# A bioinformatics screen reveals hox and chromatin remodeling factors at the *Drosophila* histone locus

Lauren J. Hodkinson[1†], Connor Smith[2†], H. Skye Comstra[2], Bukola A. Ajani[2], Eric H. Albanese[2], Kawsar Arsalan[2], Alvaro Perez Daisson[2], Katherine B. Forrest[2], Elijah H. Fox[2], Matthew R. Guerette[2], Samia Khan[2], Madeleine P. Koenig[2], Shivani Lam[2], Ava S. Lewandowski[2], Lauren J. Mahoney[2], Nasserallah Manai[2], JonCarlo Miglay[2], Blake A. Miller[2], Olivia Milloway[2], Nhi Ngo[2], Vu D. Ngo[2], Nicole F. Oey[2], Tanya A. Punjani[2], HaoMin SiMa[2], Hollis Zeng[2], Casey A. Schmidt[2*†] and Leila E. Rieder[2*†]

## Abstract

**Background**  Cells orchestrate histone biogenesis with strict temporal and quantitative control. To efficiently regulate histone biogenesis, the repetitive *Drosophila melanogaster* replication-dependent histone genes are arrayed and clustered at a single locus. Regulatory factors concentrate in a nuclear body known as the histone locus body (HLB), which forms around the locus. Historically, HLB factors are largely discovered by chance, and few are known to interact directly with DNA. It is therefore unclear how the histone genes are specifically targeted for unique and coordinated regulation.

**Results**  To expand the list of known HLB factors, we performed a candidate-based screen by mapping 30 publicly available ChIP datasets of 27 unique factors to the *Drosophila* histone gene array. We identified novel transcription factor candidates, including the *Drosophila* Hox proteins Ultrabithorax (Ubx), Abdominal-A (Abd-A), and Abdominal-B (Abd-B), suggesting a new pathway for these factors in influencing body plan morphogenesis. Additionally, we identified six other factors that target the histone gene array: JIL-1, hormone-like receptor 78 (Hr78), the long isoform of female sterile homeotic (1) (fs(1)h) as well as the general transcription factors TBP associated factor 1 (TAF-1), Transcription Factor IIB (TFIIB), and Transcription Factor IIF (TFIIF).

**Conclusions**  Our foundational screen provides several candidates for future studies into factors that may influence histone biogenesis. Further, our study emphasizes the powerful reservoir of publicly available datasets, which can be mined as a primary screening technique.

†Lauren J. Hodkinson and Connor Smith contributed equally; Casey A. Schmidt and Leila E. Rieder contributed equally

*Correspondence:
Casey A. Schmidt
casey.schmidt@emory.edu
Leila E. Rieder
leila.rieder@emory.edu

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

## Introduction

Cells rely on strict temporal and quantitative orchestration of gene expression. One way the nucleus accomplishes coordinated gene regulation is through the establishment of nuclear bodies (NBs), membraneless concentrations of proteins and RNAs. The NB microenvironment facilitates processes such as efficient gene expression through transcription and RNA processing [1–3].

The histone locus body (HLB) is a conserved NB that regulates histone gene expression and forms at the loci of the replication-dependent histone genes [4] in many different organisms, including humans and *Drosophila*. The HLB is characterized by a set of factors that collectively regulate the uniquely organized histone genes. The *Drosophila melanogaster* histone locus is a cluster of ~100 tandemly repeated arrays, in which each 5 Kb array includes the 5 canonical histone genes along with their respective promoters and regulatory elements [4–6]. Each array contains two TATA-box containing promoters, one for *H3* and *H4* and one for *H2A* and *H2B* (Fig. 1A). Additionally, the *H1* gene has its own unique promoter that lacks a TATA-box. The promoters contain some known transcription factor motifs [7–9], but

overall little is known about how the locus is transcriptionally controlled. The clustered, repetitive organization of the locus allows for precise HLB formation at a single genomic location and highly coordinated histone biogenesis linked to S-phase of the cell cycle [10, 11].

The *Drosophila* HLB is a well-characterized NB that includes several known components that play a role in both the cell cycle regulation of histone gene transcription and the unique processing of histone mRNA transcripts. Several proteins are involved in the initiation and regulation of histone gene transcription including Chromatin Linked Adaptor for MSL proteins (CLAMP [9, 12]); Multi Sex combs (Mxc [11, 13]), the *Drosophila* ortholog of human Nuclear Protein mapped to the Ataxia-Telangiectasia locus (NPAT [14]); FLICE-associated huge protein (FLASH [2]); and Muscle wasted (Mute [15]). Histone mRNA processing is distinct from that of other mRNAs because histone pre-mRNAs lack polyA tails and introns [4]. Several known factors are involved in histone mRNA processing and target the histone gene locus, including the U7 snRNP [16], Stem Loop Binding Protein (SLBP [17]), and Lsm11 [4].

Other than CLAMP, the above-mentioned factors target the histone locus but do not interact directly with
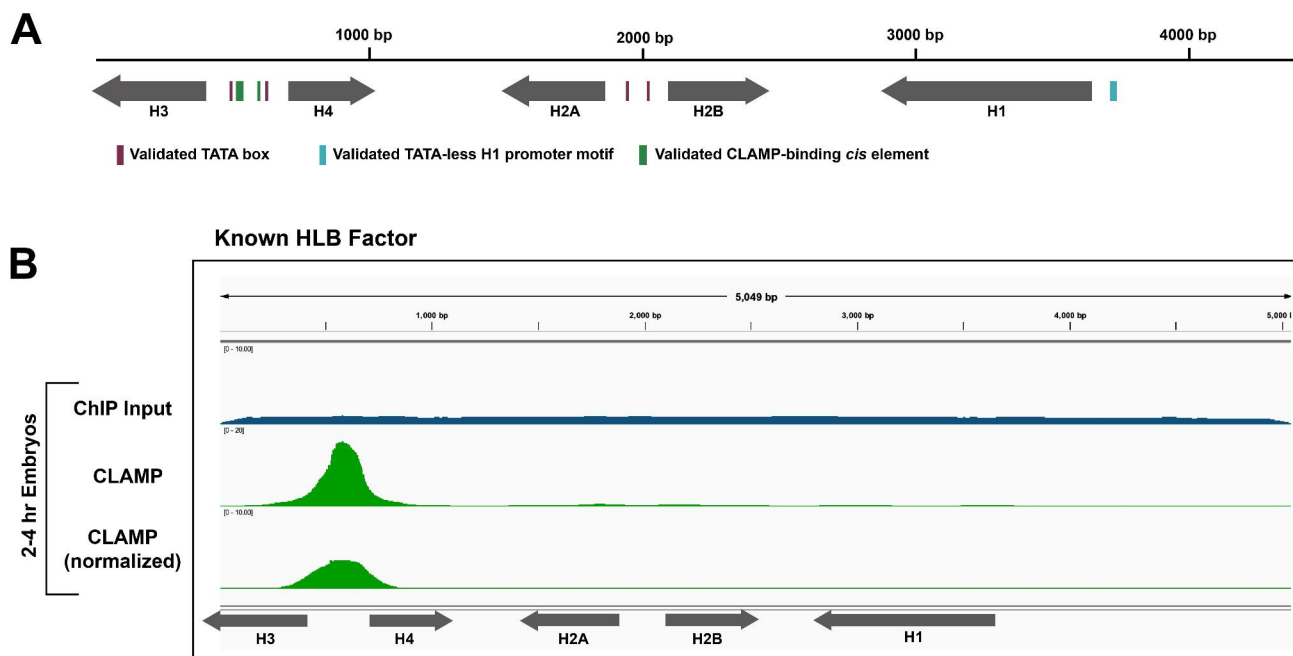


**Fig. 1** Known HLB factor CLAMP localizes to the GA-repeat *cis* elements in the *H3/H4* promoter. **(A)** A diagram detailing the validated *cis* elements in the histone array including the TATA-box elements (maroon boxes), the TATA-less motif (teal box), and the CLAMP binding GA-repeat elements (green boxes). **(B)** We mapped ChIP-seq data for the known HLB factor CLAMP (green) from 2–4 h embryos [12]. The ChIP signal was normalized to its respective ChIP input signal (blue)

DNA sequence. Since CLAMP is found at locations genome-wide, it is currently unclear how non-DNA binding factors identify and target the histone locus. The presence of histone mRNA is likely to play a role [18] as is the presence of *cis* elements within the histone gene array [9, 19]. One critical interaction involves CLAMP recognizing GA-repeat sequences within the *H3/H4* promoter [9] (Fig. 1). Although the presence of CLAMP is critical for the localization of HLB-specific factors such as Mxc [9], the interaction between CLAMP and GA-repeats is not strictly necessary for HLB formation [20] and CLAMP is not sufficient for HLB formation [9]. Therefore, it is likely that other DNA-interacting proteins participate in defining the histone locus. We still lack a comprehensive list of factors associated with histone biogenesis and therefore our model of the mechanisms of histone gene regulation remains incomplete.

Historically, novel HLB factors are often discovered by chance through immunofluorescence, such as CLAMP [9], Myc [21], Mute [15], and Abnormal oocyte [22]. To discover novel DNA-binding proteins that target the histone locus, we first screened the literature for likely candidates and then funneled these into a secondary bioinformatics screen. We leveraged publicly available *Drosophila* ChIP-seq datasets and knowledge of histone gene regulation to curate and analyze a list of candidate DNA-binding factors. We used a bioinformatics pipeline on Galaxy [23, 24] to map candidate ChIP-seq data to a single copy of the histone gene array. The ~100 histone gene arrays are nearly identical in sequence [6] and we can collapse -omics data from the entire locus onto a single array [5, 9, 20]. Supervised undergraduate students conducted much of the initial screen as part of a course-based undergraduate research experience (CURE; [25]), demonstrating the simplicity and versatility of the pipeline design. Using our qualitative analysis criteria (Supplemental Fig. 1), we discovered several DNA-interacting proteins that pass our initial bioinformatics screen. Our novel candidates that target the histone gene array include developmental transcription factors such as Hox factors, which may provide a mechanistic link between segment identity and cell division.

Future wet lab studies are required to confirm the presence of these candidates at the histone locus, determine any tissue and temporal specificity, and describe the precise roles of candidates in HLB formation and histone biogenesis. As a whole, our screen establishes mining of existing -omics data as a tool to identify new candidate HLB factors. Although we are limited by the factors, tissues, treatments, and timepoints interrogated by the dataset generators, our pipeline is an inexpensive and rapid tool to screen candidate factors for future wet-lab studies.

## Methods

### GEO datasets

All datasets were downloaded from the NCBI SRA Run Selector through the Gene Expression Omnibus (GEO). See Table 1 for Accession numbers and references.

### Bioinformatic analysis and data visualization

We directly imported individual FASTQ datasets into the web-based platform Galaxy [23, 24] through the NCBI SRA Run Selector by selecting the desired runs and utilizing the computing Galaxy download feature. We retrieved the FASTQ files from SRA using the "Faster Download and Extract Reads in FASTQ format from NCBI SRA" Galaxy command. Because the ~100 histone gene arrays are extremely similar in sequence [6], we do not utilize the dm6 or dm3 genomes and instead can collapse ChIP-seq data onto a single histone array [5, 6, 20]. We used a custom "genome" that includes a single *Drosophila melanogaster* histone array similar to that in Mckay et al. 2015, which we directly uploaded to Galaxy using the "upload data" feature, and normalized using the Galaxy command "NormalizeFasta" specifying an 80 bp line length for the output .fasta file. We aligned ChIP reads to the normalized histone gene array using Bowtie2 [26] to create .bam files using the user built-in index and "very sensitive end-to-end" parameter settings. We converted the .bam files to .bigwig files using the "bamCoverage" Galaxy command in which we set the bin size to 1 bp and set the effective genome size to user specified: 5000 bp (approximate size of 1 histone array). We also mapped relevant input or IgG datasets. If an input dataset was available, we normalized ChIP datasets to input using the "bamCompare" Galaxy command in which we set the bin size to 1 bp. We visualized the .bigwig files using the Integrative Genome Viewer (IGV) [27].

### Criteria for positive vs. negative candidates

Because we focused our analysis on a single 5 Kb sequence and condensed data from ~100 identical histone arrays onto a single array, we were unable to use quantitative peak calling programs. We instead utilized the following qualitative criteria to determine positive and negative candidates (Supplemental Fig. 1). We only considered the candidate as positive if a peak emerged in the ChIP data that was not present in the input. We considered the following false positives: (1) obvious overrepresentation of gene bodies (e.g. Su(z)12, Supplemental Fig. 2), (2) underrepresentation of intergenic regions (e.g. CP190 input, Fig. 3C), and (3) if the input coverage and ChIP coverage peaks looked identical (e.g. MSL1, Fig. 3B). Datasets with the above-mentioned characteristics cause peaks to emerge in the normalized data that do not represent the binding of the factor but rather a bias in the amplification of the ChIP library or alignment. We

**Table 1** DNA-binding factor candidate datasets

| Candidate | GEO Accession # | SRA Run Selector # | Paper citation |
|---|---|---|---|
| Abd-A<br>Abdominal-A | GSE69796 | **anti-GFP ChIP DNA from Kc167 cells expressing AbdA-GFP**<br>**1**- SRR2060648 **2 -**SRR2060649<br>**Input**<br>**1 -** SRR2060652 **2 -** SRR2060653 | [63] |
| Abd-B<br>Abdominal-B | GSE69796 | **anti-GFP ChIP DNA from Kc167 cells expressing AbdB-GFP**<br>**1**- SRR2060650 **2 -**SRR2060651<br>**Input**<br>**1 -** SRR2060652 **2 -** SRR2060653 | [63] |
| ANTP<br>Antennapedia | GSE125604 | **anti-GFP (Invitrogen) from ANTP-GFP genotype**<br>**1 -**<u>SRR8483063</u><br>**Input**<br>**1 -** SRR8483064 | [61] |
| CP190<br>Centrosomal protein 190kD | GSE118699 | **CP190 rabbit (Pai et al. 2004)**<br>**1 -** SRR7706256 **2 -** SRR7706258<br>**Input**<br>**1 -** SRR7706251 **2 -** SRR7706252 | [40] |
| CTCF | GSE175402 | **CTCF**<br>**1** - SRR14631231 2 - SRR14631232<br>**Input**<br>**1** - SRR14631233 2 - SRR14631234 | [42] |
| Exd<br>Extradenticle | GSE125604 | **anti-V5 (Invitrogen) on exd-V5 transgene genotype**<br>**1 -**<u>SRR848305</u>5<br>**Input**<br>**1 -** SRR8483056 | [61] |
| Fs(1)h<br>Female sterile (1) homeotic | GSE42086 | **Female late embryo-derived cell line, ChIP of Fs(1)h long isoform**<br>**1-** SRR611533<br>**Female late embryo-derived cell line, ChIP of both isoforms of Fs(1)h**<br>**1 -** SRR611535<br>**Input**<br>**1 -** SRR611537 | [60] |
| Gcn5 | GSE83408 | **Gcn5 rabbit polyclonal antibody (5 ug/IP)**<br>**1 -** SRR3671294 **2 -** SRR3671295 **3 -** SRR3671298<br>**Input**<br>**1 -** SRR3671296 **2 -** SRR3671297 **3 -** SRR3671299 | [41] |
| Hr78<br>Hormone-receptor-like 78 | GSE50370 | **Hr78-GFP_8–16_embryonic_ChIP-seq_ChIP**<br>**1 -** SRR1198798 **2 -** SRR1198799<br>**Input**<br>**1 -** SRR1198796 **2 -** SRR1198797 | [73] |
| Hnf4<br>Hepatocyte nuclear factor 4 | GSE73675 | **rat anti-dHNF4 3600**<br>**1 -** SRR2548371 **2 -** SRR2548372<br>**3 -** SRR2548373 **4 -** SRR2548374<br>**Inputs**<br>**1 -** SRR2548367 **2 -** SRR2548368<br>**3 -** SRR2548369 **4 -** SRR2548370 | [53] |
| HTH<br>Homothorax | GSE125604 | **anti-Hth (gp52, N-terminal)**<br>**1 -** SRR8483065<br>**Input**<br>**1 -** SRR8483066 | [61] |
| JIL-1 | GSE54438 | **JIL-1 monoclonal antibody 5C9**<br>**1 -** SRR1145605 **2 -** SRR1145606<br>**Input**<br>**1 -** SRR1145612 **2 -** SRR1145613 | [57] |
| M1BP<br>Motif 1 Binding Protein | GSE97841 | **M1BP_Antibody**<br>**1 -** SRR10759878<br>**Input**<br>**1 -** SRR10759877 | [30] |
| MSL-1<br>Male-specific Lethal 1 | GSE37864 | **polyclonal rabbit MSL1, crude serum**<br>**1 -** SRR495366 **2 -** SRR495367<br>**Input**<br>**1 -** SRR495378 **2 -** SRR495380 | [39] |

Hodkinson *et al. BMC Genomic Data*        (2023) 24:54

Page 5 of 18

**Table 1** (continued)

| Candidate | GEO Accession # | SRA Run Selector # | Paper citation |
|---|---|---|---|
| Ndf/CG4747 <br> Nucleosome-destabilizing factor | GSE42025 | **PAP antibody (Sigma P1291)** <br> **1 -** SRR611192 **2 -** SRR611194 <br> **3 -** SRR611196 **4 -** SRR611198 <br> **Input** <br> **1 -** SRR611193 **2 -** SRR611195 <br> **3 -** SRR611197 **4 -** SRR611199 | [29] |
| Nej (S2 cells) <br> Nejire | GSE72666 | **anti-CBP, custom-made antibodies** <br> **1**- SRR2232434 <br> **Input** <br> **1 -** SRR2232432 | [74] |
| Nej (Embryos) <br> Nejire | GSE68983 | **Nej** <br> **1 -** SRR4044401 <br> **Input** <br> **1 -** SRR2031906 | [75] |
| Opa <br> Odd Paired | GSE140722 | **In-house anti-Opa antibody** <br> **1 -** SRR10502454 **2 -** SRR10502455 <br> **3 -** SRR10502458 **4 -** SRR10502459 <br> **Input** <br> **1 -** SRR10502456 **2 -** SRR10502457 <br> **3 -** SRR10502460 **4 -** SRR10502461 | [51] |
| Pan <br> Pangolin | GSE50340 | **Pan** <br> **1 -** SRR1198824 **2 -** SRR1198825 <br> **Input** <br> **1 -** SRR1198822 **2 -** SRR1198823 | [73] |
| Pnt <br> Pointed | GSE114092 | **Pnt** <br> **1 -** SRR7126165 <br> **Input** <br> **1 -** SRR7126164 | [55] |
| Psc <br> Posterior sex combs | GSE38166 | **Psc Mitotic S2** <br> **1** - SRR 500,149 **2** - SRR 500,150 <br> **Psc Control S2** <br> **1** - SRR500151 **2** - SRR500152 <br> **Psc Mitotic S2 Input** <br> **1** - SRR 500,153 **2** - SRR 500,154 <br> **Psc Control S2 Input** <br> **1** - SRR 500,155 **2** - SRR 500,156 | [43] |
| Scm (S2 cells) <br> Sex comb on midleg | GSE66183 | **BioTAP-N-Scm** <br> **1 -** SRR1813243 **2 -** SRR1813245 <br> **Input** <br> **1 -** SRR1813244 **2 -** SRR1813246 | [76] |
| Scm (embryos) <br> Sex comb on midleg | GSE66183 | **BioTAP-N-Scm** <br> **1 –** SRR1813233 <br> **Input** <br> **1 -** SRR1813234 | [76] |
| su(z)12 <br> suppressor of zeste 12 | GSE36039 | **Su(z)12 ChIP** <br> **1 -** SRR363407 **2 -** SRR363408 <br> **Input** <br> **1 -** SRR363409 **2 -** SRR363410 | [44] |
| TAF1 <br> TBP-Associated Factor 1 | GSE97841 | **TAF1 Antibody** <br> 1 - SRR5452843 2 - SRR5452844 <br> **Inputs** <br> 1 - SRR5452847 2 - SRR5452848 | [30] |
| TFIIB <br> Transcription Factor II B | GSE120152 | **anti-TFIIB rabbit polyclonal, custom** <br> **1 -** SRR7874066 **2 -** SRR7874067 <br> **Inputs** <br> **1 -** SRR7874069 **2 -** SR7874070 | [31] |
| TFIIF <br> Transcription Factor II F | GSE120152 | **anti-TFIIF rabbit polyclonal, custom** <br> **1 -** SRR7874068 <br> **Inputs** <br> **1 -** SRR7874069 | [31] |

Hodkinson *et al. BMC Genomic Data*          (2023) 24:54

Page 6 of 18

**Table 1**  (continued)

| Candidate | GEO Accession # | SRA Run Selector # | Paper citation |
|---|---|---|---|
| TRF2<br>TBP protein-related factor 2 | GSE97841 | **TRF2 Antibody**<br>**1 -** SRR5452845 **2 -** SRR5452846<br>**Inputs**<br>**1 -** SRR5452847 **2 -** SRR5452848 | [30] |
| Ubx (Kc cells)<br>Ultrabithorax | GSE69796 | **anti-GFP ChIP DNA from Kc167 cells expressing Ubx-GFP**<br>1 - SRR2060646 2 - SRR2060647<br>**Inputs**:<br>**1 -** SRR2060652 **2 -** SRR2060653 | [63] |
| Ubx (embryos)<br>Ultrabithorax | GSE64284 | **Anti-V5 ChIP, Ubx-V5**<br>**1 -** SRR1721317 **2 -** SRR1721321<br>**Inputs**<br>**1 -** SRR1721316 **2 -** SRR1721320 | [77] |
| Ubx (larva)<br>Ultrabithorax | GSE184454 | **Anti-FLAG monoconal, 3xFLAG-Ubx**<br>**1 -** SRR15972582 **2 -** SRR15972584<br>**Inputs**<br>**1-** SRR15972583 **2 -** SRR15972585 | [78] |

also checked spot length (read length) and considered peaks over the GA-repeat *cis* elements in the *H3/H4* promoter (Fig. 1A) found in datasets with read lengths $\leq$ 50 bp false positive peaks (e.g. Psc, Supplemental Fig. 2).

## Results
### Validating the bioinformatics pipeline by mapping TATA-associated factors to the histone gene array
We first sought to validate our bioinformatics pipeline through analysis of known histone locus proteins and associated factors. Isogai et al. (2007) used immuno-fluorescence and cell culture ChIP-qPCR assays to demonstrate that the TATA binding protein (TBP)/TFIID complex selectively binds to the *H3/H4* promoter and the *H2A/H2B* promoter, but TBP-related factor 2 (TRF2) targets the promoter of the TATA-less *H1* promoter. We identified a publicly available TRF2 ChIP-exo dataset from Baumann et al. (2017) for TRF2 and used our pipeline to map the data to the histone gene array. ChIP-exo is similar to ChIP-seq but identifies a more complete set of binding locations for a factor with higher resolution than standard ChIP-seq [28]. We validated that TRF2 localizes to the *H1* promoter (Fig. 2A). Because we were unable to normalize to an input dataset, we compared the TRF2 alignment to an IgG control. The localization of TRF2 to the TATA-less *H1* promoter is consistent with Isogai et al. (2007) and is consistent with where a TBP-related factor (TRF) would be expected to bind as they are known to target TATA-less promoters [29]. Baumann et al. (2017) demonstrated that Motif 1 binding protein (M1BP) interacts with TRF2, but that this interaction is mostly restricted to the ribosomal protein genes [30]. We mapped ChIP-exo data for M1BP and observed that it did not localize to the *H1* promoter under our qualitative criteria (Supplemental Fig. 1) as we saw with TRF2, nor to any other part of the histone array (Fig. 2A), further validating our pipeline.

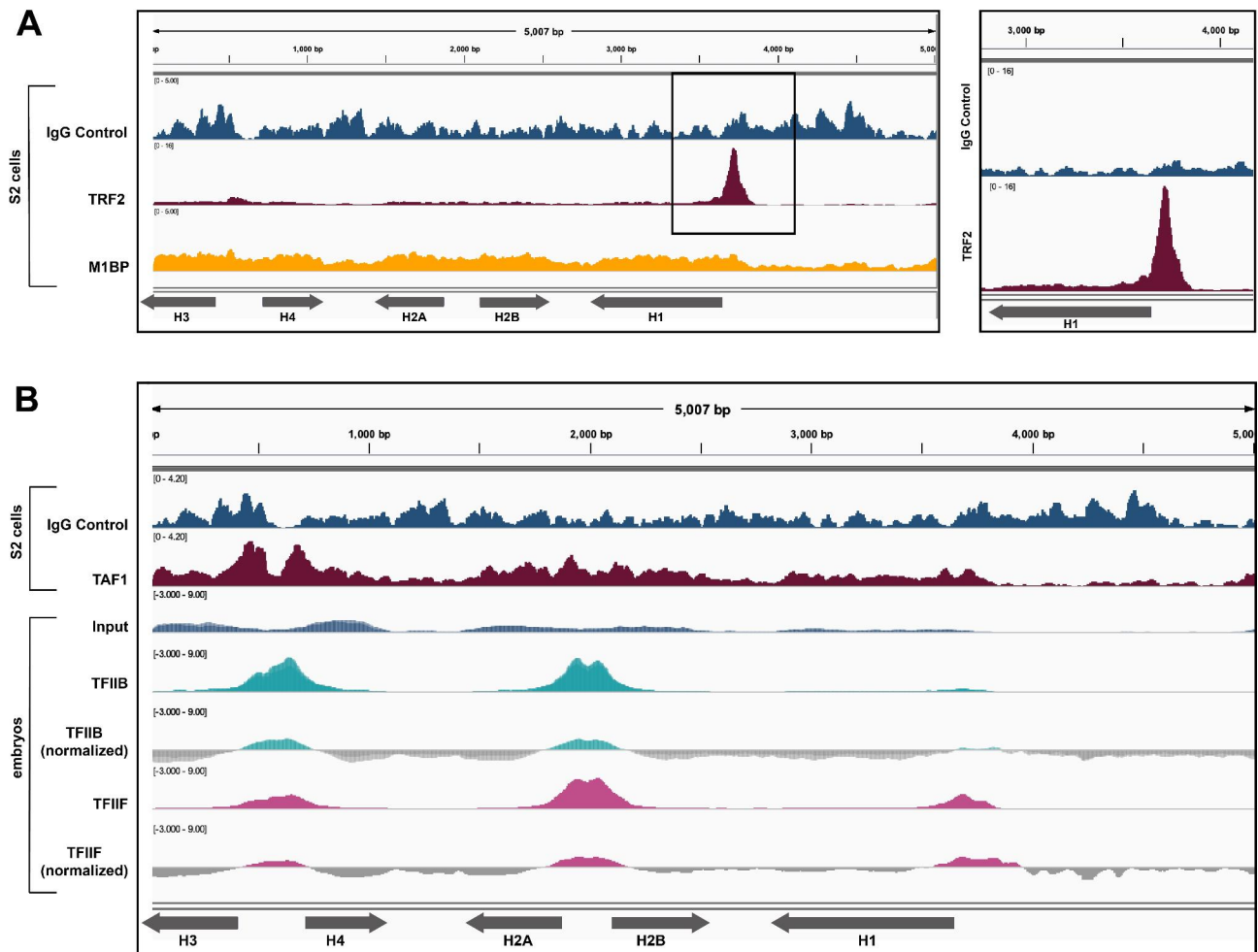### Novel general transcription factors that target the histone locus
To expand the list of general transcription factors that target the histone locus, we mapped an additional ChIP-exo dataset from Baumann et al. (2017) for TAF1 (TBP associated factor 1). TAF1 is a member of the Transcription Factor IID (TFIID) complex which Isogai et al. (2007) also suggested localized to the same regions of the histone gene array as TBP. When we mapped the TAF1 ChIP-exo data, we observed that TAF1 targets the TATA-box regions of the *H3/H4* promoter and, less robustly, the TATA-box regions of the *H2A/H2B* promoter (Fig. 2B, elements annotated in Fig. 1A). Again, we compared this alignment to an IgG control because we were unable to normalize to an input, but because TAF1 associates with TBP which binds to AT-rich (TATA box) regions [30], the localization of TAF1 to the TATA-box regions of the core histone genes is expected.

To test the ability of our pipeline to identify novel factors that localize to the histone gene array, we investigated the relationships of additional general transcription factors to the histone array. We identified ChIP-seq datasets for both TFIIB and TFIIF. Both TFIIB and TFIIF are associated with TBP [31] and therefore we would expect them to localize to the *H3/H4* and *H2A/H2B* promoters, similar to TBP [8]. We observed both TFIIB and TFIIF localization to the *H3/H4* and *H2A/H2B* promoters while, surprisingly, TFIIF also localized to the *H1* promoter (Fig. 2B).

### Candidate DNA-binding factors that did not pass the bioinformatics screen
After verifying our bioinformatics pipeline, we curated a list of candidate DNA-binding factors (Table 1, Supplemental Table 1) that we hypothesized would target the histone gene array. To create this candidate list, we prioritized factors that meet at least one of the following

**Fig. 2** Expected general transcription factors localize to the histone array. **(A)** We mapped ChIP-exo data for TRF2 (maroon, [30]) from S2 cells to the histone gene array, which recapitulates results from Isogai et al. 2007 showing localization specifically to the *H1* promoter, validating our bioinformatics pipeline. We also mapped ChIP-exo data for M1BP (yellow, [30]) which did not localize to the histone gene array, further validating our pipeline. We compared ChIP-exo data to an IgG control (blue, [30]. **(B)** We aligned ChIP-exo data for TAF-1 (maroon, [30]) from S2 cells to the histone gene array and compared to a corresponding IgG control. We aligned ChIP-seq datasets for TFIIB (teal, two replicates overlayed, [31]) and TFIIF (pink, one replicate, [31]) from OregonR mixed population embryos to the histone gene array and normalized to the provided input (blue). TFIIB shows localization to the *H3/H4* promoter and the *H2A/H2B* promoter, and TFIIF shows localization to both core promoters and the *H1* promoter, confirming that our bioinformatics pipeline can be used to identify novel factors that localize to the histone gene array

criteria: (1) DNA-binding factors with a relationship to a validated HLB factor; (2) DNA-binding factors involved in dosage compensation, because CLAMP, a non-sex specific dosage compensation factor, targets the histone locus [9, 20] (Fig. 1B); (3) chromatin remodeling or histone-interacting factors, since the epigenetic landscape of the histone locus is largely undefined; (4) early developmental transcription factors, since histone gene regulation is critical during early development and synchronized cell division [32]. We also utilized the online platform STRING [33] that provides the known and inferred interactomes of a given protein to identify candidates that met the above criteria. Out of the 27 candidates, we rejected 19 as likely not targeting the histone

gene array based on our qualitative analysis of the datasets we investigated (Supplemental Fig. 1).

### HLB factor-associated candidates

We investigated the DNA-binding factor Sex comb on midleg (Scm), because of its suspected interaction with the known HLB factor Multi-sex combs (Mxc; [11, 13]). Based on STRING, Scm is predicted to interact with Mxc, as determined by a genetic interference assay in which a double Mxc/Scm mutant resulted in enhanced mutant sex comb phenotypes [34, 35]. Despite possible interaction with Mxc, neither Scm ChIP-seq data from S2 cells (Fig. 3A) nor from 12 to 24 h embryos (Supplemental Fig. 2) gave meaningful signal over the histone gene array. This result was surprising because the human
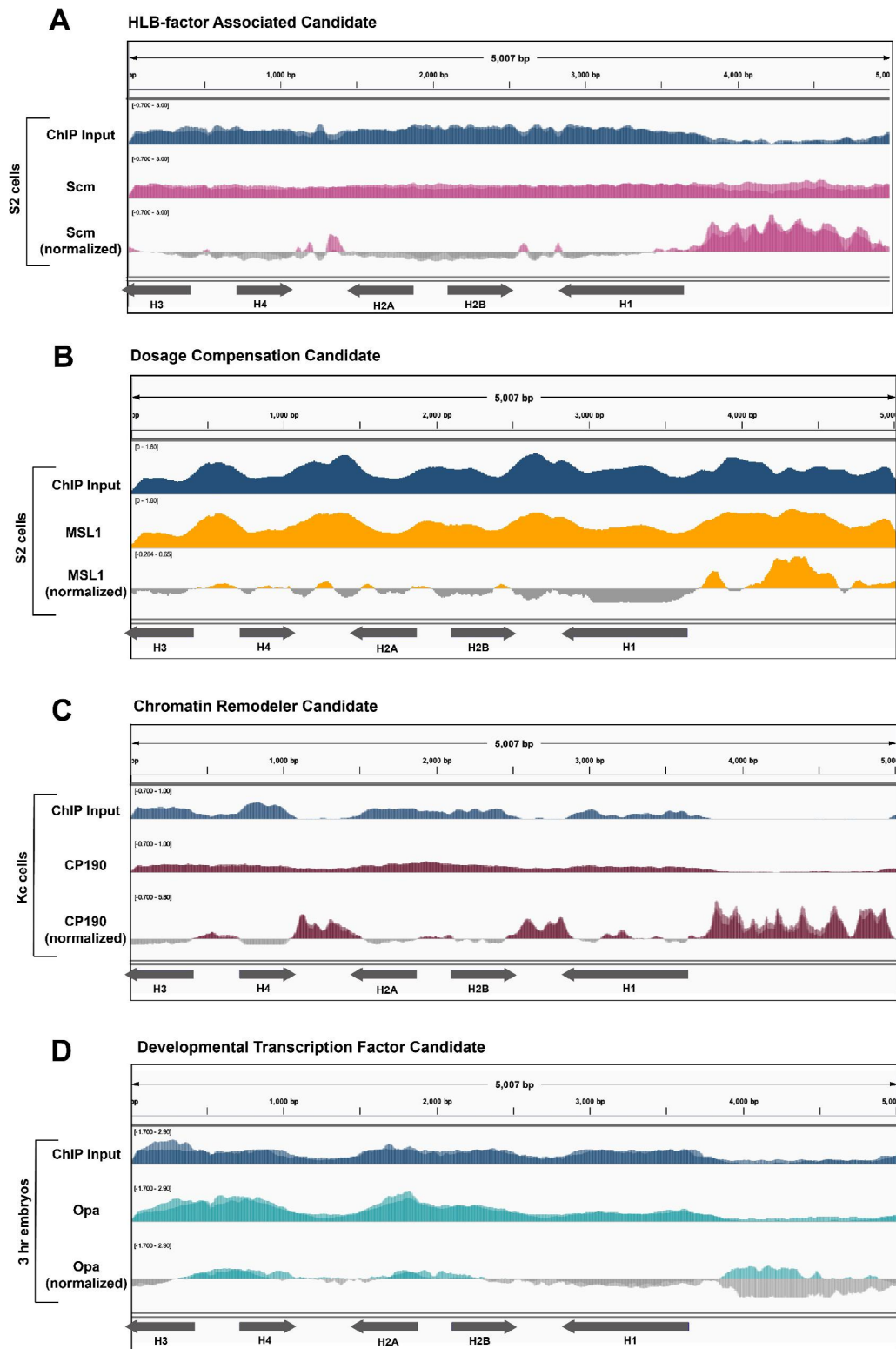
**Fig. 3** DNA-binding factors from different categories that did not pass the bioinformatics screen. We aligned ChIP-seq datasets for **(A)** Scm (pink, two replicates overlayed, [76]) from S2 cells, **(B)** MSL1 (yellow, one replicate, [39]) from S2 cells, **(C)** CP190 (maroon, two replicates overlayed, [40]) from Kc cells, and **(D)** Opa (teal, two replicates overlayed, [51]) from 3 h mixed population embryos to the histone array. We normalized each ChIP signal to its respective ChIP input signal (blue)

Hodkinson *et al. BMC Genomic Data*        (2023) 24:54

Page 9 of 18

ortholog of Mxc (NPAT) associates exclusively with the histone promoters [36], and Mxc is only found at the histone locus [14].

### Dosage compensation candidates

The HLB factor CLAMP targets the *H3/H4* promoter (Fig. 1B) and regulates histone gene expression [9], but also plays additional roles in *Drosophila* male dosage compensation: it binds to GA-rich elements along the male X-chromosome and recruits the Male Specific Lethal complex (MSLc). Further, MSL2, the male specific component of MSLc, also emerged from a cell-based HLB factor screen [11], and we recently discovered that MSL2 targets one histone gene locus in *Drosophila virilis* [37]. We therefore hypothesized that dosage compensation factors target the histone gene array along with CLAMP. We chose the following DNA-binding factors for our candidate screen because of their relationship to dosage compensation: MSL1, a protein that scaffolds MSLc [38, 39], and nucleosome destabilizing factor (Ndf, CG4747), a putative H3K36me3-binding protein that is important for MSLc localization [29]. When we mapped ChIP-seq datasets from these factors, we found that neither gave meaningful signal over the histone gene array (MSL1: Fig. 3B, Ndf/CG4747: Supplemental Fig. 2). This result is not surprising as we previously determined that MSL2 does not target the histone locus in *Drosophila melanogaster* by polytene chromosome immunofluorescence [37].

### Chromatin remodeling candidates

One of the lesser-studied characteristics of the histone locus is the regional chromatin environment. The endogenous histone locus is located on chromosome 2 L, proximal to pericentric heterochromatin. Despite this proximity, histone expression rapidly increases at the start of G1 in preparation for DNA synthesis during S phase, and quickly ceases upon G2 [4], indicating that chromatin remodeling is likely critical in precisely controlling histone gene expression. We therefore hypothesized that chromatin remodeling factors target the histone locus. We chose the following candidates because of their association with chromatin or role in chromatin remodeling: centrosomal 190 kDa protein (CP190), an insulator protein that impacts enhancer-protein interactions and stops the spread of heterochromatin [40]; Gcn5, a lysine acetyltransferase critical for oogenesis and morphogenesis [41]; CCCTC-binding factor (CTCF), a genome architectural protein [42]; Posterior sex combs (Psc), a polycomb-group gene [43]; and Suppressor 12 of zeste 12 (su(z)12), a subunit of polycomb repressive complex 2 [44].

After identifying relevant ChIP-seq datasets (Table 1), we used our analysis pipeline to map data to the histone gene array. We observed that none of the above chromatin remodeling candidates gave meaningful signal over the histone gene array (CP190: Fig. 3C, all others: Supplemental Fig. 2). We were especially surprised that CP190 did not target the histone array. CP190 binds promoter regions, aids enhancer-promoter interactions, and halts the spreading of heterochromatin. Because the histone locus is proximal to pericentric heterochromatin, we hypothesized the presence of CP190 could explain how centromeric heterochromatin does not expand into the histone locus. In addition, CP190 is a member of the Late Boundary Complex (LBC) [45], which also contains the CLAMP protein [46]. We discovered that the LBC binds to the *H3/H4* promoter region in vitro [37]. We were therefore surprised that CP190 does not appear to target the histone gene array, based on the ChIP-seq datasets we analyzed. These data underscore the requirement for visualizing both ChIP and input datasets, rather than just the final normalized trace: although CP190 ChIP-seq does not show enrichment over the histone gene array, bias in the input dataset leads to misleading peaks in the normalized data (Fig. 3C, Supplemental Fig. 2).

### Developmental transcription factor candidates

Zygotic histone biogenesis is critical for the constantly dividing embryo; increased histone expression can lengthen the cell cycle whereas decreased histone levels can shorten the cell cycle [32, 47]. Histone biogenesis is tightly coupled to DNA replication, and excess histones are buffered so as not to interfere with zygotic chromatin [48–50]. We therefore hypothesized that early embryonic transcription factors target the histone locus. We chose the following DNA-binding factors based on their roles in the early embryo: Odd paired (Opa), a pair ruled gene that contributes to morphogenesis [51]; Motif 1 binding protein (M1BP), a transcriptional pausing factor that interacts with the Hox proteins [30, 52]; Hepatocyte nuclear factor 4 (Hnf4), a general developmental transcription factor [53]; Pangolin (Pan), a component of the Wingless signaling pathway [54]; and Pointed (Pnt), a factor that regulates cell proliferation and differentiation during development [55, 56]. When we mapped appropriate ChIP-seq datasets from these factors, none gave meaningful signal over the histone array (Opa: Fig. 3D, M1BP: Fig. 2A, all others: Supplemental Fig. 2).

### Candidates that passed the bioinformatics screen

We found several factors that exhibited distinct, meaningful localization patterns to the histone gene array and therefore warrant further investigation (Fig. 4). First, we used our bioinformatics pipeline to map a ChIP-seq dataset for the kinase JIL-1, which is responsible for phosphorylating serine 10 on histone 3 [57, 58]. We observed JIL-1 localizing to the histone gene array, specifically to
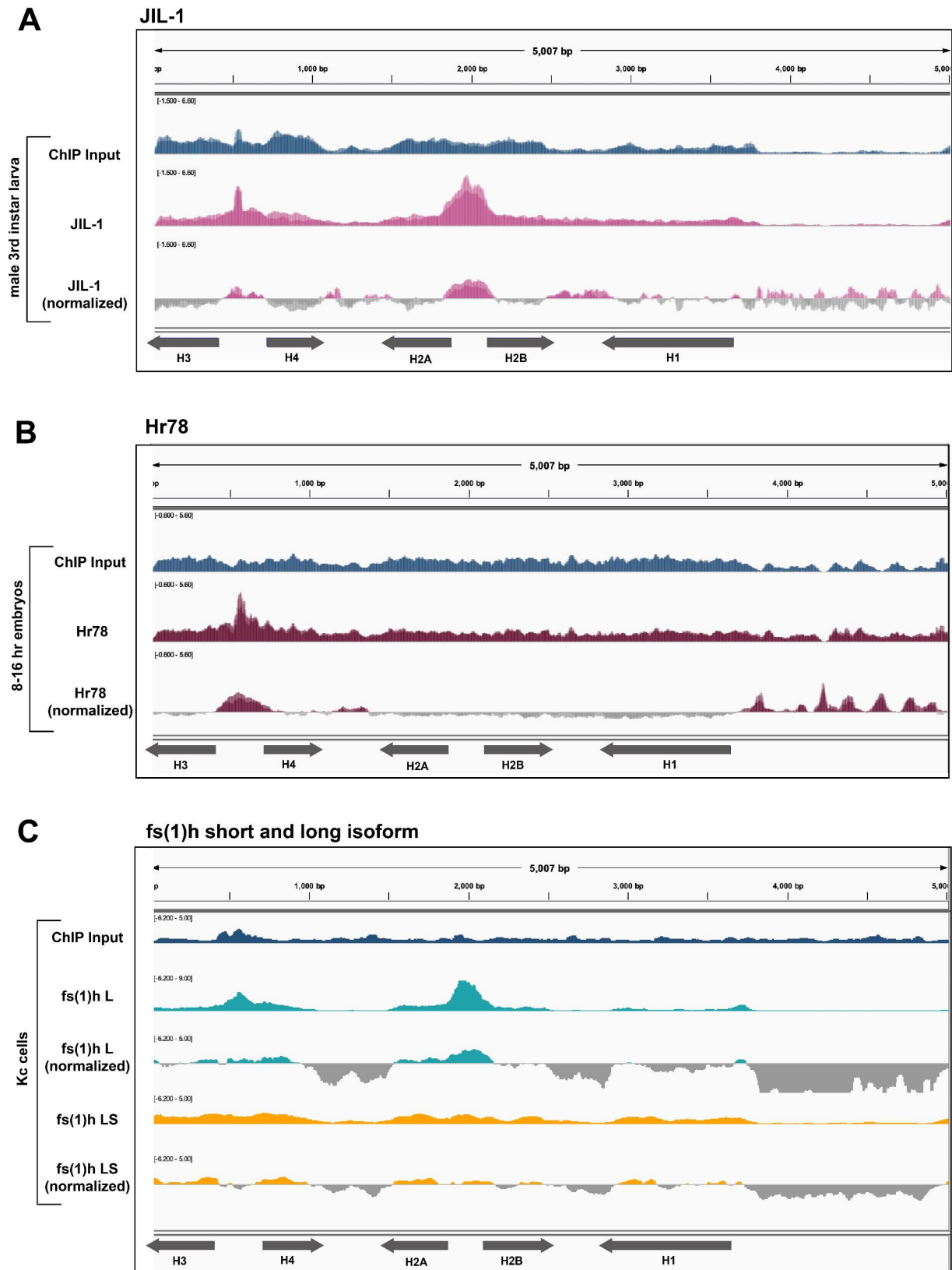
**Fig. 4** JIL-1, Hr78, and Fs(1)hL localize to the histone gene array. We mapped ChIP datasets for **(A)** JIL-1 (pink, two replicates overlayed, [57]) from male third instar larvae, **(B)** Hr78 (maroon, two replicates overlayed, [73]) from 8–16 h mixed population embryos, and **(C)** the long (L, teal) and short (S, yellow) isoforms of fs(1)h from Kc cells [60] to the histone gene array. We normalized each ChIP-seq dataset to its respective input (blue)

the *H2A/H2B* promoter (Fig. 4A). We observed an additional sharp peak at the *H3/H4* promoter, but this peak is likely an artifact of short read lengths from the dataset and overlaps with a perfect, long GA-repeat sequence in the *H3/H4* promoter (Fig. 1A, Supplemental Fig. 1). JIL-1 is a DNA-binding factor that associates with the Maleless helicase and MSL1, two members of MSLc [58]. In addition to CLAMP performing a role in histone biogenesis, it also plays a role in dosage compensation and associates with MSLc [59].

We also observed hormone-like receptor 78 (Hr78) localize to the *H3/H4* promoter (Fig. 4B). Finally, we mapped two isoforms of female sterile (1) homeotic (fs(1)h; the *Drosophila* homolog of BRD4). The long and short isoforms of fs(1)h have distinct binding profiles, but both are assumed to have a role in chromatin architecture [60]. We observed that the long isoform, but not the short isoform, localizes to both the *H2A/H2B* and the *H3/H4* promoters (Fig. 4C). Interestingly, Kellner et al. (2013) inferred that the fs(1)h long isoform has a unique role in chromatin remodeling by interacting with specific insulator proteins, including CP190, which did not pass our screen (Fig. 3C).

### Hox factors localize to the *Drosophila* histone gene array when overexpressed in cell culture

Hox factors (Fig. 5A) are critical for developmental processes like morphogenesis, in which cells are constantly dividing and therefore require a near constant supply of histones [4]. Histone biogenesis is critical within the first few hours of *Drosophila* development [32, 47]. We therefore investigated histone array localization patterns of transcription factors that act during early development, including Hox factors. We identified a publicly available dataset (Table 1) in which Beh et al. (2016) individually expressed the three Bithorax complex Hox proteins, Ultrabithorax (Ubx), Abdominal-A (Abd-A), and Abdominal-B (Abd-B), in Kc167 cells and performed ChIP-seq. We used our analysis pipeline to map the Ubx, Abd-A, and Abd-B ChIP-seq datasets to the histone gene array and observed striking localization to the *H3/H4* promoter (Fig. 5B-C). We conclude that when overexpressed in cultured cells, Ubx, Abd-A, and Abd-B all target the histone gene array by ChIP-seq.

Because our Hox factor observation (Fig. 5) could be an artifact of overexpression in cultured cells, we identified two additional Ubx ChIP-seq datasets from 0 to 16 h embryos and third instar larval imaginal discs (Table 1). We used our pipeline to map these data to the histone gene array and observed that Ubx targets the *H3/H4* promoter and, to a lesser extent, the *H2A/H2B* promoter (Fig. 6). We conclude that Ubx targets the histone gene array at various developmental stages and in various tissues and is therefore a promising candidate for future

wet-lab research designed to validate these bioinformatic observations.

To further investigate the relationship between Hox factors and the histone locus, we identified three additional datasets for Hox proteins and Hox cofactors. There are two Hox gene complexes in *Drosophila*: the Bithorax complex (which includes Ubx, Abd-A, and Abd-B) and the Antennapedia complex. We first mapped ChIP-seq data for Antennapedia (Antp) [61] but did not observe robust localization to the histone gene array (Supplemental Fig. 2). We next mapped ChIP-seq datasets for the Hox cofactors extradenticle (Exd) and Homothorax (Hth) [61]. Exd and Hth associate with the hexapeptide motif in Hox proteins and form heterodimers to impact Hox binding specificity to their gene targets [62, 63]. We observed that neither Exd nor Hth gave meaningful ChIP signal over the histone gene array (Supplemental Fig. 2).

### Power and limitations of the screen

The range of results from our candidate screen demonstrates both the power and limitations of our bioinformatics pipeline. In total, we analyzed datasets for 27 different DNA-binding factors and produced 9 candidates that warrant further wet lab investigation. Despite the power of this screen, we are limited by the availability of public datasets. Characteristics of these datasets, such as quality of reads, read length, and inclusions of controls such as inputs are based on the original experimental design and research. Furthermore, we are also restricted by the tissues or genotypes investigated in the original study, limiting the scope of our investigation.

For example, we analyzed several datasets for Nejire (Nej; homolog of mammalian CREB-binding protein (CBP)) and Pointed (Pnt). A previous screen in S2 cells identified Nej and Pnt as potential HLB factors [11]. We investigated two Nej ChIP-seq datasets (Table 1) in which we obtained disparate results. The Nej ChIP-seq dataset from S2 cells did not yield meaningful signal over the histone gene array (Fig. 7, center). In contrast, we investigated a Nej ChIP-seq dataset from early *Drosophila* embryos and observed robust localization to the *H3/H4* promoter, *H2A/H2B* promoter and, to a lesser extent, the *H1* promoter (Fig. 7, top). From these observations, we conclude that Nej likely targets the histone gene array in embryos and would therefore be a strong candidate for future wet-lab studies to validate this observation. Additionally, we mapped a Pnt ChIP-seq dataset from Stage 11 embryos (Table 1) and observed that Pnt does not give meaningful signal over the histone gene array (Fig. 7, bottom).

Our Pnt and Nej observations demonstrate how our screening approach is powerful but limited by data availability and experimental variables.
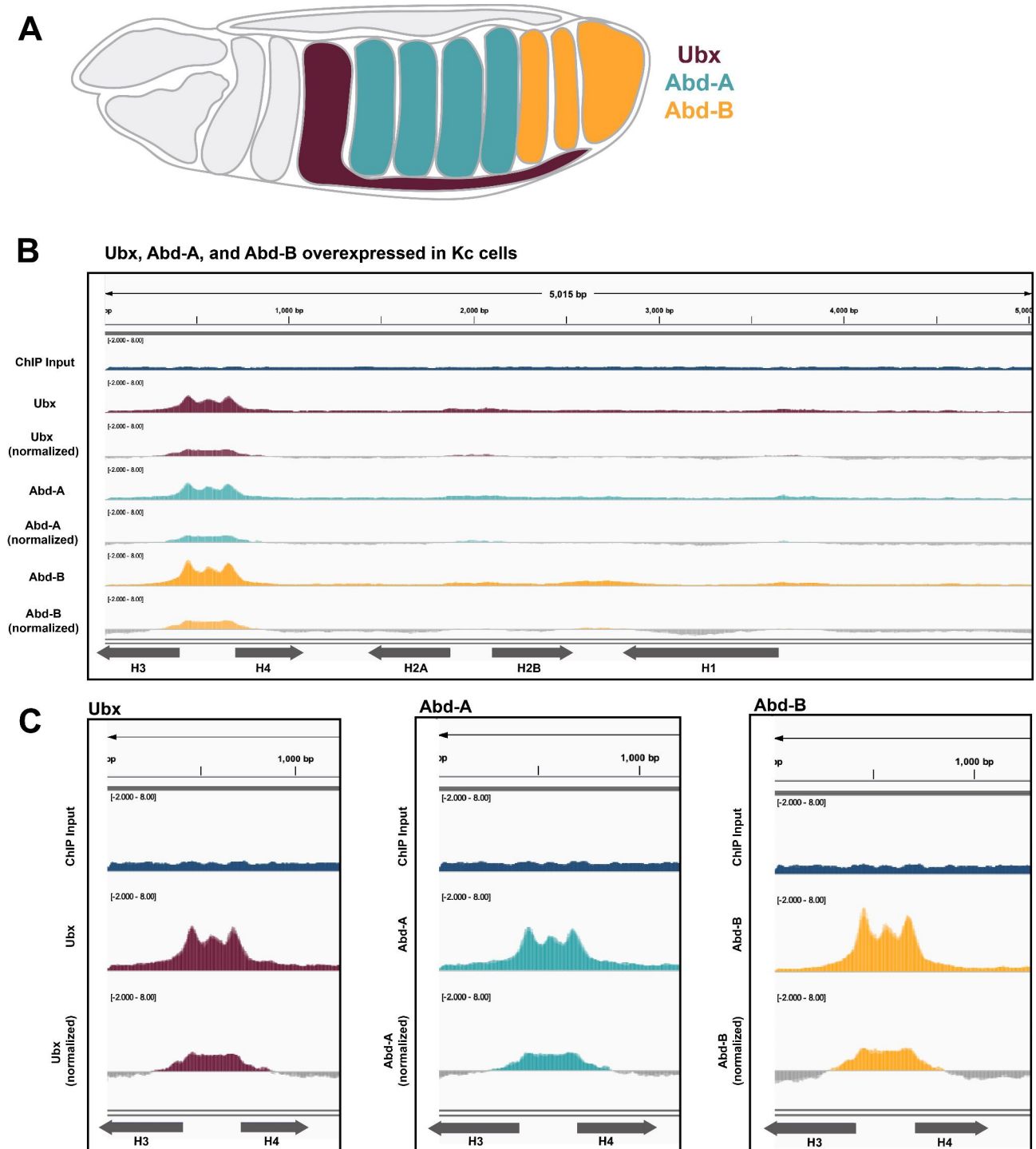
**Fig. 5** Hox factors Ubx, Abd-A, and Abd-B localize to the histone array. **(A)** Diagram of relative tissue expression patterns for Ubx (maroon), Abd-A (teal) and Abd-B (yellow). **(B)** We aligned ChIP-seq datasets from Kc cells expressing Ubx (maroon, two replicates overlayed, [63] ), Abd-A (teal, two replicates overlayed, [63] ), and Abd-B (yellow, two replicates overlayed, [63] ) to the histone gene array. We normalized each ChIP-seq dataset to the provided input (blue, two replicates overlayed, [63] ). **(C)** Enlarged signal from (B) of Ubx (maroon), Abd-A (teal), and Abd-B (yellow) over the *H3/H4* promoter

## Discussion

To broaden our understanding of factors that impact histone biogenesis in *Drosophila melanogaster*, we conducted a candidate-based bioinformatics screen for DNA-binding factors that localize to the histone gene array. Although many HLB factors are known, it is likely that there are many other factors critical for histone biogenesis that have yet to be identified, since several have

**A**  **Embryos**



**B**  **Imaginal Wing Discs - 3rd Instar Larva**
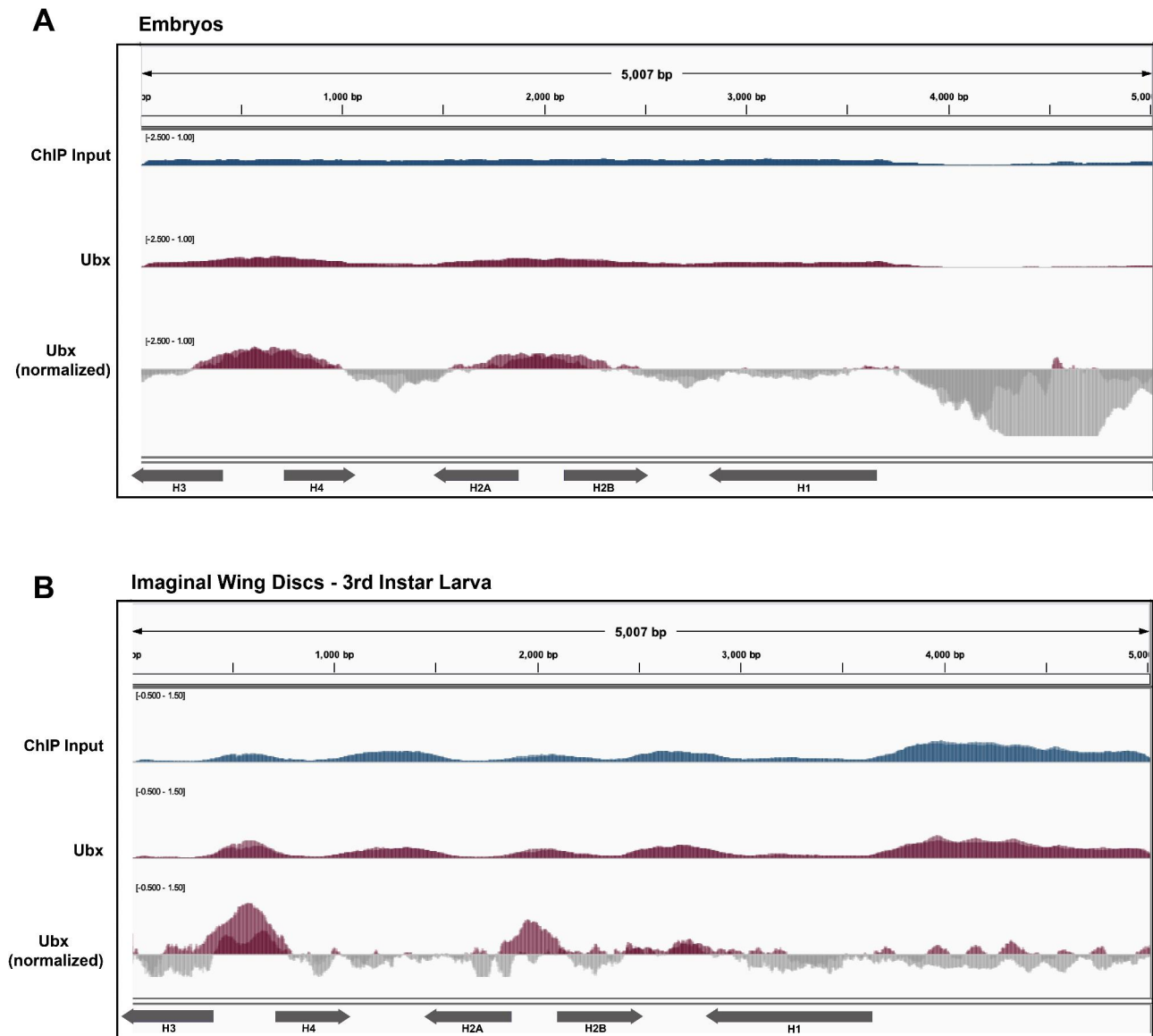


**Fig. 6** Ubx localizes to the *H3/H4* promoter in embryos and 3rd instar larva. We mapped Ubx ChIP-seq datasets from **(A)** mixed population embryos (maroon, two replicates overlayed, [77] ) and **(B)** imaginal wing discs in third instar larva (maroon, two replicates overlayed, [78] ) to the histone gene array. We normalized ChIP-seq datasets to the provided inputs (blue, two replicates overlayed)

been discovered by chance in the past few years including CLAMP [9], Winged-Eye (WGE; [64]), and Myc [21]. To begin to close this gap in knowledge, we chose 27 factors based on their roles in chromatin remodeling, dosage compensation, development, and interaction with known HLB factors, hypothesizing that these represent strong candidates for novel HLB factors. As our screen is limited by availability of relevant datasets, it will likely produce both false positives and negatives. Additionally, because we used a targeted screening approach by investigating factors with relevant functions and at relevant developmental timepoints to histone gene expression, we expected more positive hits than we would find using completely unbiased screen. Given our starting pool of

27 factors, we were pleased to produce 9 candidates for potential HLB factors. We envision that the final 9 candidates that passed our qualitative bioinformatics analysis will be investigated through future wet lab experiments [9, 19, 65].

We validated our bioinformatics pipeline by investigating TRF2, a general transcription factor known to target the histone genes [8], and confirmed that TRF2 binds to the TATA-less *H1* promoter. Isogai et al. (2007) determined that TBP, another general transcription factor, targets the TATA-containing *H3/H4* and *H2A/H2B* promoters. We expanded this observation by investigating TBP-associated factors TAF1, TFIID, and TFIIF. We discovered that all of these general transcription

Hodkinson *et al. BMC Genomic Data*      (2023) 24:54

Page 14 of 18
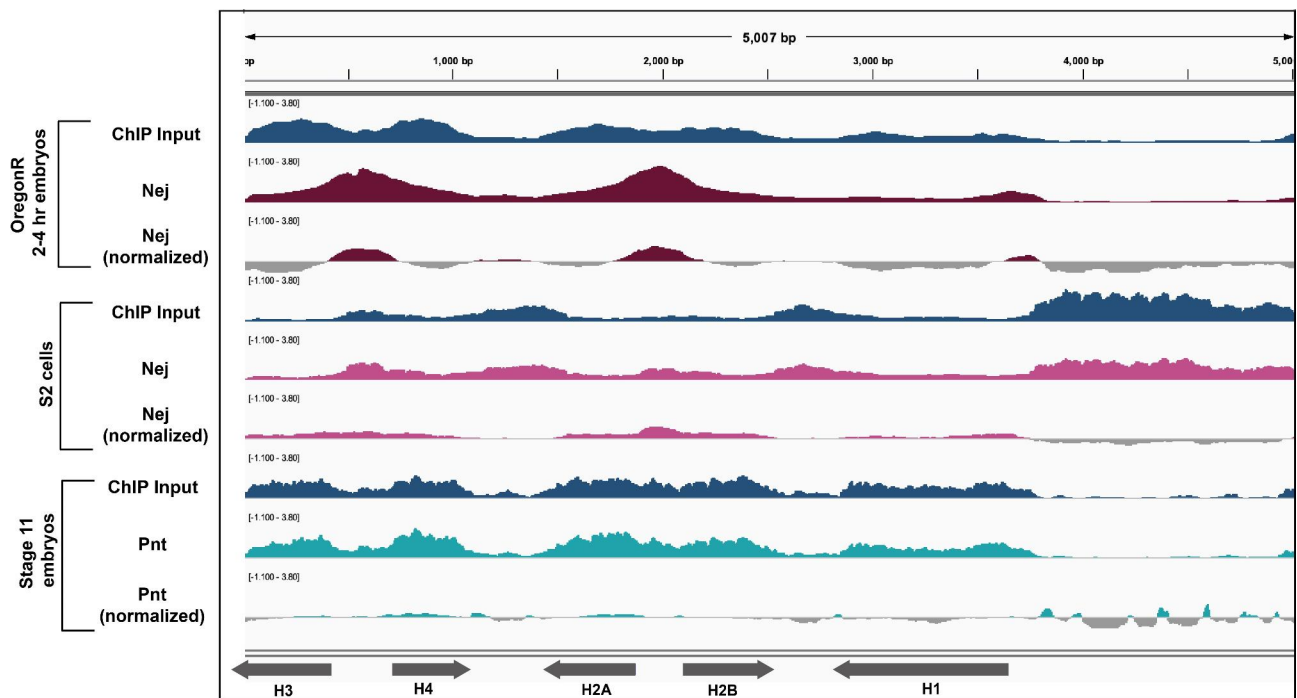


**Fig. 7** ChIP-seq datasets from different tissues can show different alignment results. We mapped two different ChIP-seq datasets for Nejire (Nej) to the histone gene array. ChIP data from 2–4 h embryos (maroon, one replicate, [74]), showed localization to the *H3/H4* promoter and the *H2A/H2B* promoter, while ChIP-seq data from S2 cells (pink, one replicate, [75] ) showed no localization to the histone gene array. We also aligned ChIP-seq data for Pnt from stage 11 embryos [55] to the histone gene array. We normalized the ChIP-seq signals to their respective input signals (blue)

factors target the histone gene array, further validating our pipeline.

We also discovered that the localization of some factors, such as Nej, to the histone gene array is tissue specific. Nej emerged from a proteomic screen for factors involved in HLB activation in cultured cells [11]. However, Nej ChIP-seq from cultured cells did not give meaningful signal over the histone gene array, whereas embryo ChIP-seq showed Nej at histone promoters. These observations denote limitations of our screening technique: we are hindered by the availability and quality of datasets for candidate proteins in specific tissues, genotypes, and conditions.

We initially identified several categories of candidate factors, some of which produced positive hits whereas some did not. For example, Scm, which may interact with the confirmed HLB scaffolding factor Mxc [34, 35, 66], did not show meaningful signal over the histone gene array and therefore we determined that it likely does not target the histone genes.

We also investigated factors involved in dosage compensation, including MSL1, Ndf/CG4747, and JIL-1, because the HLB factor CLAMP plays a key role in male X-chromosome activation. MSL2 was identified in an unbiased proteomics-based HLB candidate screen in cultured cells [11], and we recently discovered that MSLc targets one of the two histone loci in *Drosophila virilis*

in salivary gland polytene chromosomes [37]. Although neither MSL1 nor Ndf localized to the histone gene array, JIL-1 robustly localized to the histone gene array.

Of note, the ChIP-seq datasets for MSL1 were produced from S2 cells, the Ndf datasets were from both male and female larvae, and the JIL-1 dataset came specifically from male third instar larvae. MSL1 and Ndf may target the histone gene array in other tissues or only in embryos, representing potential false negatives in our bioinformatics screen. However, JIL-1 is a more generalized kinase that is responsible for phosphorylating serine 10 on histone 3 across the genome, not just on the male X-chromosome [57, 58, 67]. JIL-1 may therefore be present at the histone locus independent of its role in dosage compensation by contributing to the epigenetic landscape of the locus. Taken together, our results indicate that dosage compensation and histone gene expression are likely distinct regulatory events, and the majority of factors are not shared between these processes in *Drosophila melanogaster*.

One of the lesser studied characteristics of the histone locus is the local chromatin environment and how epigenetic marks influence histone gene expression. We chose CP190, Gcn5, Psc, Pangolin, and su(z)12 as chromatin remodeling candidates that might target the histone genes, but after mapping relevant datasets, none of these candidate chromatin remodelers targets the

histone gene array. We did, however, discover that the long isoform of fs(1)h (fs(1)hL) robustly localizes to the histone gene array. Fs(1)hL has a unique role in chromatin remodeling that differs from the short isoform, as it associates with insulator proteins, including CP190 [60]. Since the histone locus is situated near heterochromatin, it is possible that insulators prevent spreading of heterochromatin into the histone locus. CP190 was also a strong candidate for histone locus association. CLAMP and CP190 share binding profiles at many promoters and each is important for the other's localization [40]. However, when we mapped a CP190 ChIP-seq dataset from female embryos, we did not observe histone array localization. Based on these observations, we conclude that fs(1)hL is a strong candidate for future wet lab studies. Fs(1)hL and CLAMP may interact with CP190 at the histone locus, in specific tissues, or at precise developmental timepoints that were not captured in the datasets we investigated.

Finally, we explored several developmental transcription factors, because histone biogenesis is critical in the first few hours of *Drosophila* development during rapid zygotic cell divisions. We chose Opa, M1BP, and Hnf4 as candidates. Despite their roles in early development and patterning, these factors did not target the histone gene array. However, we identified Nej as a candidate that targets the histone gene array, specifically in *Drosophila* embryos but not in S2 cells. Nej was previously identified as an HLB candidate through a cell-based proteomics screen [11]. Nej is a histone acetyltransferase, but it has roles in cell proliferation and developmental patterning. Nej could influence the chromatin environment of the histone locus during key times in development or in tissues that are constantly dividing where histone proteins would be needed. Because of the roles Nej plays in general developmental processes, it is a strong candidate for future wet lab studies.

We were surprised to discover that the Hox proteins Ubx, Abd-A and Abd-B, all localize to the histone array when overexpressed in Kc cells. Specifically, these factors all target the *H3/H4* promoter. This ~ 300 bp promoter is unique within the 5 Kb histone gene array; it is the minimal sequence required for Mxc localization and HLB formation [19] and contains critical GA-repeat *cis*-elements targeted by CLAMP [9]. The CLAMP-GA-repeat interaction promotes recruitment of histone-locus specific transcription factors [9, 20]. To confirm that our observations were not a byproduct of overexpression, we also investigated independent Ubx ChIP-seq datasets prepared from early embryos (0–16 h) and from third instar larval imaginal wing discs. These data confirm that Ubx targets the histone gene array, although the distribution across the array varies between tissues. Ubx, as well as Abd-A and Abd-B, is highly active in the early embryo when histone

proteins are needed to organize newly synthesized DNA. Therefore Ubx, Abd-A, and Abd-B could provide a spatial and temporal link between histone biogenesis, cell division, and embryo morphogenesis.

With 9 out of 27 hits from our screen emerging as strong candidates for future studies, our screen has proven to be a powerful tool to identify candidates for DNA-binding factors that target the histone gene array. Controls are specifically important to our pipeline because relative peaks at a given location do not always represent true localization. Our negative hits show a range of different negative signals displayed in Fig. 3. In some cases, we saw clear enrichment for open chromatin regions over promoters and/or gene bodies, but did not characterize these factors as hits based on our qualitative analysis criteria. These regions can be overrepresented in the ChIP sequencing experiment as a whole and, therefore, do not reflect where the DNA-binding factor is truly localizing. This phenomenon is best demonstrated when looking at inputs that also show enrichment over open chromatin or gene bodies as shown in Supplemental Fig. 2. Inputs between datasets can be highly variable and, because they are used in the normalization process, can bias the final visualization.

The HLB was discovered by Liu and Gall only seventeen years ago [68]. Since then, novel HLB factors have largely been discovered one at a time by chance. Proteomic screens identified several new candidates but also failed to identify known factors, including CLAMP [11], indicating the screens are far from saturated. A comprehensive inventory of HLB factors is necessary to establish a thorough mechanism of histone biogenesis. Histone regulation is especially critical in the early animal embryo: excess histones drive extra, asynchronous mitotic cycles, whereas depletion of maternal histones lengthens cell divisions in *Drosophila* embryos [32]. The timing of important early developmental events such as the mid-blastula transition is influenced by histone to DNA ratios [47]. Histone levels also affect pre-mRNA splicing in human cells [69], and *H1* isoform loss-of-function mutations are associated with B cell lymphomas [70]. Factors that influence histone biogenesis likely contribute to these developmental and disease phenotypes.

## Conclusions

Here we present a candidate-based screen for novel histone locus-associating factors. Our screen was largely driven by the undergraduate student coauthors in two stages: first, we identified strong candidates based on their established or inferred roles; second, we identified and mapped relevant ChIP-seq datasets to the histone gene array. A similar recent bioinformatic screen searched through thousands of datasets and hundreds of hematopoietic transcription factors for those associated

with the repetitive mammalian rDNA array. This analysis identified numerous candidate transcription factors but required intensive computational pairwise comparisons and thresholding [71]. Another recent screen searched through 1200 chromatin proteins and post-translational modifications to identify those associated with repetitive human centromeres [72]. We instead chose an informed, narrow list of initial candidates and identified 9 out of 27 that we will prioritize for future wet lab studies. Our results not only identify factors that may be involved in histone biogenesis, but also demonstrate the power of a candidate-based bioinformatics screen driven by students.

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12863-023-01147-0.

Supplementary Material 1

## Data Availability
The authors affirm that all datasets used in the screen are available on GEO (Gene Expression Omnibus). All GEO accession numbers and runs from the SRA run selector are specified in **Table 1**.

## Declarations

### Ethics approval and consent to participate
Not applicable.

### Consent for publication
Not applicable.

### Competing interests
The authors declare no conflicts of interest.

### Author details
¹Genetics and Molecular Biology graduate program, Emory University, Atlanta, GA 30322, USA
²Department of Biology, Emory University, 1510 Clifton Road, Atlanta, GA 30322, USA

## References
1. Matera AG, Izaguire-Sierra M, Praveen K, Rajendra TK. Nuclear Bodies: Random Aggregates of Sticky Proteins or Crucibles of Macromolecular Assembly? Dev Cell. 2009;17(5):639–47.
2. Tatomer DC, Terzo E, Curry KP, Salzler H, Sabath I, Zapotoczny G, et al. Concentrating pre-mRNA processing factors in the histone locus body facilitates efficient histone mRNA biogenesis. J Cell Biol. 2016;213(5):557–70.
3. Arias Escayola D, Neugebauer KM. Dynamics and function of Nuclear Bodies during Embryogenesis. Biochemistry. 2018;57(17):2462–9.
4. Duronio RJ, Marzluff WF. Coordinating cell cycle-regulated histone gene expression through assembly and function of the Histone Locus Body. RNA Biol. 2017/01/07 ed. 2017;14(6):726–38.
5. McKay DJ, Klusza S, Penke TJ, Meers MP, Curry KP, McDaniel SL et al. Interrogating the function of metazoan histones using engineered gene clusters. Dev Cell. 2015/02/12 ed. 2015;32(3):373–86.
6. Bongartz P, Schloissnig S. Deep repeat resolution—the assembly of the Drosophila histone complex. Nucleic Acids Res. 2018;47(3):e18–8.
7. Crayton ME, Ladd CE, Sommer M, Hampikian G, Strausbaugh LD. An organizational model of transcription factor binding sites for a histone promoter in D. melanogaster. Silico Biol. 2004;4(4):537–48.
8. Isogai Y, Keles S, Prestel M, Hochheimer A, Tjian R. Transcription of histone gene cluster by differential core-promoter factors. Genes Dev. 2007;21(22):2936–49.
9. Rieder LE, Koreski KP, Boltz KA, Kuzu G, Urban JA, Bowman SK et al. Histone locus regulation by the Drosophila dosage compensation adaptor protein CLAMP. Genes Dev. 2017/08/26 ed. 2017;31(14):1494–50.
10. Marzluff WF, Gongidi P, Woods KR, Jin J, Maltais LJ. The human and mouse replication-dependent histone genes. Genomics. 2002/11/01 ed. 2002;80(5):487–98.
11. White AE, Burch BD, Yang XC, Gasdaska PY, Dominski Z, Marzluff WF et al. Drosophila histone locus bodies form by hierarchical recruitment of components. J Cell Biol. 2011/05/18 ed. 2011;193(4):677–94.
12. Duan J, Rieder L, Colonnetta MM, Huang A, Mckenney M, Watters S et al. eLife. eLife Sciences Publications Limited; 2021 [cited 2023 Jan 9]. CLAMP and Zelda function together to promote Drosophila zygotic genome activation. Available from: https://elifesciences.org/articles/69937.
13. Yang X, Sabath I, Kunduru L, van Wijnen AJ, Marzluff WF, Dominski Z. A conserved interaction that is essential for the biogenesis of histone locus bodies. J Biol Chem. 2014;289(49):33767–82.
14. Terzo EA, Lyons SM, Poulton JS, Temple BRS, Marzluff WF, Duronio RJ. Distinct self-interaction domains promote Multi Sex Combs accumulation in and formation of the Drosophila histone locus body. Mol Biol Cell. 2015;26(8):1559–74.
15. Bulchand S, Menon SD, George SE, Chia W. Muscle wasted: a novel component of the Drosophila histone locus body required for muscle integrity. J Cell Sci. 2010;123(16):2697–707.
16. Godfrey AC, White AE, Tatomer DC, Marzluff WF, Duronio RJ. The Drosophila U7 snRNP proteins Lsm10 and Lsm11 are required for histone pre-mRNA processing and play an essential role in development. RNA. 2009;15(9):1661–72.
17. Jaeger S, Martin F, Rudinger-Thirion J, Giegé R, Eriani G. Binding of human SLBP on the 3'-UTR of histone precursor H4-12 mRNA induces structural rearrangements that enable U7 snRNA anchoring. Nucleic Acids Res. 2006;34(17):4987–95.
18. Shevtsov SP, Dundr M. Nucleation of nuclear bodies by RNA. Nat Cell Biol. 2011;13(2):167–73.
19. Salzler HR, Tatomer DC, Malek PY, McDaniel SL, Orlando AN, Marzluff WF et al. A sequence in the Drosophila H3-H4 Promoter triggers histone locus body assembly and biosynthesis of replication-coupled histone mRNAs. Dev Cell. 2013/03/30 ed. 2013;24(6):623–34.
20. Koreski KP, Rieder LE, McLain LM, Marzluff WF, Duronio RJ. Drosophila histone locus body assembly and function involves multiple interactions. bioRxiv. 2020;2020.03.16.994483.
21. Daneshvar K, Khan A, Goodliffe JM. Myc localizes to histone locus bodies during replication in Drosophila. PLoS ONE. 2011;6(8):e23928.
22. Berloco M, Fanti L, Breiling A, Orlando V, Pimpinelli S. The maternal effect gene, abnormal oocyte (abo), of Drosophila melanogaster encodes

a specific negative regulator of histones. Proc Natl Acad Sci U S A. 2001;98(21):12126–31.

23. Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. 2016;44(W1):W3–10.

24. The Galaxy Community. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. Nucleic Acids Res. 2022;50(W1):W345–51.

25. Schmidt CA, Hodkinson LJ, Comstra HS, Rieder LE. A cost-free CURE: Using bioinformatics to identify DNA-binding factors at a specific genomic locus [Internet]. bioRxiv; 2022 [cited 2022 Dec 16]. p. 2022.10.21.513244. Available from: https://www.biorxiv.org/content/https://doi.org/10.1101/2022.10.21.513244v1.

26. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.

27. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integr Genomics Viewer Nat Biotechnol. 2011;29(1):24–6.

28. Rhee HS, Pugh BF. ChIP-exo method for identifying genomic location of DNA-Binding proteins with Near-Single-nucleotide accuracy. Curr Protoc Mol Biol. 2012;100(1):21241–14.

29. Wang CI, Alekseyenko AA, LeRoy G, Elia AEH, Gorchakov AA, Britton LMP, et al. Chromatin proteins captured by ChIP-mass spectrometry are linked to dosage compensation in Drosophila. Nat Struct Mol Biol. 2013;20(2):202–9.

30. Baumann DG, Gilmour DS. A sequence-specific core promoter-binding transcription factor recruits TRF2 to coordinately transcribe ribosomal protein genes. Nucleic Acids Res. 2017;45(18):10481–91.

31. Ramalingam V, Natarajan M, Johnston J, Zeitlinger J. TATA and paused promoters active in differentiated tissues have distinct expression characteristics. Mol Syst Biol. 2021;17(2):e9866.

32. Chari S, Wilky H, Govindan J, Amodeo AA. Histone concentration regulates the cell cycle and transcription in early development. Development. 2019;146(19):dev177402.

33. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47(D1):D607–13.

34. Docquier F, Saget O, Forquignon F, Randsholt NB, Santamaria P. The multi sex combs gene of Drosophila melanogaster is required for proliferation of the germline. Rouxs Arch Dev Biol. 1996;205(5–6):203–14.

35. Saget O, Forquignon F, Santamaria P, Randsholt NB. Needs and targets for the multi sex combs gene product in Drosophila melanogaster. Genetics. 1998;149(4):1823–38.

36. Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun. 2019;10(1):1930.

37. Xie M, Hodkinson LJ, Comstra HS, Diaz-Saldana PP, Gilbonio HE, Gross JL et al. MSL2 targets histone genes in Drosophila virilis [Internet]. bioRxiv; 2022 [cited 2023 Jan 3]. p. 2022.12.14.520423. Available from: https://www.biorxiv.org/content/https://doi.org/10.1101/2022.12.14.520423v1.

38. Larschan E, Alekseyenko AA, Lai WR, Park PJ, Kuroda MI. MSL complex associates with clusters of actively transcribed genes along the Drosophila male X chromosome. Cold Spring Harb Symp Quant Biol. 2007/03/27 ed. 2006;71:385–94.

39. Straub T, Zabel A, Gilfillan GD, Feller C, Becker PB. Different chromatin interfaces of the *Drosophila* dosage compensation complex revealed by high-shear ChIP-seq. Genome Res. 2013;23(3):473–85.

40. Bag I, Dale RK, Palmer C, Lei EP. The zinc-finger protein CLAMP promotes gypsy chromatin insulator function in Drosophila. J Cell Sci. 2019;132(5):jcs226092.

41. Ali T, Krüger M, Bhuju S, Jarek M, Bartkuhn M, Renkawitz R. Chromatin binding of Gcn5 in Drosophila is largely mediated by CP190. Nucleic Acids Res. 2017;45(5):2384–95.

42. Kyrchanova O, Klimenko N, Postika N, Bonchuk A, Zolotarev N, Maksimenko O, et al. Drosophila architectural protein CTCF is not essential for fly survival and is able to function independently of CP190. Biochimica et Biophysica Acta (BBA) -. Gene Regul Mech. 2021;1864(10):194733.

43. Follmer NE, Wani AH, Francis NJ. A polycomb group protein is retained at specific sites on chromatin in mitosis. PLoS Genet. 2012;8(12):e1003135.

44. Herz HM, Mohan M, Garrett AS, Miller C, Casto D, Zhang Y, et al. Polycomb repressive complex 2-dependent and -independent functions of Jarid2 in transcriptional regulation in Drosophila. Mol Cell Biol. 2012;32(9):1683–93.

45. Wolle D, Cleard F, Aoki T, Deshpande G, Schedl P, Karch F. Functional requirements for Fab-7 Boundary activity in the Bithorax Complex. Mol Cell Biol. 2015;35(21):3739–52.

46. Kaye EG, Booker M, Kurland JV, Conicella AE, Fawzi NL, Bulyk ML et al. Differential Occupancy of Two GA-Binding Proteins Promotes Targeting of the Drosophila Dosage Compensation Complex to the Male X Chromosome. Cell Rep. 2018/03/22 ed. 2018;22(12):3227–39.

47. Amodeo AA, Jukam D, Straight AF, Skotheim JM. Histone titration against the genome sets the DNA-to-cytoplasm threshold for the Xenopus midblastula transition. Proceedings of the National Academy of Sciences. 2015;112(10):E1086–95.

48. Li Z, Thiel K, Thul PJ, Beller M, Kühnlein RP, Welte MA. Lipid droplets control the maternal histone supply of Drosophila embryos. Curr Biol. 2012;22(22):2104–13.

49. Li Z, Johnson MR, Ke Z, Chen L, Welte MA. Drosophila lipid droplets buffer the H2Av supply to protect early embryonic development. Curr Biol. 2014;24(13):1485–91.

50. Stephenson RA, Thomalla JM, Chen L, Kolkhof P, White RP, Beller M, et al. Sequestration to lipid droplets promotes histone availability by preventing turnover of excess histones. Development. 2021;148(15):dev199381.

51. Koromila T, Gao F, Iwasaki Y, He P, Pachter L, Gergen JP et al. Odd-paired is a pioneer-like factor that coordinates with Zelda to control gene expression in embryos. Struhl K, Hobert O, Clark E, editors. eLife. 2020;9:e59610.

52. Bag I, Chen S, Rosin LF, Chen Y, Liu CY, Yu GY, et al. M1BP cooperates with CP190 to activate transcription at TAD borders and promote chromatin insulator activity. Nat Commun. 2021;12(1):4170.

53. Barry WE, Thummel CS. The Drosophila HNF4 nuclear receptor promotes glucose-stimulated insulin secretion and mitochondrial function in adults. Elife. 2016;5:e11183.

54. Ravindranath A, Cadigan KM. Structure-function analysis of the C-clamp of TCF/Pangolin in Wnt/ß-catenin signaling. PLoS ONE. 2014;9(1):e86180.

55. Webber JL, Zhang J, Massey A, Sanchez-Luege N, Rebay I. Collaborative repressive action of the antagonistic ETS transcription factors pointed and Yan fine-tunes gene expression to confer robustness in Drosophila. Development. 2018;145(13):dev165985.

56. Vivekanand P. Lessons from Drosophila pointed, an ETS family transcription factor and key nuclear effector of the RTK signaling pathway. Genesis. 2018;56(11–12):e23257.

57. Cai W, Wang C, Li Y, Yao C, Shen L, Liu S, et al. Genome-wide analysis of regulation of gene expression and H3K9me2 distribution by JIL-1 kinase mediated histone H3S10 phosphorylation in Drosophila. Nucleic Acids Res. 2014;42(9):5456–67.

58. Albig C, Wang C, Dann GP, Wojcik F, Schauer T, Krause S, et al. JASPer controls interphase histone H3S10 phosphorylation by chromosomal kinase JIL-1 in Drosophila. Nat Commun. 2019;10(1):5343.

59. Larschan E, Soruco MM, Lee OK, Peng S, Bishop E, Chery J et al. Identification of chromatin-associated regulators of MSL complex targeting in Drosophila dosage compensation. PLoS Genet. 2012/07/31 ed. 2012;8(7):e1002830.

60. Kellner WA, Van Bortle K, Li L, Ramos E, Takenaka N, Corces VG. Distinct isoforms of the Drosophila Brd4 homologue are present at enhancers, promoters and insulator sites. Nucleic Acids Res. 2013;41(20):9274–83.

61. Kribelbauer JF, Loker RE, Feng S, Rastogi C, Abe N, Rube HT, et al. Context-dependent gene regulation by Homeodomain transcription factor complexes revealed by shape-readout deficient proteins. Mol Cell. 2020;78(1):152–167e11.

62. Rezsohazy R, Saurin AJ, Maurel-Zaffran C, Graba Y. Cellular and molecular insights into hox protein action. Development. 2015;142(7):1212–27.

63. Beh CY, El-Sharnouby S, Chatzipli A, Russell S, Choo SW, White R. Roles of cofactors and chromatin accessibility in hox protein target specificity. Epigenetics Chromatin. 2016;9:1.

64. Ozawa N, Furuhashi H, Masuko K, Numao E, Makino T, Yano T, et al. Organ identity specification factor WGE localizes to the histone locus body and regulates histone expression to ensure genomic stability in Drosophila. Genes Cells. 2016;21(5):442–56.

65. Xie M, Comstra S, Schmidt C, Hodkinson L, Rieder LE. Max is likely not at the Drosophila histone locus [Internet]. bioRxiv; 2022 [cited 2023 Jan 5]. p. 2022.09.11.507040. Available from: https://www.biorxiv.org/content/https://doi.org/10.1101/2022.09.11.507040v1.

66. Kemp JP, Yang XC, Dominski Z, Marzluff WF, Duronio RJ. Superresolution light microscopy of the Drosophila histone locus body reveals a core–shell organization associated with expression of replication–dependent histone genes. MBoC. 2021;32(9):942–55.

67. Regnard C, Straub T, Mitterweger A, Dahlsveen IK, Fabian V, Becker PB. Global analysis of the relationship between JIL-1 kinase and transcription. PLoS Genet. 2011;7(3):e1001327.
68. Liu JL, Murphy C, Buszczak M, Clatterbuck S, Goodman R, Gall JG. The Drosophila melanogaster Cajal body. J Cell Biol. 2006;172(6):875–84.
69. Jimeno-González S, Payán-Bravo L, Muñoz-Cabello AM, Guijo M, Gutierrez G, Prado F et al. Defective histone supply causes changes in RNA polymerase II elongation rate and cotranscriptional pre-mRNA splicing. Proceedings of the National Academy of Sciences. 2015;112(48):14840–5.
70. Yusufova N, Kloetgen A, Teater M, Osunsade A, Camarillo JM, Chin CR, et al. Histone H1 loss drives lymphoma by disrupting 3D chromatin architecture. Nature. 2021;589(7841):299–305.
71. Antony C, George SS, Blum J, Somers P, Thorsheim CL, Wu-Corts DJ, et al. Control of ribosomal RNA synthesis by hematopoietic transcription factors. Mol Cell. 2022;82(20):3826–3839e9.
72. Corless S, Pratap-Singh N, Benabdallah NS, Böhm J, Simon AM, Dolejš V et al. The bromodomain inhibitor JQ1 is a molecular glue targeting centromeres [Internet]. bioRxiv; 2023 [cited 2023 Jul 5]. p. 2023.03.15.532673. Available from: https://www.biorxiv.org/content/https://doi.org/10.1101/2023.03.15.532673v2.
73. THE MODENCODE CONSORTIUM, Roy S, Ernst J, Kharchenko PV, Kheradpour P, Negre N, et al. Identification of functional elements and Regulatory Circuits by Drosophila modENCODE. Science. 2010;330(6012):1787–97.
74. Doiguchi M, Nakagawa T, Imamura Y, Yoneda M, Higashi M, Kubota K, et al. SMARCAD1 is an ATP-dependent stimulator of nucleosomal H2A acetylation via CBP, resulting in transcriptional regulation. Sci Rep. 2016;6:20179.
75. Koenecke N, Johnston J, Gaertner B, Natarajan M, Zeitlinger J. Genome-wide identification of Drosophila dorso-ventral enhancers by differential histone acetylation analysis. Genome Biol. 2016;17(1):196.
76. Kang H, McElroy KA, Jung YL, Alekseyenko AA, Zee BM, Park PJ, et al. Sex comb on midleg (scm) is a functional link between PcG-repressive complexes in Drosophila. Genes Dev. 2015;29(11):1136–50.
77. Shlyueva D, Meireles-Filho ACA, Pagani M, Stark A. Genome-wide Ultrabithorax binding analysis reveals highly targeted genomic loci at developmental regulators and a potential connection to polycomb-mediated regulation. PLoS ONE. 2016;11(8):e0161997.
78. Feng S, Rastogi C, Loker R, Glassford WJ, Tomas Rube H, Bussemaker HJ, et al. Transcription factor paralogs orchestrate alternative gene regulatory networks by context-dependent cooperation with multiple cofactors. Nat Commun. 2022;13(1):3808.

## Publisher's Note