


RESEARCH

Open Access



Role of epigenetics in unicellular to multicellular transition in *Dictyostelium*

Simon Yuan Wang^{1,2}, Elizabeth Ann Pollina^{3†}, I-Hao Wang^{4†}, Lindsay Kristina Pino⁵, Henry L. Bushnell^{1,2}, Ken Takashima^{1,2}, Colette Fritsche¹, George Sabin¹, Benjamin Aaron Garcia⁵, Paul Lieberman Greer⁴ and Eric Lieberman Greer^{1,2*} 

* Correspondence: eric.greer@childrens.harvard.edu

[†]Elizabeth Ann Pollina and I-Hao Wang contributed equally to this work.

¹Division of Newborn Medicine, Boston Children's Hospital, 300 Longwood Avenue, Boston, MA 02115, USA

²Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA

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

Abstract

Background: The evolution of multicellularity is a critical event that remains incompletely understood. We use the social amoeba, *Dictyostelium discoideum*, one of the rare organisms that readily transits back and forth between both unicellular and multicellular stages, to examine the role of epigenetics in regulating multicellularity.

Results: While transitioning to multicellular states, patterns of H3K4 methylation and H3K27 acetylation significantly change. By combining transcriptomics, epigenomics, chromatin accessibility, and orthologous gene analyses with other unicellular and multicellular organisms, we identify 52 conserved genes, which are specifically accessible and expressed during multicellular states. We validated that four of these genes, including the H3K27 deacetylase *hdaD*, are necessary and that an SMC-like gene, *smc1*, is sufficient for multicellularity in *Dictyostelium*.

Conclusions: These results highlight the importance of epigenetics in reorganizing chromatin architecture to facilitate multicellularity in *Dictyostelium discoideum* and raise exciting possibilities about the role of epigenetics in the evolution of multicellularity more broadly.

Keywords: Epigenetics, Multicellularity, *Dictyostelium discoideum*, Methylation, Acetylation, *hdaD*, *smc1*, HDAC, *srfA*

Background

A critical evolutionary event, the transition from unicellularity to multicellularity, requires the acquisition of cell-cell adhesion and cooperation, and frees the organism for growth and specialization [1]. However, the molecular adaptations that facilitated this transition remain incompletely understood. Multicellularity is believed to have evolved due to both aggregation of non-clonal single cells cooperating to form a multicellular organism [2] and to clonal development where the individual cells never dissociate [3, 4]. However, how this evolution from primitive amoeba or fungal unicellular organisms occurred ~900 million years ago is still incompletely understood [5]. Earlier studies proposed that the acquisition of new genes or new transcription factor networks drove

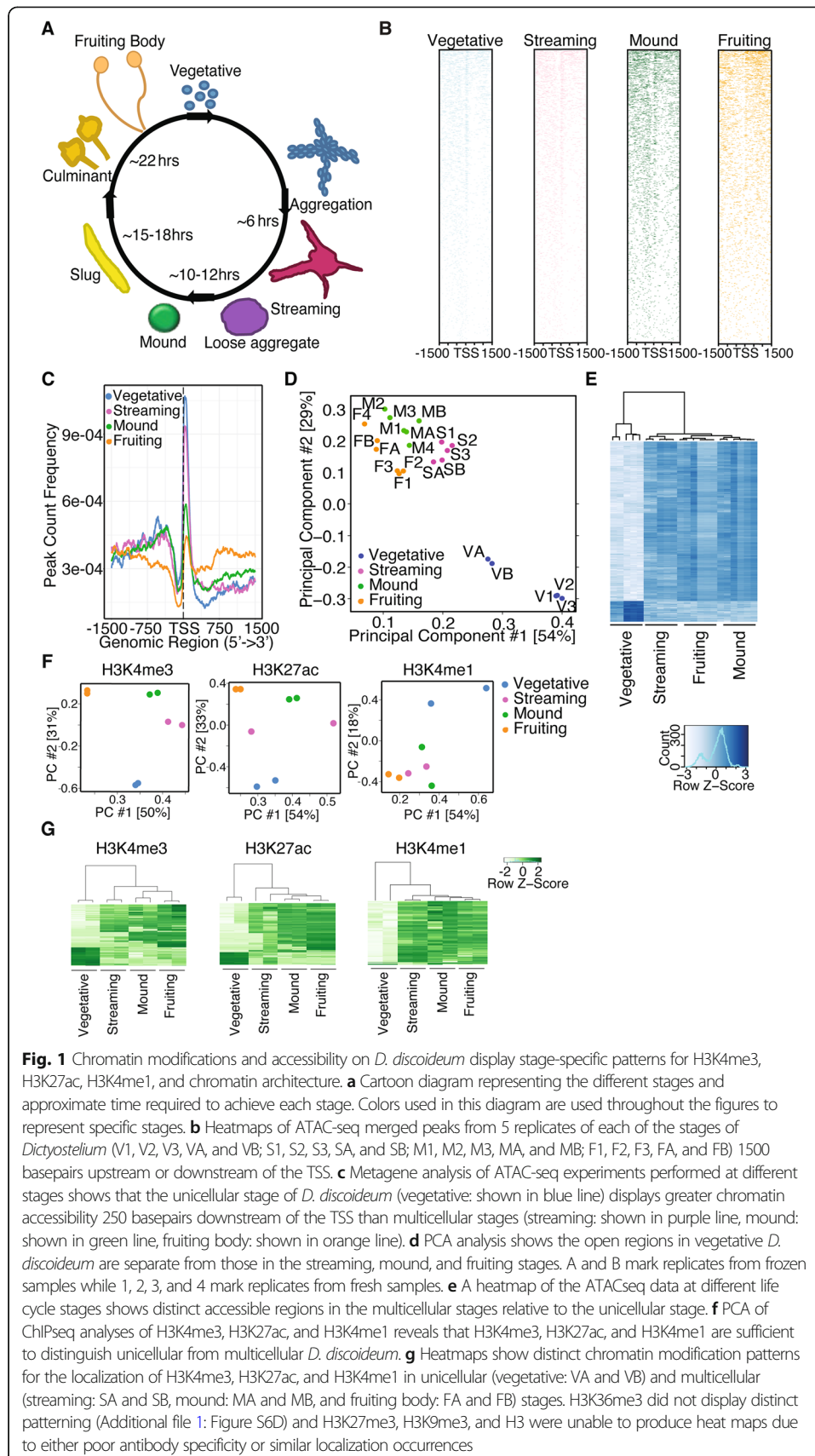


© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

the evolution of multicellularity [6]. However, the emergence of multicellularity does not appear to be explained exclusively by the appearance of new genes as unicellular ancestors contain complex gene repertoires required for multicellular functions [7, 8]. Likewise, the acquisition of new transcription factor networks does not always correlate with multicellularity [9]. These observations suggest a potential role for chromatin regulation in the transition to multicellularity. Indeed, recent work in the unicellular *Capsaspora owczarzaki* suggested that cellular state changes in this organism were due to non-genetic factors rather than the conserved transcription factor networks present in this unicellular eukaryote [10]. Although many of the individual components for modern epigenetic regulatory systems are present in more ancient ancestors, their integration into a coherent system of gene regulation did not occur until eukaryotes. Moreover, epigenetic diversity expanded rapidly upon the transition to multicellularity [11–14]. However, whether the increased complexity of epigenetic regulation may have helped to drive the transition to multicellularity is unknown.

We sought to examine multicellularity using *Dictyostelium discoideum* because it is one of the rare organisms that readily transforms back and forth through both unicellular and multicellular states. When bacterial food is plentiful, *Dictyostelium* are single-celled amoebae. However, when the food supply is exhausted, *Dictyostelium* produce a chemoattractant, which signals to neighboring cells to aggregate into a multicellular organism of ~215,000 cells [15, 16]. Within 6–8 h, unicellular *Dictyostelium* aggregate into a multicellular mound that develops to a migrating facultative slug state [17]. After locating an appropriate environment, the slug will subsequently develop into a mature fruiting body, complete with both a stalk and spores that disseminate to distant locations (Fig. 1a). Once food becomes available, spores quickly germinate into single-celled amoebae [18]. The rapidity with which this organism undergoes such dramatic and plastic physiological changes, while maintaining the same DNA code, makes it an ideal model organism for studying the role of epigenetics in multicellularity. Epigenetic changes have been extensively cataloged as totipotent zygotes unidirectionally develop into multicellular organisms [19, 20]. *Dictyostelium* represents an inherently epigenetic model organism, as *Dictyostelium* transitions from unicellular to multicellular states and back again and therefore do not represent a developmental progression but rather an adaptive cellularity transition. Additionally, *Dictyostelium* has a relatively small genome (34 megabases [21]), is easy to culture in a lab and to manipulate genetically [22–24], and can be induced to develop using simple environmental manipulations [18]. Importantly, *Dictyostelium* contain the conserved epigenetic regulating systems of other eukaryotes [25], including histone modifications and modifying enzymes that regulate gene expression [26–28].

Here we provide the first comprehensive map of epigenetic states (RNA expression, histone modifications, and chromatin accessibility) across four life cycle stages of *Dictyostelium*. By comparing 2473 orthologues gene expression patterns and chromatin accessibility between unicellular and multicellular *Dictyostelium* with that of fission yeast *S. pombe* and *C. elegans*, we identify candidate genes that could play important roles in the transition to multicellularity. We take advantage of the epigenetic plasticity of *Dictyostelium* to examine how epigenetic modifications can regulate transitions from unicellular to multicellular states. Furthermore, knockout of four multicellular genes caused delays in multicellularity while overexpression of one of these multicellular



genes increased multicellularity. Our findings identify critical, novel regulators of multicellularity, and reveal that epigenomic changes correlate with the transition from unicellularity to multicellularity in *Dictyostelium*.

Results

Dictyostelium display unicellular and multicellular specific chromatin patterns

To begin to determine whether chromatin modifications might contribute to *D. discoideum* life cycle changes, overall histone methylation and acetylation levels were assessed by western blots at different cellularity stages. These analyses revealed that histone H3 lysine 4 trimethylation (H3K4me3) levels decreased and H3K4me1 increased significantly upon the transition to multicellularity suggesting that chromatin modification might help regulate the transition to multicellularity in *Dictyostelium* (Additional file 1: Figure S1A). To examine, in a more unbiased manner, how histone modifications change during different cellularity stages we performed comparative quantitative mass spectrometry at 4 different stages. This analysis confirmed an increase in H3K4me1 upon the transition to multicellularity, as well as identified increases in H2Ax9me2 which was often accompanied by an increase in H2AxS1 acetylation (H2AxS1ac), H2AxK17ac, and H2AxK18me1 (Additional file 1: Figure S1B). This transition to multicellularity was also correlated with an increase in H3K14ac, H4K9ac, and H4K13acK17me1K21ac triply modified histone tails (Additional file 1: Figure S1B). Some of these changes in histone modifications were previously found in *Dictyostelium* [27] while others were identified for the first time here (Additional file 1: Figure S1C and S1D). These methods of quantifying histone modification changes are by no means comprehensive, but do reveal that many different chromatin modifications change in overall abundance upon the transition from unicellular to multicellular *Dictyostelium discoideum* raising the possibility that these changes in chromatin modifications could help to drive multicellularity in *Dictyostelium*. To more precisely ascertain where within the genome specific changes in chromatin state occur during this transition, we performed Assay for Transposase-Accessible Chromatin using sequencing (ATACseq) [29], which broadly identifies which regions of the chromatin are open and accessible for transcription and which regions are closed and quiescent at four stages of *D. discoideum* life cycle: the vegetative, streaming, mound, and fruiting stages. We performed ATACseq on both frozen and fresh samples with highly reproducible results (Additional file 1: Figure S2). Similar to most eukaryotes [30], the chromatin is compacted directly preceding the transcriptional start sites (TSS) (Fig. 1b). By performing a metagene analysis of the ATACseq data, we found that there was a reduction in the number of genes with accessible promoters immediately after the TSS and an increase in the number of genes with accessible gene bodies as *Dictyostelium* transition from the unicellular to multicellular forms (Fig. 1c and Additional file 1: Figure S3A). A principal component analysis (PCA) and clustering analyses of the ATACseq datasets clearly separated the vegetative unicellular stage from the multicellular stages, indicating that the accessible chromatin might reveal genomic patterns unique to unicellular and multicellular states of *D. discoideum* (Fig. 1d, e, and Additional file 2: Table S1). This was accompanied by a progressive increase of accessibility at more gene promoters and a progressive increase of compaction at more gene exons as *Dictyostelium* proceed from vegetative toward

fruiting body stage (Additional file 1: Figure S3B). A motif analysis of accessible regions revealed a motif, CCATATATGG, that was enriched in all multicellular stages but was absent in the unicellular stage (Additional file 1: Figure S4). This motif provides a binding site for MADS-box transcription factors [31], which, interestingly, have been implicated in the evolution of multicellularity [32]. Deletion of two *Dictyostelium* MADS-box transcription factors, *mef2A* and *srfa*, has been shown to slow progression to multicellular states [33–35]. These results raise the possibility that to transition to multicellularity, *Dictyostelium* reorganizes its chromatin structure to make binding sites for multicellular specific transcription factors *mef2A* and *srfa* accessible for binding, likely to facilitate the transcription of their target genes. Alternatively, these transcription factors could act as pioneering transcription factors, opening up the chromatin to facilitate their binding and transcriptional activity. Together these data reveal unicellular and multicellular *D. discoideum* have different patterns of chromatin accessibility regions throughout the genome, which could expose different transcription factor binding sites at specific stages.

To characterize chromatin modifications, which might explain these alterations in chromatin accessibility, ChIP-seq was performed in duplicate at the four stages analyzed by ATACseq using antibodies that recognize Histone H3, Histone H3 lysine 4 monomethylation (H3K4me1), H3K4 trimethylation (H3K4me3), H3K9me3, H3K27me3, H3K36me3, or H3K27 acetylation (H3K27ac) (Fig. 1f, g and Additional file 1: Figures S5 and S6). A high correlation in ChIPseq peaks was observed across biological replicates indicating good specificity and reproducibility (Additional file 1: Figure S5). A principal component analysis (PCA) of observed ChIPseq patterns revealed that certain chromatin modifications accurately distinguished between the life cycle stages of *Dictyostelium*. For instance, H3K4me3 and H3K27ac marks clearly separated the unicellular and multicellular stages from each other (Fig. 1f, g). H3K4me1 modifications were able to separate the unicellular stage from the multicellular stages but were not as powerful at distinguishing the different multicellular stages (Fig. 1f, g). H3K36me3 had some capacity to distinguish stages at the 2nd principal component, but H3K27me3, H3K9me3, and Histone H3 did not discern distinct stages even by the 3rd principal component (Additional file 1: Figure S6B, Additional file 3: Tables S2, and Additional file 4: Table S3), suggesting that these marks may not change consistently with life cycle stage and that histone occupancy is not sufficient to distinguish different stages. The chromatin modifications that were capable of distinguishing unicellular from multicellular stages were similar to those marks that we observed to have different overall levels across the life cycle of *Dictyostelium* (Additional file 1: Figure S1). Additionally, chromatin marks with the capacity to distinguish stages correlate with gene activation, while marks that correlate with gene repression are unable to distinguish *Dictyostelium* stages. This observation suggests that activating chromatin marks may be more important for determining cellular states than repressive marks. Together, these data demonstrate that chromatin modifications change as *D. discoideum* transitions through different life cycle stages and that the most dramatic chromatin modification changes occur in H3K4 methylation and H3K27 acetylation upon the transition from unicellular to multicellular states raising the possibility that regulation of these specific chromatin marks could play a critical role in the transition to multicellularity in *Dictyostelium*.

Histone deacetylation and methylation inhibitors delay multicellularity

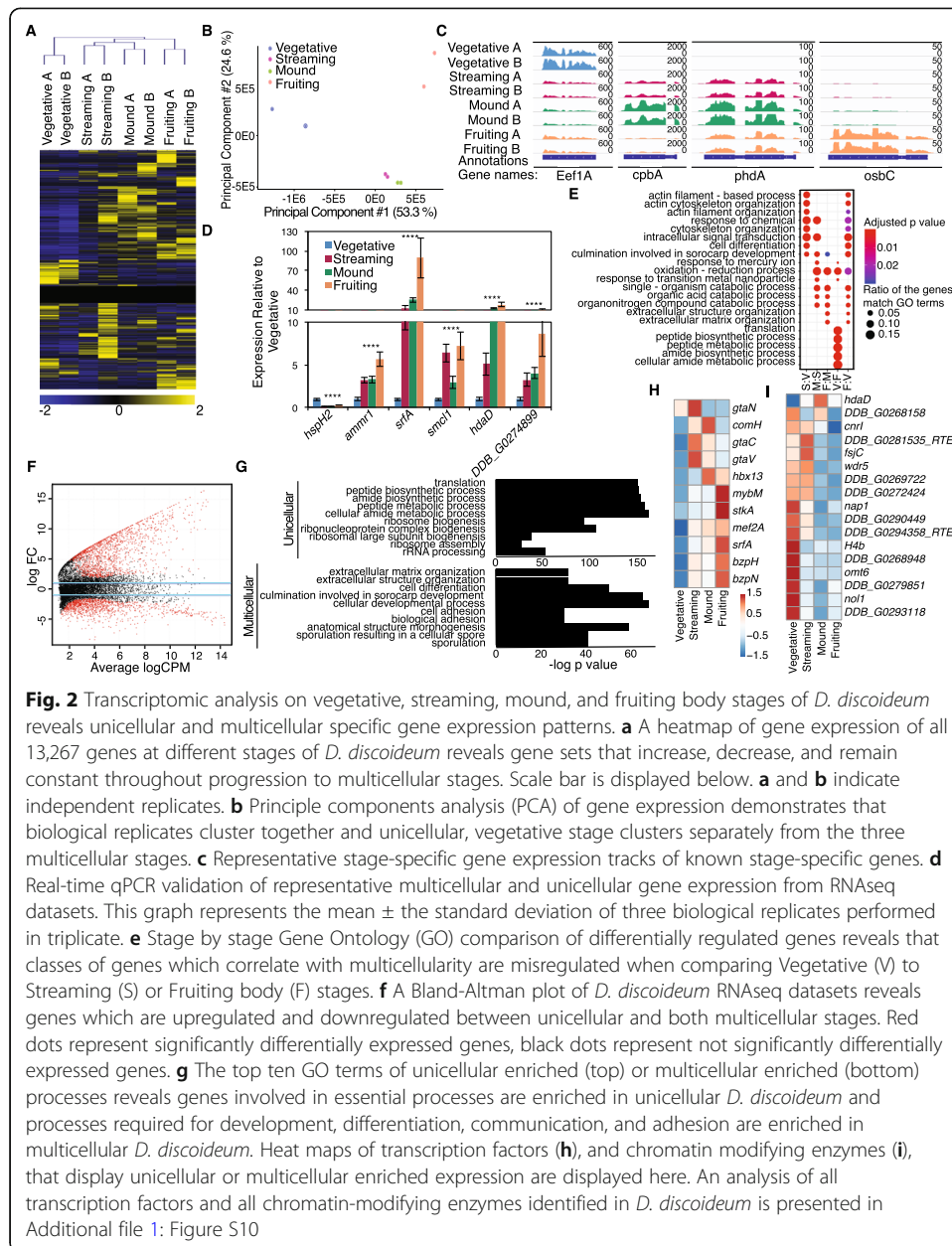
To begin to test whether changes in H3K4 methylation and H3K27 acetylation affect the transition to multicellularity, we treated vegetative *D. discoideum* with chemical inhibitors of histone deacetylation and methylation enzymes and then induced multicellularity by starvation. The histone deacetylation inhibitor Trichostatin A (TSA), the histone methylation enzyme inhibitors sinefungin and BIX 01294 all caused a concentration-dependent delay in the initiation of multicellularity, defined by the amount of time required to achieve the mound stage, as well as a delay in the completion of multicellularity transition, defined by the time required to produce fruiting bodies (Additional file 1: Figure S7). We confirmed that TSA caused an increase in H3K27ac while sinefungin caused a decrease in H3K4me3 and H3K9me3. BIX01294, although reported to be an H3K9me3 inhibitor, led to a decrease in H3K4me3 and H3K27ac without any measurable effects on H3K9me3 in 293 T cells (Additional file 1: Figure S7A). These inhibitors delayed the initiation of multicellularity from ~12 h in control conditions to ~17 h in 5 μ M sinefungin ($p=0.001$), ~15 h in 10 μ M BIX 01294 ($p=0.047$), and ~17 h in 1 μ M TSA ($p=0.0024$) (Additional file 1: Figure S7B and S7C). Histone modification chemical inhibition similarly delayed the completion of multicellularity transition (Additional file 1: Figure S7B and S7C). While *D. discoideum* took ~22 h to reach the fruiting body stage under basal conditions, 10 μ M BIX 01294, 5 μ M sinefungin, and 1 μ M TSA significantly increased this time to 27, 37, and 29 h, respectively (Additional file 1: Figure S7B and S7C, BIX 01294 $p=0.0126$; sinefungin $p < 0.0001$; TSA $p=0.0009$). Treatment with sinefungin also caused a reduction in the number of cells making up the multicellular *D. discoideum* (Additional file 1: Figure S7D). The diameter of the fruiting body, which correlates with the number of cells composing the multicellular form [15, 16], was ~9 μ m without treatment and ~6 μ m after 5 μ M sinefungin treatment ($p < 0.05$). To determine whether these inhibitors specifically delayed multicellularity or were more generally toxic, the effects of the lowest concentrations of TSA, sinefungin, and BIX 01294 that delayed multicellularity were measured for effects on chemotaxis and cell viability. Although 1 μ M TSA and 5 μ M sinefungin significantly reduced chemotaxis towards folate (Additional file 1: Figure S8A and S8B, $p < 0.0001$ and $p=0.007$, respectively), 10 μ M BIX 01294 had no significant effect on chemotaxis (Additional file 1: Figure S8A and S8B, $p=0.9978$). Additionally, 10 μ M BIX 01294 and 5 μ M sinefungin had no effect on cell viability as assessed by trypan blue staining (Additional file 1: Figure S8C), propidium iodide (PI) staining (Additional file 1: Figure S8D), or annexin V staining (Additional file 1: Figure S8E) while 1 μ M TSA caused a modest decrease in cell viability when assessed by annexin V staining (Additional file 1: Figure S8E). Thus, chemicals which regulate histone methylation inhibit multicellularity without having dramatic effects on other properties of *D. discoideum*. Interestingly, while both sinefungin and BIX 01294 affected other histone methylation marks, they both caused a decrease in H3K4me3. Confirming the importance of H3K4 methylation in multicellularity, previous work has demonstrated that deletion of the H3K4me3 methyltransferase *set1*, causes unusually rapid multicellularity [27]. Taken together, these results suggest that changes to the amounts or locations of chromatin modifications might regulate multicellularity.

Stage-specific gene expression patterns highlight important transcription factors and chromatin modifiers

To begin to investigate whether the chromatin changes we observed correlate with gene expression changes, we performed RNA-seq analysis on vegetative, streaming, mound, and fruiting stages of *D. discoideum*. Biological replicates were reproducible (Additional file 1: Figure S9A), individual stages were readily distinguished from each other by correlation analysis in a heatmap and Bland-Altman plots (Additional file 1: Figure S9B-F), and our results correlated well with previously generated RNAseq data (Additional file 1: Figure S9G-J) [36]. As expected, the expression of some genes remained constant, some declined, and others increased at each transition (Fig. 2a and Additional file 5: Table S4). A principal component analysis (PCA) readily distinguished individual stages. The similar mound and streaming stages clustered close to each other, while the unicellular vegetative and the final fruiting body stage clustered far from each other (Fig. 2b). Importantly, this approach recapitulates previously reported stage-specific patterns of gene expression [36], (Fig. 2c). We confirmed by qRT-PCR the stage-specific gene expression patterns of 5 multicellular and 1 unicellular enriched genes (Fig. 2d). Together, these data suggest that the RNAseq analysis is reliable. Gene ontology (GO) analysis revealed stage-specific enrichment of gene expression in distinct processes and pathways (Fig. 2e). Using a threshold FDR of < 0.05, 1216 genes were enriched in the unicellular state and 1649 genes were enriched in the multicellular states (Fig. 2f). By GO analysis, unicellular genes were enriched for genes involved in protein synthesis (Fig. 2g, upper panel and Additional file 5: Table S4). Conversely, multicellular enriched genes were selectively involved in cell differentiation, development, and adhesion and cellular components required for extracellular elements, such as the cell surface and the plasma membrane (Fig. 2g, lower panel and Additional file 5: Table S4). These changes reflect the shifting cellular priorities during multicellular stages.

As transcription factors play an important role in the transition to multicellularity [37, 38], we next identified transcription factors that had stage-specific expression (Additional file 1: Figure S10A). Ten transcription factors were expressed predominantly in multicellular stages, but none was found to be predominant in the vegetative stage (Fig. 2h). Two of these transcription factors were the MADS-box transcription factors *srfA* and *mef2A*, which were previously shown to play a role in *D. discoideum* progression to multicellular states [33–35]. We found that genes which were upregulated in multicellular states were enriched for MADS-box transcription factor binding motifs (Additional file 1: Figure S10C $p = 4.08e-25$). These results suggest that not only does the chromatin reorganize to expose *srfA* and *mef2A* binding sites (Additional file 1: Figure S4), but *srfA* and *mef2A* expression is coordinately upregulated (Fig. 2d, h) to facilitate the transcription of *SrfA* and *Mef2A* targets.

We also examined stage-specific expression of all putative chromatin modifying and binding proteins, identified based on homology (Additional file 1: Figure S10B). Seventeen chromatin-modifying enzymes or binding proteins displayed unicellular or multicellular specific expression patterns (Fig. 2i). More transcription factors whose expression varied with cellularity stages were upregulated in the multicellular stages, whereas more chromatin-modifying enzymes were upregulated in the unicellular stage (Additional file 1: Figure S10A and S10B, $p < 0.05$). Notably, one of the chromatin



regulating enzymes, the putative histone deacetylase *hdaD*, was selectively enriched in multicellular states (Fig. 2d,i), raising the possibility that *hdaD* could be activated in multicellular states to deacetylate histones and reorganize chromatin to facilitate the transition to multicellularity.

Comparative epigenomic analysis of *S. pombe*, unicellular and multicellular *D. discoideum*, and *C. elegans* identifies unicellular and multicellular signatures

To begin to determine whether chromatin modifications or accessibility could be used to identify critical regions whose chromatin changed upon multicellularity, we first investigated whether chromatin modifications in *D. discoideum* are interpreted in a

similar manner as they are in other eukaryotes. We therefore characterized how gene expression correlated with the ChIP-seq and ATACseq data we generated in *D. discoideum*. As has previously been demonstrated in mammalian cell lines [39, 40], chromatin accessibility correlated with gene expression in *Dictyostelium*; the top 1000 most lowly and highly expressed genes displayed lower and higher chromatin accessibility, respectively (Fig. 3a and Additional file 6: Table S5). Similarly, the 1216 unicellular enriched genes, based on the RNAseq analysis, were significantly enriched for increased chromatin accessibility at the unicellular stage while the 1649 multicellular enriched genes, were significantly enriched for increased chromatin accessibility at the multicellular stages (Additional file 1: Figure S11A, $\chi^2 = 511.56$, $p < 0.00001$). Histone modifications of specific genes also correlated with their expression in similar ways in *D. discoideum* as they do in other eukaryotic organisms. H3K4me3, which typically correlates with active transcription [41, 42], displayed increased occurrence on unicellular enriched genes while largely being absent from multicellular enriched genes at the unicellular stage and increased occurrence on multicellular enriched genes while largely being absent from unicellular enriched genes at the multicellular stage (Additional file 1: Figure S11B, $\chi^2 = 391.33$, $p < 0.00001$). For instance, *srfA*, which turns on during multicellular stages (Fig. 2d), exhibited increased H3K4me3, decreased H3K27me3, and increased chromatin accessibility at its promoter at multicellular stages relative to the unicellular stage, which is when it is more highly expressed (Fig. 3b).

To investigate further whether epigenomic changes could correlate with unicellular or multicellular states, we compared the epigenomic signatures of unicellular and multicellular *D. discoideum* to two other eukaryotes whose epigenomes are well defined, the unicellular *Schizosaccharomyces pombe* and the multicellular *Caenorhabditis elegans*, which are both ~1480 million years evolutionarily separated from *D. discoideum* [43, 44] (Additional file 1: Figure S12A). We identified 2473 conserved regions by predicting orthology and paralogy relationships in all three species by metaphors using pre-computed phylogenetic trees [45] (Fig. 3c and Additional file 7: Table S6). The majority of these highly conserved genes (79.1%) were evenly expressed across all *Dictyostelium* stages while 15.6% were unicellular enriched and 5% were multicellular enriched (Additional file 1: Figure S12B). Interestingly, a PCA analysis of the orthologous genes showed that by principal component 3 the expression patterns of unicellular *Dictyostelium* genes were more similar to that of *S. pombe* than to multicellular *Dictyostelium* genes (Fig. 3d).

To determine whether epigenetic features could underlie these unicellular and multicellular gene expression signatures, we examined whether accessible chromatin regions correlated with unicellular or multicellular signatures. We examined the ATACseq profiles of *C. elegans* [46] and *S. pombe* [47] at the 2473 evolutionarily conserved regions of the genome and found that of the conserved genes, 1352 and 804 were accessible, respectively. Of the *D. discoideum* conserved genes, 922 were accessible in the unicellular stage while 720 were accessible in the multicellular stages. Among those 922 accessible regions in unicellular *D. discoideum*, 132 were accessible in multicellular *Dictyostelium* or *C. elegans* (14.3%), but 289 sites were exclusively accessible in *S. pombe* (31.3%). Similarly, among the 720 accessible regions in multicellular *D. discoideum*, 94 were accessible in unicellular *Dictyostelium* or *S. pombe* (13.1%), but 333 sites were exclusively accessible in *C. elegans* (46.3%). We further reduced these lists of 289 unicellular

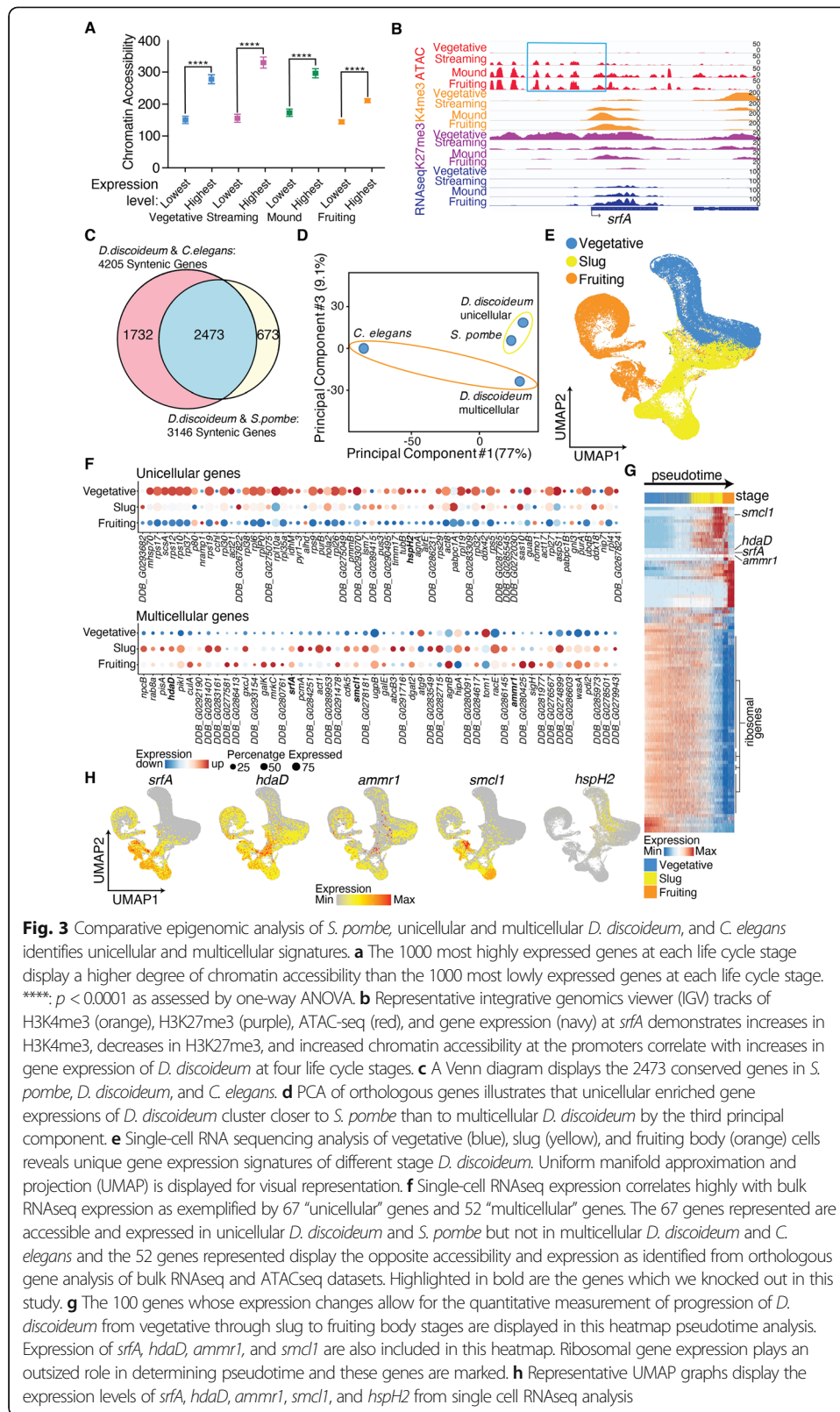


Fig. 3 Comparative epigenomic analysis of *S. pombe*, unicellular and multicellular *D. discoideum*, and *C. elegans* identifies unicellular and multicellular signatures. **a** The 1000 most highly expressed genes at each life cycle stage display a higher degree of chromatin accessibility than the 1000 most lowly expressed genes at each life cycle stage. ****. $p < 0.0001$ as assessed by one-way ANOVA. **b** Representative integrative genomics viewer (IGV) tracks of H3K4me3 (orange), H3K27me3 (purple), ATAC-seq (red), and gene expression (navy) at *srfA* demonstrates increases in H3K4me3, decreases in H3K27me3, and increased chromatin accessibility at the promoters correlate with increases in gene expression of *D. discoideum* at four life cycle stages. **c** A Venn diagram displays the 2473 conserved genes in *S. pombe*, *D. discoideum*, and *C. elegans*. **d** PCA of orthologous genes illustrates that unicellular enriched gene expressions of *D. discoideum* cluster closer to *S. pombe* than to multicellular *D. discoideum* by the third principal component. **e** Single-cell RNA sequencing analysis of vegetative (blue), slug (yellow), and fruiting body (orange) cells reveals unique gene expression signatures of different stage *D. discoideum*. Uniform manifold approximation and projection (UMAP) is displayed for visual representation. **f** Single-cell RNAseq expression correlates highly with bulk RNAseq expression as exemplified by 67 “unicellular” genes and 52 “multicellular” genes. The 67 genes represented are accessible and expressed in unicellular *D. discoideum* and *S. pombe* but not in multicellular *D. discoideum* and *C. elegans* and the 52 genes represented display the opposite accessibility and expression as identified from orthologous gene analysis of bulk RNAseq and ATACseq datasets. Highlighted in bold are the genes which we knocked out in this study. **g** The 100 genes whose expression changes allow for the quantitative measurement of progression of *D. discoideum* from vegetative through slug to fruiting body stages are displayed in this heatmap pseudotime analysis. Expression of *srfA*, *hdaD*, *ammr1*, and *smc1* are also included in this heatmap. Ribosomal gene expression plays an outsized role in determining pseudotime and these genes are marked. **h** Representative UMAP graphs display the expression levels of *srfA*, *hdaD*, *ammr1*, *smc1*, and *hspH2* from single cell RNAseq analysis

accessible and 333 multicellular accessible regions to genes which were expressed exclusively in unicellular or multicellular stages. This analysis left us with 67 genes which were accessible and expressed in unicellular state or organisms and inaccessible and not expressed in multicellular state or organisms and 52 genes which were accessible and expressed in multicellular state or organisms and inaccessible and not expressed in unicellular state or organisms (Additional file 8: Table S7), suggesting that these accessible chromatin regions correlate with unicellularity or multicellularity, respectively across evolution. Interestingly, *strfA*, the MADS-box transcription factor that had previously been shown to play a role in progression to multicellularity in *D. discoideum* [33], was one of the genes present in the multicellular list suggesting that our method of identifying critical genes involved in multicellularity based on their chromatin states was successful.

Because bulk RNAseq experiments may mask the gene expression of specific cell types that are critical for transition states, we performed single-cell RNAseq experiments. This analysis clearly distinguished the unicellular state from the multicellular stages (Fig. 3e). We found that the expression of the 67 unicellular genes and 52 multicellular genes, identified based on the orthologous gene analysis of their expression in bulk RNAseq and ATACseq experiments, correlated well with the single-cell RNAseq experiments (Fig. 3f). We next used the single-cell RNAseq data to examine which genes' expression signatures could account for the progression of *D. discoideum* from vegetative through slug to fruiting body stages. This lineage tracing analysis revealed that of the top 100 genes whose expression changes helped determine this temporal progression, 54 of them are ribosomal genes (Fig. 3g). This finding demonstrates that the ribosomal heterogeneity that was initially discovered and reported in *Dictyostelium* [48] is predominantly driven by gene expression changes in ribosomal protein genes. Interestingly, 13 of these ribosomal protein genes were identified as part of our unicellular signature (Fig. 3f and Additional file 8: Table S7), indicating that changes in chromatin accessibility at ribosomal protein genes could be a critical driver of the evolution of multicellularity. We further confirmed that 4 genes that were accessible and expressed in *C. elegans* and multicellular *Dictyostelium* and quiescent in unicellular state or *S. pombe* were highly enriched or only expressed in multicellular cells while a gene that was accessible and expressed in *S. pombe* and unicellular *Dictyostelium* and quiescent in multicellular state or *C. elegans* was highly enriched in the unicellular stage (Fig. 3h).

Deletion of epigenomically identified multicellular genes delays or prevents multicellularity in *D. discoideum*

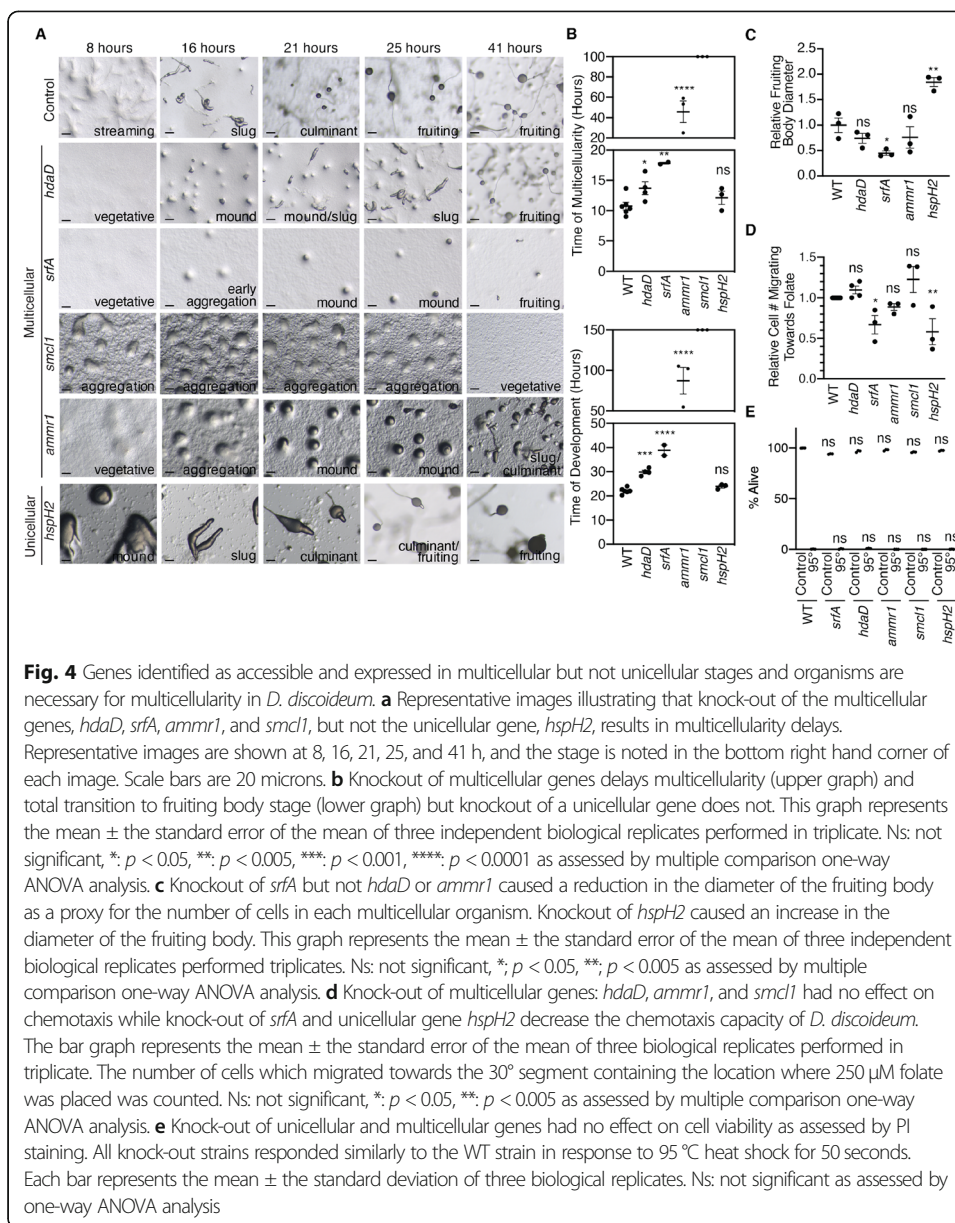
Together, these approaches identified a total of 67 “unicellular” genes, which were accessible and expressed in unicellular state or organisms and inaccessible and not expressed in multicellular state or organisms and 52 “multicellular” genes, which were accessible and expressed in multicellular state or organisms and inaccessible and not expressed in unicellular state or organisms. To begin to test whether these identified genes contribute to the transition between unicellularity and multicellularity, we analyzed knockouts of a subset of these genes. We obtained a previously generated knockout strain of *strfA* [49] and attempted to delete 4 additional “multicellular” genes and 3

“unicellular” genes individually to generate single mutation knockouts, and examined their effects on rates of multicellularity. Despite repeated attempts, one “multicellular” gene (*DDB_G0274899*) and two “unicellular” gene (*DDB_G0267824* and *DDB_G0277333*) knock-outs were inviable, suggesting that these genes are essential for *D. discoideum* survival. Successful knock-outs of the remaining 4 genes were validated by sequencing (Additional file 1: Figure S13).

Individual deletion of each of the four “multicellular” genes: *srfA*; *hdaD*; *DDB_G0283437*, a gene which we renamed *ammr1* based on its homology to the uncharacterized Alport syndrome gene *Ammerc1*; and *DDB_G0288013*; a gene which we renamed *smc11* based on its similarity to a structural maintenance of chromatin (SMC) gene, delayed the transition to multicellularity and extended the entire progression to multicellular state (Fig. 4a,b: *srfA* $p=0.0014$ and $p < 0.0001$; *hdaD* $p=0.0165$ and $p=0.0011$; *ammr1* $p < 0.0001$ and $p < 0.0001$; *smc11* $p < 0.0001$ and $p < 0.0001$, respectively). For instance, knock-out of *hdaD* caused the initiation of multicellularity to be delayed from 10.75 h in control cells to 13.6 h and the time required to achieve a fruiting body to be delayed from 22.11 h to 29.88 h (Fig. 4a, b). Similarly, *srfA* knock-out caused a delay of the initiation of multicellularity (16.96 h, $p=0.0022$) and a delay in complete transition to fruiting body stage (38.83 h, $p < 0.0001$) and a reduction in the number of cells in the multicellular state to 45% of the wildtype *Dictyostelium* (Fig. 4c, $p=0.0405$). Likewise, deletion of *ammr1* dramatically slowed multicellularity to 45.8 h ($p < 0.0001$) and caused complete transition to fruiting body stage to take 87.33 h ($p < 0.0001$), while deletion of *smc11* caused a complete blockage in the progression past the aggregation stage to any multicellular stages (Fig. 4a, b). By contrast, knock-out of the “unicellular” gene *DDB_G0293674*, a gene which we renamed *hspH2* based on its homology to the heat shock protein *hsc70-3*, had no effect on the transition to multicellularity or the transition to fruiting body stage (Fig. 4a, b, $p=0.5166$ and $p=0.2585$, respectively) but did cause an increase in the fruiting body diameter (Fig. 4c, $p < 0.01$).

To determine whether deletion of these “multicellular” genes was simply reducing overall fitness, or was more specifically affecting *D. discoideum* multicellularity, we examined the effect of gene knock-outs on chemotaxis and cell viability. While the “unicellular” gene *hspH2* displayed no effect on *D. discoideum* transition to multicellularity, it did result in chemotaxis delay (Fig. 4d, $p < 0.0001$), suggesting that this gene is essential for unicellular phenotypes. Of the “multicellular” genes only *srfA* displayed a decrease in chemotaxis ($p=0.0144$ by one-way ANOVA) while *hdaD*, *ammr1*, and *smc11* were indistinguishable from control cells (Fig. 4d, $p=0.7128$, $p=0.06836$, and $p=0.1199$, respectively). Additionally, knock-outs of all “multicellular” genes had no effect on cell viability under basal conditions or in response to a 95 °C heat shock as assessed by PI staining (Fig. 4d), suggesting that these deletions are not simply making the *Dictyostelium* sick and are more specifically affecting the transition to multicellularity.

To begin to determine whether the function of the genes that we identified as being necessary for multicellularity was also sufficient to induce a multicellular state in *Dictyostelium*, we overexpressed two of them, *hdaD* and *smc11*, in WT *Dictyostelium*. We confirmed the overexpression of these genes by real-time RT PCR (Additional file 1: Figure S14A). While overexpression of neither *hdaD* nor *smc11* accelerated the timing of multicellularity (in fact *smc11* overexpression caused a slowing of multicellularity timing) (Fig. 5a, b, and Additional file 1: Figure S14B and S14C), overexpression of



smc1 but not *hdaD*, did cause a ~3 fold increase in the number of cells which make up a multicellular *Dictyostelium* organism (Fig. 5c–e and Additional file 1: Figure S14D), suggesting that *smc1* enhances multicellularity. We found that overexpression of *smc1* had no effect on chemotaxis (Additional file 1: Figure S14E, $p=0.1207$). Together, these findings suggest that a gene which we identified based on its epigenomic status, the SMC-like gene *smc1*, was both necessary and sufficient to regulate multicellularity.

Together, these experiments identified four genes that regulate the transition from unicellularity to multicellularity. To begin to investigate the mechanisms by which these genes control multicellularity, we focused on *hdaD*, which sequence alignment suggested might encode a histone deacetylase. HDACs are attractive candidates to mediate these processes because of their capacity to deacetylate histones

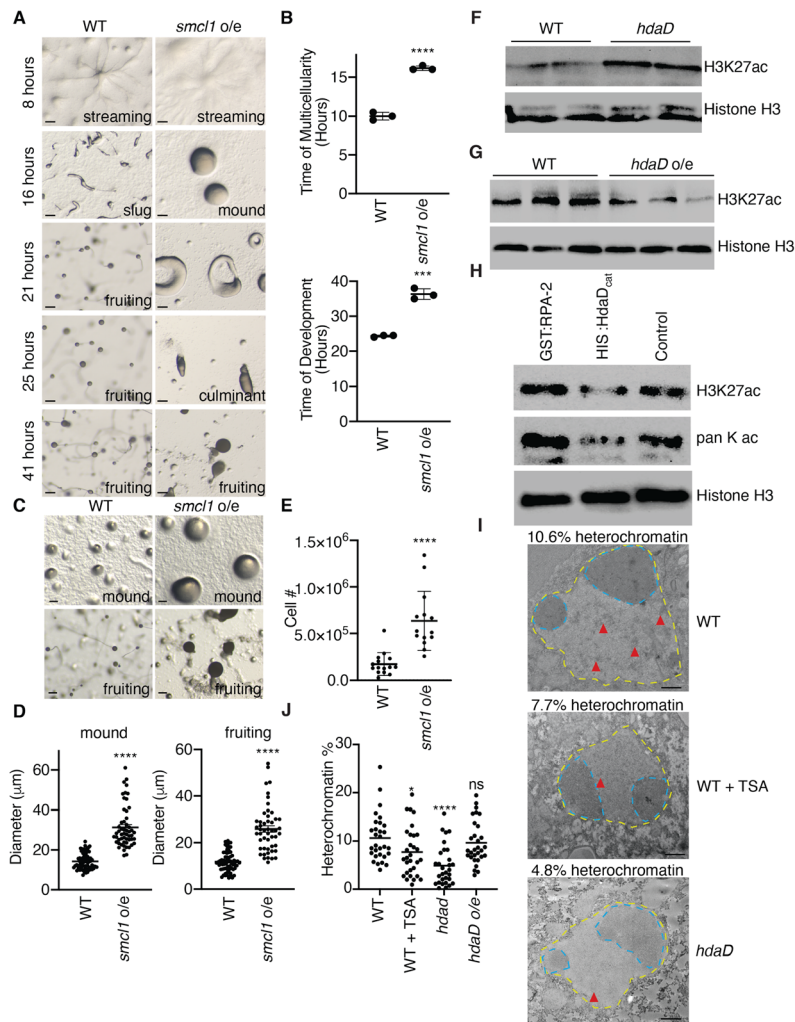


Fig. 5 *smc11* overexpression is sufficient for multicellularity in *D. discoideum* and HdaD is an H3K27 deacetylase. **a** Representative pictures of *D. discoideum* cellular stages on KK2 plates show multicellularity delay after overexpression of *smc11* at 8, 16, 21, 25, and 41 h. Stage is displayed in the lower right hand corner. Scale bars are 20 microns. **b** Overexpression of *smc11* delays *D. discoideum* multicellularity, time to mound stage is shown in the graph on top and complete progression to fruiting body stage is shown in the graph on the bottom. Each column represents the mean \pm the standard error of the mean of three biological replicates performed in triplicate. ***, $p < 0.0005$, ****, $p < 0.0001$, as assessed by unpaired *t* test. **c** Representative pictures of *D. discoideum* show increased size of mound (top) and fruiting body (bottom) after overexpression of *smc11*. Scale bars are 20 microns. **d** Overexpression of *smc11* was sufficient to increase the size of mound (top) and fruiting body (bottom) diameter. Each column represents the mean \pm the standard error of the mean of 54–67 *D. discoideum* multicellular organisms. ****, $p < 0.0001$ as assessed by unpaired *t* test. **e** Overexpression of *smc11* was sufficient to increase the number of cells in each fruiting body. Each column represents the mean \pm the standard error of the mean of 14–15 *D. discoideum* fruiting bodies. ****, $p < 0.0001$ as assessed by unpaired *t* test. **f** Representative western blots demonstrate an increase in H3K27ac in *hdaD* knock-out strains. Two independent biological replicates are displayed with the upper blot probed with H3K27ac and the bottom blot probed with Histone H3 antibodies. **g** *hdaD* overexpression causes a decrease in H3K27ac as assessed by western blot. **h** His-tagged HdaD catalytic domain (His:HdaD_{catalytic}) but not GST-tagged RPA-2 is able to deacetylate lysine 27 of histones purified from *Dictyostelium*. This western blot is representative of 3 independent experiments. **i, j** Electron microscopy reveals that TSA treatment and *hdaD* knock-out strains have a decrease in heterochromatin. Representative images are shown (**i**) and quantification of 30 images (**j**). The scale bars represent 500 nm. % heterochromatin was calculated by examining the condensed chromatin which appears darker in the nucleus (outlined in yellow) while excluding the nucleolus (outlined in blue) from calculations. Examples of heterochromatin are marked by a red arrow. Representative images of *hdaD* overexpression and quantification of the nucleolus diameter in all 30 images are displayed in Additional file 1: Figure S15

and therefore cause a reorganization of the chromatin. Western blotting of *hdaD* knockouts revealed an increase in H3K27ac when compared to wild type *Dictyostelium* (Fig. 5f and Additional file 1: Figure S14F), suggesting that HdaD is a H3K27 deacetylase in *D. discoideum*. Consistent with this hypothesis, overexpression of *hdaD* caused a decrease in H3K27 acetylation (Fig. 5g and Additional file 1: Figure S14G). To determine whether HdaD is a direct H3K27 deacetylase we expressed a polyhistidine (His)-tagged catalytic domain of HdaD in bacteria, purified His-HdaD_{catalytic domain} to a single band, and analyzed its ability to deacetylate histones purified from *D. discoideum*. We found that His-HdaD_{catalytic domain}, but not control proteins, was able to deacetylate H3K27 in purified histones (Fig. 5h and Additional file 1: Figure S14H and I), suggesting that HdaD is an H3K27 deacetylase in vitro and in vivo. Reinforcing the hypothesis that HdaD is regulating multicellularity through an alteration in chromatin, we found that *hdaD* knockouts, similar to the deacetylase inhibitor TSA, had decreased heterochromatin as assessed by electron microscopy (Fig. 5i, j and Additional file 1: Figure S15).

The finding that *hdaD* deletion inhibits multicellularity highlights the importance of H3K27 acetylation regulation that we had observed correlating on a global level with multicellularity in our ChIPseq experiments (Fig. 1f, g). Together these results indicate that the epigenomic status of genes can be used to identify critical genes necessary for multicellularity. Interestingly histone deacetylases of the *hdaD* family have been shown to bind to MADS box transcription factors on DNA [50, 51], raising the possibility of a coordinated evolution of transcription factor networks and epigenetic machinery for regulating multicellularity.

Discussion

Dictyostelium: a new epigenetic model organism for studying the evolution of multicellularity

By comparing the chromatin accessibility at orthologous genes of *Dictyostelium* to *S. pombe* and *C. elegans*, we were able to identify a set of candidate genes whose chromatin status correlated with unicellularity or multicellularity. This analysis identified one gene, *srfA*, which had previously been shown to play a role in timing of progression to multicellular states [33, 34], as well as several uncharacterized genes. We found that knocking out three novel “multicellular” genes delayed or inhibited multicellularity without making the *Dictyostelium* less fit. Similarly, the overexpression of one of these “multicellular” genes was sufficient to triple the number of cells integrated into a multicellular state. Together the approaches employed in this study have identified several genes, based on their epigenetic status, which are necessary and sufficient for the transition to multicellularity. This analysis also suggests that a deeper comparison of not only the chromatin accessibility at promoters, as was performed in this study, but also the chromatin accessibility in the gene body or further upstream of the TSS along with specific chromatin modifications across different unicellular and multicellular organisms and *Dictyostelium* stages could reveal additional genes essential for multicellularity. Additionally, a similar epigenetic comparative analysis of more closely related Amoebozoa, those which can and cannot transition to multicellular forms [52], could help to further refine this multicellular epigenetic signature. It will be important, in future experiments, to also examine whether changing the epigenome at these critical genes is sufficient to induce multicellularity.

It is interesting to note that the uncharacterized genes which we identified as being important for determining multicellularity elicited unique cellular responses in addition to their effect on multicellularity. This highlights how several independent characteristics, including cell to cell communication, adhesion, cellular specialization, cooperation, and cellular altruism all had to develop for multicellularity to have evolved [32]. It will be interesting, in future studies, to examine how altering one of these characteristics elicits a trade-off in other cellular traits. While increasing the amount of cells which create a multicellular *Dictyostelium* could provide advantages in facilitating increased cellular specialization, this would have to come at the expense of either a shortening of or increased cell number of the stalk to accommodate the larger sorocarp. This would elicit either a decrease in the distance from the immediate environment which lacks sufficient food that spores would be able to travel or an increase in the number of cells which sacrifice themselves for the survival of the remaining cells in the organism.

Since *Dictyostelium* rapidly transitions between physiologically and morphologically distinct states, this organism is ideally situated to be utilized as an epigenetic model organism for examining some of the critical outstanding questions in the field of epigenetics. Here we have displayed the ease and capacity to not only perform comparative observational science but also to manipulate the genome and epigenome of *Dictyostelium* to test theories. Therefore, we feel this will be a useful organism to study a host of epigenetic questions beyond the role of epigenetics in the evolution of multicellularity. Our study has provided a comprehensive epigenomic sequencing of critical stages of *Dictyostelium*. It also demonstrates how this comprehensive dataset can be used to study the role of epigenetics in multicellularity. This comprehensive dataset opens the possibility of identifying additional processes that epigenetic modifications and drugs can regulate.

Proof-of-principle that epigenetics plays a role in regulating multicellularity (genetic vs epigenetic)

In this study, we used *Dictyostelium discoideum* as a model organism to investigate the role of epigenetics in the evolution of multicellularity. Through a combination of chemical, comparative genomic, and epigenomic approaches, we identified several key components of multicellular progression. The histone deacetylase inhibitor, TSA, had previously been demonstrated to delay multicellularity in *Dictyostelium* [34], suggesting that histone acetylation could play a critical role in the transition to multicellularity. We confirmed these findings and found that chemical inhibitors of histone methylation, Sinefungin and BIX 01294, also delayed induction of multicellularity, suggesting that chromatin regulation is important for multicellularity. We also examined the distribution of several histone modifications; H3K4me3, H3K4me1, H3K27ac, H3K36me3, H3K27me3, and H3K9me3. Principal component analysis (PCA) demonstrated that specific histone modifications, such as H3K4me3 and H3K27ac, were sufficient to distinguish unicellular from multicellular forms of *Dictyostelium*. Because our ChIPseq experiments did not include an exogenous spike-in, we can only assess the relative changes in abundance rather than identify the absolute quantifications of histone modification changes. The importance of H3K4me3 in regulating multicellularity in *Dictyostelium* is reinforced by the demonstration that deletion of the H3K4me3

methyltransferase *set1*, causes an unusually rapid progression to multicellular states [27]. Our results further demonstrated the importance of H3K27ac by our unbiased identification that the histone deacetylase *hdaD*, which we demonstrate is an H3K27 deacetylase (Fig. 5f–h), increases expression upon transition to multicellularity (Fig. 2d) and is necessary for an appropriately timed induction of multicellularity (Fig. 4a, b). Together these results suggest that chromatin modifications are important and could help to drive multicellularity in *Dictyostelium*.

To gain a global view of chromatin accessibility of *Dictyostelium*, we performed ATAC-seq at specific life cycle stages. We found dramatic differences in chromatin accessibility in the unicellular versus multicellular stages, suggesting that chromatin reorganization during this transition to multicellularity is critical. This analysis revealed that MADS-box transcription factor motifs become accessible upon transition to multicellularity (Additional file 1: Figure S4). The importance of these binding sites becoming accessible is reinforced by previously published work demonstrating that two MADS-box transcription factors, *mef2A* and *srfA*, are necessary for appropriate multicellularity timing [33–35]. These results suggest that to transition to multicellularity, *Dictyostelium* reorganizes its chromatin structure to make binding sites for multicellular specific transcription factors *mef2A* and *srfA* accessible for binding, to facilitate the transcription of their target genes. Excitingly, class II histone deacetylases, of which *hdaD* is a member, have been shown to bind to MADS box transcription factors on DNA [50, 51], raising the possibility of a coordinated evolution of transcription factor networks and epigenetic machinery for regulating multicellularity. It will be interesting, in future studies, to examine whether the “multicellular” genes that we identified here, functionally evolved during the transition to multicellularity or whether it is simply a question of when these genes are activated that facilitates multicellularity. While our work, and the work of many other labs, has demonstrated that genetic changes helped contribute to the evolution of multicellularity, this study adds to this work and suggests that these genetic changes were accompanied by contemporaneous epigenetic changes and suggests that the co-evolution of new transcription factor networks and chromatin remodeling coalesced to facilitate multicellularity.

Methods

Strain

The Axenic strain (AX2) of *Dictyostelium discoideum* was generously given by Peter Devreotes at Johns Hopkins University. This strain was used as the main material for RNA-seq, ChIP-seq and ATAC-seq experiments. The other axenic strain (AX4) and the *srfA* knock out mutant was obtained from Dicty Stock Center. We cultured the strain at 22 °C shaken in the incubator at 180 rpm in HL5 medium with 300 µg/ml streptomycin.

Antibodies

Antibodies used for histone chromatin immunoprecipitation and western blots were: rabbit anti-H3 (Abcam Ab1791), rabbit anti-H3K27ac (Abcam Ab4729), rabbit anti-H3K4me3 (Abcam Ab8580), rabbit anti-H3K9me3 (Abcam Ab8898), rabbit anti-H3K27me3 (Abcam Ab6002), rabbit anti-H3K36me3 (Abcam Ab 9050), and rabbit

anti-H3K4me1 (Abcam Ab8895). These antibodies have demonstrated specificity in eukaryotes [53–59].

Quantitative mass spectrometry

Histone chemical derivatization and peptide preparation for mass spectrometry

Histones were propionylated as previously described [60]. Briefly, proteins were diluted with acetonitrile and neutralized with ammonium hydroxide. Chemical derivatization of lysines was performed with propionic anhydride twice, proteins were digested with trypsin, then two additional rounds of derivatization were performed. Propionylated peptides were desalted using an MCX protocol (Oasis MCX cartridge 1 cc/30 mg LP, Waters Corporation).

Orbitrap liquid chromatography-mass spectrometry

Peptides were analyzed with a Thermo Dionex UPLC coupled with a Thermo Q-Exactive HF tandem mass spectrometer. We used an in-house pulled column created from 75 μm inner diameter fused silica capillary packed with 3 μm ReproSil-Pur C18 beads (Dr. Maisch) to 30 cm. Solvent A was 0.1% formic acid in water, while solvent B was 0.1% formic acid in 80% acetonitrile. For each injection, we loaded approximately 1 μg peptides and separated them using a 47-min gradient from 2 to 35% B, followed by a 13 min washing gradient.

For data dependent acquisition (DDA) analysis, a top 20 method was used (default charge state 2, minimum AGC target 5e4, charge exclusion 1 and > 6, and dynamic exclusion 4 s) with full MS (resolution 120,000; AGC 3e6, maximum IT 50 ms) and data dependent-MS2 (resolution 15,000; AGC 1e6, max IT 40 ms, isolation window 2.0 m/z, NCE 27).

For data-independent acquisition (DIA) analysis, the Thermo Q-Exactive HF was configured to acquire 25 \times 24 m/z (covering 300–900 m/z) precursor isolation window DIA spectra (30,000 resolution, AGC target 1e6, maximum inject time 60 ms, 27 NCE) using an optimized staggered window pattern. Precursor spectra (target range \pm 15 m/z at 60,000 resolution, AGC target 1e6, maximum inject time 60 ms) were interspersed every 51 MS/MS spectra. The isolation window scheme is detailed on the Panorama-Web project folder, https://panoramaweb.org/dictyostelium_histones.url

For parallel reaction monitoring (PRM), a Thermo Fusion tribrid mass spectrometer was configured to acquire target m/z using Orbitrap tMS² (30,000 resolution, AGC target Standard, maximum inject time 60 ms, 30% CID CE) with precursor spectra (target range 290–910 m/z at 60,000 resolution, AGC target 1e6, maximum inject time 60 ms) interspersed every 50 MS/MS spectra. The isolation list is detailed on the Panorama-Web project folder, https://panoramaweb.org/dictyostelium_histones.url

DDA data analysis

RAW files were converted to MZML using MSConvert (version 3.0.18). Byonic (version 3.10.2) was used to search a focused *Dictyostelium* FASTA with histone proteins and contaminants (on the PanoramaWeb project folder, https://panoramaweb.org/dictyostelium_histones.url) within Proteome Discoverer (version 2.4) using the following parameters: arg-c digestion, 2 max missed cleavages, minimum peptide length 7; 8

maximum modifications; fixed carbamidomethyl on C and U, variable oxidation on M, variable acetylation on protein N-terminus and K, methylation on R, dimethylation on K and R, trimethylation on protein N-terminal A and K, phosphorylation on S, T, and Y, propionylation on peptide N-terminus and K, propionyl-methylation on K, and GG-propionylation on K; precursor mass tolerance 10 ppm, product mass tolerance 20 ppm; HCD fragmentation.

DIA data analysis

Mass spectrometry data files were demultiplexed and converted to MZML using MSConvert (version 3.0.18). Database search results from DDA data analysis were built into a spectral library using Bibliospec [61] in Skyline-daily (version 20.1) [62]. DIA files were imported into Skyline using MS/MS IDs in the spectral library (all settings default except Peptide Settings: digestion enzyme ArgC [R|P], background proteome from Uniprot *Dictyostelium* FASTA (accessed 21 Dec 2020, 12742 entries); Transition Settings: Filter for precursor charges 1–3, ion charges 1–2, ion types y, b, p; from ion 1 to last ion; Full-Scan MS1 filtering with isotope peaks included (Count) and MS/MS filtering for DIA acquisition, centroid product mass analyzer, and Results only isolation scheme; use only scans within 5 min of MS/MS IDs).

PRM data analysis

A target list was built from the Cartesian product combination of known modifications and the *Dictyostelium* histone proteins (Uniprot, accessed 2020 Dec 31, 7 entries) using a custom Python script. Skyline-daily (version 20.1) was configured with these target peptides (all settings default except peptide settings: digestion enzyme ArgC [R|P], background proteome from Uniprot *Dictyostelium* histone FASTA (accessed 31 Dec 2020, 7 entries); transition settings: filter for precursor charges 1–3, ion charges 1–2, ion types y, b, p; from ion 1 to last ion; Full-Scan MS1 filtering with isotope peaks included (count) and MS/MS filtering for targeted acquisition, centroid product mass analyzer; include all matching scans). Data was imported and peaks were manually curated using isotope dot product, dot product when spectral library entries were available, and retention time information.

Statistical analysis of mass spectrometry data

Skyline Group Comparison feature was used to calculate differential abundances using the processed DIA data. Parameters included setting the Control group value to “vegetative”; value to compare against as either mound, fruiting, or streaming; Normalization method: Total Ion current; confidence level: 95%; scope: peptide; summary method: Tukey’s median polish; using zero for missing peaks.

RNA-seq

mRNA was extracted from AX2 cells stored at -80°C using Dynabeads mRNA direct micro purification kit (Invitrogen). Samples were lysed using tissue homogenizer (Kontes) in the lysis buffer. Two biological replicates were assigned for each vegetative, streaming, mound, and fruiting stage respectively. In total 8 groups were prepared for RNA-seq libraries with 500 ng mRNA as starting materials using NEXTflex Illumina

qRNA-Seq Library Prep Kit (Bioo Scientific). In short, mRNA was fragmented in a cationic buffer and underwent first and second strand synthesis, adenylation, molecular indexed adapter ligation and PCR amplification for 11 cycles. The amplified products were cleaned up by Agencourt AmPure XP Magnetic Beads (Beckman Coulter) and quality checked by Agilent 2200 TapeStation D1000. The optimal cluster density was determined by KAPA library quantification kit before pooling the samples into a pooled library (5 nM) and sequencing through Nextseq 500 platform.

The fastq files were filtered with Q score 30 or above and were trimmed for the sequencing adaptors using CutAdapt (version 2.5). The analysis pipeline started with pseudo-alignment of the reads using kallisto [63], followed by building an index and quantification process, which generated h5 files containing the raw binary files for preliminary differential analysis in sleuth [64]. Differential analysis was conducted in EdgeR [65]; briefly, the DEGList object was created, and genes were kept that achieved at least one count per million (cpm) in two biological replicate samples. The effective library size was computed using Trimmed Mean of M values (TMM) normalization. The significant genes were deduced by computing NB exact genewise tests with corrected $p < 0.05$. The gene ontology (GO term Gene Ontology Consortium 2015) for over-representation analysis and diagrams were generated by R package clusterProfiler (version 3.18.0) [66]. Sequencing QC matrix is shown in Additional file 9: Table S8.

ChIP-seq

10 million *D. discoideum* cells were harvested and dissociated at different life cycle stages as described in [67] with the following modifications. After dissociation cells were cross-linked with 1x linking buffer (11 mM HEPES-NaOH, 110 mM NaCl, 1.1 mM EDTA, and 1.1 mM EGTA) with 1% formaldehyde for 10 min at RT and quenched by 2 M glycine. The samples were fragmented by applying a biorupter for 25 cycles in the cold room with 30 seconds on and 30 seconds off per cycle. The precleared chromatin was incubated with antibody-coupled bead slurries and rotated at 4 °C overnight. DNA was eluted from the beads by 200 μ l TE with 1% SDS.

Histone ChIP-Seq libraries were generated using the Ovation Ultralow V2 kit (Nugen, 0344-32) according to the manufacturer's instructions. Libraries were PCR amplified for 13-16 cycles depending on the antibody, and library quality was assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies). Seventy-five bp reads were generated on the Illumina Nextseq 500 (Illumina). ChIP-seq were sequenced to a depth of at least 8 million total reads/sample.

The raw fastq files were trimmed for the first and last 8 nucleotides, the multiqc (version 1.5) was used to summarize the fastQC (version 0.11.3) results [68], and the trimmed files were mapped to the *Dictyostelium discoideum* genome (version 2.5) [21] using bowtie2 (version 2.3.4.3) [69]. The duplicated reads were removed by picard and sorted bam files were subject to peak calling with MACS2 (version 2.1.1.20160309) using H3 as controls [70, 71]. Both DNA input and H3 were used as controls. The peaks were visualized by Integrative Genomics Viewer (IGV 2.3.44). Differential binding analyses were performed by R package Diffbind (version 3.0.14) using log₂ normalized ChIP read counts with FDR < 0.05 [72]. The metagene analysis and genomic feature distribution were conducted by ChIPSeeker (version 1.26.2) using anotatePeak function

with narrowpeak files as input [73]. Transcription start site (TSS), exon, intron, and transcription termination site (TTS) were determined by using `makeTxDbFromBiomart` function from the Ensemble BioMart database (version 2.7). Promoters were defined by the regions 1500 bp upstream to the TSS, and downstream was defined as the 3000 bp downstream of gene end (TTS). Since some annotations may overlap, we adopt the default principle in `ChIPSeeker` to prioritize genomic annotation in the following order: promoter, exon, intron, downstream and intergenic. Sequencing QC matrix is shown in Additional file 9: Table S8.

ATAC-seq

To assess regions of open chromatin, frozen pellets of *D. discoideum* cells from the various stages were washed twice with cold PBS and lysed with 10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM MgCl₂, 0.1% Igepal-630. Nuclei were transposed using the Nextera DNA Library Prep Kit (Illumina, FC-121-1030). Transposed libraries were amplified for 11 cycles and size selected for fragments ranging 200-1000 bp. ATAC-seq experiments were sequenced to a minimum depth of 10 million reads/sample. For preparation of fresh samples, *Dictyostelium* were physically dissociated and washed twice with PBS with 0.04% BSA and lysed with 10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM MgCl₂, 0.5% Tween 20, 0.5% NP-50, 0.05% digitonin (Promega G944A), 1% BSA, and 0.5% cellulase on ice or room temperature for 30 min to fully lyse the cells. Validation of equivalent ATACseq lysine of different stages was performed by trypan blue staining and optimized by real-time PCR detection of ribosomal RNA release.

The raw fastq files were trimmed for the first and last 8 nucleotides, the `multiqc` was used to summarize the fastQC results [68], and the trimmed files were mapped to the *Dictyostelium discoideum* genome [21] using `bowtie2` [69], and the mitochondria genome was removed in the alignment process. The duplicated reads were removed by `picard` and sorted bam files were subject to peak calling with `MACS2` [70, 71]. The peaks were visualized in Integrative Genomics Viewer (IGV 2.3.44). Likewise, open chromatin regions and peak annotation were deducted with `Diffbind` [72] and `ChIPSeeker`, respectively [73]. Sequencing QC matrix is shown in Additional file 9: Table S8.

Motif analysis

The stage/unicellular/multicellular specific open chromatin regions, enriched histone binding sites, and the promoters of significant enriched genes were extracted as the input sequence for `homer` (version 4.11.1) [74]. All the promoter sequences were obtained from `dictyBase` as the background sequence [75]. The motif discovery was performed using “`findMotif.pl`” command from `homer`.

Orthologous gene analysis

The phylogenetic tree was constructed on interactive tree of life (iTOL version 6) [76]. The orthologous genes among *D. discoideum*, *C.elegans* and *S.pombe* were obtained using `metaphOrs` (version 1.0) based on phylogeny orthology prediction at genomic scale [45]. The comparison among the species and intersection regions were extracted using `bedtools` (version 2.27.1) [77]. The graphs were generated on R (version 4.0.2), `excel` or `prism` (version 9).

Multicellularity Assay

The *D. discoideum* cells were harvested in mid-log phase for 5×10^8 cells. The cells were suspended in developmental buffer (5 mM Na_2HPO_4 , 5 mM KH_2PO_4 , 1 mM CaCl_2 , and 2 mM MgCl_2) and washed 3 times. The multicellularity assays were performed on KK2 agar plates. 100 nM, 500 nM, and 1000 nM chemical inhibitors (Trichostatin A, T8552; BIX 01294 trihydrochloride hydrate, B9311; (R)-PFI-2, SML1408p; Sinefungin, ab144518) were added to the KK2 agar plates. The picture capture time points for the multicellularity assay with chemical treatment were set at 8 h, 16 h, 19 h, and 24 h. The picture capture time points for the multicellularity assay with mutant strains were set at 8 h, 16 h, 21 h, 25 h, and 41 h.

Cell number quantification and morphological measurement

To quantify the germ cell number in the fruiting bodies, the assay was performed as in [78]. Germ cell number from individual fruiting bodies were collected by pipette and were placed in 20 μl PBS. The cells were counted with a hemocytometer (Hausser Scientific Cat.#3520). The diameter of the individual mound and fruiting bodies were measured using stereo Discovery V8 microscope.

Chemotaxis assay

Chemotaxis assays were performed as in [79]. Briefly, the protocol was modified as follows. The *D. discoideum* cells were cultured in a density of 2×10^6 cells/ml for the purpose of the experiment. The cells were pelleted in a 1X KK2 buffer (2.2 g KH_2PO_4 , 0.7 g K_2HPO_4). The cells were starved for 5–8 h in a shaker and resuspended at a concentration of 1×10^8 cells/ml. For making the chemotaxis plates, 4 wells were made in the 1X KK2 agar plates (1X KK2 buffer and 1.5 g Agar). 4 small dots at 6 mm from the edge of each well were marked on the agar plates for showing the location of cells deposition. 20 μl of 250 μM of folic acid (Millipore Sigma CAS 59-30-3) dissolved in 100 mM NaOH were injected into the wells by a P10 pipette for 20 min absorbing by the agar. 1 μl of *D. discoideum* were loaded on the dots and the plates were incubated at room temperature in a humidified chamber. The movement of cells were captured after 5 h using Zeiss stereoREO Discovery V8 microscope. The number of cells were counted in each direction (north, south, east, and west) in a 30° angle range and plotted in Prism 8 (Version 8.1.1).

Annexin V, propidium iodide, and trypan blue staining

Cell viability was examined by trypan blue (VWR), propidium iodide (BD Biosciences), and Annexin V staining (BD Biosciences). For Annexin V staining, cells were washed twice with cold PBS and suspended in 1X Binding Buffer (10 mM Hepes pH 7.4, 140 mM NaCl, and 2.5 mM CaCl_2) at a concentration of 1 million cells/ml. One hundred microliters of the solution was transferred to a new tube and 5 μl of FITC Annexin V was added to the cells and incubated for 15 min at room temperature in the dark. The samples were imaged using a fluorescent microscope (Nikon Eclipse E600) with 485/535 nm (FITC) filter and images were analyzed by ImageJ (1.0).

For trypan blue staining, 10 μ l 0.4% solution of trypan blue was mixed with 10 μ l of cells at 5E5 cells/ml. The mixture was counted on a hemacytometer immediately after mixture.

For PI staining, 1E6 *D. discoideum* cells were treated with 1 μ M TSA, 5 μ M sinefungin, or 10 μ M BIX 01294 for 6 h. Cells were subsequently washed and resuspended with cold PBS supplemented with 5% FBS and stained with 2 mg/ml propidium iodide (BD Biosciences) at room temperature for 10 min. Stained cells were fixed with 1% PFA. Samples were analyzed by BD LSR Fortessa (BD Biosciences) and data analysis was performed using Flow Jo (Tree Star).

CRISPR-Cas9 knock-out experiment

The all-in-one sgRNA and Cas9 expression vector pTM1285 were obtained from Quantitative Biology Center (QBIC) Riken, Japan. The sgRNAs were designed on CRISPOR tool [80]. sgRNA were ligated to predigested pTM1285 with BbsI (NEB), and stable competent cells were transformed with the resulting construct (NEB, C3040). Ligation was confirmed by PCR and Sanger sequencing. Primers used were sense oligo for the target 5'-AGCAN(1)NNNNNNNNNNNNNNNNNNNN(20)-3' and tracrRNA 5'-AAGCTTAAAAAAGCACCGACTCGGTGCC-3'. The sequencing primer was 5'-AAGCTTAAAAAAGCACCGACTCGGTGCC-3'. The validated plasmids were transformed into the cells using calcium phosphate precipitation method [81]. Briefly, 10 μ g plasmid DNA is co-precipitated with 60 μ l of 1.25 M CaCl₂ in HBS buffer (4.0 g NaCl, 0.18 g KCl, 0.05 g Na₂HPO₄, 2.5 g HEPES and 0.5 g D-glucose, pH adjusted to 7.1). The DNA solution was added to the plates for 30 min and then 12.5 ml Bis-Tris (2.1 g Bis-Tris, 10 g proteose peptone, 5 g yeast extract, and 10 g D-glucose in distilled water, pH adjusted to 7.1 with HCl) for 4–8 h. The condensed DNA on cell surface was transformed into the cells by 15–18% glycerol for exact 5 min and incubated with 12.5 ml Bis-Tris solution overnight to allow the expression of the selection marker. The medium was replaced by HL5 with 10–15 μ g/ml of G418 after 16 h, and cultured for another 3 days. The cells were plated on SM agar plates with *K. aerogenes* and incubated for 3–4 days until plaque formed. The cells were transferred and expanded in the HL5 medium with 300 μ g/ml streptomycin. Knock-out strains were validated by sanger sequencing.

Overexpression *Dictyostelium* strains

cDNA of *hdaD* and *smc11* was cloned into pJSK123 over-expression plasmid [82] between BamHI and AvrII sites. The recombinant plasmids were transformed into cells using calcium phosphate precipitation method [81]. Briefly, 10 μ g plasmid DNA is co-precipitated with 60 μ l of 1.25 M CaCl₂ in HBS buffer (4.0 g NaCl, 0.18 g KCl, 0.05 g Na₂HPO₄, 2.5 g HEPES and 0.5 g D-glucose, pH adjusted to 7.1). The DNA solution was added to the plates for 30 min and then 12.5 ml Bis-Tris (2.1 g Bis-Tris, 10 g proteose peptone, 5 g yeast extract, and 10 g D-glucose in distilled water, pH adjusted to 7.1 with HCl) was added for 4–8 h. The condensed DNA on the cell surface was transformed into the cells by 15–18% glycerol for 5 min and incubated with 12.5 ml Bis-Tris solution overnight to allow the expression of the selection marker. The medium was replaced with HL5 with 10–15 μ g/ml of G418 after 16 h, and cultured for another 3 days.

The cells were then plated on SM agar plates with *K. aerogenes* and incubated for 3–4 days until plaques formed. The cells were transferred and expanded in the HL5 medium with 300 µg/ml streptomycin. Over-expression strains were validated by real-time PCR.

Quantitative real-time PCR (qRT-PCR)

The differential expression of candidate genes were analyzed using qRT-PCR. Total RNA was extracted from *D. discoideum* cells using Invitrogen PureLink RNA (Invitrogen, 12183025). First strand cDNA synthesis was conducted using SuperScript III First-Strand Synthesis (Invitrogen, 18080051). qRT-PCR of cDNA was performed using iTaq Universal SYBR Green Supermix (Bio-Rad, 172-5122) on a CFX96 Real-Time System (Bio-Rad). The PCR assay contained 1:4 diluted reverse transcription products, 1X master mix, and 200 nM of each primer, and was performed with 5-min initial denaturation at 95 °C followed by 40 cycles of 95 °C for 5 s and 60 °C for 45 s. The mtRNA of *D. discoideum* was used for internal normalization for each gene. The comparative Ct method ($\Delta\Delta C_t$) was used to calculate gene expression. The primer sequences of the selected genes were designed using Macvector (15.5.0). The following gene-specific primers were used: mtRNA-forward, 5'-gggtagtttgactggggcgg-3', mtRNA-reverse, 5'-cactttaatgggtgaacacc-3'; DDB_G0283437-forward, 5'-gagggtgtataggtacatttcagag-3', DDB_G0283437-reverse, 5'-tgggtccaattccaatcccaaac-3'; DDB_G02813787-forward, 5'-tccctaaccaaatcaaccaacga-3', DDB_G02813787-reverse, 5'-tatacatgaccagtttcggaggcaac-3'; DDB_G0288013-forward, 5'-acaagaaatcgattgggagatgt-3', DDB_G0288013-reverse, 5'-tggcattgttcttgattttct-3'; DDB_G0279267-forward, 5'-tggctaccaataaccatcacctccat-3', DDB_G0279267-reverse, 5'-cgagttcgaactattgctgttg-3'; DDB_G0274899-forward, 5'-ccagagggaataatagtctcgtg-3', DDB_G0274899-reverse, 5'-ctccaattcgaaccacaac-3'; DDB_G0267824-forward, 5'-gcagctgtagtagttcatcca-3', DDB_G0267824-reverse, 5'-tgtgcatattctttaccctcaa-3'; DDB_G0277333-forward, 5'-tggcatttctcaggcagtttcgacca-3', DDB_G0277333-reverse, 5'-ccaacatcagtgagggtattca-3'; DDB_G0293674-forward, 5'-accaacgtgctctctcgtcttgag-3', DDB_G0293674-reverse, 5'-ggaacgatgtttgtcattacaccac-3'; DDB_G0294934-forward, 5'-gggtagtttgactggggcgg-3', DDB_G0294934-reverse, 5'-cactttaatgggtgaacacc-3'; DDB_G0279267_OE-forward, 5'-tatggatccatgtcaacaattcatc-3', DDB_G0279267_OE-reverse, 5'-taccctaggttaatcgtcttcgctta-3'; DDB_G0288013_OE-forward, 5'-tatggatccatggatattgcacaatcaattca-3', DDB_G0288013_OE-reverse, 5'-taccctaggtataaagttttaagaaatttg-3', DDB_G0279267_Cat-forward, 5'-caacaatcatcacaattacc-3', DDB_G0279267_Cat-reverse, 5'-taccctaggttaatcgtcttcgctta-3'.

Drug assays

HEK 293 T cells (ATCC CRL-3216) were seeded on 6 well plates overnight, then treated with 0.25 µM, 0.5 µM, 1 µM, or 5 µM Trichostatin A (T8552, Millipore Sigma); or with 0.5 µM, 1 µM, 5 µM, or 10 µM Sinefungin (ab144518); or with 1 µM, 10 µM, 20 µM, or 30 µM BIX 01294 (B0311) for 16 h. Each concentration was seed with three replicates of 1 million cells before collection and analysis of effects on histone methylation and acetylation by western blotting.

Western blot analysis

Cells were lysed with RIPA buffer and quantified by Bradford assay (Biorad 5000006). Five to 10 μg of proteins were loaded for gel electrophoresis and transferred to nitrocellulose membrane (Bio rad 1620097) with 100 V for 1.5 h. The membranes were incubated in 5% milk for 1 h at room temperature and incubated with primary antibodies (1:1000 dilution). After washing with TBST 3 times, each for 10 min, the membranes were incubated with secondary antibody (Millipore 3136001, Rabbit IgG, 1:3000 dilution). The blots were reacted with chemiluminescent HRP substrate (Millipore WBKLS0500) and imaged in ChemiDoc Touch Imaging System (Bio rad).

Deacetylation assays

The coding sequence of the catalytic domain of *hdaD* was cloned as an in-frame fusion to the His tagged vector pET28a. The recombinant protein was expressed in *E. coli* BL21. Overnight induction of protein expression was carried out with 1 mM IPTG at 18 °C. Bacteria were harvested at 4000 rpm, 4 °C, and resuspended in 10 mL protein purification lysis buffer (50 mM Tris-HCl pH 7.5, 0.25 M NaCl, 0.1% Triton-X, 1 mM PMSF, 1 mM DTT, 20 mM imidazole, and protease inhibitors). After freezing the pellet at - 80 °C for 1 h, the lysate was sonicated with a Bioruptor for 5 min on high level with 30 s on and 30 s off. Proteins were purified with Ni-NTA beads (Qiagen). Proteins and beads were washed 3 times with protein purification lysis buffer before incubating the beads with elution buffer (400 mM imidazole in protein purification lysis buffer, pH 8.0) for 30 min. Eluates were dialyzed overnight at 4 °C with dialysis buffer (50 mM Tris-HCl pH 8.0, 1 mM EDTA, 1 mM DTT, and 20% glycerol). Bradford assays and SDS-page gel electrophoresis followed by coomassie staining were performed to determine the integrity and quantity of purified proteins. Histone proteins were purified from *D. discoideum* using the EZExtract core histone isolation kit (BioVision). In vitro deacetylation reactions were performed with 10 μg of histone proteins and 5 μg of *hdaD* protein or 5 μg GST-RPA-2 protein as a control. The deacetylation reaction was performed in a buffer solution (50 mM Tris-HCl, 4 mM KCl, 4 mM MgCl₂, 0.2 mM DTT, 1 mM NAD⁺) and incubated at 30 °C for 30 min. Western blot analysis was performed as previously described.

Single-cell RNA sequencing data analysis

D. discoideum was dissociated into single cells from different stages by physical pipetting with PBS + 0.04% BSA. For each sample, the cells were counted and diluted to 1000 cell/ μl in 1X PBS with 0.04% Bovine Serum Albumin (BSA) prior loading to 10X Genomics Chromium machine. 10X Chromium Single Cell 3' v3 reagent kits were used according to the manufacturer's instructions to generate scRNAseq libraries. Libraries were sequenced paired-end 150 bp on Illumina Hiseq4000/Novoseq (Novogene).

Single-cell RNA sequencing data analysis

Fastq files were processed with Cell Ranger (version 3.1.0) (<https://support.10xgenomics.com>) on the DolphinNext pipeline builder [83] to generate the

merged gene-barcode matrix. Matrix was loaded into R (version 3.6.3) and analyzed mainly with Seurat package ([84], version 3.1.5). Low-quality cells or outliers were filtered out from the data by the following threshold: $200 < \text{total Gene count} < 7000$, $500 < \text{total UMI count} < 30000$. Filtered matrices were normalized with SCTransform [85] and reduced the dimension with principal component analysis. The batch effect was corrected by harmony [86]. The significant principal components (PCs) are used to generate the UMAP plot for visualization. Clustering was performed by the Louvain algorithm based on K-nearest neighbor graph. Hurdle model tailored was used to perform differential expression analysis of the individual clusters [87]. Gene Ontology (GO) analysis was performed with clusterProfiler (version 3.18.0) [66].

For trajectory inference analysis, the cells from vegetative, slug, and fruiting stages were down-sampled and merged. The uninformative genes were filtered out by keeping the genes expressing 3 UMIs per cell for at least 10 cells. The filtered matrices were normalized with full quantile normalization and reduced the dimension with PCA. The first 3 PCs were used for k-means clustering and subsequent analysis. Slingshot was performed to predict the multicellularity lineage [88]. The generalized additive model was used to identify the temporally expressed gene through the pseudotime.

Electron microscopy

Dictyostelium cells were synchronized according to previously published methods [89, 90]. Briefly, the cells were grown to stationary phase with a density of 1.2×10^7 /ml. These cells were diluted into fresh HL5 medium at a density of 10^6 /ml. Then the cells were shifted to 11.5°C incubation for 20 h in 6-well plates. At this time $1 \mu\text{M}$ TSA was added to the positive control cells. The synchronized cells were fixed in fixative (2.5% glutaraldehyde 1.25% paraformaldehyde and 0.03% picric acid in 0.1 M sodium cacodylate buffer (4.28 g sodium cacodylate, 25 g calcium chloride, 2.5 ml 0.2 N hydrochloric acid in 200 ml of distilled water, pH 7.4) with 1:1 dilution of cell medium for 16 h at 4°C . Then, the cells were washed in 0.1 M cacodylate buffer and postfixed with 1% osmium tetroxide (OsO_4)/1.5% potassium ferrocyanide (KFeCN_6) for 1 h. The cells were washed twice in water, and once in maleate buffer (MB)(50 mM maleate acid, 0.2 N NaOH) and incubated in 1% uranyl acetate for 1 h followed by 2 washes in water and subsequent dehydration in grades of alcohol (10 min each; 50%, 70%, 90%, 2×10 min 100%). The samples were then put in propyleneoxide for 1 h and infiltrated overnight in a 1:1 mixture of propyleneoxide and TAAB Epon (TAAB Laboratories Equipment Ltd, <https://taab.co.uk>). The following day the samples were embedded in TAAB Epon and polymerized at 60°C for 48 h. Ultrathin sections (about 60 nm) were cut on a Reichert Ultracut-S microtome, picked up on to copper grids, stained with lead citrate, and examined in a JEOL 1200EX transmission electron microscope or a TecnaiG² Spirit BioTWIN, and 30 images of each condition were recorded with an AMT 2 k CCD camera.

The heterochromatin ratios were quantified with FIJI (Version 1.0). A 15% threshold was set for all the images which were then binarized to reveal heterochromatin regions for quantification. The heterochromatin regions were calculated for each of 30 nuclei (excluding the nucleolus) and calculated as a percentage.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-021-02360-9>.

Additional file 1: Supplementary figures S1-S15.

Additional file 2: Table S1. Differential ATAC peak regions.

Additional file 3: Table S2. ChIP-seq differential binding sites.

Additional file 4: Table S3. ChIP-seq annotated peak regions.

Additional file 5: Table S4. RNA-seq differentially expressed genes.

Additional file 6: Table S5. ATAC-seq and RNA-seq correlation.

Additional file 7: Table S6. Ortholog pairs.

Additional file 8: Table S7. Orthologous gene list.

Additional file 9: Table S8. Sequencing QC summary.

Additional file 10. Review history.

Acknowledgements

General: We thank Peter Devreotes, Marc Edwards, Budri Sharif, and other members of the Devreotes lab for the *Dictyostelium* strain and experimental advice about *Dictyostelium* culturing conditions. We thank J. Lieberman and A. Brunet for critical reading of the manuscript.

Review history

The review history is available as Additional file 10.

Peer review information

Barbara Cheifet was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Authors' contributions

S.Y.W., E.A.P., and E.L.G. conceived and planned the study and wrote the paper. All analysis and chemical and genetic manipulation of *Dictyostelium* experiments were performed by S.Y.W. E.A.P. performed ChIPseq and ATACseq experiments. I-H.W. performed single-cell RNAseq analysis. L.K.P. performed quantitative mass spectrometry and analysis. H.B. helped with chemical manipulation of *Dictyostelium* and purification of HdaD experiments. K.T. performed cell viability experiments. C.F. optimized chemotaxis assays and helped produce Figure S8. G.S. helped produce Fig. 4d and S1. B.A.G. advised L.K.P. P.L.G. advised I-H.W. and helped write the paper. E.L.G. generated staged *Dictyostelium* for RNA-seq, ChIPseq, and ATACseq experiments and performed chromatin immunoprecipitations. All authors discussed the results and commented on the manuscript. The authors read and approved the final manuscript.

Authors' information

Twitter handles: @simonyuanwang (Simon Yuan Wang); @IHaoWang1 (I-Hao Wang); @lkpino (Lindsay Kristina Pino); @hlbushnell (Henry L. Bushnell); @colettefritsche (Colette Fritsche); @GarciaLab (Benjamin Aaron Garcia); @Greer_Lab (Paul Lieberman Greer); @EricGreerLab (Eric Lieberman Greer).

Funding

S.Y.W. was supported by Croucher Foundation and Charles King fellowships. E.A.P. was supported by an LSRF fellowship and by an NIH grant (K99AG064042). L.K.P. was supported by an NIH grant (T32CA009140). K.T. was supported by a JSPS fellowship. B.A.G. was supported by an NIH grant (R01GM110174). P.L.G. was supported by grants from the Searle Scholars Program, Rita Allen Foundation, and by an NIH grant (DP2AG067490). This research was supported by NIH grants (R00AG043550 and DP2AG055947) to E.L.G.

Availability of data and materials

All ChIPseq, RNAseq, and ATACseq datasets are available in the GEO database accession # GSE137604 [91]. The codes for all bioinformatic analyses are available at GitHub at <https://github.com/BCHGreerlab/Dictyostelium> [92]. Unprocessed and uncompressed imaging data is available at <https://data.mendeley.com/datasets/zfhh6xzck4/3>. Mass spectrometry data is available in the PanoramaWeb project folder, https://panoramaweb.org/dictyostelium_histones.url. Mass spectrometry data is also available at ProteomeXchange repository ID: PXD024225 [93].

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors have no competing interests.

Author details

¹Division of Newborn Medicine, Boston Children's Hospital, 300 Longwood Avenue, Boston, MA 02115, USA.

²Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA. ³Department of Neurobiology, Harvard

Medical School, Boston, MA 02115, USA. ⁴Program in Molecular Medicine, University of Massachusetts Medical School,

Worcester, MA 01605, USA. ⁵Department of Biochemistry and Biophysics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA 19104, USA.

Received: 17 February 2021 Accepted: 22 April 2021

Published online: 04 May 2021

References

- Herron MD, Rashidi A, Shelton DE, Driscoll WW. Cellular differentiation and individuality in the 'minor' multicellular taxa. *Biol Rev Camb Philos Soc.* 2013;88:844–61.
- Grosberg RK, Strathmann RR. The evolution of multicellularity: a minor major transition? *Annu Rev Ecol Evol Syst.* 2007; 38:621–54.
- King N. The unicellular ancestry of animal development. *Dev Cell.* 2004;7:313–25.
- Rokas A. The origins of multicellularity and the early history of the genetic toolkit for animal development. *Annu Rev Genet.* 2008;42:235–51.
- Olson BJ. Multicellularity: from brief encounters to lifelong unions. *Elife.* 2013;2:e01893.
- Suga H, Chen Z, de Mendoza A, Sebe-Pedros A, Brown MW, Kramer E, et al. The *Capsaspora* genome reveals a complex unicellular prehistory of animals. *Nat Commun.* 2013;4:2325.
- King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, et al. The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature.* 2008;451:783–8.
- Fairclough SR, Chen Z, Kramer E, Zeng Q, Young S, Robertson HM, et al. Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate *Salpingoeca rosetta*. *Genome Biol.* 2013;14:R15.
- Lang D, Weiche B, Timmerhaus G, Richardt S, Riano-Pachon DM, Correa LG, et al. Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and correlation with complexity. *Genome Biol Evol.* 2010;2:488–503.
- Sebe-Pedros A, Ballare C, Parra-Acero H, Chiva C, Tena JJ, Sabido E, et al. The dynamic regulatory genome of *Capsaspora* and the origin of animal multicellularity. *Cell.* 2016;165:1224–37.
- Iyer LM, Anantharaman V, Wolf MY, Aravind L. Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int J Parasitol.* 2008;38:1–31.
- Koonin EV, Aravind L, Kondrashov AS. The impact of comparative genomics on our understanding of evolution. *Cell.* 2000;101:573–6.
- Aravind L, Abhiman S, Iyer LM. Natural history of the eukaryotic chromatin protein methylation system. *Prog Mol Biol Transl Sci.* 2011;101:105–76.
- Anantharaman V, Iyer LM, Aravind L. Comparative genomics of protists: new insights into the evolution of eukaryotic signal transduction and gene regulation. *Annu Rev Microbiol.* 2007;61:453–75.
- Raper KB. Development patterns in simple slime molds. *Growth.* 1941;5:41–76.
- Bonner JT. A note on the number of cells in a slug of *Dictyostelium discoideum*. <http://DictyBase.org>. 2012;88:253–4.
- Bonner JT, Davidowski TA, Hsu W-L, Lapeyrolerie DA, Suthers HLB. The role of surface water and light on differentiation in the cellular slime molds. *Differentiation.* 1982;21:123–6.
- Fey P, Kowal AS, Gaudet P, Pilcher KE, Chisholm RL. Protocols for growth and development of *Dictyostelium discoideum*. *Nat Protoc.* 2007;2:1307–16.
- Pedersen MT, Helin K. Histone demethylases in development and disease. *Trends Cell Biol.* 2010;20:662–71.
- Eckersley-Maslin MA, Alda-Catalinas C, Reik W. Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. *Nat Rev Mol Cell Biol.* 2018;19:436–50.
- Eichinger L, Pachebat JA, Glockner G, Rajandream MA, Sucgang R, Berriman M, et al. The genome of the social amoeba *Dictyostelium discoideum*. *Nature.* 2005;435:43–57.
- Wiegand S, Kruse J, Gronemann S, Hammann C. Efficient generation of gene knockout plasmids for *Dictyostelium discoideum* using one-step cloning. *Genomics.* 2011;97:321–5.
- Kuwayama H, Yanagida T, Ueda M. DNA oligonucleotide-assisted genetic manipulation increases transformation and homologous recombination efficiencies: Evidence from gene targeting of *Dictyostelium discoideum*. *J Biotechnol.* 2008; 133:418–23.
- Faix J, Linkner J, Nordholz B, Platt JL, Liao X-H, Kimmel AR. The Application of the Cre-loxP System for Generating Multiple Knock-out and Knock-in Targeted Loci. *Dictyostelium Discoideum Protoc.* 2013;983:249–67.
- Kaller M, Nellen W, Chubb JR. Epigenetics in *Dictyostelium*. *Methods Mol Biol.* 2006;346:491–505.
- Stevense M, Chubb JR, Muramoto T. Nuclear organization and transcriptional dynamics in *Dictyostelium*. *Dev Growth Differ.* 2011;53:576–86.
- Chubb JR, Bloomfield G, Xu Q, Kaller M, Ivens A, Skelton J, et al. Developmental timing in *Dictyostelium* is regulated by the Set1 histone methyltransferase. *Dev Biol.* 2006;292:519–32.
- Muller-Taubenberger A, Bonisch C, Furbringer M, Witteck F, Hake SB. The histone methyltransferase Dot1 is required for DNA damage repair and proper development in *Dictyostelium*. *Biochem Biophys Res Commun.* 2011;404:1016–22.
- Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods.* 2013;10:1213–8.
- Bai L, Morozov AV. Gene regulation by nucleosome positioning. *Trends Genet.* 2010;26:476–83.
- Shore P, Sharrocks AD. The MADS-box family of transcription factors. *Eur J Biochem.* 1995;229:1–13.
- Niklas KJ. The evolutionary-developmental origins of multicellularity. *Am J Bot.* 2014;101:6–25.
- Escalante R, Moreno N, Sastre L. *Dictyostelium discoideum* developmentally regulated genes whose expression is dependent on MADS box transcription factor SrfA. *Eukaryot Cell.* 2003;2:1327–35.
- Sawarkar R, Visweswariah SS, Nellen W, Nanjundiah V. Histone deacetylases regulate multicellular development in the social amoeba *Dictyostelium discoideum*. *J Mol Biol.* 2009;391:833–48.
- Galardi-Castilla M, Fernandez-Aguado I, Suarez T, Sastre L. Mef2A, a homologue of animal Mef2 transcription factors, regulates cell differentiation in *Dictyostelium discoideum*. *BMC Dev Biol.* 2013;13:12.

36. Parikh A, Miranda ER, Katoh-Kurasawa M, Fuller D, Rot G, Zagar L, et al. Conserved developmental transcriptomes in evolutionarily divergent species. *Genome Biol.* 2010;11:R35.
37. Brunet T, King N. The Origin of animal multicellularity and cell differentiation. *Dev Cell.* 2017;43:124–40.
38. Grau-Bove X, Torruella G, Donachie S, Suga H, Leonard G, Richards TA, et al. Dynamics of genomic innovation in the unicellular ancestry of animals. *Elife.* 2017;6:e26036.
39. Schick S, Fournier D, Thakurela S, Sahu SK, Garding A, Tiwari VK. Dynamics of chromatin accessibility and epigenetic state in response to UV damage. *J Cell Sci.* 2015;128:4380–94.
40. Cao J, Cusanovich DA, Ramani V, Aghamirzaie D, Pliner HA, Hill AJ, et al. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science.* 2018;361:1380–5.
41. Bernstein BE, Humphrey EL, Erlich RL, Schneider R, Bouman P, Liu JS, et al. Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc Natl Acad Sci U S A.* 2002;99:8695–700.
42. Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NC, et al. Active genes are tri-methylated at K4 of histone H3. *Nature.* 2002;419:407–11.
43. Kuspa A, Loomis WF. The genome of dictyostelium discoideum. *Methods Mol Biol.* 2006;346:15–30.
44. Kumar S, Stecher G, Suleski M, Hedges SB. timetree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol.* 2017;34:1812–9.
45. Pyszczyk LP, Huerta-Cepas J, Gabaldon T. MetaPhOrs: orthology and paralogy predictions from multiple phylogenetic evidence using a consistency-based confidence score. *Nucleic Acids Res.* 2011;39:e32.
46. Daugherty AC, Yeo RW, Buenrostro JD, Greenleaf WJ, Kundaje A, Brunet A. Chromatin accessibility dynamics reveal novel functional enhancers in *C. elegans*. *Genome Res.* 2017;27:2096–107.
47. Schep AN, Buenrostro JD, Denny SK, Schwartz K, Sherlock G, Greenleaf WJ. Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. *Genome Res.* 2015;25:1757–70.
48. Ramagopal S, Ennis HL. Regulation of synthesis of cell-specific ribosomal proteins during differentiation of *Dictyostelium discoideum*. *Proc Natl Acad Sci U S A.* 1981;78:3083–7.
49. Escalante R, Sastre L. A Serum Response Factor homolog is required for spore differentiation in *Dictyostelium*. *Development.* 1998;125:3801–8.
50. Lemerrier C, Verdel A, Galloo B, Curtet S, Brocard MP, Khochbin S. mHDA1/HDAC5 histone deacetylase interacts with and represses MEF2A transcriptional activity. *J Biol Chem.* 2000;275:15594–9.
51. Han A, He J, Wu Y, Liu JO, Chen L. Mechanism of recruitment of class II histone deacetylases by myocyte enhancer factor-2. *J Mol Biol.* 2005;345:91–102.
52. Schaap P. Evolutionary crossroads in developmental biology: *Dictyostelium discoideum*. *Development.* 2011;138:387–96.
53. Rakhimova A, Ura S, Hsu DW, Wang HY, Pears CJ, Lakin ND. Site-specific ADP-ribosylation of histone H2B in response to DNA double strand breaks. *Sci Rep.* 2017;7:43750.
54. Sun Z, Xu X, He J, Murray A, Sun MA, Wei X, et al. EGR1 recruits TET1 to shape the brain methylome during development and upon neuronal activity. *Nat Commun.* 2019;10:3892.
55. Greer EL, Maures TJ, Hauswirth AG, Green EM, Leeman DS, Maro GS, et al. Members of the H3K4 trimethylation complex regulate lifespan in a germline-dependent manner in *C. elegans*. *Nature.* 2010;466:383–7.
56. Klose RJ, Gardner KE, Liang G, Erdjument-Bromage H, Tempst P, Zhang Y. Demethylation of histone H3K36 and H3K9 by Rph1: a vestige of an H3K9 methylation system in *Saccharomyces cerevisiae*? *Mol Cell Biol.* 2007;27:3951–61.
57. Beltran M, Tavares M, Justin N, Khandelwal G, Ambrose J, Foster BM, et al. G-tract RNA removes Polycomb repressive complex 2 from genes. *Nat Struct Mol Biol.* 2019;26:899–909.
58. Tian W, Yan P, Xu N, Chakravorty A, Liefke R, Xi Q, et al. The HRP3 PWWP domain recognizes the minor groove of double-stranded DNA and recruits HRP3 to chromatin. *Nucleic Acids Res.* 2019;47:5436–48.
59. Weiterer SS, Meier-Soelch J, Georgomanolis T, Mizi A, Beyerlein A, Weiser H, et al. Distinct IL-1alpha-responsive enhancers promote acute and coordinated changes in chromatin topology in a hierarchical manner. *EMBO J.* 2020;39:e101533.
60. Sidoli S, Bhanu NV, Karch KR, Wang X, Garcia BA. Complete Workflow for Analysis of Histone Post-translational Modifications Using Bottom-up Mass Spectrometry: From Histone Extraction to Data Analysis. *J Vis Exp.* 2016;111:5411.
61. Frewen B, MacCoss MJ. Using BiblioSpec for creating and searching tandem MS peptide libraries. *Curr Protoc Bioinformatics.* 2007;Chapter 13(Unit 13):17.
62. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics.* 2010;26:966–8.
63. Bray NL, Pimentel H, Melsted P, Pachter L. Erratum: Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* 2016;34:888.
64. Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat Methods.* 2017;14:687–90.
65. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26:139–40.
66. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16:284–7.
67. Santhanam B, Cai H, Devreotes PN, Shauly G, Katoh-Kurasawa M. The GATA transcription factor GtaC regulates early developmental gene expression dynamics in *Dictyostelium*. *Nat Commun.* 2015;6:7551.
68. Ewels P, Magnusson M, Lundin S, Kaller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics.* 2016;32:3047–8.
69. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9:357–9.
70. Liu T. Use model-based Analysis of ChIP-Seq (MACS) to analyze short reads generated by sequencing protein-DNA interactions in embryonic stem cells. *Methods Mol Biol.* 2014;1150:81–95.
71. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 2008;9:R137.
72. Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature.* 2012;481:389–93.

73. Yu G, Wang LG, He QY. CHIPseeker: an R/Bioconductor package for CHIP peak annotation, comparison and visualization. *Bioinformatics*. 2015;31:2382–3.
74. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38:576–89.
75. Kreppel L, Fey P, Gaudet P, Just E, Kibbe WA, Chisholm RL, et al. dictyBase: a new Dictyostelium discoideum genome database. *Nucleic Acids Res*. 2004;32:D332–3.
76. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics*. 2007;23:127–8.
77. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26:841–2.
78. Stenhouse FO, Williams KL. Patterning in Dictyostelium discoideum: the proportions of the three differentiated cell types (spore, stalk, and basal disk) in the fruiting body. *Dev Biol*. 1977;59:140–52.
79. Hadwiger JA, Srinivasan J. Folic acid stimulation of the Galpha4 G protein-mediated signal transduction pathway inhibits anterior prestalk cell development in Dictyostelium. *Differentiation*. 1999;64:195–204.
80. Concordet JP, Haeussler M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res*. 2018;46:W242–5.
81. Gaudet P, Pilcher KE, Fey P, Chisholm RL. Transformation of Dictyostelium discoideum with plasmid DNA. *Nat Protoc*. 2007;2:1317–24.
82. King J, Keim M, Teo R, Weening KE, Kapur M, McQuillan K, et al. Genetic control of lithium sensitivity and regulation of inositol biosynthetic genes. *PLoS One*. 2010;5:e11151.
83. Yukselen O, Turkyilmaz O, Ozturk AR, Garber M, Kucukural A. DolphinNext: a distributed data processing platform for high throughput genomics. *BMC Genomics*. 2020;21:310.
84. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol*. 2018;36:411–20.
85. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol*. 2019;20:296.
86. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods*. 2019;16:1289–96.
87. Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol*. 2015;16:278.
88. Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*. 2018;19:477.
89. Maeda Y. A New Method for Inducing Synchronous Growth of Dictyostelium discoideum Cells Using Temperature Shifts. *J Gen Microbiol*. 1986;132:1189–96.
90. Weijer CJ, Duschl G, David CN. A revision of the Dictyostelium discoideum cell cycle. *J Cell Sci*. 1984;70:111–31.
91. Wang SY, Pollina EA, Wang I-H, Pino LK, Bushnell HL, Takashima K, Fritsch C, Sabin G, Garcia BA, Greer PL, Greer EL. Role of epigenetics in unicellular to multicellular transition in Dictyostelium. Datasets. Gene Expression Omnibus. 2021. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE137604>.
92. Wang SY, Pollina EA, Wang I-H, Pino LK, Bushnell HL, Takashima K, Fritsch C, Sabin G, Garcia BA, Greer PL, Greer EL. Role of epigenetics in unicellular to multicellular transition in Dictyostelium. Github. 2021. <https://github.com/BCHGreerlab/Dictyostelium>.
93. Wang SY, Pollina EA, Wang I-H, Pino LK, Bushnell HL, Takashima K, Fritsch C, Sabin G, Garcia BA, Greer PL, Greer EL. Role of epigenetics in unicellular to multicellular transition in Dictyostelium. Datasets. ProteomeXchange. 2021. <http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PX024225>.

Publisher's Note

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

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

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

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

