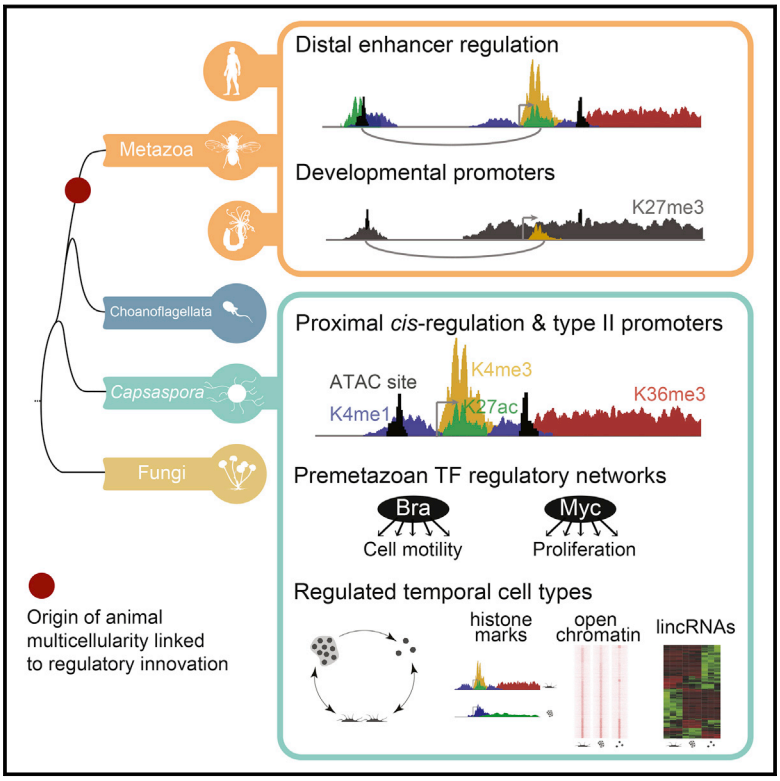
Article

Cell

The Dynamic Regulatory Genome of *Capsaspora* and the Origin of Animal Multicellularity

Graphical Abstract



Authors

Arnau Sebé-Pedrós, Cecilia Ballaré, Helena Parra-Acero, ..., José Luis Gómez-Skarmeta, Luciano Di Croce, Iñaki Ruiz-Trillo

Correspondence

arnau.sebe-pedros@weizmann.ac.il (A.S.-P.), inaki.ruiz@ibe.upf-csic.es (I.R.-T.)

In Brief

Analysis of the regulatory genome in one of our closest unicellular relatives suggests that the appearance of developmental promoters and distal enhancer elements, rather than of gene innovations, may have been the critical events underlying the origin of multicellular organisms.

Highlights

- Dynamic chromatin states and *cis*-regulatory sites in a unicellular context
- Elaborate lincRNA regulation associated with a unicellular life cycle
- Premetazoan origin of core metazoan developmental transcription-factor networks
- Distal enhancer elements are a metazoan innovation

Accession Numbers

PXD002342
GSE71131

Sebé-Pedrós et al., 2016, Cell *165*, 1224–1237
May 19, 2016
http://dx.doi.org/10.1016/j.cell.2016.03.034

CellPress

Cell

Article

The Dynamic Regulatory Genome of *Capsaspora* and the Origin of Animal Multicellularity

Arnau Sebé-Pedrós,[1,7,*] Cecilia Ballaré,[2,3] Helena Parra-Acero,[1] Cristina Chiva,[2,3] Juan J. Tena,[4] Eduard Sabidó,[2,3] José Luis Gómez-Skarmeta,[4] Luciano Di Croce,[2,3,5] and Iñaki Ruiz-Trillo[1,5,6,*]

[1]Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Spain
[2]Center for Genomic Regulation, Doctor Aiguader 88, 08003 Barcelona, Spain
[3]Universitat Pompeu Fabra (UPF), Doctor Aiguader 88, 08003 Barcelona, Spain
[4]Centro Andaluz de Biología del Desarrollo (CABD), CSIC-Universidad Pablo de Olavide-Junta de Andalucía, Carretera de Utrera Km1, 41013 Sevilla, Spain
[5]Institució Catalana de Recerca i Estudis Avançats, Pg Lluis Companys 23, 08010 Barcelona, Spain
[6]Departament de Genètica, Universitat de Barcelona, 08028 Barcelona, Spain
[7]Present address: Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel
*Correspondence: arnau.sebe-pedros@weizmann.ac.il (A.S.-P.), inaki.ruiz@ibe.upf-csic.es (I.R.-T.)
http://dx.doi.org/10.1016/j.cell.2016.03.034

SUMMARY

The unicellular ancestor of animals had a complex repertoire of genes linked to multicellular processes. This suggests that changes in the regulatory genome, rather than in gene innovation, were key to the origin of animals. Here, we carry out multiple functional genomic assays in *Capsaspora owczarzaki*, the unicellular relative of animals with the largest known gene repertoire for transcriptional regulation. We show that changing chromatin states, differential lincRNA expression, and dynamic *cis*-regulatory sites are associated with life cycle transitions in *Capsaspora.* Moreover, we demonstrate conservation of animal developmental transcription-factor networks and extensive network interconnection in this premetazoan organism. In contrast, however, *Capsaspora* lacks animal promoter types, and its regulatory sites are small, proximal, and lack signatures of animal enhancers. Overall, our results indicate that the emergence of animal multicellularity was linked to a major shift in genome *cis*-regulatory complexity, most notably the appearance of distal enhancer regulation.

INTRODUCTION

A defining feature of multicellular animals is their capacity to generate multiple specialized cell types through temporally and spatially regulated developmental programs. These programs of individual cell differentiation involve the generation of cell-specific transcriptional profiles. Recent genomic analyses, however, have shown that the unicellular ancestor of Metazoa already had a complex gene repertoire involved in multicellular functions, including specific differentiation programs (Fairclough et al., 2013; King et al., 2008; de Mendoza et al., 2015; Sebé-Pedrós et al., 2013b; Srivastava et al., 2010; Suga et al., 2013).

Since the origin of animals was not solely dependent on the appearance of new genes, it is likely that animal evolution involved a shift in the genome regulatory capabilities required to generate cell-type-specific transcriptional profiles during animal development. In animals, these profiles are established and maintained by a complex combination of chromatin regulatory dynamics, distal *cis*-regulatory elements, and transcription factor networks (Bernstein et al., 2007; Buecker and Wysocka, 2012; Ho et al., 2014; de Laat and Duboule, 2013; Levine, 2010; Levine and Tjian, 2003). Interestingly, a recent analysis of an early branching and morphologically simple animal, the cnidarian *Nematostella vectensis*, has shown that cnidarians and bilaterians share a conserved gene regulatory landscape (Schwaiger et al., 2014). However, it is unclear whether these ancient genome regulatory features are animal innovations or whether they were already present in the unicellular ancestor of Metazoa.

To determine the timing and importance of regulatory changes in the origin of Metazoa, we need to unravel the genomic regulation of the extant animal relatives. Among the closest extant unicellular relatives of Metazoa, the amoeboid filasterean *Capsaspora owczarzaki* (herein *Capsaspora*), has the richest repertoire of transcription factors described to date (Sebé-Pedrós et al., 2011). These include genes, such as *Brachyury*, *Myc*, and *Runx*, that are essential for animal development. Moreover, *Capsaspora* is known to differentiate into three temporal life stages that are transcriptionally tightly regulated (Sebé-Pedrós et al., 2013b). These temporal cell types include (1) a filopodiated amoeba, which corresponds to the proliferative trophic stage, (2) an aggregative multicellular stage, in which the cells produces an extracellular matrix, and (3) a cystic resistance form without filopodia (see an schematic representation of the life cycle in Figure 3). Its key phylogenetic position as the sister group of animals and choanoflagellates, its rich gene repertoire, and the observed regulatory capabilities of *Capsaspora*, therefore, make it an ideal candidate to explore the origin of animal genome regulation.

The advent of functional genomics assays based on next-generation sequencing (NGS) has revolutionized the study of the regulatory genome. These techniques have shown that different chromatin biochemical signatures and accessibility are associated with *cis*-regulatory elements (Creyghton et al., 2010; Rada-Iglesias et al., 2011; Thurman et al., 2012), promoter types (Lenhard et al., 2012), ncRNAs (Marques et al., 2013), and gene transcriptional states (Dunham et al., 2012; Schwaiger et al., 2014). To date, however, this new paradigm has only been systematically applied to a handful of model species (Ho et al., 2014), and our understanding of most eukaryotic genomes remains limited to primary sequence. These techniques hold the potential to go beyond genome content description and systematically explore genome regulation in non-model systems like *Capsaspora*. Here, we apply these principles to study the dynamic *Capsaspora* regulatory genome in a comparative evolutionary framework and demonstrate that a major change in genome regulation was linked to the origin and the subsequent diversification of animal body plans.

RESULTS

Histone Modifications in *Capsaspora*

Posttranslational modifications of histone tails (hPTMs) are important components of the regulatory genomic landscape in eukaryotes. hPTMs play a crucial role in maintaining and transmitting on-off transcriptional signals (Zhou et al., 2011) by modifying the chromatin structure, and they are associated with specific regulatory elements in animals (Creyghton et al., 2010; Rada-Iglesias et al., 2011). To determine whether hPTMs are conserved between animals and their closest relatives or across all eukaryotes, we first analyzed the hPTMs of *Capsaspora* by chemical derivatization coupled to mass spectrometry and compared those with eukaryotes for which hPTMs are known (Figures 1 and S1). We found that H3 and H4 modifications are largely conserved across the eukaryotes analyzed. In contrast, we identified several novel *Capsaspora*-specific modifications in H2B and H2AZ and a *Capsaspora*-specific H2A variant, indicating that H2AZ and H2B histones and histone variants are the fastest evolving components of the histone code. Additionally, there was a correspondence between hPTMs and histone-modifying enzymes in the genome of *Capsaspora* (Figure 1). An example is the lack of H3K9me3 and H3K27me3, the two best-characterized animal repressive marks, co-occurring with the absence of the enzymes responsible for writing and erasing them (Suv3/9, G9a, and SETD1B for H3K9me3 and EZH2 (PRC2 complex) for H3K27me3). Despite some linage-specific changes, H3 and H4 hPTMS are mostly conserved across eukaryotes, and thus, informative comparative analyses can be performed across distant taxa.

Dynamic Chromatin States in *Capsaspora*

To investigate the genome-wide distribution of *Capsaspora* hPTMs across temporally segregated cell types, we selected those marks that have been widely used in animals to characterize chromatin states (Ho et al., 2014). Chromatin immunoprecipitation sequencing (ChIP-seq) was carried out for H3 lysine 4 trimethylation and monomethylation (H3K4me3 and H3K4me1), H3 lysine 27 acetylation (H3K27ac), and H3 lysine 36 trimethylation (H3K36me3). Deep-sequencing reads were mapped in the *Capsaspora* genome, and their correlation with different genomic features and gene expression was analyzed (Figures 2, S2, and S3). Additionally, we undertook transposase-accessible chromatin sequencing (ATAC-seq) (Buenrostro et al., 2013) in each cell stage in order to interrogate nucleosome positioning and accessible chromatin as a proxy for active *cis*-regulatory elements. Normalized ChIP-seq read coverage around the transcription start site (TSS) reveals a unimodal H3K4me3 peak well positioned after the TSS of active genes that strongly colocalizes with H3K27ac (Figure 2A). In contrast, two sharp H3K4me1 peaks appear flanking H3K4me3/H3K27ac peaks, both before and after the TSS. Finally, H3K36me3 spreads through the gene bodies of active genes. All these marks correlate with the level of expression of active genes (Figure 2A), in a pattern similar to that observed in human cells (van Galen et al., 2016). It must be noted, though, that histone modifications might also be related to other regulatory processes; e.g., H3K36me3 has been linked to splicing (Kolasinska-Zwierz et al., 2009). Nucleosomes appear in highly ordered positions after the TSS of expressed genes, while, in contrast, nucleosomal fuzziness (which measures the deviation of each nucleosome position in the cell population) increases in weakly expressed and silent genes (Figures 2B and 2C). ATAC nucleosome-free reads are preferentially distributed in the surroundings of the TSS (Figure 2B). Finally, we also analyzed the distribution of RNApolII in *Capsaspora* genes (Figure S2), showing a strong peak around the TSS. In contrast, C-terminal domain (CTD) S2 phosphorylated RNA polymerase II (RNA Pol II) is distributed along the gene body, consistent with the known association of this S2 phosphorylated RNA Pol II form with transcriptional elongation (Egloff et al., 2012; Eick and Geyer, 2013; Schwer and Shuman, 2011). RNA Pol II coverage is associated with increased gene expression (Figure S2B) and changes dynamically between life stages (Figure S2C).

Next, we integrated these hPTM maps and ATAC nucleosome-free reads in order to predict chromatin states and their genome-wide distribution in *Capsaspora*, using a hidden Markov model (ChromHMM) (Ernst and Kellis, 2012) (Figure 2D). Overall, we defined seven different chromatin states that preferentially associated with specific genomic features (Figure 2E). For example, state one (defined by H3K36me3) is the most abundant and associates with coding regions and non-first introns (Figure 2E), consistent with the function of H3K36me3 as a transcriptional elongation mark (Dunham et al., 2012). In contrast, state seven corresponds to ATAC nucleosome-free signal, together with H3K4me1, and is strongly enriched around TSS (Figure 2E), corresponding to potential regulatory sites.

Given the absence of known repressive marks in *Capsaspora* (see Figure 1), we asked whether strongly repressed genes show any particular biochemical signature. Thus, we compared lowly expressed genes (<2 FKPMs) with active genes (Figures 2E and 2F) and observed a particular profile in which H3K4me1 shifts from two flanking peaks to a single post-TSS peak, H3K27ac is spread across the gene body, and both H3K4me3 and H3K36me3 are absent (Figures 2F, S2, and S3). Similarly, we observe a strong enrichment of state four across the gene

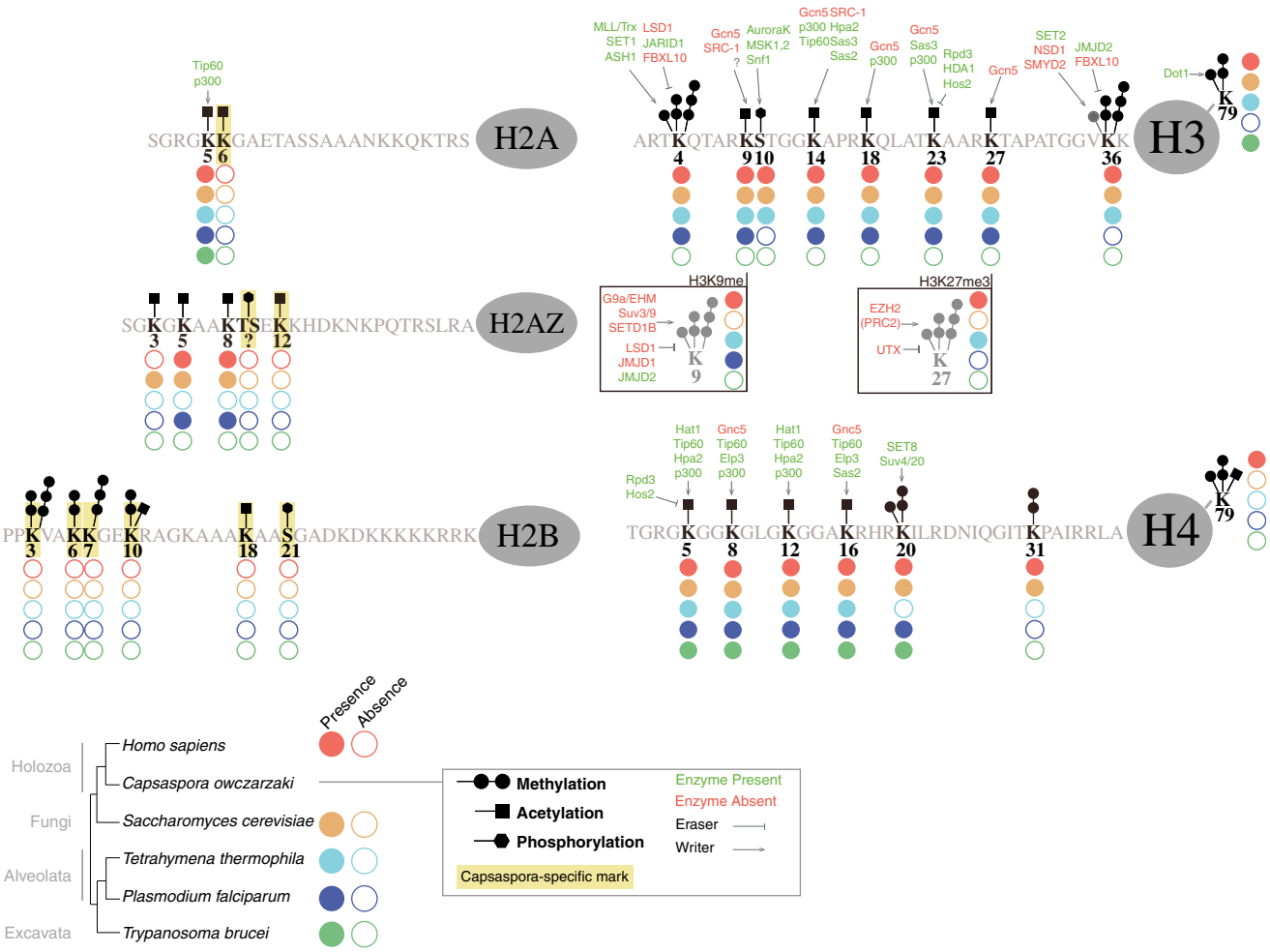


Figure 1. Histone Modifications in *Capsaspora*

Histone N-terminal tail sequences of *Capsaspora* with the identified posttranslational modifications are shown. Below: filled or empty circles indicate whether the particular histone mark is present or absent, respectively, in the different eukaryotic species represented in the phylogenetic tree (left). Above: the presence (green) or absence (red) of specific histone modifiers in the *Capsaspora* genome is shown; both enzymes that add the mark (writers) and enzymes that remove it (erasers) are indicated. *Capsaspora*-specific marks are highlighted in yellow. The repressive marks H3K9me and H3K27me3 are absent in *Capsaspora* and indicated separately in a box below the corresponding position.

See also Figure S1.

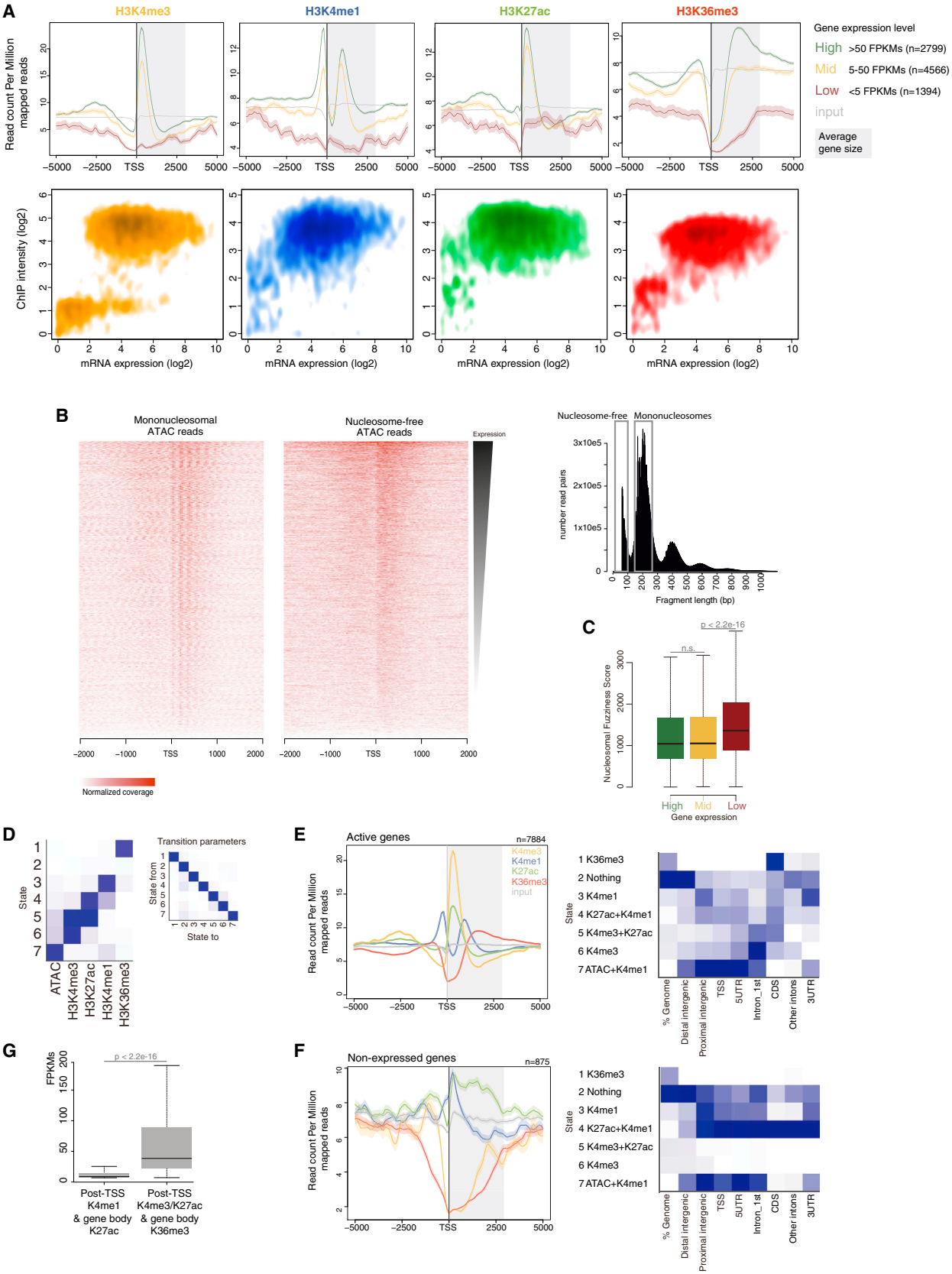
body and of state three around TSS. If we specifically select genes with H3K27ac across the gene body (>800 bp from TSS) and post-TSS H3K4me1 peaks (TSS+800 bp), we recover the population of repressed genes (Figure 2G). This signature of repression has never been described in any other organism and might represent a *Capsaspora*-specific mechanism.

Finally, we evaluated how changes in chromatin features correlate with life stage transitions in *Capsaspora*. First, we observed that chromatin marks change between life stages, correlating with changes in genes expression (Figures 3A, 3B, and S3). Second, we treated *Capsaspora* cells with Trichostatin A (TSA), a widely used histone deacetylase (HDAC) inhibitor (Simola et al., 2016), in order to study the role of histone acetylation in the life cycle of *Capsaspora*. Treatment with 3 μM TSA blocked life cycle transitions, e.g., from cystic to filopodial stage (Figure 3C). As expected when blocking HDACs, TSA induced an increase in histone acetylation levels (Figure 3D). Using RNA sequencing (RNA-seq), we also observed that TSA caused a generalized activation of gene expression (Figure 3E). These observations directly link histone modifications with life cycle transitions and gene expression in *Capsaspora*.

Overall, we obtained high-coverage linear maps of multiple epigenomic features, which show consistent patterns of association with expression states, specific genomic regions and temporal cell-type transitions. These maps allowed us to further systematically dissect functional elements in *Capsaspora* genome.

The Origin of Animal Promoter Types

To understand the evolution of proximal promoter chromatin regulatory signatures, we compared TSS profiles of *Capsaspora* with different metazoan taxa and *Saccharomyces cerevisiae* using publicly available ChIP-seq datasets (Figure 4). All species



(legend on next page)

show well-positioned post-TSS H3K4me3/H3K27ac peaks correlated with active gene expression. *Homo sapiens* show a strong bimodal peak, whereas a *Drosophila melanogaster* and *Caenorhabditis elegans* show weak biomodality. In comparison, the cnidarian *Nematostella vectensis*, as well as *Capsaspora* and *Saccharomyces*, present sharp unimodal post-TSS H3K4me3/H3K27ac peaks. This difference between bilaterians and others could be related to the presence (bimodality) or absence (unimodality) of anti-sense transcript production from some TSS (Ho et al., 2014). Moreover, H3K36me3 is present in the gene bodies of active genes in all species. Interestingly, H3K4me1 is enriched on both sides of the TSS in animals and *Capsaspora*. The signal is weaker and less sharp in animals, whereas it is sharp and complementary to the H3K4me3/H3K27ac peak in *Capsaspora*. In contrast, *Saccharomyces* has only one post-TSS H3K4me1 peak after the H3K4me3/H3K27ac peak.

The distribution of histone modifications around TSS has been used to define three different promoter types in metazoans (Lenhard et al., 2012). Type I promoters are associated with tissue-specific expression in terminal-differentiated cell types, and they are characterized by fuzzy nucleosomes, strongly positioned H3K4me3 and H3K27ac peaks, and no H3K4me1 and H3K27me3 marks. Type II promoters are found in ubiquitously expressed genes and show strongly positioned nucleosomes and flanking H3K4me1 marks (in addition to post-TSS H3K4me3 and K27ac). Finally, type III promoters, also called bivalent promoters, are associated with developmentally regulated genes and present both activation (H3K4me3) and repression (H3K27me3) marks (Lenhard et al., 2012). Thus, the different configurations observed here are likely to reflect different promoter specification modes. Interestingly, *Capsapora* TSS signatures strongly resemble those of animal type II promoters (also called "ubiquitous"), including highly ordered nucleosome positioning (Figure 2). In contrast, no type I (without flanking H3K4me1 and fuzzy nucleosomes) or type III (H3K27me3-regulated developmental promoters) promoters could be identified in *Capsaspora*. This indicates that type I and type III promoters are animal innovations and related to the emergence of cell-type-specific (type I) and developmental regulation of gene expression (type III).

Premetazoan Long Intergenic Non-coding RNAs Regulation

Long intergenic non-coding RNAs (lincRNAs) are an important component of animal genome regulation (Marques and Ponting, 2014; Ulitsky and Bartel, 2013). lincRNAs exert multiple developmental and cell-type-specific regulatory functions, and their number is greatly expanded in multicellular animals and plants (Gaiti et al., 2015; Kapusta and Feschotte, 2014; Ulitsky and Bartel, 2013). In order to understand the evolution of lincRNAs in the lineage leading to metazoan multicellularity, we used deep strand-specific polyA-enriched RNA-sequencing data to annotate lincRNAs in *Capsaspora*. After applying multiple filters, we predicted 632 lincRNAs and validated 17 of them by RT-PCR (Figures S4 and S5). This is less than those identified in multicellular animals, but more than those found in yeast (Kapusta and Feschotte, 2014). *Capsaspora* lincRNAs show dynamic expression (Figure S4A) and have multiple features that differentiate them from coding genes (Figure S4B). Interestingly, we found that predicted *Capsaspora* lincRNAs can be separated into two populations based on their association with H3K4me1 and H3K4me3 (Figures S4C and S4E), resembling those found in mouse lincRNAs (Marques et al., 2013). Moreover, similar to mouse, these two lincRNA populations show only slight differences in length, expression level, and expression variation (Figure S4D), so the functional significance of these two populations remains unclear. Thus, our data reveal that elaborate lincRNA genome regulation was already present in unicellular premetazoans.

Accessible Chromatin Landscape of Capsaspora

Transcription factors and other regulatory proteins bind to discrete DNA sequences, creating nucleosome-depleted areas of high-nuclease/transposase accessibility. We therefore used high-coverage nucleosome-free ATAC reads to identify all active regulatory sites in the *Capsaspora* genome and study their distribution and cell-type dynamics (Figures 5 and S6). In *Capsaspora*, 95% of the genome lies within 6.4 kb of one of the 11,927 discrete regulatory sites identified, and 63% of genes are associated with at least one site (Figure 5A). As an estimation of the number of regulatory inputs, we calculated the number of

Figure 2. Genome-Wide Chromatin Annotation in *Capsaspora*

(A) Top: TSS-centered average normalized read coverage plots of hPTMs in the filopodial stage for genes with high (green), intermediate (yellow), and low (red) expression levels. The x axis spans −5 to +5 kb around the TSS. The shaded gray area represents the average size of *Capsaspora* genes. Bottom: scatterplots of hPTMs coverage (log2 normalized reads) compared to mRNA expression levels (log2 fragments per kilobase of transcript per million mapped reads [FPKMs]).

(B) Heatmaps of ATAC mononucleosome-associated (left) and nucleosome-free (right) reads centered around the TSS of genes sorted by level of expression in the filopodial stage. Right: histogram showing an example of the distribution of ATAC-seq fragment sizes obtained.

(C) Boxplot representing the mean fuzziness score of the first four post-TSS nucleosomes of genes grouped by the level of expression in the filopodial stage. The p value is indicated for the Wilcoxon rank-sum test.

(D) Heatmaps representing the emission (left) and transition (right) parameters of a seven-state hidden Markov model. In the left heatmap, the white-blue (0–1) scale represents the frequency with which a given mark is found at genomic positions corresponding to the chromatin state. In the right heatmap, the white-blue (0–1) scale represents the frequency with which a given state changes into another state at the neighboring location.

(E) Chromatin signatures in active genes (>2 FPKMs) in the filopodial stage. The plot (left) represents the average normalized read coverage of histone modifications around the TSS of these active genes, and the heatmap (right) indicates the relative percentage of the genome represented by each chromatin state (first column) and relative fold enrichment for different genome features (other columns).

(F) Chromatin signatures in silent genes in the filopodial stage (heatmap and plot as in C).

(G) Boxplot representing the expression levels in the filopodial stage of genes (left) selected for having a significant peak of H3K27ac in the gene body (more than 800 bp from the TSS) and a significant peak of H3K4me1 after the TSS (within 800 bp), and vice versa (right). The p value is indicated for Wilcoxon the rank-sum test.

See also Figures S2 and S4 and Data S1.

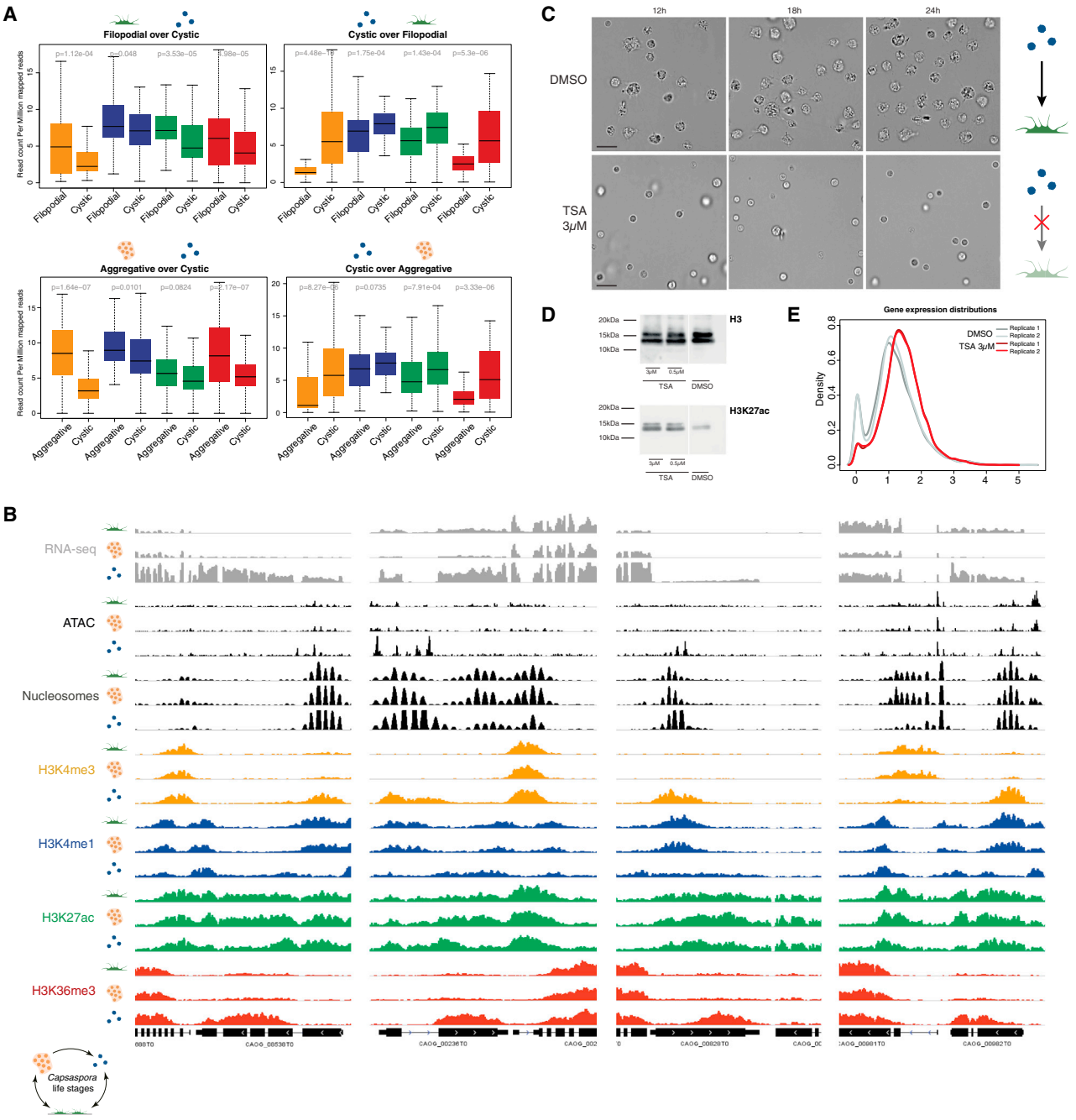


Figure 3. Dynamic Chromatin Modifications

(A) Boxplots showing hPTMs coverage levels in differentially expressed genes between stages, as indicated above each boxplot. The p value is indicated for the Wilcoxon signed-rank test.

(B) Illustrative examples of dynamic chromatin modifications in *Capsaspora*. Different genomic windows show normalized coverage for different chromatin features and their dynamic association with gene expression. For each feature, the top track corresponds to the filopodial stage, the middle track to the aggregative stage, and the bottom track to the cystic stage.

(C) Histone deacetylase inhibition experiments. Pictures of *Capsaspora* cells at different time points of incubation with DMSO (negative control) and TSA 3 μM. Transition from cystic to filopodial stage is blocked in the TSA-treated cells. Scale bar, 10 μm.

(D) Western blot against total H3 and H3K27ac on histone extracts from control cells (DMSO) and cells treated with 0.5 and 3 μM TSA. White line indicates a lane was removed.

(E) Gene expression distributions from biological replicates of control (DMSO, gray colors) and TSA-treated (red colors) cells. Notice the decrease in the fraction of non-expressed genes and the general shift in the distribution of TSA-treated cells.

See also Figures S2 and S3.

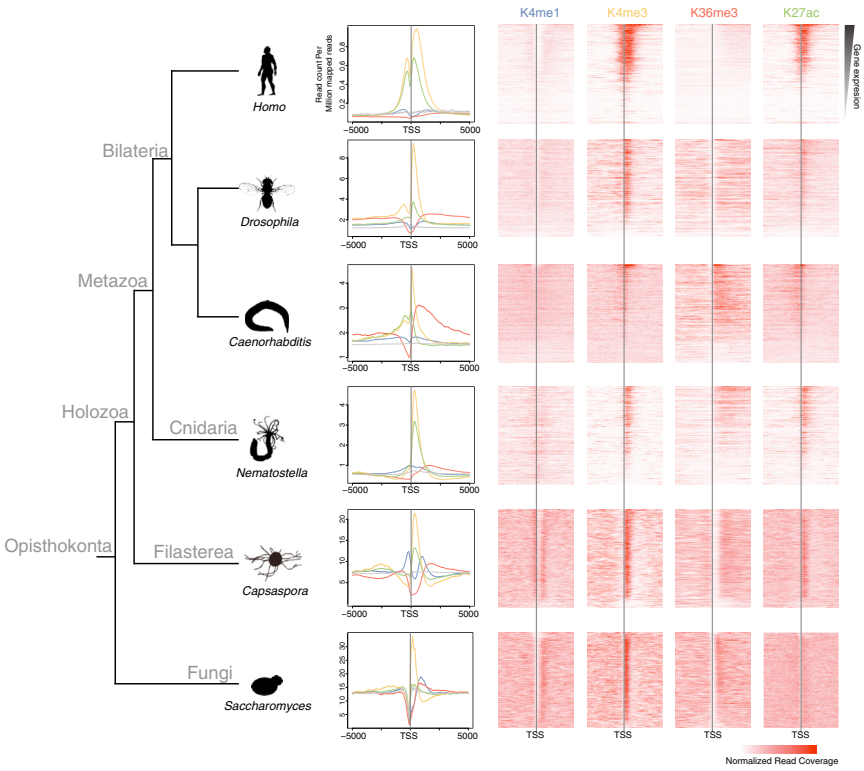


Figure 4. Comparative Proximal Distribution of Chromatin Marks across Opisthokonta Species

For each species, a plot shows the average normalized read coverage of four different histone modifications around the TSS (±5 kb), and heatmaps represent the same coverage for all genes sorted by level of expression. ChIP-seq data were obtained from publicly available datasets: *Homo sapiens*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Nematostella vectensis*, and *Saccharomyces cerevisiae*.

ATAC-defined regulatory sites per gene. Comparison of the number of associated sites across different types of genes revealed that transcription factors (TFs) are significantly enriched in regulatory sites (Figure 5B). In particular, T-box, bHLH, and bZIP TFs have the largest number of regulatory sites (Figure 5C). In contrast to previous predictions (Sebé-Pedrós and De Mendoza, 2015), this indicates the presence of intricate TF networks in *Capsaspora*. In addition, these regulatory sites were strongly enriched around TSS, in particular at proximal intergenic regions, first introns, and 5′ UTRs (Figure 5D), and depleted at gene bodies and distal intergenic regions. Interestingly, many of these regulatory sites show dynamic changes in ATAC-seq signal across life stages in *Capsaspora* (Figure 5E). In particular, 36% are stage specific and only 22% are constitutive in all three stages. Therefore, this specific and primarily proximal regulatory lexicon supports temporal cell-type transitions in *Capsaspora* and very likely also in the unicellular ancestors of animals.

Ancient Transcription Factor Networks

Capsaspora has a rich repertoire of metazoan-like TFs that are enriched in regulatory sites; however, it is unclear which specific genes are regulated by these TFs. To gain insights into premetazoan TF networks, we used motif analysis of the ATAC-defined regulatory sites. First, we looked for sites potentially bound by *Capsaspora-Brachyury*, an essential gene for animal gastrulation and mesoderm differentiation and the only TF whose binding site has been experimentally validated in *Capsaspora* (Sebé-Pedrós et al., 2013a). We found approximately 900 instances of this motif in the regulatory sites, all of them consistently displaying a similar tag density profile (Figure 6A). When compared with the whole population of *cis*-regulatory elements, these inferred *Bra* sites are preferentially located at the first intron and 5′ UTR and are predominantly associated with the filopodial amoeba and aggregative stages (Figure 6B). Accordingly, these *Capsaspora-Bra* sites are also more strongly correlated with the activating marks H3K4me3 and H3K27ac in these two stages (Figure 6C) compared with the cystic stage, and they are also enriched in these active histone marks compared with random *Bra* motifs found outside ATAC-defined regions (Figure 6D).

In order to validate some of our *Bra* downstream target predictions, we developed an anti-*Capsaspora-Bra* (CoBra) antibody (Figures 6E and 6F) and performed ChIP-qPCR experiments. We selected 20 ATAC-defined regulatory sites with *Bra* motifs (e.g., Figure 6H), including several with our lower limit selection threshold (0.80 Matscan cutoff [Blanco et al., 2006]), and compared them with ten random regions in the genome with strongly conserved Bra motifs (>0.90 Matscan cutoff). The ATAC-defined Bra regulatory sites were strongly enriched in CoBra compared with random motifs (Figure 6G), validating our *Bra* target prediction approach.

The *Capsaspora-Bra* downstream target network includes genes involved in establishment of cell polarity, phagocytosis, metabolism, transcription factors, and GPCR signaling genes (Figure 6I). Moreover, we identified 63 shared orthologs between inferred *Capsaspora-Bra* targets and those known for mouse *Brachyury* (Lolas et al., 2014). Interestingly, those shared orthologs are enriched in actin cytoskeleton and amoeboidal cell-motility functions (Figure 6J). This suggests that there was a conserved *Brachyury* downstream target network already present in premetazoan lineages and involved in cell migration, an essential cellular function later used in animal gastrulation.

Next, we performed a blind motif-enrichment analysis of all ATAC-defined sites in order to gain additional information on other TFs. Among the 29 significantly enriched nucleotide motifs, three of them strongly resemble (~90% similarity) known motifs for animal *Runx*, *NFAT/NFkappaB*, and *Myc* TFs. *Capsaspora* has clear orthologs of these three TFs (Sebé-Pedrós et al., 2011). Assuming that the motifs represent the consensus motifs for these *Capsaspora* orthologs, this provides evidence of

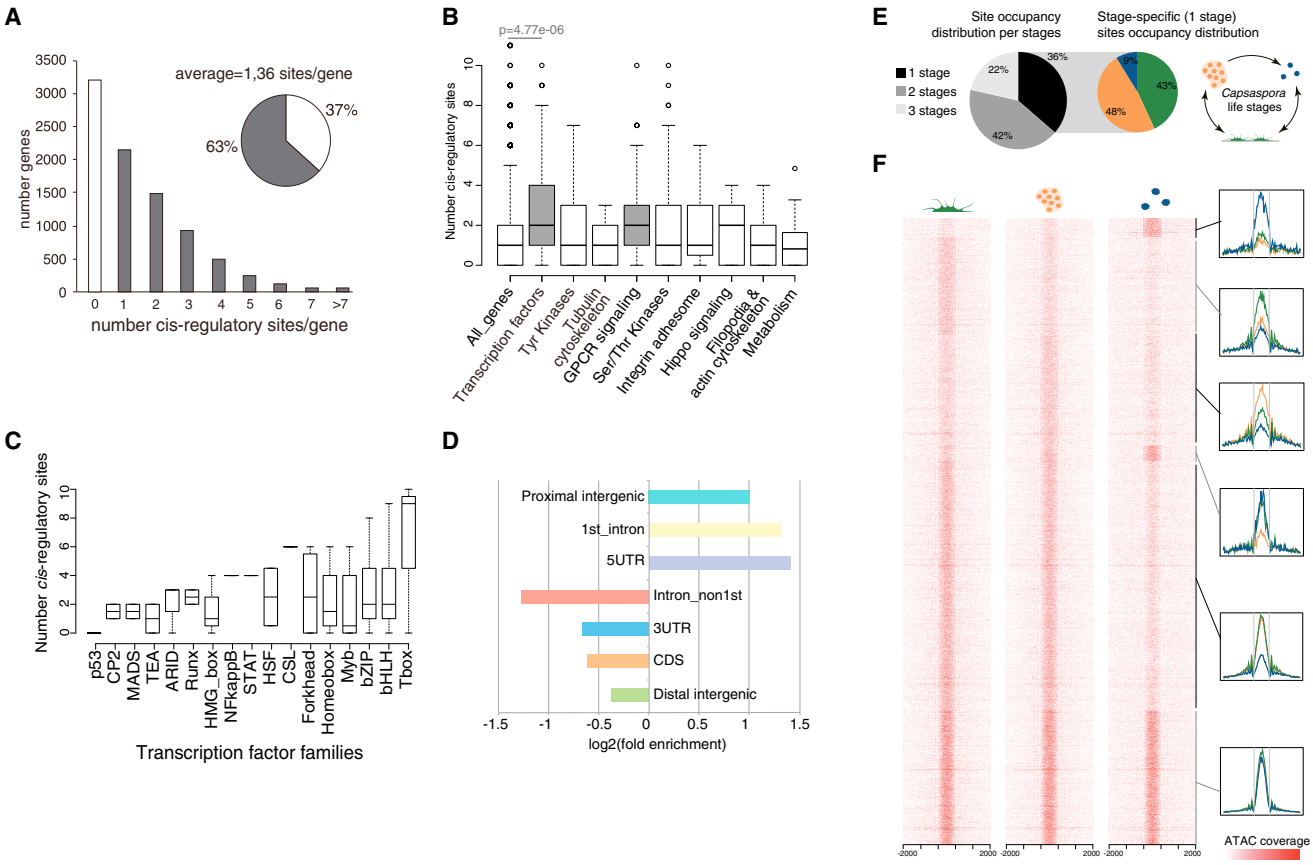


Figure 5. The Genomic Landscape of *cis*-Regulatory Elements in *Capsaspora*

(A) Distribution of the number of regulatory sites per gene.

(B) Number of *cis*-regulatory elements associated with different gene categories. Highlighted in gray are those with a significant enrichment (Wilcoxon rank-sum test p value < 0.01) compared with all genes.

(C) *Capsaspora* transcription factor families sorted by the number of *cis*-regulatory elements associated per gene.

(D) Preferential distribution of *cis*-regulatory sites across genomic features.

(E) Pie charts showing the distribution of the number of stages in which each site is occupied (left) and the stage distribution of the stage-specific fraction of regulatory sites (right).

(F) Heatmaps of clustered *cis*-regulatory elements (±2 kb) showing dynamic normalized ATAC nucleosome-free read coverage between stages. Plots show the associated average coverage profiles of each cluster.

See also Figure S6.

associations with genomic features and hPTMs (Figures 6 and S7). In particular, *Capsaspora-Myc*, a well-studied proto-oncogene in animals, appears to be strongly associated with regulatory sites that show higher ATAC-seq signal in the filopodial stage (Figure 6L), the proliferative stage in *Capsaspora* (Sebé-Pedrós et al., 2013b). These *Capsaspora-Myc* sites are more strongly correlated with the activating marks H3K4me3 and H3K27ac in filopodial and aggregative stages (Figure 6M) compared with the cystic stage, and they are also enriched in these active histone marks compared with random *Myc* motifs found outside ATAC-defined regions (Figure 6N). Moreover, *Myc* regulates genes mainly involved in ribosome biogenesis and translation (Figure 6O), similar to what is known for animal *Myc* networks (van Riggelen et al., 2010).

Interestingly, all TFs analyzed here show an enrichment of other TFs in their inferred downstream networks, reinforcing the idea of relatively complex TF-TF regulatory interactions in *Capsaspora*. The expansion of the TF repertoire at the stem of Metazoa (Sebé-Pedrós and De Mendoza, 2015), both in the total number of genes and of TF families, was probably associated with an increase in complexity of these TF networks. Remarkably, however, the inferred *Capsaspora* TF downstream targets suggest that at least some TF downstream regulatory networks were already conserved in the unicellular ancestor of metazoans and then subsequently remodeled within the animal lineage.

Distal Enhancers Are Animal Specific

To address whether there are potential distal enhancer elements in the genome of *Capsaspora*, we compared the regulatory sites defined by ATAC between *Capsaspora* and animals. Regulatory sites in *Capsaspora* are significantly smaller and more uniformly distributed than are sites in *Drosophila* and *Homo sapiens*

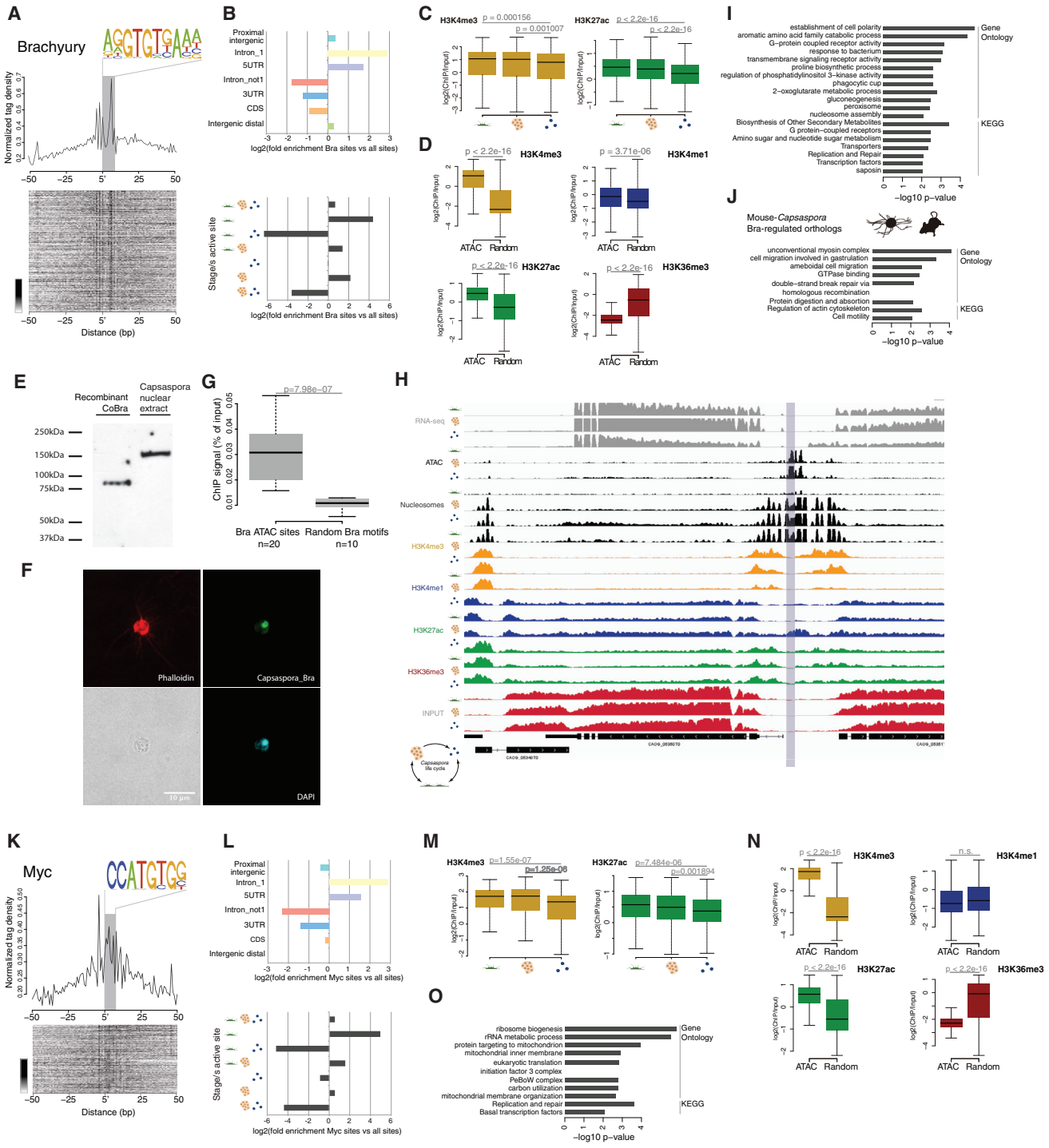


Figure 6. *Capsaspora* Brachyury and Myc Regulation

(A) Plot of ATAC-seq nucleosome-free reads average density around *Bra* motifs (top) and heatmap of the signal around the individual sites (bottom).

(B) Differential distribution of regulatory sites containing *Bra* motif compared with all sites according to genomic feature (top) and stage/s in which the site is active (bottom).

(C) Enrichment of different histone modifications (ChIP versus input) at *Bra* sites across stages.

(D) Enrichment of different histone modifications (ChIP versus input) at *Bra* motifs in ATAC-defined sites compared with motifs occurring randomly in the genome.

(E) Western blot of recombinant *Capsaspora-Brachyury* protein and *Capsaspora* nuclear protein extract, using *Capsaspora-Brachyury* affinity-purified antibody from guinea pig.

(legend continued on next page)

(Figure 7A). This means that regulatory sites in *Capsaspora* are bound by small numbers of proteins, whereas in animals large assemblies of regulatory factors bind active sites, allowing more complex combinatorial regulation. Comparison of the distribution of regulatory sites across genomic features revealed that distal sites, located at non-first introns and at intergenic regions, are extremely abundant in *Homo* and *Drosophila* but rare in *Capsaspora* (Figure 7B), and even rarer in *Saccharomyces cerevisiae* (Figure S7I) (Bulger and Groudine, 2011). Distal regulatory sites in animals, called enhancer elements, have unique chromatin signatures (Creyghton et al., 2010; Rada-Iglesias et al., 2011), including the presence of H3K4me1 constitutively and H3K27ac when activated. Using ATAC-defined proximal and distal intergenic *cis*-regulatory sites, we analyzed the read coverage for different histone marks for those sites in *Capsaspora*. At distal intergenic sites, *Capsaspora* shows no enrichment of H3K4me1 and H3K27ac compared with H3K4me3, whereas this enrichment is evident in *Homo sapiens* and *Drosophila* (Figure 7C). At proximal intergenic sites (800 bp upstream of TSS), H3K4me3 is significantly enriched over H3K4me1, although in *Capsaspora* this is less marked due to the proximal intergenic H3K4me1 enrichment described above. In all examined sites, H3K36me3 is depleted, as is expected outside gene bodies. Thus, regulatory sites in *Capsaspora* are mostly proximal, in contrast with the distal regulation observed in animals. Moreover, none of the regulatory sites in *Capsaspora* have biochemical signatures typical of animal enhancers. This indicates that distal regulation by enhancer elements is an animal evolutionary innovation and, probably, the most important difference in the genomic regulatory capabilities between premetazoans and metazoans.

DISCUSSION

In order to understand the evolution of the metazoan regulatory genome, we have here performed the first integrative analysis of the genome regulatory biology of a close unicellular relative of metazoans, the amoeboid filasterean *Capsaspora owczarzaki*. Indeed, this is the first such analysis of temporal chromatin dynamics in any non-model eukaryote.

We show that histone postranslational modifications, particularly those in H3 and H4, are highly conserved between *Capsaspora* and animals and also in other eukaryotes. Furthermore, good correspondence exists between these modifications and the presence/absence of known histone-modifying enzymes in the *Capsaspora* genome. For example, *Capsaspora* lacks H3K27me3 Polycomb repression marks, and it also lacks the PRC2 complex proteins, including EZH2, the central methyltransferase of the complex (Margueron and Reinberg, 2011).

We observe that 91.7% of the compact *Capsaspora* genome (28 Mb) includes regions producing transcripts (protein coding or lincRNAs) and/or regions with particular chromatin signatures and *cis*-regulatory sites. These signatures and regulatory sites are dynamically associated with life cycle progression and gene expression in *Capsaspora* and some, particularly active chromatin states, are shared with metazoans.

Our results indicate that *Capsaspora* has more numerous potential TF-TF regulatory connections than was previously thought, suggesting complex regulatory networks exist. Transcription factor networks tend to be quickly rewired during evolution (Li and Johnson, 2010; Sorrells and Johnson, 2015), and despite this, we find a remarkable degree of conservation between *Capsaspora* and animals in the downstream networks of orthologous TF that are key to animal multicellularity and development, such as *Brachyury* and *Myc*. These findings suggest that core downstream target networks of some developmental TF evolved long before the advent of animal multicellularity (Davidson and Erwin, 2006), controlling behaviors, such as proliferation and cell motility, in the first animal cells. These core conserved TF networks were subsequently integrated into complex developmental programs during animal evolution (Peter and Davidson, 2011).

Capsaspora also has a large repertoire of polyadenylated and, in some cases, alternatively spliced lincRNAs. These lincRNAs have temporal, cell-type-specific expression patterns, and they are associated with chromatin signatures similar to those found in metazoans (Marques et al., 2013). These *Capsaspora* lincRNAs show no homology with any known metazoan lincRNA, due to the fast evolution of lincRNA genes (Hezroni et al., 2015; Kapusta and Feschotte, 2014), and their functions are currently unknown. Despite this, our results indicate that elaborate genome regulation by long non-coding RNAs is not exclusive to multicellular organisms and was likely present in the protistan ancestors of Metazoa.

In contrast, the most important difference observed between *Capsaspora* and animal genome regulation is the marginal presence of distal *cis*-regulatory sites in *Capsaspora*, together with the absence of particular chromatin signatures associated with animal enhancers. This is in line with what is known in yeast, where regulation is proximal to the TSS (Bulger and Groudine, 2011) and no distal regulatory loops have been identified in genome 3D structure studies (Duan et al., 2010; Tanizawa et al., 2010). This result strongly indicates that distal enhancer elements are a major animal evolutionary innovation and constitute the basis of the sophisticated and highly evolvable gene

(F) *Capsaspora* filopodial stage cell stained with phalloidin (red, actin cytoskeleton), DAPI (blue, nucleus), and *Capsaspora-Brachyury* antibody (green). Notice *Bra* localization in the nucleus.
(G) Boxplot showing the *Capsaspora-Brachyury* ChIP-qPCR signal for predicted *Bra* regulatory sites versus random *Bra* motifs in the genome.
(H) Illustrative case example of a predicted *Bra* regulatory site (highlighted in blue). For each feature, the top track corresponds to the filopodial stage, the middle track to the aggregative stage, and the bottom track to the cystic stage. Notice the decreased ATAC signal in the putative Bra-regulatory site in the cystic stage.
(I) Enriched gene ontology (GO) terms and KEGG pathways among genes associated with *Bra* regulatory sites.
(J) Enriched GO terms and KEGG pathways among genes associated with *Bra* regulatory sites with shared orthologs regulated by *Bra* in mouse.
(K–O) Same as (A–D) and (I) for *Capsaspora Myc*.
See also Figure S7.

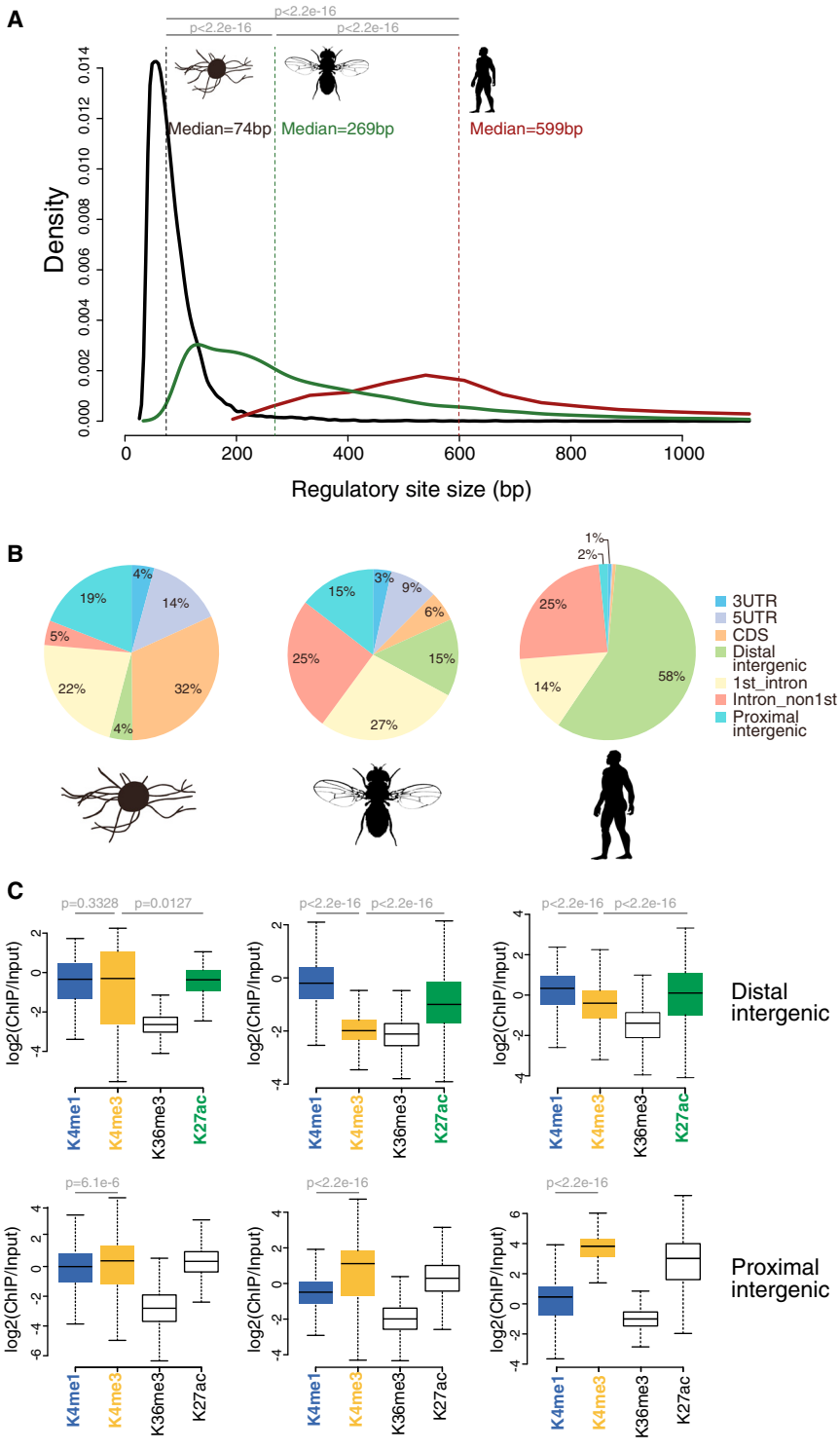


Figure 7. Comparative Analysis of Regulatory Sites between *Capsaspora* and Animals

(A) Distribution of ATAC-defined regulatory site sizes (bp) in *Capsaspora*, *Drosophila* and *Homo*.

(B) Genomic feature distribution of regulatory sites in *Capsaspora*, *Drosophila* and *Homo*.

(C) Enrichment of different histone modifications (ChIP versus input) at regulatory sites in distal (top) and proximal (bottom) intergenic regions in *Capsaspora*, *Drosophila* and *Homo*. In each boxplot, p values are indicated for Wilcoxon signed rank tests between H3K4me3 and H3K4me1 and between H3K4me3 and K27ac (only for distal intergenic).

See also Figure S7.

regulatory landscapes observed in animals (Andersson et al., 2014; Schwaiger et al., 2014; Villar et al., 2014). The emergence of these long-range *cis*-regulatory elements could also explain the pervasiveness of conserved syntenic regulatory blocks in animal genomes (Irimia et al., 2013) and the absence of these blocks in animal unicellular relatives like *Capsaspora* (Irimia et al., 2012). Moreover, the observation that *cis*-regulatory sites in *Capsaspora* are much smaller than those of animals further indicates that complex combinatorial TF binding appeared after the divergence of animals. Thus, not only did TF numbers expand (de Mendoza et al., 2013) and TF interaction networks became more intricate (Reinke et al., 2013; Sebé-Pedrós et al., 2013a) at the stem of Metazoa, but also the combinatorial binding of multiple TFs increased the potential number of developmental regulatory states in animals (Erwin, 2009; Erwin and Davidson, 2009; Peter and Davidson, 2011). Finally, specific promoter types for cell-type-specificity and developmental regulation, defined by chromatin signatures, appear also to be an animal innovation, since *Capsaspora* only has type II promoters. The emergence of additional promoter architectures in animals allowed distinct groups of genes to be controlled in different ways (Lenhard et al., 2012).

Overall, we reconstruct an evolutionary scenario in which the emergence of specific enhancer and promoter features at the onset of Metazoa, together with the expansion and remodeling of TF networks and non-coding RNA systems, allowed for fine-tuned spatiotemporal control of gene expression. Thus, the increase in regulatory genome complexity was probably a crucial step for the integration of cell types associated with the emergence of animal multicellularity. The precise molecular basis for this regulatory change remains to be determined. However, we hypothesize that it could be associated with the emergence of new chromatin modifying and remodeling enzymes and/or linked to the evolution of mechanisms for long-range genomic interaction and compartmentalization (Tanay and

Cavalli, 2013). Future analyses in other unicellular holozoans and in early branching animals, together with the study of the three-dimensional genome architecture of these taxa, will be crucial to further delineate the early evolution of the animal regulatory genome.

EXPERIMENTAL PROCEDURES

Capsaspora Cultures

Capsaspora strain ATCC30864 cells were grown axenically in ATCC medium 1034 at 23°C and differentiated as described in the Supplemental Experimental Procedures.

Histone Mass Spectrometry

Capsaspora histones were isolated by acid extraction, derivatized with propionic anhydride, and digested as described in Garcia et al. (2007). Tryptic peptides were analyzed via liquid chromatography-tandem mass spectrometry on an LTQ-Orbitrap Velos Pro mass spectrometer. Peptides were identified using the Mascot search engine.

Chromatin Immunoprecipitation

ChIP-seq and ChIP-qPCR were performed at three different life stages using antibodies against H3K4me3, H3K4me1, H3K27ac, H3K36me3, RNApolII, and CoBra as detailed in the Supplemental Experimental Procedures. 50 bp single-end Illumina sequencing reads were aligned to the Capsaspora genome (v.2) using Bowtie (Langmead et al., 2009), and regions of enrichment were determined using MACS2 (Zhang et al., 2008), correcting for genome mappability. Chromatin state definition and genomic feature enrichment was performed using ChromHMM (Ernst and Kellis, 2012). *Capsaspora* genome was reannotated as described in the Supplemental Experimental Procedures.

HDAC Inhibition Experiments

Capsaspora cystic stage cells were transferred to fresh medium and treated with 3 μM TSA and DMSO (negative control), and stage transition to the filopodial stage was monitored every 6 hr. Histones were isolated from *Capsaspora* cells incubated with DMSO or TSA by acid extraction, and the levels of histone acetylation were measure by western blot. Total RNA from treated cells was also extracted for RNA-seq. Further details are provided in the Supplemental Experimental Procedures.

ATAC-Seq

ATAC-seq was performed as originally described in Buenrostro et al. (2013), using 500,000 cells per cell stage. 50 bp paired-end sequencing reads were aligned to the Capsaspora genome (v.2) using Bowtie. Nucleosomal-free reads were used to define *cis*-regulatory sites using MACS2. The blind TF motif enrichment analysis was performed in these sites using HOMER (Heinz et al., 2010). Mononucleosomal reads were used to define nucleosome positions and fuzziness using Danpos2 (Chen et al., 2013).

lincRNA Annotation

High-coverage RNA-seq data were used for de novo annotation Capsaspora lincRNAs as detailed in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

The accession number for the mass spectrometry proteomics data reported in this paper has been uploaded to PRIDE repository: PXD002342. The accession number for the ChIP-seq and ATAC-seq data reported in this paper has been uploaded to GEO: GSE71131.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, one table, and three data files and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2016.03.034.

AUTHOR CONTRIBUTIONS

A.S.-P., C.B., and H.P.-A. performed the experiments. A.S.-P., C.C., and E.S. planed, acquired, and analyzed the proteomics data. J.J.T. was involved in ATAC-seq data analysis. A.S.-P., J.L.G.-S., I.R.-T., and L.D. were involved in the study design. A.S.-P. analyzed the data and generated the figures. A.S.-P. and I.R.-T. wrote the paper. All authors discussed the results and commented on the manuscript.

ACKNOWLEDGMENTS

We thank Alex de Mendoza, Xavier Grau-Bové, Ignacio Maeso, and Manuel Irimia for comments on the manuscript and figures. This work was supported by an Institució Catalana de Recerca i Estudis Avançats contract, a European Research Council Consolidator Grant (ERC-2012-Co-616960), and a grant from Ministerio de Economía y Competitividad (MINECO) (BFU-2011-23434) (to I.R.-T.). We also acknowledge financial support from Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya (project 2014 SGR 619). The work in L.D.'s laboratory was supported by grants from the Spanish "Ministerio de Educación y Ciencia" (SAF2013-48926-P), AGAUR, and the European Commission's 7th Framework Program 4DCellFate (277899). A.S.-P. is supported by an EMBO Long-Term Fellowship (ALTF 841-2014). J.L.G.-S. was funded by grants from Ministerio de Economía y Competitividad (BFU2013-41322-P) and the Andalusian Government (BIO-396). J.J.T. has a postdoctoral grant from the University Pablo de Olavide. The CRG/UPF Proteomics Unit is part of the "Plataforma de Recursos Biomoleculares y Bioinformáticos (ProteoRed)" supported by a grant from Instituto de Salud Carlos III (ISCIII) (PT13/0001). We thank Guadalupe Espadas for her support with the histone derivatization protocol and Núria Ros and Meritxell Antó for technical support. Finally, we thank the CRG Genomics Unit for helping with ChIP-seq and RNA-seq sequencing.

Received: July 21, 2015
Revised: February 3, 2016
Accepted: March 18, 2016
Published: April 21, 2016

REFERENCES

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al.; FANTOM Consortium (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455–461.

Bernstein, B.E., Meissner, A., and Lander, E.S. (2007). The mammalian epigenome. Cell *128*, 669–681.

Blanco, E., Messeguer, X., Smith, T.F., and Guigó, R. (2006). Transcription factor map alignment of promoter regions. PLoS Comput. Biol. *2*, e49.

Buecker, C., and Wysocka, J. (2012). Enhancers as information integration hubs in development: lessons from genomics. Trends Genet. *28*, 276–284.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Bulger, M., and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. Cell *144*, 327–339.

Chen, K., Xi, Y., Pan, X., Li, Z., Kaestner, K., Tyler, J., Dent, S., He, X., and Li, W. (2013). DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing. Genome Res. *23*, 341–351.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. USA *107*, 21931–21936.

Davidson, E.H., and Erwin, D.H. (2006). Gene regulatory networks and the evolution of animal body plans. Science *311*, 796–800.

de Laat, W., and Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. Nature *502*, 499–506.

de Mendoza, A., Sebé-Pedrós, A., Šestak, M.S., Matejcic, M., Torruella, G., Domazet-Loso, T., and Ruiz-Trillo, I. (2013). Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. Proc. Natl. Acad. Sci. USA *110*, E4858–E4866.

de Mendoza, A., Suga, H., Permanyer, J., Irimia, M., and Ruiz-Trillo, I. (2015). Complex transcriptional regulation and independent evolution of fungal-like traits in a relative of animals. eLife *4*, e08904.

Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y.J., Lee, C., Shendure, J., Fields, S., Blau, C.A., and Noble, W.S. (2010). A three-dimensional model of the yeast genome. Nature *465*, 363–367.

Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., et al.; ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Egloff, S., Dienstbier, M., and Murphy, S. (2012). Updating the RNA polymerase CTD code: adding gene-specific layers. Trends Genet. *28*, 333–341.

Eick, D., and Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. Chem. Rev. *113*, 8456–8490.

Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat. Methods *9*, 215–216.

Erwin, D.H. (2009). Early origin of the bilaterian developmental toolkit. Philos. Trans. R. Soc. Lond. B Biol. Sci. *364*, 2253–2261.

Erwin, D.H., and Davidson, E.H. (2009). The evolution of hierarchical gene regulatory networks. Nat. Rev. Genet. *10*, 141–148.

Fairclough, S.R., Chen, Z., Kramer, E., Zeng, Q., Young, S., Robertson, H.M., Begovic, E., Richter, D.J., Russ, C., Westbrook, M.J., et al. (2013). Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate Salpingoeca rosetta. Genome Biol. *14*, R15.

Gaiti, F., Fernandez-Valverde, S.L., Nakanishi, N., Calcino, A.D., Yanai, I., Tanurdzic, M., and Degnan, B.M. (2015). Dynamic and widespread lncRNA expression in a sponge and the origin of animal complexity. Mol. Biol. Evol. *32*, 2367–2382.

Garcia, B.A., Mollah, S., Ueberheide, B.M., Busby, S.A., Muratore, T.L., Shabanowitz, J., and Hunt, D.F. (2007). Chemical derivatization of histones for facilitated analysis by mass spectrometry. Nat. Protoc. *2*, 933–938.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Hezroni, H., Koppstein, D., Schwartz, M.G., Avrutin, A., Bartel, D.P., and Ulitsky, I. (2015). Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. Cell Rep. *11*, 1110–1122.

Ho, J.W.K., Jung, Y.L., Liu, T., Alver, B.H., Lee, S., Ikegami, K., Sohn, K.-A., Minoda, A., Tolstorukov, M.Y., Appert, A., et al. (2014). Comparative analysis of metazoan chromatin organization. Nature *512*, 449–452.

Irimia, M., Tena, J.J., Alexis, M.S., Fernandez-Miñan, A., Maeso, I., Bogdanovic, O., de la Calle-Mustienes, E., Roy, S.W., Gómez-Skarmeta, J.L., and Fraser, H.B. (2012). Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. Genome Res. *22*, 2356–2367.

Irimia, M., Maeso, I., Roy, S.W., and Fraser, H.B. (2013). Ancient cis-regulatory constraints and the evolution of genome architecture. Trends Genet. *29*, 521–528.

Kapusta, A., and Feschotte, C. (2014). Volatile evolution of long noncoding RNA repertoires: mechanisms and biological implications. Trends Genet. *30*, 439–452.

King, N., Westbrook, M.J., Young, S.L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I., et al. (2008). The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. Nature *451*, 783–788.

Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X.S., and Ahringer, J. (2009). Differential chromatin marking of introns and expressed exons by H3K36me3. Nat. Genet. *41*, 376–381.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Lenhard, B., Sandelin, A., and Carninci, P. (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nat. Rev. Genet. *13*, 233–245.

Levine, M. (2010). Transcriptional enhancers in animal development and evolution. Curr. Biol. *20*, R754–R763.

Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. Nature *424*, 147–151.

Li, H., and Johnson, A.D. (2010). Evolution of transcription networks–lessons from yeasts. Curr. Biol. *20*, R746–R753.

Lolas, M., Valenzuela, P.D.T., Tjian, R., and Liu, Z. (2014). Charting Brachyury-mediated developmental pathways during early mouse embryogenesis. Proc. Natl. Acad. Sci. USA *111*, 4478–4483.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343–349.

Marques, A.C., and Ponting, C.P. (2014). Intergenic lncRNAs and the evolution of gene expression. Curr. Opin. Genet. Dev. *27*, 48–53.

Marques, A.C., Hughes, J., Graham, B., Kowalczyk, M.S., Higgs, D.R., and Ponting, C.P. (2013). Chromatin signatures at transcriptional start sites separate two equally populated yet distinct classes of intergenic long noncoding RNAs. Genome Biol. *14*, R131.

Peter, I.S., and Davidson, E.H. (2011). Evolution of gene regulatory networks controlling body plan development. Cell *144*, 970–985.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283.

Reinke, A.W., Baek, J., Ashenberg, O., and Keating, A.E. (2013). Networks of bZIP protein-protein interactions diversified over a billion years of evolution. Science *340*, 730–734.

Schwaiger, M., Schönauer, A., Rendeiro, A.F., Pribitzer, C., Schauer, A., Gilles, A.F., Schinko, J.B., Renfer, E., Fredman, D., and Technau, U. (2014). Evolutionary conservation of the eumetazoan gene regulatory landscape. Genome Res. *24*, 639–650.

Schwer, B., and Shuman, S. (2011). Deciphering the RNA polymerase II CTD code in fission yeast. Mol. Cell *43*, 311–318.

Sebé-Pedrós, A., and de Mendoza, A. (2015). Transcription factors and the origin of animal multicellularity. In Evolutionary Transitions to Multicellular Life, I. Ruiz-Trillo and A.M. Nedelcu, eds. (Springer), pp. 379–394.

Sebé-Pedrós, A., de Mendoza, A., Lang, B.F., Degnan, B.M., and Ruiz-Trillo, I. (2011). Unexpected repertoire of metazoan transcription factors in the unicellular holozoan Capsaspora owczarzaki. Mol. Biol. Evol. *28*, 1241–1254.

Sebé-Pedrós, A., Ariza-Cosano, A., Weirauch, M.T., Leininger, S., Yang, A., Torruella, G., Adamski, M., Adamska, M., Hughes, T.R., Gómez-Skarmeta, J.L., and Ruiz-Trillo, I. (2013a). Early evolution of the T-box transcription factor family. Proc. Natl. Acad. Sci. USA *110*, 16050–16055.

Sebé-Pedrós, A., Irimia, M., Del Campo, J., Parra-Acero, H., Russ, C., Nusbaum, C., Blencowe, B.J., and Ruiz-Trillo, I. (2013b). Regulated aggregative multicellularity in a close unicellular relative of metazoa. eLife *2*, e01287–e01287.

Simola, D.F., Graham, R.J., Brady, C.M., Enzmann, B.L., Desplan, C., Ray, A., Zwiebel, L.J., Bonasio, R., Reinberg, D., Liebig, J., and Berger, S.L. (2016). Epigenetic (re)programming of caste-specific behavior in the ant Camponotus floridanus. Science *351*, aac6633.

Sorrells, T.R., and Johnson, A.D. (2015). Making sense of transcription networks. Cell *161*, 714–723.

Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M.E., Mitros, T., Richards, G.S., Conaco, C., Dacre, M., Hellsten, U., et al. (2010). The

Amphimedon queenslandica genome and the evolution of animal complexity. Nature *466*, 720–726.

Suga, H., Chen, Z., de Mendoza, A., Sebé-Pedrós, A., Brown, M.W., Kramer, E., Carr, M., Kerner, P., Vervoort, M., Sánchez-Pons, N., et al. (2013). The Capsaspora genome reveals a complex unicellular prehistory of animals. Nat. Commun. *4*, 2325.

Tanay, A., and Cavalli, G. (2013). Chromosomal domains: epigenetic contexts and functional implications of genomic compartmentalization. Curr. Opin. Genet. Dev. *23*, 197–203.

Tanizawa, H., Iwasaki, O., Tanaka, A., Capizzi, J.R., Wickramasinghe, P., Lee, M., Fu, Z., and Noma, K. (2010). Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. Nucleic Acids Res. *38*, 8164–8177.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82.

Ulitsky, I., and Bartel, D.P. (2013). lincRNAs: genomics, evolution, and mechanisms. Cell *154*, 26–46.

van Galen, P., Viny, A.D., Ram, O., Ryan, R.J., Cotton, M.J., Donohue, L., Sievers, C., Drier, Y., Liau, B.B., Gillespie, S.M., et al. (2016). A multiplexed system for quantitative comparisons of chromatin landscapes. Mol. Cell *61*, 170–180.

van Riggelen, J., Yetil, A., and Felsher, D.W. (2010). MYC as a regulator of ribosome biogenesis and protein synthesis. Nat. Rev. Cancer *10*, 301–309.

Villar, D., Flicek, P., and Odom, D.T. (2014). Evolution of transcription factor binding in metazoans - mechanisms and functional implications. Nat. Rev. Genet. *15*, 221–233.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhou, V.W., Goren, A., and Bernstein, B.E. (2011). Charting histone modifications and the functional organization of mammalian genomes. Nat. Rev. Genet. *12*, 7–18.