of the Drosophila hsp70 promoter by a subset of basal transcription factors diminishes upon transcriptional activation. Proc Natl Acad Sci USA 102:18087–18092
61. Clayton AL, Mahadevan LC (2003) MAP kinase-mediated phosphoacetylation of histone H3 and inducible gene regulation. FEBS Lett 546:51–58
62. Sterner DE, Grant PA, Roberts SM, Duggan LJ, Belotserkovskaya R, Pacella LA, Winston F, Workman JL, Berger SL (1999) Functional organization of the yeast SAGA complex: distinct components involved in structural integrity, nucleosome acetylation, and TATA-binding protein interaction. Mol Cell Biol 19:86–98