Cellular and Molecular Life Sciences (2022) 79:353
https://doi.org/10.1007/s00018-022-04383-2

ORIGINAL ARTICLE

Comparative interactome analysis of the PRE DNA-binding factors: purification of the Combgap-, Zeste-, Psq-, and Adf1-associated proteins

Darya Chetverina[1] · Nadezhda E. Vorobyeva[2] · Marina Yu. Mazina[3] · Lika V. Fab[4] · Dmitry Lomaev[1] · Alexandra Golovnina[1] · Vladic Mogila[5] · Pavel Georgiev[5] · Rustam H. Ziganshin[6] · Maksim Erokhin[4]

Received: 15 November 2021 / Revised: 14 April 2022 / Accepted: 8 May 2022 / Published online: 9 June 2022

Abstract

The Polycomb group (PcG) and Trithorax group (TrxG) proteins are key epigenetic regulators controlling the silenced and active states of genes in multicellular organisms, respectively. In *Drosophila*, PcG/TrxG proteins are recruited to the chromatin via binding to specific DNA sequences termed polycomb response elements (PREs). While precise mechanisms of the PcG/TrxG protein recruitment remain unknown, the important role is suggested to belong to sequence-specific DNA-binding factors. At the same time, it was demonstrated that the PRE DNA-binding proteins are not exclusively localized to PREs but can bind other DNA regulatory elements, including enhancers, promoters, and boundaries. To gain an insight into the PRE DNA-binding protein regulatory network, here, using ChIP-seq and immuno-affinity purification coupled to the high-throughput mass spectrometry, we searched for differences in abundance of the Combgap, Zeste, Psq, and Adf1 PRE DNA-binding proteins. While there were no conspicuous differences in co-localization of these proteins with other functional transcription factors, we show that Combgap and Zeste are more tightly associated with the Polycomb repressive complex 1 (PRC1), while Psq interacts strongly with the TrxG proteins, including the BAP SWI/SNF complex. The Adf1 interactome contained Mediator subunits as the top interactors. In addition, Combgap efficiently interacted with AGO2, NELF, and TFIID. Combgap, Psq, and Adf1 have architectural proteins in their networks. We further investigated the existence of direct interactions between different PRE DNA-binding proteins and demonstrated that Combgap–Adf1, Psq–Dsp1, and Pho–Spps can interact in the yeast two-hybrid assay. Overall, our data suggest that Combgap, Psq, Zeste, and Adf1 are associated with the protein complexes implicated in different regulatory activities and indicate their potential multifunctional role in the regulation of transcription.

Keywords Polycomb · Trithorax · Mediator · TFIID · NELF · SWI/SNF

Introduction

The Polycomb group (PcG) repressor and trithorax group (TrxG) activator proteins are key epigenetic factors responsible for maintaining gene expression states of many developmentally regulated genes [1–5]. Both groups of proteins are evolutionary well conserved in multicellular organisms, including *Drosophila* and humans. Malfunctions in PcG/TrxG systems are well known to lead to various diseases, including cancer [6–10]. Majority of the known PcG proteins are organized into the multi-subunit complexes: the Polycomb repressive complex 1 (PRC1) and Polycomb repressive complex 2 (PRC2). PRC1 complex has four core subunits: Polycomb (Pc), Polyhomeotic (Ph), Sex combs extra (Sce, aka. dRing), and Posterior sex combs (Psc) [11–13]. PRC2 contains the Enhancer of zeste (E(z)), Suppressor of zeste 12 (Su(z)12), Extra sex combs (Esc) and Chromatin assembly factor 1 (Caf1) [14, 15]. The E(z) PRC2 subunit is a histone methyltransferase (HMT) that creates H3K27me3 modification [14, 15] known to be specific for chromatin regions repressed by the PcG system [16, 17]. The TrxG members are more heterogeneous, and, in particular, include subunits

✉ Darya Chetverina
daria.chetverina@gmail.com

✉ Maksim Erokhin
yermaxbio@yandex.ru

Extended author information available on the last page of the article

of the SWI/SNF chromatin remodelers, CBP, Trx, Mediator, and cohesin subunits [2, 5, 18].

In *Drosophila*, PcG and TrxG proteins are recruited to specialized DNA sequences termed the Polycomb response elements (PREs). PREs are more known for their repressive activity; however, PREs are switchable elements that upon certain conditions can activate transcription as well [1, 19–21]. The outcome of the PRE activity is proposed at least in part to be the result of direct competition between PcG and TrxG, and the level of these factors binding to DNA element [1, 19–21].

Importantly, recent studies indicate that besides PREs and the H3K27me3 silent domains, PRC1 can be recruited to the promoters of active genes and to developmental enhancers [22–24]. Moreover, there is evidence that at least some of the developmental enhancers can function as PREs at later development stages, indicating a close functional connection between these groups of regulatory elements [25].

The core PREs sequences are composed of the binding sites for different DNA-binding proteins that together are suggested to account for the PcG proteins recruitment [19–21]. Currently, identified factors implicated in the PRE-repressive function include Pleiohomeotic (Pho) [26, 27], Pleiohomeotic-like (Phol) [28], GAGA factor (GAF) [29, 30], Combgap (Cg) [31], Pipsqueak (Psq) [32, 33], Zeste (Z) [12, 34], Adh transcription factor 1 (Adf1) [35], Dorsal switch protein 1 (Dsp1) [36], Sp1-like factor for pairing-sensitive silencing (Spps) [37], and Grainy head (Grh) [38].

The role of the Pho protein in the PcG function is the most studied. The mutants of the *pho* gene die as pharate adults with the sex combs on the second and third legs, demonstrating a classical PcG phenotype, identifying Pho as a genuine PcG member [39]. The Pho factor (as well as its homolog Phol) binds PREs in a complex with the methyl-lysine-binding protein Sfmbt [40]. Additional direct contacts between Pho/Phol and PRC2/PRC1 subunits [41, 42], as well as between Sfmbt and PRC1 interacting protein Scm [43] are suggested to be important for a stable PRC1 and PRC2 recruitment to PRE.

At the same time, genome-wide experiments indicate that Pho requires collaboration with other PRE DNA-binding proteins for the PcG recruitment [44]. Accordingly, Pho sites alone are not sufficient for the recruitment of PcG to chromatin [19]. Consistently with the model that Pho requires assistance of other DNA-binding factors, several direct interactions between proteins of this group have been established: Pho interacts with Grh [38] and with GAF [45], GAF interacts with Psq [46, 47], and Scm with Spps [45]. Moreover, Psq [48], Combgap [31], Grh [49], Zeste [12], and Adf1 [35] were shown to co-purify with the PRC1 subunits, and GAF with both the PRC1 and PRC2 [50, 51]. However, the detailed information about the PRE DNA-binding protein interactomes is currently lacking.

One exception is the GAF multifunctional pioneering protein implicated in the activity of different regulatory elements, including besides PREs, promoters, boundaries, and enhancers [52–54]. In agreement with its multifunctional role, GAF interacts with a wide range of factors implicated in distinct regulatory functions, including PcG/TrxG proteins, promoter-associated, and architectural (a.k.a. insulator, or boundary) proteins [50, 55, 56]. Similar to GAF, there are indications that the activities of other PRE DNA-associated factors are not restricted to PREs and to the PcG repressor function, since proteins with the established genome-wide profiles were reported to also bind outside of the PcG domains [35, 44, 57, 58]. However, the lack of the detailed data on their protein partners limits our understanding of their role in the PcG complex recruitment and in other aspects of transcription regulation.

In current study, we selected four PRE DNA-binding proteins for detailed analysis: Combgap, Zeste, Psq, and Adf1.

Combgap has been recently characterized as a PRE DNA-binding protein by the Judith Kassis' lab [31]. Combgap contains a cluster of eight zinc-finger C2H2-type motifs in the central part of the protein, which are suggested to be required for its DNA-binding activity. This factor binds to many characterized PREs and the loss of Combgap leads to a decreased Ph recruitment at a subset of PREs [31].

Zeste contains the Myb/SANT-like DNA-binding domain, and was characterized as a minor component of the PRC1 complex isolated via the Ph subunit [12]. Accordingly, Zeste is partially co-localized with the PRC1 (Ph) peaks genome-wide [58], and was shown to be involved in the *Fab-7*PRE (a.k.a. "iab-7 PRE") mediated silencing [29].

The DNA-binding domain of Psq consists of four helix–turn–helix (HTH) 50-amino acid sequence motifs (Psq domain). In addition, Psq carries a conserved BR-C, ttk, and bab (BTB) domain involved in protein–protein interactions [59]. Mutations in *psq* gene enhance homeotic phenotypes of *ph* and *Pc*, indicating its role in PcG-dependent silencing of Hox genes [32, 33]. Psq was shown to be partially co-localized on the chromatin with Psc, Pc [33, 57], and GAF [47].

The Adf1 protein contains an N-terminal DNA-binding MADF-domain and a C-terminal BESS domain involved in self-dimerization and protein–protein interactions [60]. Mutation in the *Adf1* gene was demonstrated to enhance homeotic phenotypes of *Pc*, indicating its role in the PcG repression. In N-ChIP studies, Adf1 was reported to bind to PREs genome-wide [35].

However, for each of these proteins, there are indications that their function can extend the repressive activity of PREs and PcG. For example, Combgap was shown to participate in the Ph recruitment to a number of promoters outside of PREs, independently of Pho protein, PRC1 subunits (Psc, Pc) and H3K27me3 [31]. Several studies indicated that Zeste can counteract repression and is necessary

for the PRE-mediated inheritance of the active chromatin state of small *Fab-7*PRE fragment [61], and plays activating role in the case of the *bxd*PRE element [62]. Moreover, in addition to PREs, Zeste binds to enhancers and promoters of a number of developmentally regulated genes, and is implicated in their activity [63–65]. While Psq associates with the known H3K27me3-repressive PcG domains, the majority of the Psq and PRC1 co-localization regions seem to represent active enhancer domains co-bound with CBP, GAF, and enriched in the H3K27Ac chromatin modification [57]. In addition, Psq was shown to partially co-localize with boundaries, ISWI ATP-dependent nucleosome-remodeling factor and EcR [57]. Adf1 was first identified as a promoter-binding factor activating the alcohol dehydrogenase (Adh) gene [66, 67], and later was reported to bind to promoters genome-wide [35].

Here, using ChIP-seq experiments and a combination of immuno-affinity purification with liquid chromatography-tandem mass spectrometry (IP/LC–MS), we analyzed the whole-genome distribution of the Combgap, Zeste Psq, and Adf1 factors, and purified protein complexes associated with them. As a result, we were able to demonstrate that the interactions of these factors are not limited to the PcG/TrxG members. While they demonstrated a relatively similar whole-genome co-localization with the analyzed transcription factors, each of the proteins possessed a unique genome-wide binding profile. Moreover, the obvious differences came from the interactome data analysis, and each of the proteins displayed different degrees of association with proteins of distinct functional classes. We further explored the direct mutual interactions between PRE associated DNA-binding factors by yeast two-hybrid (Y2H) analysis and identified several new contacts between them. Our results extensively expand current knowledge of the Combgap, Psq, Zeste, and Adf1 interactomes and uncover their connections with distinct chromatin regulatory complexes.

Materials and methods

Antibodies

Antibodies against the following proteins were raised in rabbits: Psq [amino acid (a.a.) residues 44–1065, PsqA, UniProtKB: Q9V5N1, GenBank: AAC47153.1]; Zeste (full-length, isoform PA); Sce (full-length, isoform PA); Crp (a.a. 245–631, isoform PA); Brm (a.a. 1470–1604, isoform PE). Antigens for antibody production were expressed as 6×His-tagged fusion proteins in *Escherichia coli*, affinity-purified on Ni Sepharose 6 Fast Flow (GE Healthcare), according to the manufacturer's protocol, and injected into rabbits following standard immunization procedures. Antibodies were affinity-purified from serum on the same antigen as were used for the immunization and tested by RNAi/Western blotting (RNAi/WB) to confirm their specificity (Supplementary File 1).

The rabbit polyclonal antibodies against the following proteins were described previously: Adf1 (full-length, isoform PC) [50]; Combgap (a.a. 31–269, isoform PF) [68]; Ph (a.a. 86–520, ph-p, isoform PA) [68–70]; NELF-A (a.a. 943–1248, isoform PA) [71, 72]; GAF (a.a. 1–519, isoform PC) [50, 70]; MED17 (a.a. 459–643, isoform PA), MED26 (a.a. 1–135, isoform PA) [73]; Mor (antigene plasmid was a gift from C. Peter Verrijzer [74]) [75, 76]; Snr1 (a.a. 2–370, isoform PA) [75]; Osa (a.a. 109–330, isoform PA) [77]; NELF-B (a.a. 150–594, isoform PA) [71]; Taf8 (full-length PA) [78].

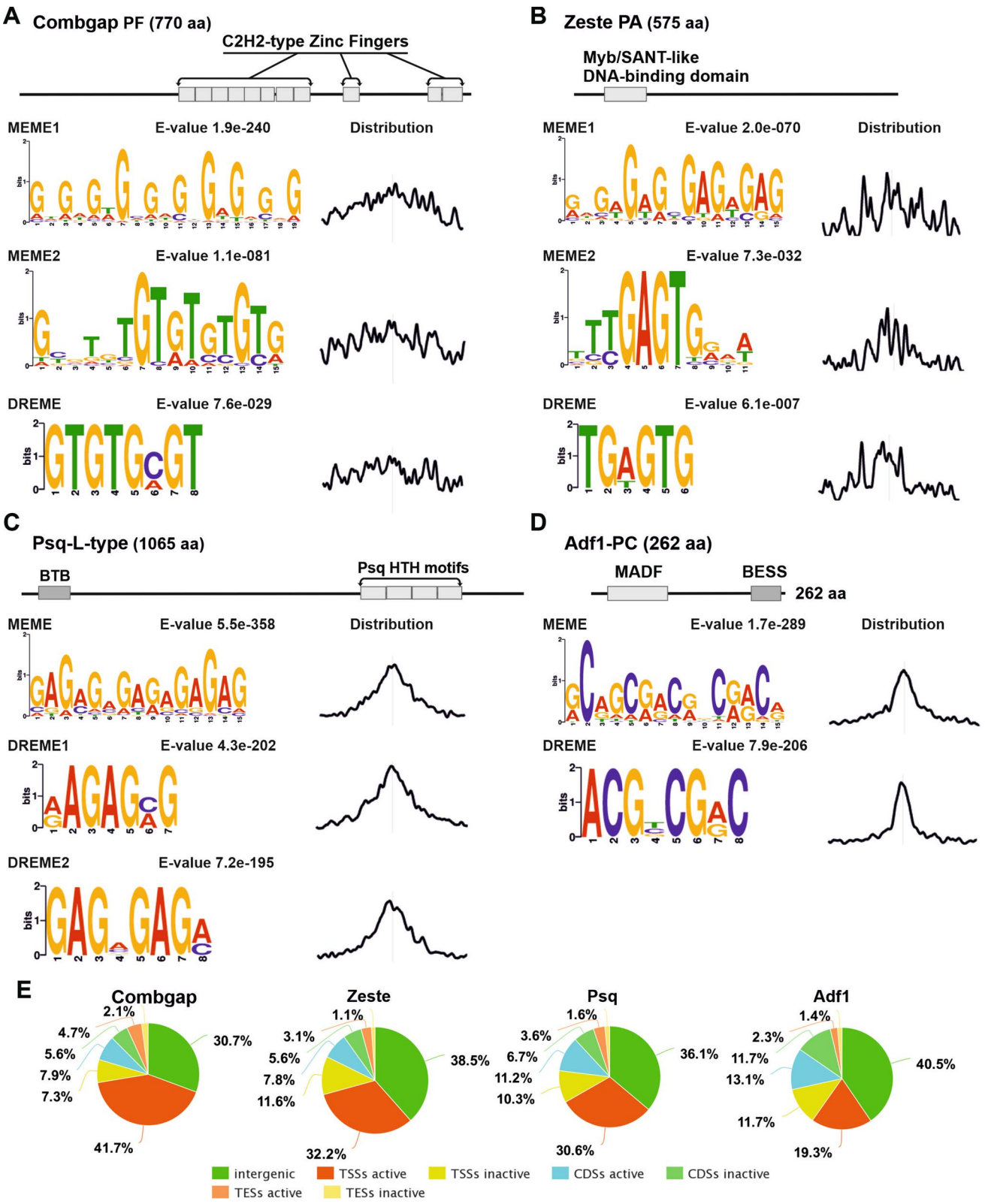
ChIP-seq

The chromatin immunoprecipitation (ChIP) was performed and analyzed exactly as previously described [79, 80]. ChIP-Seq libraries were obtained using the NEBNext Ultra™ II DNA library preparation kit (New England Biolabs). Only the library fragments of 250–500 bp were subjected to NGS sequencing. New-generation sequencing was performed by Genetico (genetico.ru) with the Illumina NovaSeq6000 sequencer. For each of the ChIP-Seq libraries, approximately 7–16 millions of unique paired-end or single-end reads were obtained. The reads in FastQ format were mapped to the dm6 *Drosophila* genome assembly using Bowtie2 (Galaxy Version 2.3.4.3) [81], and filtered (with a minimum MAPQ quality score = 5).

BigWig files were generated using bamCoverage (Galaxy Version 3.5.1.0.0) with scores representing number of reads normalized by the size of the library (the protein-binding levels were normalized to the genome content, calculated as RPGC: number of reads per bin/(total number of mapped reads * fragment length/effective genome size) [82]. The final BigWig files (representing the protein-binding profiles) were obtained using bigwigcompare tool (Galaxy Version 3.5.1.0.0) as a ratio of ChIP signal to Input. The peaks of Combgap, Adf, Psq, and Zeste protein binding were defined by MACS2 callpeak (Galaxy Version 2.1.1.20160309.6) with the following parameters: -gsize '120,000,000' -keep-dup '1'-qvalue '0.01' -mfold '5' '50' --bw '350' 2 > &1 > macs2_stderr [83]. Corresponding input DNA was used as a control for peak calling.

Analysis of ChIP-Seq data was performed on the Galaxy-P platform [84]. All ChIP-Seq data obtained in the current study were deposited into the Gene Expression Omnibus, GSE200213.

The definition of DNA motifs for Combgap, Adf1, Psq, and Zeste binding was performed with MEME suite 5.4.1 [85]. For Adf1 and Psq, the motif search was performed using the sets of all peaks identified with MACS2. For


A
Combgap PF (770 aa)
C2H2-type Zinc Fingers
MEME1
E-value 1.9e-240
Distribution
MEME2
E-value 1.1e-081
DREME
E-value 7.6e-029
B
Zeste PA (575 aa)
Myb/SANT-like
DNA-binding domain
MEME1
E-value 2.0e-070
Distribution
MEME2
E-value 7.3e-032
DREME
E-value 6.1e-007
C
Psq-L-type (1065 aa)
BTB
Psq HTH motifs
MEME
E-value 5.5e-358
Distribution
DREME1
E-value 4.3e-202
DREME2
E-value 7.2e-195
D
Adf1-PC (262 aa)
MADF
BESS
262 aa
MEME
E-value 1.7e-289
Distribution
DREME
E-value 7.9e-206
E
Combgap
2.1%
4.7%
5.6%
7.9%
7.3%
30.7%
41.7%
Zeste
1.1%
3.1%
5.6%
7.8%
11.6%
38.5%
32.2%
Psq
1.6%
3.6%
6.7%
11.2%
10.3%
36.1%
30.6%
Adf1
1.4%
2.3%
11.7%
13.1%
11.7%
40.5%
19.3%
intergenic
TSSs active
TSSs inactive
CDSs active
CDSs inactive
TESs active
TESs inactive

◂**Fig. 1** DNA-binding motifs and whole-genome localization of Combgap, Zeste, Psq, and Adf1 defined by ChIP-seq in the *Drosophila* S2 cells. **A–D** Top: structure of proteins. Domains are shown according to InterPro https://www.ebi.ac.uk/interpro/. Below: DNA-binding motifs identified by MEME suite 5.4.1. for the corresponding protein. (**A**) Combgap, PF 770 aa isoform is shown. Gray boxes—C2H2-type Zinc Fingers. **B** Zeste, PA 575 aa isoform is shown. Gray box—Myb/SANT-like DNA-binding domain. **C** Psq, 1065 aa large (Psq-L-type) isoform is shown. Dark gray box—BTB domain. Light gray boxes—Psq HTH DNA-binding motifs. **D** Adf1, 262 aa PC isoform is shown. Light gray box—MADF DNA-binding domain. Dark gray box—BESS-domain. **E** The percentages of Combgap, Zeste, Psq, and Adf1 peaks located at the intergenic genome regions, and active/inactive TSSs, CDSs, TESs

Combgap, the motif search was performed on TOP-20% of the peaks showing the strongest Combgap binding, for Zeste—on TOP-5% of the peaks showing the strongest Zeste binding.

The following previously described ChIP-Seq data for the S2 *Drosophila* cell line were used: H3K4me3, H3K27ac, H3K27me3 (GSE41440) [86], Polycomb (GSE24521) [87], CP190 (GSE139316) [72], CBP/Nejire, NELF-A, NELF-B (GSE156847) [71], Ph (GSE60686) [88]. MED1 and MED30 ChIP-Seq from the BG3 *Drosophila* cell line were used (from GSE118484) [89]. We have not used any figures or text from the previously published manuscripts—only analyzed the data deposited at the free public access databases.

Group of PREs was determined as H3K27me3-enriched Ph sites using clustering analysis by plotHeatmap tool of deepTools package [82]. K-means clustering algorithm was selected and the number of clusters to compute was set to 2. Group of active enhancers (En) was defined as an intersection of H3K27ac peaks, localized outside TSSs, with CBP/Nejire peaks. Active genes (active TSSs, active CDSs, active TESs) in Fig. 1 were selected on the basis of co-localization of their TSSs with the H3K4me3 peaks. All sets of elements presented on Venn diagrams were defined as the regions of 250 bp around the summits of corresponding protein ChIP-Seq peaks. TSSs and TESs were also defined as areas of 250 bp around the position of the corresponding site according to the UCSC annotation for the dm6 genome (http://genome-euro.ucsc.edu). The CDSs were defined as the regions between the corresponding TSSs and TESs.

X-ChIP

The ChIP was performed and analyzed exactly as previously described [73, 77]. Chromatin was isolated from the whole late larvae (L3 stage) of wild-type or cg^{A22} mutant lines and immuno-precipitated with the specific antibodies (against Combgap, NELF-A, and NELF-B), or with nonimmune IgG that was used as a non-specific antibody control. The enrichment of specific DNA fragments was analyzed by real-time qPCR, using a C1000™ Thermal Cycler with CFX96 real-time PCR detection module (Bio-Rad). Primers used in the ChIP/real-time PCR analyses are listed in Supplementary File 2.

IP/liquid chromatography/tandem mass spectrometry (LC–MS)

Preparation of the nuclear extracts and immuno-affinity purification of protein complexes for LC–MS

The nuclear extracts were obtained from the *Drosophila* S2 cells. For this purpose, 10^9 cells were washed twice in 10 mL of ice-cold PBS and re-suspended in 10 mL of ice-cold IP-Sucrose buffer (10 mM Tris, pH 7.5; 10 mM NaCl, 10 mM $MgCl_2$, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 250 mM sucrose, EDTA-free protease inhibitor cocktail), homogenized with a Dounce loose pestle and incubated on ice for 10 min. The nuclei were then pelleted by centrifugation at 3000 g, 4˚C for 10 min. The pellet was re-suspended in 1 mL of IP-500 buffer (10 mM Tris, pH 7.5; 500 mM NaCl, 10 mM $MgCl_2$, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 0.1% NP-40, 10% glycerol, EDTA-free protease inhibitor cocktail), homogenized with a Dounce tight pestle and incubated for 1 h at + 4 °C on rotator. Lysates were cleared by centrifugation at 18,000*g*, + 4 °C for 10 min. The Combgap, Psq, Zeste, and Adf1 protein complexes were isolated from the obtained nuclear extracts by the immuno-affinity purification following protocols described previously [50]. For these, the Combgap, Psq, Zeste, and Adf1 antibody or the IgG of non-immunized rabbits (negative control) were separately coupled to the protein A Sepharose beads (Pierce) using DMP (Sigma) according to a published protocol [50], and loaded onto columns. The nuclear extracts in the amount of 10 mg each (protein content) were diluted with HEMG buffer (25 mM HEPES–KOH pH 7.6, 12.5 mM $MgCl_2$, 0.1 mM EDTA, 10% glycerol, 1 mM DTT, 0.1% Nonidet P-40) to 150 mM final concentration of NaCl, and loaded onto the columns containing the antibody–protein A Sepharose beads and incubated overnight at 4 °C. Each column was washed twice with HEMG-500 (HEMG buffer with 500 mM NaCl) and once with HEMG-150 (HEMG buffer with 150 mM NaCl), and then proteins were eluted with buffer contained 2% SDS, 100 mM Tris–HCl pH 8.0, 0.5 mM EDTA. Purification procedures for each antibody or IgG of non-immunized rabbits were performed in triplicate. At the next step, the probes were precipitated by TCA and followed by liquid chromatography/tandem mass spectrometry (LC–MS) step procedures.

LC–MS procedures

Reduction, alkylation, and digestion of proteins

Sodium deoxycholate (SDC) reduction and alkylation buffer, pH 8.5 (20 µL) contained 100 mM Tris, 1% (w/v) SDC, 10 mM TCEP and 20 mM 2-chloroacetamide were added to a 20 µg of each protein sample. Each sample was sonicated in an ultrasonic water bath for 1 min, heated at 95 °C for 10 min, cooled to a room temperature and the equal volume of trypsin solution in 100 mM Tris pH 8.5 was added in a 1:50 (w/w) ratio. After overnight digestion at 37 °C, peptides were acidified by 40 µL of 2% trifluoroacetic acid (TFA) mixed with 80 µL of ethyl acetate and loaded on SDB-RPS StageTips containing two 14-gage SDB-RPS plugs, and the StageTip was centrifuged at 300 g until all the solution went through the StageTip (typically 5 min). After washing the StageTips with a 100 µL of 1% TFA/ethyl acetate 1:1 mixture (2 times) and 100 µL of 0.2% TFA, peptides were eluted into a clean tube by 50 µL of 50% acetonitrile/5% ammonia mixture using centrifugation at 300 g. The collected material was vacuum-dried and stored at − 80 °C. Before analyses, the peptides were dissolved in 2% acetonitrile/0.1% TFA buffer at a concentration of 0.5 µg/µL and sonicated for 1 min.

Liquid chromatography and mass spectrometry

Samples were loaded to a home-made trap column 20 × 0.1 mm, packed with Inertsil ODS3 3 µm sorbent (GLSciences), in the loading mobile phase (2% acetonitrile (ACN), 98% H_2O, 0.1% TFA) at 10 µL/min flow and separated at RT in a home-packed fused-silica column 300 × 0.1 mm packed with Reprosil PUR C18AQ 1.9 (Dr. Maisch) into an emitter prepared with P2000 Laser Puller (Sutter, USA). Reverse-phase chromatography was performed with an Ultimate 3000 Nano LC System (Thermo Fisher Scientific), which was coupled to the Orbitrap Lumos Tribrid mass spectrometer (Thermo Fisher Scientific) via a nanoelectrospray source (Thermo Fisher Scientific). Water containing 0.1% (*v*/*v*) formaldehyde (FA) was used as mobile phase A and ACN containing 0.1% FA (*v*/*v*), 20% (*v*/*v*) H_2O as mobile phase B. Peptides were eluted from the trap column with a linear gradient: 3–6% B for 3 min; 6–25% B for 30 min, 25–40% B for 25 min, 40–60% B for 4 min, 60% B during 3 min, 60–99% B for 0.1 min, 99% B during 10 min, 99–2% B for 0.1 min at a flow rate of 500 nL/min. After each gradient, the column was re-equilibrated with A for 10 min. MS data were collected in DDA mode. MS1 parameters were as follows: 120 K resolution, 350–1500 scan range, max injection time—auto, AGC target—standard. Ions were isolated with 1.2 *m*/*z* window, preferred peptide match and isotope exclusion. Dynamic exclusion was set to 30 s. MS2 fragmentation was carried out in HCD mode at 7.5 K resolution with HCD collision energy 30%, max injection time—32 ms, AGC target—standard. Other settings: charge exclusion—unassigned, 1, > 7.

Data analysis

MS raw files were analyzed by Peaks studio 10.0 (Bioinformatics Solutions Inc.). Identification of proteins was made by searching against the *Drosophila melanogaster* Uniprot FASTA database version 25.07.2019 (44,072 entries) with a carbamidomethyl Cys as a fixed modification and deamidation Asn/Gln and Met oxidation as variable modifications. False discovery rate for peptide-spectrum matches was set to 0.01 and was determined by searching a reverse database. Enzyme specificity was set as C-terminal to arginine and lysine, and a maximum of two missed cleavages were allowed in the database search. Peptide identification was performed with an allowed initial precursor mass deviation of up to 15 ppm and an allowed fragment mass deviation of 0.05 Da. For gene ontology data, the Uniprot (https://www.uniprot.org/) database was used. The Power Query Microsoft Excel Tool was used to combine the data from different files. Statistical significance of the identified peptides (spectral count scores) was analyzed using the Student's *t* test and expressed as a *P* value.

Two-hybrid analysis

Two-hybrid assays were carried out using yeast strain pJ69-4A, plasmids, and protocols obtained from Clontech. For the growth assays, plasmids were transformed into pJ69-4A by the lithium acetate method, as described by the manufacturer and were plated on the nonselective media lacking tryptophan and leucine. After 3 days of growth at 30 °C, plates were replicated on selective media lacking tryptophan, leucine, histidine, and adenine. Each assay was repeated at least thrice and growth was compared after 2, 4, and 7 days. Based on the extent of growth, the interactions are scored as strong (detected on 2nd day, "+++"), intermediate (detected on 4th day, "++") or weak (detected on 7th day, "+").

Plasmids for Y2H assay

The cDNA fragments encoding tested proteins were cloned into pGBT9 vector (Clontech) to make fusions with the GAL4 DNA-binding domain, and into pGAD24 vector to make fusions with the GAL4 activating domain. The full-length cDNAs were used in the case of Zeste (CG7803) PA/PB a.a. 1–575; ADF1 (CG15845) PC a.a. 1–262; Dsp1

(CG12223) PD/PG a.a. 1–385; Spps (CG5669) PA/PB a.a. 1–968. The cDNA fragments were used for Psq (CG2368, PsqA) a.a. 2–1065; Pho (CG17743) PA/PB a.a. 1–518; Combgap (CG8367) PF a.a. 31–467.

Results

Overview of Combgap, Zeste, Psq, and Adf1 genomic landscapes in the *Drosophila* S2 cells

To discern the molecular functions of Combgap, Zeste, Psq, and Adf1 proteins, we mapped the genome-wide distribution of these factors by performing ChIP followed by next-generation sequencing (ChIP-seq) on chromatin extracted from the *Drosophila* S2 cells. We used polyclonal antibodies specific to Combgap, Zeste, Psq, or Adf1 (Methods section). The generated ChIP-seq data are available through the GEO Series accession number GSE200213. ChIP-seq peak calling against the input control data revealed 7566/2090/4357/4904 peaks for Combgap/Zeste/Psq/Adf1, respectively.

By de novo motif discovery, using MEME and DREME algorithms, we identified whole-genome Combgap-, Zeste-, Psq-, or Adf1-binding motifs (Fig. 1A–D). The identified motifs well coincided with the previously established ones for all the four proteins: Combgap [31], Zeste [90–93], Psq [59], and Adf1 [35, 94].

The Combgap motif was identified as GT-rich: MEME2 (*E* value 1.1e−081) and DREME (*E* value 7.6 e−029) (Fig. 1A). The Zeste motif with the preferential TGAGTG core was identified as MEME2 (*E* value 7.3e−032) and DREME (*E* value 6.1e−007) (Fig. 1B). The Psq (GA)n motif was identified as MEME (*E* value 5.5e−358) and two GA-enriched motifs with similar E values were identified as DREME1 and DREME2 (*E* value 4.3e−202 and *E* value 7.2e−195) (Fig. 1C). In the case of Adf1 (Fig. 1D), both, MEME (*E* value 1.7e−289) and DREME (A-C-G-T/C/G-C-G-A/G-C, *E* value 7.9e−206) motifs were in accordance with the previously described MEME motifs [35, 94], containing the core of the computationally derived high-affinity Adf-1 binding consensus (G-T/C-C-G-G/A-C) [95].

Our next step was to analyze the co-localization of Combgap/Zeste/Psq/Adf1 relative to the intergenic regions and genes, including transcription start sites (TSSs), coding sequences (CDSs), and transcription end sites (TESs). The genes were subdivided into active and inactive based on the presence or absence of the H3K4me3 histone modification, respectively (Fig. 1E). In summary, over 30% of Combgap/Zeste/Psq/Adf1 peaks fell into the intergenic regions. In addition, over 19% of peaks were enriched at the TSSs of active genes with the highest overlap (42%) in the case of the Combgap protein.

We further analyzed the overlap between the Combgap, Zeste, Psq, and Adf1 (Fig. 2A). The majority of Zeste peaks overlapped with Combgap (90%), Psq (88%), and Adf1 (65%) sites. In addition, Psq peaks overlapped with 88% of Combgap and 54% of Adf1 peaks. At the same time, each protein displayed its unique genome-wide binding profile, with only 1168 regions showing overlap for all four tested proteins.

Next, we analyzed the extend of the Combgap, Zeste, Psq, and Adf1 overlap with representative proteins and chromatin features of different DNA regulatory elements, including PREs, enhancers, and CP190 as a mark of the boundary elements. For this, we used ChIP-seq data obtained in the previous studies on the S2 *Drosophila* cells (Methods section). PREs were defined here as the Ph peaks overlapping with the H3K27me3 PcG-specific histone modification (Ph+/H3K27+). In addition, we measured the association of each protein with Ph regardless of H3K27me3, since a large number of Ph sites occurred outside of the H3K27me3 domains [24]. Enhancer-like elements were defined here as the chromatin regions enriched with CBP and H3K27Ac histone modification that are devoid of TSSs (CBP+/H3K27Ac+/TSS−).

The total overlap of the Combgap/Zeste/Psq/Adf1 with PREs was very similar for each of the proteins (Fig. 2B). About of 7%/10%/9% of Cg/Psq/Adf1 fell into the PRE category. For Zeste, twice the percentage of sites (17%) fell into PRE. At the same time, relative to PREs, Combgap/Zeste/Psq/Adf1 peaks were present in 73%/53%/66%/63% of PREs indicating that more than half of PREs were bound by each of the tested proteins. Significantly larger co-localization was observed for Ph in the absence of the H3K27me3 histone mark (Fig. 2C). Around 86% of Zeste, 81% of Combgap, 78% of Psq, and 53% of Adf1 peaks overlapped with the Ph peaks genome-wide, revealing that Adf1 shows a less extensive overlap with Ph in comparison with Zeste/Combgap/Psq.

From 8 to 13% of Combgap/Zeste/Psq/Adf1 overlapped with the enhancer-like regions (Fig. 2D), this constituted about 50% of enhancers in the case of the Combgap/ Psq/ Adf1, and 25% in the case of Zeste.

In the case of the CP190 architectural protein that marked the boundaries/insulators, about 60% of CP190 peaks overlapped with 54% of Combgap (Fig. 2E). Other factors were less co-localized with CP190: 41% of Zeste, 38% of Psq, and 22% of Adf1 peaks fell into the 12%/24%/15% of CP190 sites, respectively.

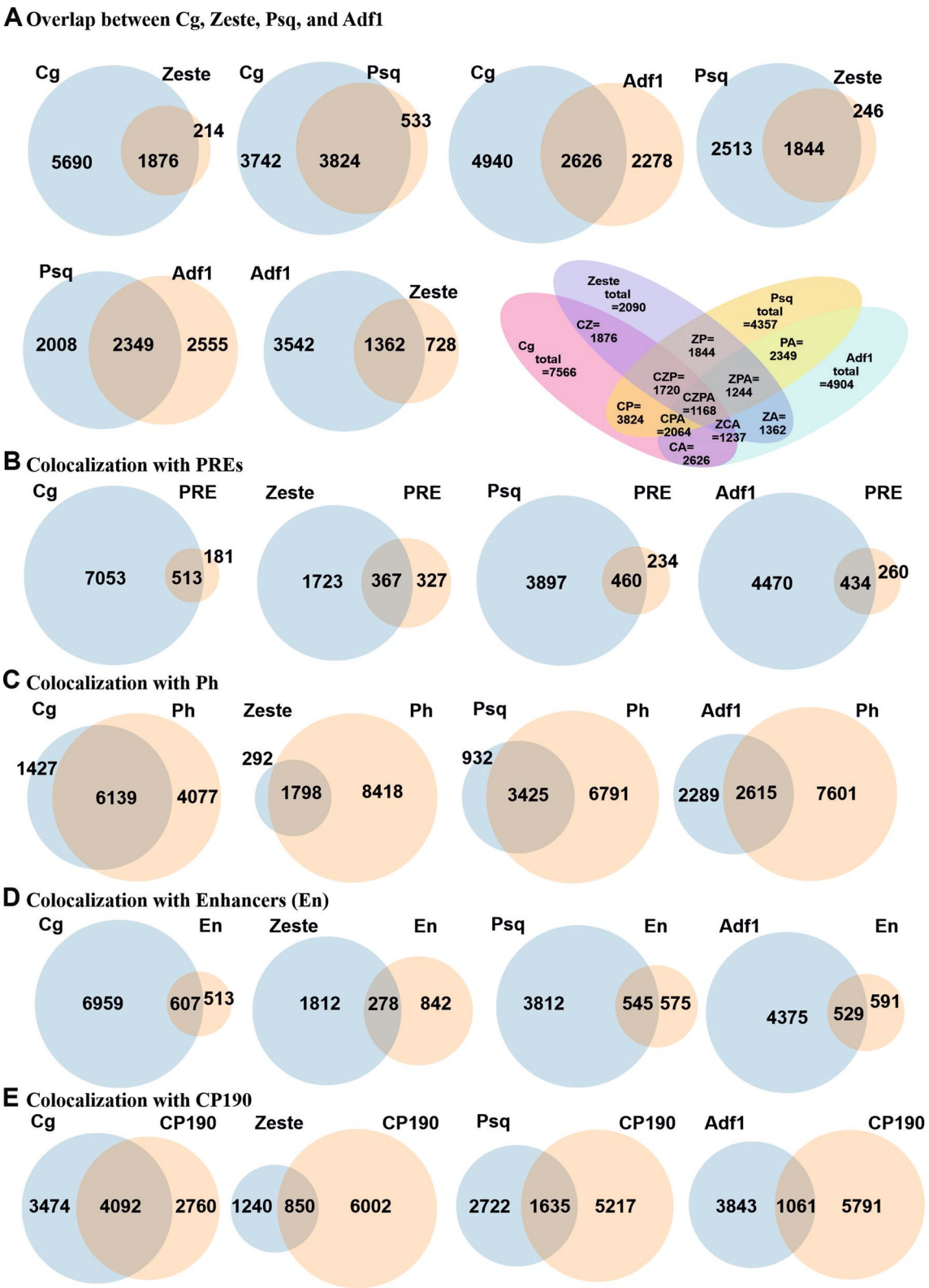


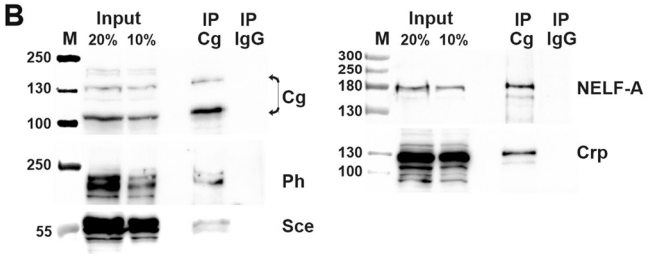
Fig. 2 Venn diagrams showing overlaps between the bound regions of the different protein profiles. **A** Overlaps between the Combgap, Zeste, Psq, and Adf1 peaks, genome-wide. **B** Overlaps between the Combgap, Zeste, Psq, Adf1 peaks and PREs (defined as H3K27me3+/Ph+ sites). **C** Overlaps between the Combgap, Zeste, Psq, Adf1 peaks and Ph. **D** Overlaps between the Combgap, Zeste, Psq, Adf1 peaks and the predicted enhancers (defined as H3K27Ac+/CBP+ peaks outside of the TSS). **E** Overlaps between Combgap, Zeste, Psq, Adf1 peaks and the boundary-associated CP190 protein

A **TOP 20 proteins associated with Combgap**

	Protein	Spectral count: IgG IP	Combgap IP	P-value	Function/complex
	Combgap	5|9|6	167|152|126	0.007	C2H2-type Zinc Finger
1	AGO2	0|0|0	185|164|189	0.0019	siRNA processing, RISC complex
2	CG5953	5|0|0	134|114|122	0.001	MADF-domain, DNA-binding
3	Ph-p	0|0|0	81|80|80	1,7E-05	PRC1 complex
4	Ph-d	0|0|0	73|70|74	0.0003	PRC1 complex
5	Mus201	0|0|0	59|88|67	0.014	UV-damage excision repair
6	Crp	6|14|14	71|71|76	0.00014	Helix-loop-helix DNA-binding
7	Uba5	2|5|7	72|65|61	0.00064	Ubiquitin-activating enzyme
8	Nup35	0|0|0	44|74|59	0.0208	Nuclear pore complex
9	Lola	1|2|13	43|58|59	0.00237	C2H2-type Zinc Finger
10	NELF-A	0|0|0	37|58|48	0.016	Pausing
11	Stwl	0|0|0	32|49|55	0.022	MADF-BESS DOMAIN
12	Ken	0|0|0	33|41|45	0.0078	C2H2-type Zinc Finger
13	Ncoa6	0|0|0	29|46|40	0.016	COMPASS-like complex
14	Chro	16|17|19	39|59|65	0.041	CP190 complex
15	Su(z)2	0|0|0	23|48|37	0.038	PRC1 complex variant
16	M1BP	3|5|5	37|43|40	0.0007	C2H2-type Zinc Finger
17	Sce	0|0|0	35|35|31	0.0016	PRC1 complex
18	Bap111	0|4|7	35|38|38	0.0008	SWI/SNF complex
19	Wee1	0|0|0	21|47|32	0.047	Tyrosine kinase/mitosis control
20	Taf7	0|0|0	41|25|33	0.019	TFIID complex

C **TOP 20 proteins associated with Zeste**

	Protein	Spectral count: IgG IP	Zeste IP	P-value	Function/complex
	Zeste	0|2|2	131|98|100	0.009	PRE-associated, DNA-binding
1	Nocte	0|1|1	124|126|133	0.0003	Circadian clock
2	How	0|4|5	102|104|105	6,35E-06	RNA-binding protein
3	Bre1	0|0|1	54|88|84	0.0197	TrxG
4	Psq	0|0|1	44|38|39	0.0016	PRE-associated, DNA-binding
5	CG17233	0|0|1	30|42|37	0.0087	Unknown
6	Tlk	4|7|12	33|45|44	0.0037	Serine/threonine kinase
7	CG32317	0|0|0	29|28|38	0.0099	Unknown
8	Rept	23|25|29	50|49|58	0.0027	Chromatin remodelling
9	Crp	6|14|14	24|39|42	0.034	Helix-loop-helix DNA-binding
10	Dsp1	2|2|6	37|22|21	0.039	PRE-associated, DNA-bindingr
11	Wmd	2|0|0	21|24|24	0.0001	WD40-repeat protein
12	Pont	22|26|27	40|48|48	0.0058	Chromatin remodelling
13	Lola	1|2|13	18|27|31	0.021	C2H2-type Zinc Finger
14	Sqd	14|8|10	28|30|33	0.0012	SPLICEOSOME COMPLEX A
15	Mocs2B	2|0|0	13|21|23	0.022	ATAC COMPLEX
16	Ph-p	0|0|0	12|20|17	0.0198	PRC1 complex
17	Srp	8|7|9	26|24|22	0.0012	Zinc finger, GATA-type
18	CG8578	4|5|8	20|20|22	0.0014	GTPase binding activity
19	Hexim	0|0|0	11|13|19	0.027	RNA binding protein
20	Prod	12|6|13	17|26|28	0.037	Myb/SANT-like DNA-binding



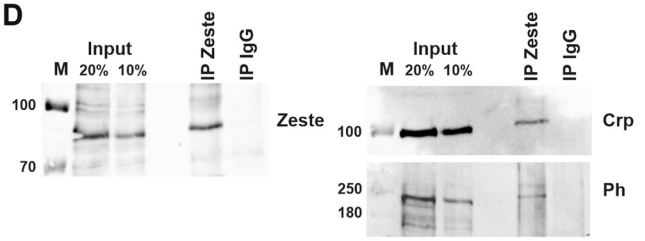


Fig. 3 Top enrichment list of Combgap- and Zeste-interacting proteins in *Drosophila* S2 cells. **A** Spectral counts for the Top nuclear proteins with p values < 0.05 identified in the Combgap IP/LC–MS are listed. The spectral counts for specific antibodies or control IgG probes in three biological replicates are presented. The proteins are sorted in descending order of the difference between the spectrum sums in the three IP/LC–MS repeats with Combgap antibodies and the spectrum sums in the three IP/LC–MS repeats with IgG. In the rightmost column, the known function of the protein, or the protein functional domain are indicated. Statistical significance was analyzed using the Student's t test and expressed as a p value. **B** IP/WB analysis of interactions between Combgap and Ph, Sce, Nelf-A, Crp proteins. Nuclear extract (Input), precipitated fractions (IP) were Western blotted with antibodies against indicated proteins. IP IgG—control immunoprecipitation with the IgG of non-immunized rabbit. **C** Top enrichment list of the Zeste-interacting proteins in *Drosophila* S2 cells. Spectral counts for the Top nuclear proteins with p values < 0.05 identified in the Zeste IP/LC–MS are listed. **D** IP/WB analysis of interactions between Zeste and Crp, Ph proteins

In summary, the ChIP-seq comparative analyses revealed that each of the Combgap, Zeste, Psq, and Adf1 proteins displayed its unique whole-genome binding profile. At the same time, while differing in the extent of the overlap, each of them demonstrated partial co-localization with the chromatin proteins implicated in distinct chromatin activities.

The Combgap, Zeste, Psq and Adf1 TOP-associated proteins

To investigate the functions of Combgap/Zeste/Psq/Adf1 factors thoroughly, we next performed a detailed analysis of their interactomes. For this, we performed immuno-affinity purifications (IP) from the *Drosophila* S2 cells. Nuclear extracts were incubated with Combgap-, Zeste-, Psq-, or Adf1-specific antibody or IgG of the non-immunized rabbits (control IgG) that were coupled to Protein A Sepharose beads. After extensive washing, the bound protein was eluted from the Protein A Sepharose beads with an SDS-containing buffer and analyzed by LC–MS. Experiments for each of the protein-specific antibody or the control IgG were done in triplicate. As a result, we identified proteins in the Combgap, Psq, Zeste, or Adf1 S2 cell pull-downs that were enriched relative to the IgG control samples. The TOP20 nuclear protein partners identified in our IP/LC–MS with the p values < 0.05 are presented for each of the PRE DNA-binding

A **TOP 20 proteins associated with Psq**

	Protein	Spectral count: IgG IP	Psq IP	P-value	Function/complex
	Psq	0|0|1	195|179|177	0.00093	DNA binding HTH domain
1	Nocte	0|1|1	447|514|469	0.0017	Circadian clock
2	CG17233	0|0|1	144|154|139	0.0009	Unknown
3	Pont	22|26|27	155|175|173	0.001	Chromatin remodelling
4	Ens	0|0|0	140|141|132	0.0004	Microtubule binding
5	Fs(1)h	0|0|0	114|142|136	0.004	DNA-binding, TrxG
6	Tlk	4|7|12	129|141|133	2,6E-05	Serine/threonine kinase
7	Rept	23|25|29	128|149|138	0.0015	Chromatin remodelling
8	Dsp1	2|2|6	108|101|98	0.0001	PRE-associated, DNA-binding
9	Srp	8|7|9	99|101|98	7,03E-07	Zinc finger, GATA-type
10	Kis	0|0|0	82|79|69	0.0026	TrxG
11	Crtc	0|0|0	88|75|65	0.0076	Transcriptional coactivator
12	Bre1	0|0|1	47|86|89	0.032	TrxG
13	CG10543	0|0|0	55|71|71	0.0065	C2H2-type Zinc Finger
14	Ct	0|0|0	45|71|68	0.0174	Homeobox transcriptional factor
15	Wmd	2|0|0	64|58|57	0.0006	WD40-repeat protein
16	CG15478	0|0|0	59|62|49	0.0047	Unknown
17	Tai	0|0|0	49|54|61	0.004	Helix-loop-helix DNA-binding
18	Bap111	0|4|7	46|62|65	0.007	SWI/SNF, TrxG
19	GAF	9|0|0	58|58|51	0.0002	PRE-associated, DNA-binding
20	Mor	13|19|25	61|71|75	0.0009	SWI/SNF, TrxG

C **TOP 20 proteins associated with Adf1**

	Protein	Spectral count: IgG IP	Adf1 IP	P-value	Function/complex
	Adf1	0|8|7	296|281|258	0.001	PRE-associated, DNA-binding
1	Chinmo	0|0|0	150|94|103	0.021	C2H2-type Zinc Finger
2	MED15	0|0|0	84|78|78	0.0006	Mediator
3	CG9795	0|0|0	77|75|88	0.0025	Unknown
4	MED27	0|0|0	64|72|87	0.008	Mediator
5	MED6	0|0|1	79|64|59	0.0078	Mediator
6	CG1815	0|0|0	71|63|65	0.0013	Zinc finger, PHD-finger
7	MED14	0|0|0	51|69|42	0.021	Mediator
8	MED25	0|0|0	62|44|52	0.0096	Mediator
9	MED1	0|0|0	64|36|52	0.025	Mediator
10	MED17	0|0|0	47|49|46	0.0003	Mediator
11	Nocte	0|1|1	46|41|55	0.007	Circadian clock
12	MED29/ix	0|0|0	54|40|45	0.0077	Mediator
13	MED7	0|0|0	40|39|59	0.019	Mediator
14	MED8	0|0|0	58|36|40	0.022	Mediator
15	MED20	0|0|0	28|49|49	0.027	Mediator
16	CG8547	0|2|0	29|39|50	0.023	Unknown
17	Combgap	5|9|6	37|50|48	0.007	PRE-associated, DNA-binding
18	MED4	0|0|0	37|41|37	0.0013	Mediator
19	MED24	0|0|0	46|30|32	0.019	Mediator
20	MED23	0|0|0	44|41|23	0.032	Mediator

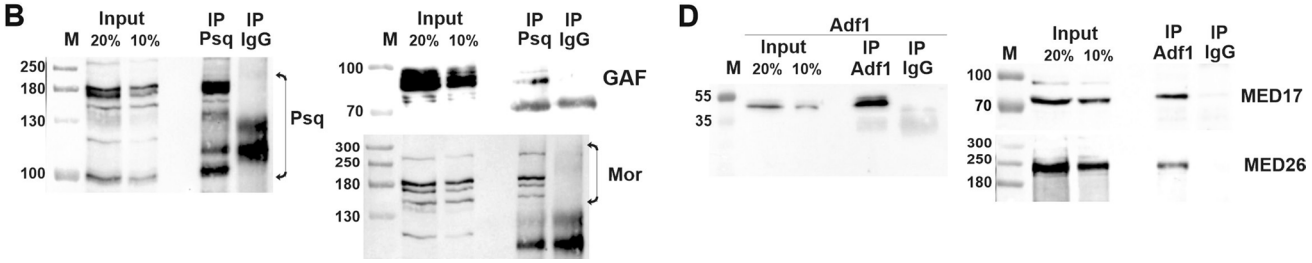


Fig. 4 Top enrichment list of Psq- and Adf1-interacting proteins in *Drosophila* S2 cells. **A** Spectral counts for the Top nuclear proteins with *p* values < 0.05 identified in the Psq IP/LC–MS are listed. **B** IP/WB analysis of interactions between Psq and GAF, Mor proteins. **C** Top enrichment list of Adf1-interacting proteins in *Drosophila* S2 cells. Spectral counts for the Top nuclear proteins with *p* values < 0.05 identified in the Adf1 IP/LC–MS are listed. **D** IP/WB analysis of interactions between Adf1 and MED17, MED26 proteins. Other designations as in Fig. 3

factors in Figs. 3, 4, 5 and 6. All purified proteins are listed in the Supplementary Files 3–6. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [96] partner repository with the dataset identifier PXD029459 and https://doi.org/10.6019/PXD029459.

Combgap TOP-interacting proteins

In the previous studies, Combgap was shown to interact with PRC1 subunits, Ph, Pc and Psc, in co-immunoprecipitation (co-IP) experiments [31]. In agreement with these data, in the current study, we detected Ph (Ph-p and Ph-d homologs) and Sce among the TOP20 Combgap-interacting proteins (Fig. 3A). In addition, Combgap effectively precipitated Su(z)2 (Fig. 3A). Su(z)2 is a homolog of Psc protein that previously was shown to co-purify with PRC1 (via Pc as a bait) [49] and can substitute for Psc in the canonical PRC1 [97]. Using co-IP coupled with Western blotting (IP/WB), we confirmed an association of Combgap with Ph and Sce in the S2 nuclear extracts (Fig. 3B). Here and below, all the main IP/Western blots were validated using reciprocal immunoprecipitation (Supplementary File 1).

At the same time, the highest enrichment in the Combgap IP was observed for the Argonaute 2 (AGO2) protein. The AGO2 is well known for its role in the RNA interference (RNAi)-mediated post-transcriptional gene silencing in the cytoplasm [98]. However, more recently, it was shown to be recruited to the chromatin and has nuclear functions that seem to occur independently of the RNAi pathway. AGO2 binds to many known PREs (bx, bxd, iab-2, Fab-6, Fab-7, and Fab-8) in the BX-C HOX locus [99]. At the same time, in addition to PREs, AGO2 was demonstrated to extensively overlap with boundaries and promoters genome-wide. [99]. The AGO2 role at promoters is currently most studies. AGO2 was found to interact physically with the Pol II [100,

101] and the Negative Elongation Factor (NELF) complex [100], which is involved in promoter pausing [102, 103]. A more recent study indicates that AGO2 attenuates Pol II elongation and affects NELF association with chromatin [104].

Identification of AGO2 as a main Combgap partner indirectly indicates that Combgap can have an AGO2-related function in the regulation of gene transcription and chromatin dynamics. In support of potential Combgap role at promoters, the TOP20 Combgap purified proteins included the NELF-A pausing subunit, Taf7 (component of TFIID complex), and DNA-binding protein M1BP, a known promoter pausing factor [105]. We further confirmed interactions between Combgap and NELF-A by IP/WB assay (Fig. 3B).

Another intriguing Combgap interactor is Cropped (Crp, a.k.a. dAP-4)—a ubiquitously expressed transcription factor, containing helix–loop–helix (bHLH) DNA-binding domain and required for activation of the Sgs-4 gene in response to the ecdysone [106]. Crp is controlled by dMyc and, in particular, it is required for the branching morphogenesis of terminal cells of the tracheal tubes [107]. Interaction between Combgap and Crp was further confirmed by IP/WB assay (Fig. 3B).

In addition, of interest is the enrichment of Combgap IP/LC–MS with proteins bearing DNA-binding motifs. These, except for M1BP and Crp, included CG5953, Lola, Stwl and Ken. It would be important in future to understand the functional relations between these proteins and Combgap, as they could co-participate in the recruitment of regulatory complexes to chromatin.

Zeste TOP-interacting proteins

A previous study has indicated that Zeste can be purified in co-IP via the Ph subunit of PRC1 complex [12]. Moreover, in case of the *Fab-7*PRE, Zeste sites were required for recruitment of the SWI/SNF TrxG Brm protein to the transgene [61] and Brm was reported to be a direct Zeste partner [108].

In line with the PcG Zeste function, Ph was present in the TOP20 Zeste-interacting proteins (Fig. 3C). In addition, the Zeste TOP-purified partners included Rept and Pont subunits of the INO80 complex [109]. The Rept was previously reported to co-purify with Ph [12], and both factors, Rept and Pont, were detected in the Pho- and dSfmbt-tandem affinity purifications (TAP) [40, 110]. Of the TrxG members, the TOP20 interactors included the Bre1 protein—an E3 ubiquitin-ligase required for the H2B mono-ubiquitylation [111] (Fig. 3C).

At the same time, the highest enrichment in Zeste IP/LC–MS was observed for Nocte (Fig. 3C), which is involved in the temperature compensation of the circadian clock [112]. The functional connection between Zeste and Nocte is currently unknown. Not less intriguing, is the presence of Crp and Lola, the DNA-binding proteins, among the TOP20 Zeste interactors, which were also present in the Combgap TOP20 list. In addition to Crp and Lola, the Zeste TOP20 proteins included Psq, Dsp1, and Prod DNA-binding proteins. Two of them, Psq and Dsp1, are known PRE DNA-binding proteins. We further validated our IP/LC–MS results by confirming Zeste interaction with Crp, and Ph using the IP/WB analysis (Fig. 3D).

Psq TOP-interacting proteins

Psq has been previously shown to co-immuno-precipitate with Pc protein (PRC1 complex) [33], GAF PRE DNA-binding protein [47, 50], and with the Mod(mdg4)2.2 [57].

In our study, the core components of the PcG complexes were absent from the TOP20 Psq interactors (Fig. 4A). Moreover, only a background amounts of the PRC1 core subunits were present in IP/LC–MS (Supplementary File 5 and 7, Fig. 5). At the same time, TOP20 Psq interactors were more represented by the TrxG factors. Two of them, Bap111 and Mor, are subunits of the ATP-dependent SWI/SNF chromatin-remodeling complexes, while Fs(1)h (a.k.a. dBrd4), Kis, Bre1, are the TrxG proteins that are not assigned to a specific complex (for reviews about these proteins see [1, 3, 5, 19]). In addition, TOP20 Psq interactors included the PRE DNA-binding factors, Dsp1 and GAF, as well as Pont and Rept remodelers. Intriguingly, as in the case of Zeste IP/LC–MS, the Nocte circadian clock protein was the most enriched protein in IP/LC–MS. In addition, among the Psq TOP20 are proteins with the DNA-binding domains: CG10543 (a C2H2-type zinc finger protein), Ct (a homeobox transcriptional factor), Tai (a helix–loop–helix protein). We further confirmed Psq-Mor and Psq-GAF interactions by the IP/WB assay (Fig. 4B).

Adf1 TOP-interacting proteins

In a previous study, Adf1 has been shown to interact with the Pc (PRC1) protein in co-IP experiments [35].

Our analysis revealed that the Adf1 interactome is strikingly different from other tested PRE DNA-binding proteins (Fig. 4C). Several findings are of interest. First, almost the entire list of the Adf1 TOP20-associated proteins is represented by the subunits of the Mediator complex. Second, the highest enrichment in Adf1 IP/LC–MS is observed for Chinmo protein. Chinmo has a BTB-domain and two Zinc finger C2H2-type motifs with the DNA-binding potential. Previous studies have linked the function of Chinmo to the

A **PcG/TrxG proteins associated with Combgap, Zeste, Psq and Adf1**

		Spectral count							
Protein	IgG IP	Combgap IP	P-value	Zeste IP	P-value	Psq IP	P-value	Adf1 IP	P-value
Polycomb Group (PcG) Proteins									
PRC1 complex									
Sce/dRing	-	35|35|31	0.0016	4|10|10	0.057	0|3|0	0.423	4|2|0	0.225
Ph-p	-	81|80|80	1.7E-05	12|20|17	0.0198	0|1|0	0.423	-	-
Ph-d	-	73|70|74	0.0003	**0|17|15**	**0.185**	0|1|0	0.423	-	-
Pc	-	9|13|8	0.023	0|0|1	0.423	-	-	1|0|0	0.423
Psc	-	0|8|6	0.1917	-	-	-	-	-	-
PRC1 associated proteins									
Su(z)2	-	23|48|37	0.038	-	-	-	-	-	-
Other PcG									
Asx	-	-	-	-	-	7|12|9	0.023	1|0|0	0.423
Sxc/Ogt	0|1|0	-	-	-	-	17|20|20	0.001	-	-
Dom	-	-	-	-	-	9|19|17	0.039	-	-
Trithorax Group (TrxG) Proteins									
SWI/SNF ATP-dependent chromatin remodelers subfamily									
Bap111	0|4|7	35|38|38	8E-04	7|6|18	0.222	46|62|65	0.007	7|13|12	0.064
Mor	13|19|25	**24|39|34**	**0.079**	**15|22|28**	**0.63**	61|71|75	0.0009	13|18|0	0.85
Bap55	9|8|10	**15|25|24**	**0.056**	9|15|24	0.248	33|39|37	0.002	9|13|14	0.179
Snr1	3|3|8	9|12|10	0.057	-	-	15|24|21	0.012	-	-
Bap60	0|2|2	1|7|6	0.207	-	-	22|40|40	0.031	-	-
Brm	-	2|2|3	0.02	-	-	**7|26|17**	**0.093**	-	-
Act5C	90|66|66	91|76|59	0.919	94|72|89	0.352	-	-	-	-
BAP specific subunits									
Osa	-	2|5|1	0.157	-	-	24|41|36	0.022	2|0|0	0.423
Other TrxG									
Kis	-	0|2|0	0.4226	4|3|6	0.039	82|79|69	0.0026	18|9|10	0.05
Bre1	0|0|1	0|1|4	0.384	54|88|84	0.02	47|86|89	0.032	-	-
Fs(1)h	-	2|7|6	0.082	12|10|16	0.019	114|142|136	0.004	0|1|0	0.432

MED12/Kto, MED13/Skd - described in Mediator complex

B

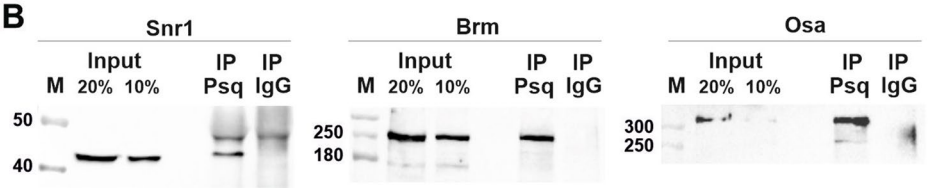


Fig. 5 PcG/TrxG proteins associated with Combgap, Zeste, Psq, and Adf1. **A** Spectral counts for the PcG/TrxG proteins in IP/LC–MSs are presented. Blue—statistically significant interactions (p value < 0.05); bold - the difference between the spectrum sums in the three IP/LC–MS repeats with specific antibodies and the spectrum sums in the three IP/LC–MS repeats with IgG ≥30. Other designations are as in Fig. 3. **B** IP/WB analysis of interactions between Psq and Snr1, Brm, Osa

JAK/STAT signaling pathway [113]. Third, the Combgap PRE DNA-binding protein is among the TOP20 enriched in the Adf1 IP/LC–MS factors. At the same time, no PcG core subunits were present among its TOP20 proteins. Using the IP/WB assay, we confirmed the interaction between Adf1 and Med17 (Fig. 4D). In addition to the Med17 subunit of the head module, we confirmed the interaction between the Med26 subunit of the middle module and Adf1 by the IP/WB analysis (Fig. 4D).

Comparative analysis of Combgap, Zeste, Psq, and Adf1 proteins interactomes

For a more comprehensive understanding of the Combgap, Zeste, Psq, and Adf1 interactomes, we performed a comparative analysis of the functional complexes present in the IP/LC–MS interactome data.

PcG/TrxG factors

We first analyzed in detail the presence of the PcG- and TrxG-related proteins in IP/LC–MSs (Fig. 5, for the full list of analyzed PcG/TrxG group proteins, please see Supplementary File 7).

Several findings were of interest. Figure 5A shows that the components of PcG proteins are most strongly precipitated by the antibodies against Combgap, which effectively purified Sce, Ph-p/Ph-d, and Pc subunits of the PRC1 (p values of < 0.05), and associated unstably with Psc. In the case of Zeste, we detected Ph (p value < 0.05) and less significant

A **Mediator proteins associated with Combgap, Zeste, Psq, Adf1**

Spectral count

Protein	IgG IP	Combgap IP	P-value	Zeste IP	P-value	Psq IP	P-value	Adf1 IP	P-value
Mediator									
Head module									
MED6	0|0|1	3|8|5	0.068	0|1|3	0.38	**5|15|15**	**0.075**	79|64|59	0.0078
MED8	-	5|7|10	0.037	0|1|1	0.1835	11|20|17	0.026	58|36|40	0.022
MED11	-	3|2|6	0.092	-	-	5|6|6	0.003	31|31|31	0
MED17	-	-	-	-	-	-	-	47|49|46	0.0003
MED18	-	0|1|2	0.225	0|0|2	0.42	4|5|7	0.026	18|13|19	0.012
MED19	-	-	-	-	-	0|9|8	0.185	**15|37|29**	**0.052**
MED20	-	-	-	-	-	0|9|8	0.185	28|49|49	0.027
MED22	-	5|4|5	0.005	0|0|1	0.42	7|4|4	0.038	17|18|17	0.0004
MED28	-	6|5|6	0.003	-	-	7|12|9	0.023	33|28|38	0.007
MED30	0|0|1	4|7|8	0.03	2|4|3	0.025	8|9|10	0.0007	37|27|28	0.01
Middle module									
MED1	0|0|1	-	-	-	-	0|0|1	0.423	64|36|52	0.025
MED4	-	5|6|7	0.009	-	-	16|15|17	0.0013	37|41|37	0.0012
MED7	-	3|5|6	0.033	-	-	7|11|14	0.034	40|39|59	0.019
MED9	-	3|0|1	0.27	-	-	2|2|2	0.014	11|10|10	0.001
MED10	-	1|0|0	0.423	-	-	5|4|6	0.013	23|20|25	0.004
MED21	-	0|2|2	0.1835	-	-	3|2|3	0.015	22|17|16	0.010
MED31	-	3|3|1	0.072	-	-	3|4|3	0.01	**5|11|17**	**0.086**
MED14	-	-	-	-	-	-	-	51|69|42	0.021
MED27	-	0|2|4	0.2254	-	-	3|13|13	0.1	64|72|87	0.008
MED26	-	0|2|0	0.423	-	-	-	-	25|31|32	0.005
Tail module									
MED29/ix	-	4|4|5	0.005	-	-	8|9|12	0.015	54|40|45	0.0077
MED24	-	-	-	-	-	0|1|0	0.423	46|30|32	0.019
MED23	-	-	-	-	-	-	-	44|41|23	0.032
MED15	-	4|9|9	0.047	-	-	12|15|19	0.017	84|78|78	0.0006
MED16	-	-	-	-	-	-	-	29|27|18	0.018
MED25	-	0|2|0	0.423	-	-	9|19|20	0.045	62|44|52	0.0096
CDK8 module									
Cdk8	-	-	-	-	-	0|1|0	0.423	8|6|8	0.008
MED12/Kto	-	-	-	-	-	1|6|5	0.120	**23|7|12**	**0.097**
MED13/Skd	-	-	-	-	-	-	-	**32|20|12**	**0.067**
CycC	-	-	-	-	-	-	-	0|2|2	0.1835

B

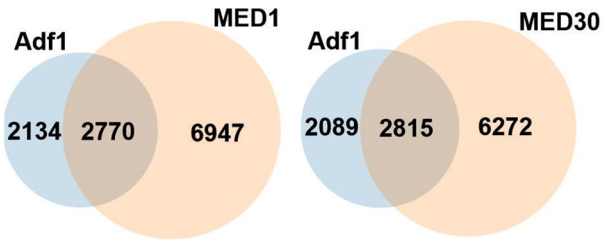


Fig. 6 Mediator proteins associated with Combgap, Zeste, Psq, and Adf1. **A** Spectral counts for the subunits of Mediator complex in IP/LC–MSs are presented. Designations are as in Figs. 3 and 5. **B** Co-localization of Adf1 peaks (S2 cells, present study) with the MED1 and MED30 proteins (ChIP-seq data from BG3 cells)

association with Sce (p value 0.057). Similar, higher affinity of Zeste to Ph than to other PRC1 subunits was observed previously in purifications of Flag-tagged PRC1 subunits [114]. Only a background non-significant association of several core PRC1 subunits was detected in the case of Psq and Adf1 IP/LC-MSs. While not associated with the PcG core complexes, the Psq displayed a stable association with Sxc (a.k.a. Ogt) and Dom PcG proteins, which are not assigned to a specific complex. Importantly, we found no association of the E(z), Esc, Su(z)12 PRC2 core subunits and PhoRC with the tested proteins, reflecting a closer linkage to PRC1 (Supplementary file 7).

At the same time, of the purified interactomes, the Psq IP/LC–MS displayed the highest association with the TrxG factors. In addition to Bap111 and Mor in the TOP20 proteins

(Fig. 4A), the Psq IP/LC–MS is enriched in Bap55, Snr1, Bap60, and Brm representing together all the core SWI/SNF subunits, except for Act5C (Fig. 5A). It should be noted that while Brm does not give significant *p* value (0.093), it has a relatively high spectral count score in IP/LC–MS analysis. Accordingly, IP/Western blot analysis confirms the Psq-Snr1 as well as Psq-Brm interactions (Fig. 5B).

Two distinct Brm ATPase SWI/SNF family complexes exist: BAP and PBAP. They are distinguished by the presence of specific subunits: Osa in the case of BAP, and Polybromo, Bap170, and SAYP, in the PBAP complex [115, 116]. In the case of Psq IP/LC–MS, we detected a high and statistically significant association with the Osa subunit of BAP complex (Fig. 5A), while the PBAP-specific subunits were absent (Supplementary File 7). To further validate our IP/LC–MS data, we confirmed interaction between Psq and Osa subunit of the BAP complex by IP/WB assay (Fig. 5B).

Three other TOP20 TrxG factors of the Psq IP/LC–MS were also detected in Zeste and Adf1 IP/LC–MS: Zeste, in addition to Bre1 in TOP20 list, interacts with Fs(1)h and weakly with Kis; Adf1 associated with Kis.

The Mediator complex

The Mediator complex consists of 30 polypeptides that can be divided into four distinct modules, termed the head, middle, tail, and CDK8 kinase module [117]. Since almost all of the TOP Adf1 IP/LC–MS interactors are represented by the Mediator subunits, we analyzed the presence of the entire complex in IP/LC–MSs. Importantly, anti-Adf1 antibodies precipitated all the components of the Mediator complex (Fig. 6A).

The mutations of genes encoding MED12/Kto and MED13/Skd Mediator subunits were originally identified as the suppressors of the *Pc* mutation [118], but, at the same time, they were shown to cause the *Ubx* derepression [119] indicating their participation in the TrxG activation, as well as in the PcG repression. However, the CDK8 kinase module in IP/LC–MS showed a less stable association with Adf1, especially the CycC subunit, which was unstably detected only at background levels (Fig. 6A). These are indications that the PcG/TrxG factors of the CDK8 kinase module are not the primary Adf1 partners of the Mediator complex. Importantly, a strong association with the Mediator complex was specific to Adf1. The interactomes of the other tested proteins showed much less enrichment in the Mediator subunits (Fig. 6A). To evaluate the physical connection of Adf1 with Mediator on the chromatin, we further estimated the degree of the Adf1 co-localization with MED1 and MED30 subunits, genome-wide (Fig. 6B), using the previously published data for Mediator subunits, obtained on the BG3 *Drosophila* cells [89]. Over 55% of the Adf1 peaks coincided with the MED1/MED30 peaks. These co-localized peaks corresponded to 29% of MED1 and to 31% of MED30 peaks, respectively.

Promoter-associated proteins

We further analyzed the presence of promoter-associated proteins in IP/LC-MSs first pointing our attention to the presence of factors implicated in promoter pausing. The promoter pausing requires at least two factors that function cooperatively: NELF and DRB sensitivity-inducing factor (DSIF). In addition to NELF and DSIF, several other factors have been implicated in pausing, most notable among them are GAF and M1BP DNA-binding factors, the RNA polymerase II-associated factor 1 complex (PAF1C), and transcription factor II D (TFIID) [103]. Importantly, only Combgap displayed a stable association with the NELF complex, M1BP, and is enriched in the TFIID subunits (Fig. 7A). None of the tested factors, displayed association with DSIF or PAF1C complexes. At the same time Combgap, Psq, and Adf1, interacted with GAF protein (Fig. 7A). However, due to the association of GAF with PREs, promoters, boundaries, and presumably, with the active enhancers, it is currently impossible to discern whether they are co-participating in the promoter functions. We further confirmed interactions between Combgap and NELF-B, and Taf8 by IP/WB assay (Fig. 7B).

Using the previously published ChIP-seq data for NELF-A and NELF-B obtained on the S2 *Drosophila* cells [71], we next assessed the extent of the Combgap co-localization with the NELF-A and NELF-B subunits, genome-wide (Fig. 7C). Intriguingly, 88% and 84% of the Combgap peaks fell into the NELF-A and NELF-B peaks, respectively. These co-localized peaks corresponded to 64% of NELF-A and to 65% of NELF-B peaks, respectively. Thus, Combgap showed a high degree of co-localization with the NELF complex.

To explore the potential functional implication of Combgap in recruitment of the NELF complex to the chromatin, we further estimated the effect of Combgap mutation, using the cg^{A22} zygotic mutant, at the third-instar larval stage. The cg^{A22} zygotic mutants were previously used to demonstrate the role of Combgap in the Ph recruitment, genome-wide [31]. In this study, the Ph binding was affected at several hundreds of sites, which constitute a relatively small proportion of the total number of Ph-binding regions. It was suggested that such a modest effect can be a result of the persistence of a low level of Combgap protein on the chromatin in cg^{A22} mutant larvae or of the combinatorial nature of recruitment of PcG proteins. Due to this, we have selected

A **Promoter and elongation factors associated with Combgap, Zeste, Psq, Adf1**

Spectral count

Protein	IgG IP	Combgap IP	P-value	Zeste IP	P-value	Psq IP	P-value	Adf1 IP	P-value
NELF-complex									
NELF-B	-	19|26|26	0.009	-	-	0|0|1	0.423	-	-
NELF-D/TH1	-	4|10|7	0.056	-	-	-	-	-	-
NELF-A	-	37|58|48	0.016	0|1|0	0.423	-	-	-	-
NELF-E	-	21|22|18	0.003	-	-	-	-	-	-
DSIF-complex									
Spt4	-	2|0|1	0.225	-	-	-	-	1|1|1	0
Spt5 - not present									
DNA-binding factors									
M1BP	3|5|5	37|43|40	0.0007	5|5|9	0.27	5|10|11	0.15	4|10|10	0.2
Trl/GAF	9|0|0	32|22|27	0.005	10|8|17	0.1	58|58|51	0.0002	29|27|28	0.012
TFIID complex									
TAF1, KAT4	-	2|10|3	0.185	-	-	-	-	-	-
TAF2	-	-	-	-	-	-	-	-	-
TAF3	-	0|2|0	0.423	-	-	-	-	-	-
TAF4	-	11|21|21	0.034	-	-	-	-	1|1|0	0.1835
TAF5	-	0|8|4	0.225	-	-	-	-	-	-
TAF6	-	3|14|11	0.105	-	-	-	-	5|1|2	0.157
TAF7	-	41|25|33	0.019	0|0|3	0.423	1|8|6	0.138	3|4|7	0.06
TAF8	-	9|15|13	0.019	-	-	0|2|2	0.1835	2|6|4	0.074
TAF9/e(y)1	-	11|12|15	0.009	0|0|1	0.423	1|6|7	0.128	3|0|4	0.19
TAF10	-	5|3|4	0.02	-	-	-	-	-	-
TAF11	-	4|9|8	0.044	-	-	-	-	1|0|0	0.423
TAF12	-	11|8|9	0.009	-	-	1|1|1	0	-	-
TAF13	-	3|5|5	0.023	-	-	-	-	-	-
TBP	-	1|1|1	0.183	-	-	-	-	1|0|1	0.581
"super elongation complex" SEC									
Su(Tpl)	-	-	-	-	-	44|46|50	0.0014	-	-
Eaf	-	-	-	-	-	26|42|31	0.02	-	-
Ear, Lilli - not present									

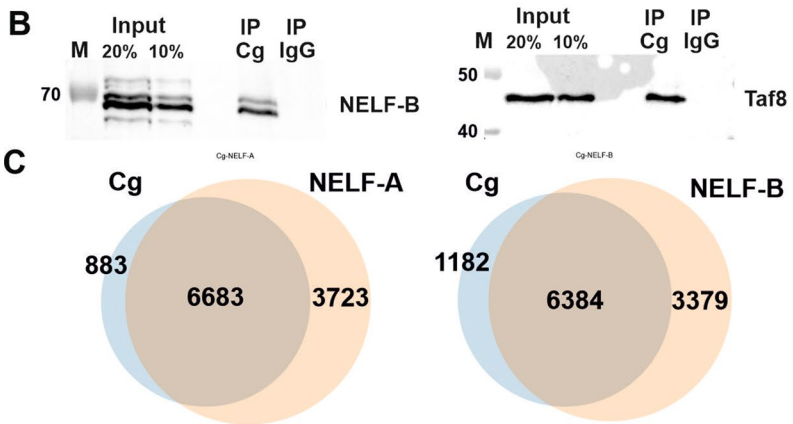


Fig. 7 Promoter and elongation proteins associated with Combgap, Zeste, Psq, and Adf1. **A** Spectral counts for the promoter and architectural proteins in IP/LC-MSs are presented. Designations are as in Figs. 3 and 5. **B** IP/WB analysis of interactions between Combgap and NELF-B, Taf8. **C**. Co-localization of Combgap peaks with the NELF-A and NELF-B proteins in S2 *Drosophila* cells

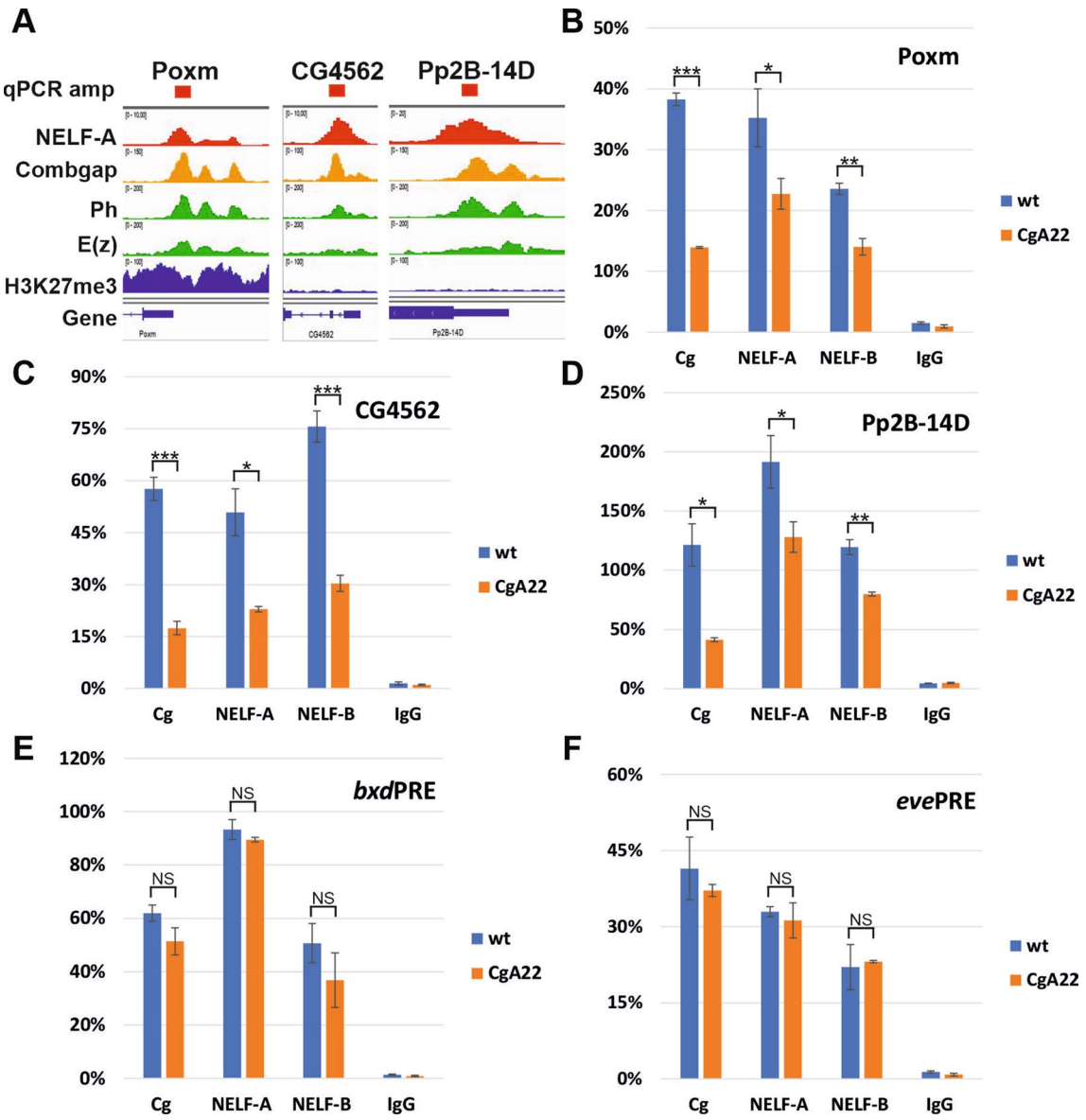


Fig. 8 Combgap is required for the NELF-A and NELF-B binding. **A** ChIP-seq profiles for NELF-A (wild-type whole third-instar larvae), Combgap, Ph, E(z), and H3K27me3 (wild-type third-instar larval imaginal disks and brains) at the promoter regions of *Poxm*, *CG4562*, and *Pp2B-14D* genes. **B**–**D** X-ChIP-qPCR analyses of the Combgap, NELF-A, and NELF-B recruitment in wild-type (wt) or cg^{A22} mutant third-instar larvae at the selected regions. **E**, **F** Two control regions, *bxd*PRE and *eve*PRE, largely not sensitive to the Combgap mutation. IgG—immunoprecipitation with the non-specific IgG. The ordinate shows the percentage of the target sequences in the immuno-precipitated material, relative to the 10% of input DNA. Vertical lines indicate SDs. * $P<0.05$, ** $P<0.01$, *** $P<0.001$

3 genome positions at which Ph was shown to be sensitive to the Combgap mutation: the promoter regions of the *Poxm*, *CG4562*, and *Pp2B-14D* genes (Fig. 8A). According to published ChIP-seq data [31, 44, 71], all three regions are co-bound by the NELF-A, by Ph PRC1 subunit and by the E(z) methyltransferase of the PRC2 complex. In addition, the *Poxm* promoter is covered by the H3K27me3 histone modification, indicating that this element can represent functional PRE. On the contrary, the promoters of the *CG4562* and *Pp2B-14D* genes are devoid of the H3K27me3 modification, suggesting that they are not under the PcG repression at the third-instar larval stage. As a control, we have selected two functional PREs, at which the binding of Ph was previously shown to be largely independent of Combgap: *bxd*PRE and *eve*PRE (NELF recruitment is shown in Supplementary File 8). The ChIP from the third-instar whole larvae followed by qPCR demonstrated that Combgap mutation affected NELF-A and NELF-B recruitment at the *Poxm*, *CG4562*, and *Pp2B-14D* genes (Fig. 8B–D), but not at the *bxd*PRE and *eve*PRE control regions (Fig. 8E, F). These data suggested

that Combgap is involved in the NELF recruitment to the chromatin. Moreover, NELF recruitment could be regulated by Combgap in both the H3K27me3^{+} and H3K27me3^{-} chromatin states.

The cyclin-dependent kinase positive transcription elongation factor b (P-TEFb) facilitates the pause release by phosphorylating Pol II, DSIF, and NELF, resulting in the dissociation of NELF from the elongation complex and the transition of DSIF from a negative elongation factor to a positive elongation factor. In its most active form, P-TEFb is part of the multi-subunit super-elongation complex (SEC), which facilitates the transition of Pol II from the paused state to productive elongation [103]. To clarify whether Combgap-NELF connection could reflect its role in elongation in addition to pausing, we further analyzed the presence of P-TEFb (Cdk9 and CycT) and SEC (Su(Tpl), Eaf, Ear and Lilli) subunits in its interactome. Both P-TEFb and SEC were not precipitated by Combgap pointing to its specific connection to NELF (Supplementary File 9). Interestingly, while Psq was not connected to promoter pausing and its interactome is devoid of P-TEFb (Supplementary File 9), we detected its strong association with two of the SEC complex subunits, Su(Tpl) and Eaf (Fig. 7A). Whether it reflected Psq participation in elongation, for example of the enhancer-associated transcripts, or pointed to a role independent of elongation for

A Architectural proteins associated with Combgap, Zeste, Psq, Adf1

Spectral count

Protein	IgG IP	Combgap IP	P-value	Zeste IP	P-value	Psq IP	P-value	Adf1 IP	P-value
CP190	16|24|16	**30|48|46**	**0.04**	9|22|29	0.85	**18|34|40**	**0.201**	-	-
Su(Hw)	3|6|2	8|16|12	0.049	-	-	-	-	13|12|8	0.022
Mod(mdg4)	-	**16|29|26**	**0.026**	2|0|0	0.423	0|14|15	0.184	**32|25|30**	**0.005**
Ibf1	-	**24|20|25**	**0.004**	0|0|10	0.423	**6|12|14**	**0.047**	**15|21|18**	**0.009**
Ibf2	-	8|8|10	0.006	0|3|4	0.192	5|9|7	0.026	8|8|11	0.012
Chro	16|17|19	**39|59|65**	**0.041**	6|27|26	0.766	26|50|50	0.089	20|21|17	0.256
BEAF-32	-	**14|16|16**	**0.002**	5|10|10	0.038	**12|14|16**	**0.007**	**14|14|15**	**0.0005**
Clamp	5|5|4	10|12|17	0.054	4|6|5	0.649	**18|15|15**	**0.004**	**28|18|27**	**0.024**
Pita	-	0|2|2	0.1835	-	-	-	-	2|2|1	0.037
ZIPIC	-	0|6|4	0.199	-	-	-	-	2|0|4	0.23
Dref	9|16|14	-	-	-	-	-	-	21|15|13	0.355

CTCF - not present

B PRE DNA binding proteins associated with Combgap, Zeste, Psq, Adf1

Spectral count

Protein	IgG IP	Combgap IP	P-value	Zeste IP	P-value	Psq IP	P-value	Adf1 IP	P-value
Combgap	5|9|6	**167|152|126**	**0.007**	4|19|14	0.327	6|15|15	0.21	**37|50|48**	**0.007**
Zeste	0|2|2	-	-	**131|98|100**	**0.009**	**17|29|32**	**0.031**	0|6|6	0.313
Psq	0|0|1	3|13|7	0.126	**44|38|39**	**0.0016**	**195|179|177**	**9E-04**	8|10|9	0.0007
Adf1	0|8|7	**30|28|27**	**0.0063**	-	-	-	-	**296|281|258**	**0.001**
Dsp1	2|2|6	0|10|8	0.487	**37|22|21**	**0.039**	**108|101|98**	**0.0001**	5|5|2	0.71
Pho	-	0|1|1	0.1835	-	-	0|2|3	0.199	2|0|0	0.423
Spps	-	-	-	-	-	-	-	1|0|0	0.423
Trl/GAF	9|0|0	**32|22|27**	**0.005**	10|8|17	0.1	**58|58|51**	**0.0002**	**29|27|28**	**0.012**

Phol, Grh not present

C Y2H analysis of Combgap, Zeste, Psq, Adf1 direct partners

Protein	BD (negative control)	BD-Combgap	BD-Zeste	BD-Psq	BD-Adf1	BD-Dsp1	Pho-BD	BD-Spps
AD (negative control)	-	-	-	-	-	-	-	-
AD-Combgap	-	+++	-	-	-	-	-	-
AD-Zeste	-	-	+++	-	-	-	-	-
AD-Psq	-	-	-	+++	-	-	-	-
AD-Adf1	-	+++	-	-	+++	-	-	-
AD-Dsp1	-	-	-	++	-	-	-	-
AD-Pho	-	-	-	-	-	-	+++	-
AD-Spps	-	-	-	-	-	-	++	+++

Fig. 9 Architectural and PRE DNA-binding proteins associated with Combgap, Zeste, Psq, and Adf1. **A** Architectural proteins associated with Combgap, Zeste, Psq, and Adf1. **B** PRE DNA-binding proteins associated with Combgap, Zeste, Psq, and Adf1. Spectral counts for the PRE DNA-binding proteins in IP/LC–MSs are presented. Designations are as in Figs. 3 and 5. **C** Y2H analysis of direct interactions between PRE DNA-binding proteins. Proteins of interest were fused with AD (activation domain of GAL4 factor) or BD (DNA-binding domain of GAL4 factor). Based on the extent of growth, the interactions were scored as strong (detected on 2nd day, "+++") or intermediate (detected on 4th day, "++")

the Su(Tpl) and Eaf factors, could be a subject of a further investigation.

Architectural proteins

Architectural proteins were initially identified as factors bound to the boundaries of chromatin domains and endowing these boundary elements with the ability to block enhancer–promoter contacts and to block spreading of the repressive chromatin [120, 121]. Further studies revealed that architectural proteins mediate interactions between distant genomic sequences and are responsible for the formation of the 3D chromatin structure [122]. Boundary proteins are frequently found juxtaposed next to the PRE elements [29, 123–125], and we previously reported that they can potentiate PREs activity [68]. In accordance, we have previously identified a Combgap binding motif as a characteristic of sites of the Su(Hw) indirect recruitment, allowing us to suggest that Combgap is a potential Su(Hw) recruiter at these sites [126]. These observations, together with our ChIP-seq data, indicate a potential interplay between these classes of regulatory elements.

We further analyzed interactions between Combgap, Zeste, Psq, Adf1 and the architectural proteins (Fig. 9A), and found that all, except for Zeste, demonstrate statistically significant strong interactions with the architectural proteins. In accordance with ChIP-seq data, Combgap more effectively precipitated the CP190 boundary protein. In addition, a statistically significant enrichment in the Combgap IP/LC–MS was observed for the Su(Hw), Mod(mdg4), Ibf1, Ibf2, Chro, and BEAF-32 architectural proteins. Psq most effectively precipitated Ibf1, BEAF-32, and Clamp; while Adf1–Mod(mdg4), Ibf1, BEAF-32 and Clamp. It should be noted that none of the tested proteins displayed interactions with the CTCF boundary protein, suggesting that specificity of interactions could exist between interacting sets of the PRE DNA-binding factors and the associated architectural proteins.

PRE DNA-binding proteins

The PRE DNA elements comprise the binding sites for different DNA-binding factors. One of the possible mechanisms for the formation of active PRE assumes the presence of multiple interactions between different DNA-binding proteins, which leads to the formation of a "platform" for the recruitment of the core PcG/TrxG complexes [19]. According to this model, the high binding affinity is achieved by simultaneous recruitment of a certain combination of DNA-binding factors to the PRE.

In accordance with this model, multiple PRE DNA-binding partners are present in the purified complexes (Fig. 9B). Antibodies against Combgap effectively precipitated Adf1 and GAF. Similarly, Adf1 interacted strongly with Combgap and GAF. Psq and Zeste interacted with each other and with the Dsp1 PRE DNA-binding protein. In addition, Psq associated with GAF.

At the next stage, using the yeast two-hybrid (Y2H) assay, we tested whether PRE DNA-binding proteins were able to form direct protein–protein contacts. For the Y2H experiments, the tested proteins were either fused to the GAL4 DNA-binding domain or to the GAL4 activation domain. All interactions were tested in direct and reciprocal experiments. Several findings were of interest. First, Combgap, Zeste, Psq, Adf1, Pho, and Spps were able to form self-contacts. Previously, self-interactions were demonstrated for Zeste [127], Psq [46], Adf1 [60], GAF [46, 128], and Grh [129] proteins. It was proposed that these factors can bind DNA as dimers or multimers, and display higher affinity for the DNA fragments containing several closely placed binding sites for the target proteins [60, 128, 130]. The results of the current study suggested that Pho and Spps may have similar di- or multimerization properties. Second, several new heterologous interactions were detected: Combgap interacted with Adf1, Psq with Dsp1, and Pho with Spps. Thus, different PRE DNA-binding factors are immuno-precipitated by each other and can establish direct contacts with each other supporting the model of their cooperative recruitment.

Discussion

Despite a long history of research on the *Drosophila* PRE associated proteins, the detailed whole-genome binding profiles and the interactomes of the majority of them remain mostly unexplored.

Here, we characterized whole-genome binding profiles and interactomes of the four PRE DNA-binding factors, Combgap, Psq, Zeste, and Adf1. We show that each of the proteins have its unique whole-genome binding profile. At the same time, the chromatin binding for all of them extends much beyond of PREs and they all co-localize with the chromatin peaks of proteins of different chromatin functions. The most obvious differences come from their interactome analysis. We show that each protein displays a different range of specificity in the interactions with proteins of various regulatory classes (Fig. 10).

We found that the Combgap displays the strongest degree of association with PcG factors. Combgap effectively purifies with the PRC1 complex subunits Sce, Ph, Pc, and Su(z)2. The Zeste protein is also strongly associated with Ph thus connecting it to PRC1. At the same time, among the

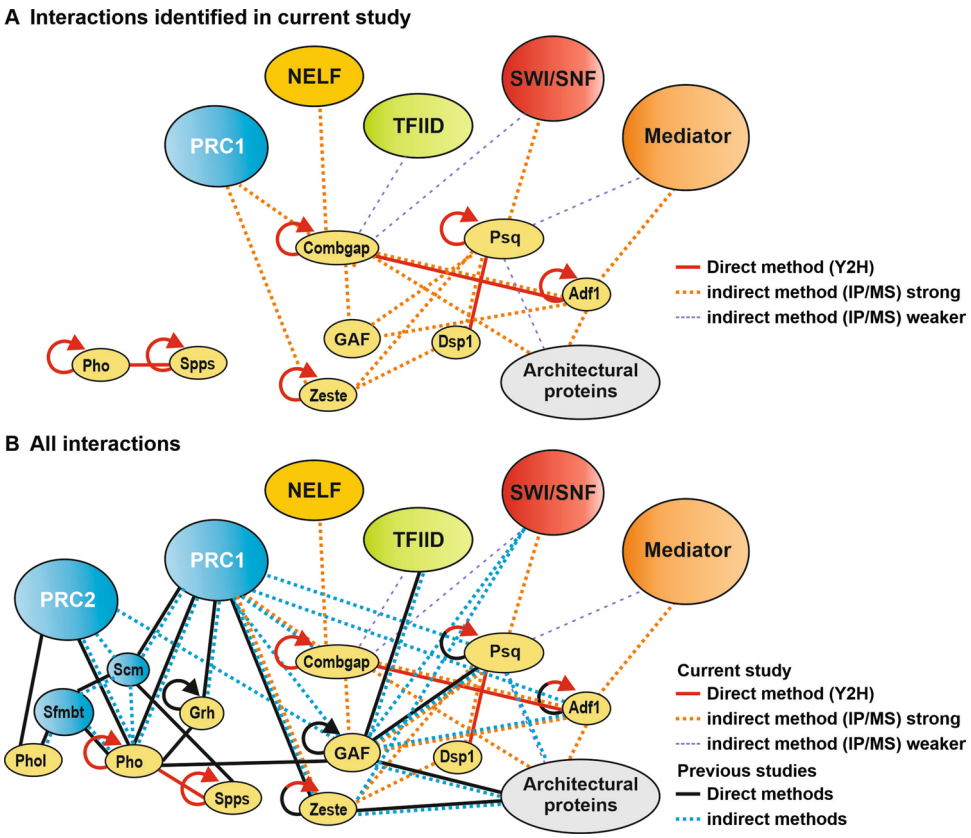


Fig. 10 PRE DNA-binding factors physical partners. A Summary of interactions identified in the current study. Contacts established by indirect methods are indicated by the dotted orange lines (strong interaction) or purple (weak interaction); direct partners are indicated by the solid red lines. Self-interactions are indicated by a circular arrow. B Summary of interactions identified in the current and earlier studies. Contacts established in current study are indicated as in (A). Contacts established in earlier studies by indirect methods are indicated by the dotted blue lines; direct partners are indicated by the solid black lines. References to direct interactions: Phol-PRC2, Pho-PRC2, Phol-Sfmbt, Pho-Sfmbt, Sfmbt-Scm, Pho-PRC1, Pho-Grh, GAF-Psq, Grh-PRC1, Zeste-PRC1, Scm-PRC1 (see refs in [19]); Scm-Spps, Pho-GAF [45]; Zeste-architectural protein (mod(mdg4)) [134]; GAF-architectural protein (mod(mdg4)), GAF-TFIID (TAF3, TAF4) (see refs in [52]). Self-interactions (indicated by self-pointing arrow): Zeste [127], Psq [46], Adf1 [60], GAF [46, 128], and Grh [129]. Other interactions were uncovered in the current study. References to interactions obtained by indirect methods: Phol-Sfmbt, Pho-Sfmbt [40]; Zeste-SWI/SNF [108]; Zeste-architectural proteins (mod(mdg4), Su(Hw)) [134]; Sfmbt-Scm, Scm-PRC1, Scm-PRC2, Scm-Pho [135]; Adf1-PRC1 [35]; Psq-architectural protein (mod(mdg4)) [57]; Psq-PRC1, Combgap-PRC1, Grh-PRC1, Zeste-PRC1, Pho-PRC2, Pho-PRC1 (see refs in [19]); GAF-TFIID (TAF3, TAF4), GAF-Psq, GAF-PRC1, GAF-PRC2, GAF-SWI/SNF, GAF-Adf1, GAF-architectural proteins (mod(mdg4), E(y)2, CP190, Su(Hw), Chro) (see refs in [52]). Other interactions were uncovered in the current study.

PcG/TrxG members, the highest degree of association Zeste displays with Bre1 TrxG factor.

In the current study, we did not identify stable association of core PcG complexes with Psq and Adf1. Previously, Psq was identified as a subunit of CRASCH complex purified from the S2 cells expressing FLAG-tagged Pc protein [33], at the same time, it was not detected in the Pc-Bio experiments [49] supporting that Psq–PRC1 contacts are unstable and depend on conditions used for the complex purification. Similarly, Adf1 was previously co-purified with Pc from the MNase-digested chromatin in S2R+ cells [35], suggesting differences in the co-purification conditions.

At the same time, while the Psq interactome is devoid of the core PcG complexes, it interacts with the PcG factors Sxc, Dom, and displays stable interactions with TrxG proteins, including the BAP chromatin-remodeling complex.

Besides the PcG/TrxG factors, the Combgap TOP interactors are AGO2, M1BP, and the subunits of NELF and TFIID, indicating potential involvement of Combgap in the promoter pausing function. In agreement with this, the Combgap protein was found at gene promoters [31, 44]. Here,

we estimated the whole-genome co-localization between Combgap and NELF and demonstrated that the majority of Combgap peaks (over 80%) overlap with the NELF-A and NELF-B subunits. Next, we explored the role of Combgap in recruitment of NELF-A and NELF-B pausing subunits on several model genes. We demonstrated that Combgap is required for an effective NELF recruitment to the promoter of the *Poxm* gene, which is marked by H3K27me3 histone modification, and to promoters of two other genes, *CG4562* and *Pp2B-14D*, which lack the repressive chromatin mark. Thus, Combgap could be important for the NELF recruitment at both H3K27me3^{+} and H3K27me3^{-} chromatin regions. A further analysis is required to estimate the impact of Combgap on the NELF recruitment genome-wide.

Except Combgap in present study, the connection of PcG proteins to promoter pausing was previously suggested for PRC1 and Pho. The PRC1 subunits were shown to be present at many active gene promoters that lack H3K27me3 [22, 24] and the depletion of PRC1 subunits by RNA interference altered phosphorylation of RNAP II and recruitment of Spt5 subunit of DSIF complex at the active genes [23]. Pho, which can directly interact with Spt5 [131, 132] was indicated as a possible connector of PRC1 with the pausing factors. Specifically, Pho, together with PRC1 (Ph), was shown to regulate heat shock response at the *hsp70* locus [131]. Due to an extensive overlap of Pho with Spt5 and NELF across the genome (up to 70%) including TSS regions [132], the participation of Pho in pausing is expected not to be limited to *hsp70* locus.

Furthermore, both Combgap and Pho can potentially target the pausing proteins to subset of PREs/enhancers. In support of this, NELF (present study and [71]) and Spt5 [23, 133] have been detected at PREs and enhancers. Moreover, the *spt5* and *nelf-A* mutants enhance the polycomb phenotype observed in *pho* mutant flies, suggesting a direct role of the pausing factors in PcG silencing [132].

Here, we found that Adf1 interactome was the most dissimilar from other tested PRE DNA-binding proteins and it was very tightly connected to the Mediator complex. The top Adf1-interacting factors are mainly represented by the subunits of «head», «body», and «tail» modules of Mediator. This suggests that the Adf1 main function could be connected to the recruitment of the Mediator complex to chromatin.

Not less intriguing was the presence of the architectural proteins in the PRE DNA-binding factor interactomes. While the boundary proteins are not directly recruited to PREs, both elements are frequently located in relatively close proximity to each other [29, 123–125]. Moreover, we have recently shown that closely placed boundary sites can initiate PRE-mediated silencing by providing recruitment of PRE DNA-binding and PcG/TrxG proteins [68]. A stimulation of PRE by nearby boundary is accompanied by alterations in the local chromatin structure, lowering the nucleosome density, and is dependent on the boundary–boundary chromatin contacts between the homologous chromosomes. Here, we show that Combgap precipitates CP190, Su(Hw), Mod(mdg4), Ibf1, Ibf2, Chro, and BEAF-32 architectural proteins. The Psq preferential interactors are Ibf1, BEAF-32, and Clamp; while Adf1 most efficiently precipitates Mod(mdg4), Ibf1, BEAF-32, and Clamp. While these interactions could be passive due to frequent juxtaposition of the regulatory elements, they could potentially reflect a direct role of the architectural proteins in recruitment of PcG/TrxG factors to PREs.

While differing significantly, the Combgap, Zeste, Psq, and Adf1 interactomes demonstrate certain similarity in the way that they can co-purify other PRE DNA-binding factors. The strongest contacts were observed between Combgap and Adf1/GAF, between Zeste and Psq/Dsp1, between Psq and GAF. Moreover, we demonstrated that Combgap directly interacts with Adf1, while Psq with Dsp1. At the same time, we have not observed stable contacts of the Combgap–Zeste–Psq–Adf1 group of proteins with the Pho–Spps–Phol–Grh group. Accordingly, GAF also seemed to be recruited independently of Pho, as we were able to detect only a background amount of Pho when we purified the GAF complex [50]. In addition, the Pho–Spps–Phol–Grh seemed to be more tightly linked together. In support of this idea, in the current study, we found that Spps directly interacted with Pho, while the Pho–Grh interactions were reported previously [38]. Moreover, genome regions showing depletion of H3K27me3 after Spps knockout are frequently accompanied by a decreased Pho binding, suggesting their cooperative recruitment [44].

Intriguingly, besides the known PRE DNA-binding factors, the Combgap, Psq, Zeste, and Adf1 TOP20 interactors included proteins with known or predicted DNA-binding activities that can potentially co-operate with them on different regulatory elements. We found that Combgap precipitated CG5953, Crp, Lola, Stwl, Ken, M1BP; Zeste—Crp, Lola and Prod; Psq—CG10543, Ct and Tai; Adf1—Chinmo. Probably, they can assist in providing the specificity in the selection of the complex that will be targeted to DNA.

In summary, our data suggest strong physical connections of all four tested proteins, Combgap, Zeste, Psq, and Adf1 PRE DNA-binding, to different regulatory elements. Their ability to co-bind each other can potentially create a platform for recruitment of distinct regulatory complexes with a unique degree of specificity of each factor for a specific regulatory chromatin complex. The presence of other DNA-binding partners and combinations of sites for distinct DNA-binding proteins at a particular DNA

regulatory element, can determine which specific regulatory complex will be actually targeted to DNA.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00018-022-04383-2.

Acknowledgements Experiments of LC–MS fingerprinting were carried out using the equipment provided by the Institute of Bioorganic Chemistry core facility (CKP IBCH). We thank the Center for Precision Genome Editing and Genetic Technologies for Biomedicine, IGB RAS for the use of their equipment. We thank Judith Kassis for providing the Cg^{A22} fly line.

Author contributions Conception and design of the project: DC, PG, ME. Plasmid cloning—DC, NEV, LVF, DL, AG, MYM, VM, ME. Antigen expression, antibodies production, purification and verification—DC, NEV, LVF, DL, MYM, ME. ChIP-seqs and X-ChIP—MYM, NEV, ME. Purification and analysis of protein complexes—ME, LVF, DC. LC–MS—RHZ. Interactome data analysis—DC and ME. IP/Western blotting—NEV, MYM, LVF, AG, ME. Y2H assay—DC, LVF, DL, AG. Data visualization—DC, NEV, MYM, ME. The initial draft of this manuscript was written by—DC and ME, reviewed and edited by—DC, NEV, MYM, VM, PG, RHZ, ME. All authors read and approved the final manuscript.

Funding The work was supported by the Russian Science Foundation (RSF) Grant no.18-74-10091.

Data availability All relevant data are within the paper and its Supporting Information. The ChIP-seq data generated for this work are accessible through the GEO Series accession number GSE200213. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [96] partner repository with the dataset identifier PXD029459 and https://doi.org/10.6019/PXD029459.

Declarations

Conflict of interest The authors declare no competing interests.

Ethical approval Animal handling for the antibody production was carried out strictly according to the procedures outlined in the NIH (USA) Guide for the Care and Use of Laboratory Animals. The protocols used were approved by the Committee on Bioethics of the Institute of Gene Biology, Russian Academy of Sciences. All procedures were performed under the supervision of a licensed veterinarian, under conditions that minimize pain and distress. Rabbits were purchased from a licensed specialized nursery, Manihino. Soviet chinchilla rabbits used in the study are not endangered or protected. Only healthy rabbits, certified by a licensed veterinarian were used. The rabbits were individually housed in standard size, stainless steel rabbit cages, and provided an ad libitum access to alfalfa hay, commercial rabbit food pellets, and water. The appetite and behavior of each rabbit was monitored daily by a licensed veterinarian. Body weight and temperature of each rabbit were evaluated prior to and daily following the immunization. No animals became ill or died at any time prior to the experimental endpoint. At the end of the study period all rabbits were euthanized by intravenous injection of barbiturate anesthetics.

Consent to participate Not applicable.

Consent to publish Not applicable.

References

1. Chetverina DA, Elizar'ev PV, Lomaev DV, Georgiev PG, Erokhin MM (2017) Control of the gene activity by polycomb and trithorax group proteins in Drosophila. Genetika 53(2):133–154
2. Chetverina DA, Lomaev DV, Erokhin MM (2020) Polycomb and Trithorax group proteins: the long road from mutations in Drosophila to use in medicine. Acta Naturae 12(4):66–85. https://doi.org/10.32607/actanaturae.11090
3. Kassis JA, Kennison JA, Tamkun JW (2017) Polycomb and Trithorax group genes in Drosophila. Genetics 206(4):1699–1725. https://doi.org/10.1534/genetics.115.185116
4. Kuroda MI, Kang H, De S, Kassis JA (2020) Dynamic competition of Polycomb and Trithorax in transcriptional programming. Annu Rev Biochem. https://doi.org/10.1146/annurev-biochem-120219-103641
5. Schuettengruber B, Bourbon HM, Di Croce L, Cavalli G (2017) Genome regulation by Polycomb and Trithorax: 70 years and counting. Cell 171(1):34–57. https://doi.org/10.1016/j.cell.2017.08.002
6. Chetverina DA, Lomaev DV, Georgiev PG, Erokhin MM (2021) Genetic impairments of PRC2 activity in oncology: problems and prospects. Russ J Genet 57(3):258–272. https://doi.org/10.1134/S1022795421030042
7. Erokhin M, Chetverina O, Gyorffy B, Tatarskiy VV, Mogila V, Shtil AA, Roninson IB, Moreaux J, Georgiev P, Cavalli G, Chetverina D (2021) Clinical correlations of Polycomb repressive complex 2 in different tumor types. Cancers. https://doi.org/10.3390/cancers13133155
8. Fagan RJ, Dingwall AK (2019) COMPASS Ascending: Emerging clues regarding the roles of MLL3/KMT2C and MLL2/KMT2D proteins in cancer. Cancer Lett 458:56–65. https://doi.org/10.1016/j.canlet.2019.05.024
9. Mittal P, Roberts CWM (2020) The SWI/SNF complex in cancer—biology, biomarkers and therapy. Nat Rev Clin Oncol 17(7):435–448. https://doi.org/10.1038/s41571-020-0357-3
10. Piunti A, Shilatifard A (2021) The roles of Polycomb repressive complexes in mammalian development and cancer. Nat Rev Mol Cell Biol 22(5):326–345. https://doi.org/10.1038/s41580-021-00341-1
11. Francis NJ, Saurin AJ, Shao Z, Kingston RE (2001) Reconstitution of a functional core polycomb repressive complex. Mol Cell 8(3):545–556
12. Saurin AJ, Shao Z, Erdjument-Bromage H, Tempst P, Kingston RE (2001) A Drosophila Polycomb group complex includes Zeste and dTAFII proteins. Nature 412(6847):655–660. https://doi.org/10.1038/35088096
13. Shao Z, Raible F, Mollaaghababa R, Guyon JR, Wu CT, Bender W, Kingston RE (1999) Stabilization of chromatin structure by PRC1, a Polycomb complex. Cell 98(1):37–46. https://doi.org/10.1016/S0092-8674(00)80604-2
14. Czermin B, Melfi R, McCabe D, Seitz V, Imhof A, Pirrotta V (2002) Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell 111(2):185–196
15. Muller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, Wild B, Miller EL, O'Connor MB, Kingston RE, Simon JA (2002) Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell 111(2):197–208
16. Schwartz YB, Kahn TG, Nix DA, Li XY, Bourgon R, Biggin M, Pirrotta V (2006) Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. Nat Genet 38(6):700–705. https://doi.org/10.1038/ng1817

17. Tolhuis B, de Wit E, Muijrers I, Teunissen H, Talhout W, van Steensel B, van Lohuizen M (2006) Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in *Drosophila melanogaster*. Nat Genet 38(6):694–699. https://doi.org/10.1038/ng1792
18. Geisler SJ, Paro R (2015) Trithorax and Polycomb group-dependent regulation: a tale of opposing activities. Development 142(17):2876–2887. https://doi.org/10.1242/dev.120030
19. Erokhin M, Georgiev P, Chetverina D (2018) Drosophila DNA-binding proteins in Polycomb repression. Epigenomes 2(1):1. https://doi.org/10.3390/epigenomes2010001
20. Kassis JA, Brown JL (2013) Polycomb group response elements in Drosophila and vertebrates. Adv Genet 81:83–118. https://doi.org/10.1016/B978-0-12-407677-8.00003-8
21. Bauer M, Trupke J, Ringrose L (2015) The quest for mammalian Polycomb response elements: are we there yet? Chromosoma. https://doi.org/10.1007/s00412-015-0539-4
22. Loubiere V, Delest A, Thomas A, Bonev B, Schuettengruber B, Sati S, Martinez AM, Cavalli G (2016) Coordinate redeployment of PRC1 proteins suppresses tumor formation during Drosophila development. Nat Genet 48(11):1436–1442. https://doi.org/10.1038/ng.3671
23. Pherson M, Misulovin Z, Gause M, Mihindukulasuriya K, Swain A, Dorsett D (2017) Polycomb repressive complex 1 modifies transcription of active genes. Sci Adv 3(8):e1700944. https://doi.org/10.1126/sciadv.1700944
24. Schaaf CA, Misulovin Z, Gause M, Koenig A, Gohara DW, Watson A, Dorsett D (2013) Cohesin and polycomb proteins functionally interact to control transcription at silenced and active genes. PLoS Genet 9(6):e1003560. https://doi.org/10.1371/journal.pgen.1003560
25. Erceg J, Pakozdi T, Marco-Ferreres R, Ghavi-Helm Y, Girardot C, Bracken AP, Furlong EE (2017) Dual functionality of cis-regulatory elements as developmental enhancers and Polycomb response elements. Genes Dev 31(6):590–602. https://doi.org/10.1101/gad.292870.116
26. Brown JL, Mucci D, Whiteley M, Dirksen ML, Kassis JA (1998) The Drosophila Polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1. Mol Cell 1(7):1057–1064
27. Fritsch C, Brown JL, Kassis JA, Muller J (1999) The DNA-binding polycomb group protein pleiohomeotic mediates silencing of a Drosophila homeotic gene. Development 126(17):3905–3913
28. Brown JL, Fritsch C, Mueller J, Kassis JA (2003) The Drosophila pho-like gene encodes a YY1-related DNA binding protein that is redundant with pleiohomeotic in homeotic gene silencing. Development 130(2):285–294
29. Hagstrom K, Muller M, Schedl P (1997) A Polycomb and GAGA dependent silencer adjoins the Fab-7 boundary in the Drosophila bithorax complex. Genetics 146(4):1365–1380
30. Strutt H, Cavalli G, Paro R (1997) Co-localization of Polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression. EMBO J 16(12):3621–3632. https://doi.org/10.1093/emboj/16.12.3621
31. Ray P, De S, Mitra A, Bezstarosti K, Demmers JA, Pfeifer K, Kassis JA (2016) Combgap contributes to recruitment of Polycomb group proteins in Drosophila. Proc Natl Acad Sci USA 113(14):3826–3831. https://doi.org/10.1073/pnas.1520926113
32. Hodgson JW, Argiropoulos B, Brock HW (2001) Site-specific recognition of a 70-base-pair element containing d(GA)(n) repeats mediates bithoraxoid polycomb group response element-dependent silencing. Mol Cell Biol 21(14):4528–4543. https://doi.org/10.1128/MCB.21.14.4528-4543.2001
33. Huang DH, Chang YL, Yang CC, Pan IC, King B (2002) pipsqueak encodes a factor essential for sequence-specific targeting of a polycomb group protein complex. Mol Cell Biol 22(17):6261–6271. https://doi.org/10.1128/mcb.22.17.6261-6271.2002
34. Hur MW, Laney JD, Jeon SH, Ali J, Biggin MD (2002) Zeste maintains repression of Ubx transgenes: support for a new model of Polycomb repression. Development 129(6):1339–1343
35. Orsi GA, Kasinathan S, Hughes KT, Saminadin-Peter S, Henikoff S, Ahmad K (2014) High-resolution mapping defines the cooperative architecture of Polycomb response elements. Genome Res 24(5):809–820. https://doi.org/10.1101/gr.163642.113
36. Dejardin J, Rappailles A, Cuvier O, Grimaud C, Decoville M, Locker D, Cavalli G (2005) Recruitment of Drosophila Polycomb group proteins to chromatin by DSP1. Nature 434(7032):533–538. https://doi.org/10.1038/nature03386
37. Brown JL, Kassis JA (2010) Spps, a Drosophila Sp1/KLF family member, binds to PREs and is required for PRE activity late in development. Development 137(15):2597–2602. https://doi.org/10.1242/dev.047761
38. Blastyak A, Mishra RK, Karch F, Gyurkovics H (2006) Efficient and specific targeting of Polycomb group proteins requires cooperative interaction between Grainyhead and Pleiohomeotic. Mol Cell Biol 26(4):1434–1444. https://doi.org/10.1128/MCB.26.4.1434-1444.2006
39. Gehring WJ (1970) A recessive lethal (l(4)29) with a homeotic effect in *D. melanogaster*. Dros Inform Serv 45:103
40. Klymenko T, Papp B, Fischle W, Kocher T, Schelder M, Fritsch C, Wild B, Wilm M, Muller J (2006) A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. Genes Dev 20(9):1110–1122. https://doi.org/10.1101/gad.377406
41. Mohd-Sarip A, Venturini F, Chalkley GE, Verrijzer CP (2002) Pleiohomeotic can link polycomb to DNA and mediate transcriptional repression. Mol Cell Biol 22(21):7473–7483. https://doi.org/10.1128/MCB.22.21.7473-7483.2002
42. Wang L, Brown JL, Cao R, Zhang Y, Kassis JA, Jones RS (2004) Hierarchical recruitment of polycomb group silencing complexes. Mol Cell 14(5):637–646. https://doi.org/10.1016/j.molcel.2004.05.009
43. Frey F, Sheahan T, Finkl K, Stoehr G, Mann M, Benda C, Muller J (2016) Molecular basis of PRC1 targeting to Polycomb response elements by PhoRC. Genes Dev 30(9):1116–1127. https://doi.org/10.1101/gad.279141.116
44. Brown JL, Sun MA, Kassis JA (2018) Global changes of H3K27me3 domains and Polycomb group protein distribution in the absence of recruiters Spps or Pho. Proc Natl Acad Sci USA 115(8):E1839–E1848. https://doi.org/10.1073/pnas.1716299115
45. Shokri L, Inukai S, Hafner A, Weinand K, Hens K, Vedenko A, Gisselbrecht SS, Dainese R, Bischof J, Furger E, Feuz JD, Basler K, Deplancke B, Bulyk ML (2019) A comprehensive *Drosophila melanogaster* transcription factor interactome. Cell Rep 27(3):955-970 e957. https://doi.org/10.1016/j.celrep.2019.03.071
46. Bonchuk A, Denisov S, Georgiev P, Maksimenko O (2011) Drosophila BTB/POZ domains of "ttk group" can form multimers and selectively interact with each other. J Mol Biol 412(3):423–436. https://doi.org/10.1016/j.jmb.2011.07.052
47. Schwendemann A, Lehmann M (2002) Pipsqueak and GAGA factor act in concert as partners at homeotic and many other loci. Proc Natl Acad Sci USA 99(20):12883–12888. https://doi.org/10.1073/pnas.202341499
48. Huang DH, Chang YL (2004) Isolation and characterization of CHRASCH, a polycomb-containing silencing complex. Methods Enzymol 377:267–282. https://doi.org/10.1016/S0076-6879(03)77016-5
49. Strubbe G, Popp C, Schmidt A, Pauli A, Ringrose L, Beisel C, Paro R (2011) Polycomb purification by in vivo biotinylation

tagging reveals cohesin and Trithorax group proteins as interaction partners. Proc Natl Acad Sci USA 108(14):5572–5577. https://doi.org/10.1073/pnas.1007916108
50. Lomaev D, Mikhailova A, Erokhin M, Shaposhnikov AV, Moresco JJ, Blokhina T, Wolle D, Aoki T, Ryabykh V, Yates JR 3rd, Shidlovskii YV, Georgiev P, Schedl P, Chetverina D (2017) The GAGA factor regulatory network: identification of GAGA factor associated proteins. PLoS ONE 12(3):e0173602. https://doi.org/10.1371/journal.pone.0173602
51. Poux S, Melfi R, Pirrotta V (2001) Establishment of Polycomb silencing requires a transient interaction between PC and ESC. Genes Dev 15(19):2509–2514. https://doi.org/10.1101/gad.208901
52. Chetverina D, Erokhin M, Schedl P (2021) GAGA factor: a multifunctional pioneering chromatin protein. Cell Mol Life Sci. https://doi.org/10.1007/s00018-021-03776-z
53. Gaskill MM, Gibson TJ, Larson ED, Harrison MM (2021) GAF is essential for zygotic genome activation and chromatin accessibility in the early Drosophila embryo. Elife. https://doi.org/10.7554/eLife.66668
54. Judd J, Duarte FM, Lis JT (2021) Pioneer-like factor GAF cooperates with PBAP (SWI/SNF) and NURF (ISWI) to regulate transcription. Genes Dev 35(1–2):147–156. https://doi.org/10.1101/gad.341768.120
55. Nakayama T, Shimojima T, Hirose S (2012) The PBAP remodeling complex is required for histone H3.3 replacement at chromatin boundaries and for boundary functions. Development 139(24):4582–4590. https://doi.org/10.1242/dev.083246
56. Shimojima T, Okada M, Nakayama T, Ueda H, Okawa K, Iwamatsu A, Handa H, Hirose S (2003) Drosophila FACT contributes to Hox gene expression through physical and functional interactions with GAGA factor. Genes Dev 17(13):1605–1616. https://doi.org/10.1101/gad.1086803
57. Gutierrez-Perez I, Rowley MJ, Lyu X, Valadez-Graham V, Vallejo DM, Ballesta-Illan E, Lopez-Atalaya JP, Kremsky I, Caparros E, Corces VG, Dominguez M (2019) Ecdysone-induced 3D chromatin reorganization involves active enhancers bound by pipsqueak and Polycomb. Cell Rep 28(10):2715-2727 e2715. https://doi.org/10.1016/j.celrep.2019.07.096
58. Schuettengruber B, Ganapathi M, Leblanc B, Portoso M, Jaschek R, Tolhuis B, van Lohuizen M, Tanay A, Cavalli G (2009) Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. PLoS Biol 7(1):e13. https://doi.org/10.1371/journal.pbio.1000013
59. Lehmann M, Siegmund T, Lintermann KG, Korge G (1998) The pipsqueak protein of *Drosophila melanogaster* binds to GAGA sequences through a novel DNA-binding domain. J Biol Chem 273(43):28504–28509. https://doi.org/10.1074/jbc.273.43.28504
60. Cutler G, Perry KM, Tjian R (1998) Adf-1 is a nonmodular transcription factor that contains a TAF-binding Myb-like motif. Mol Cell Biol 18(4):2252–2261. https://doi.org/10.1128/MCB.18.4.2252
61. Dejardin J, Cavalli G (2004) Chromatin inheritance upon Zeste-mediated Brahma recruitment at a minimal cellular memory module. EMBO J 23(4):857–868. https://doi.org/10.1038/sj.emboj.7600108
62. Horard B, Tatout C, Poux S, Pirrotta V (2000) Structure of a polycomb response element and in vitro binding of polycomb group complexes containing GAGA factor. Mol Cell Biol 20(9):3187–3197
63. Benson M, Pirrotta V (1988) The Drosophila zeste protein binds cooperatively to sites in many gene regulatory regions: implications for transvection and gene regulation. EMBO J 7(12):3907–3915
64. Pirrotta V, Bickel S, Mariani C (1988) Developmental expression of the Drosophila zeste gene and localization of zeste protein on polytene chromosomes. Genes Dev 2(12B):1839–1850. https://doi.org/10.1101/gad.2.12b.1839
65. Laney JD, Biggin MD (1992) zeste, a nonessential gene, potently activates Ultrabithorax transcription in the Drosophila embryo. Genes Dev 6(8):1531–1541. https://doi.org/10.1101/gad.6.8.1531
66. England BP, Admon A, Tjian R (1992) Cloning of Drosophila transcription factor Adf-1 reveals homology to Myb oncoproteins. Proc Natl Acad Sci USA 89(2):683–687. https://doi.org/10.1073/pnas.89.2.683
67. Heberlein U, England B, Tjian R (1985) Characterization of Drosophila transcription factors that activate the tandem promoters of the alcohol dehydrogenase gene. Cell 41(3):965–977. https://doi.org/10.1016/s0092-8674(85)80077-5
68. Erokhin M, Gorbenko F, Lomaev D, Mazina MY, Mikhailova A, Garaev AK, Parshikov A, Vorobyeva NE, Georgiev P, Schedl P, Chetverina D (2021) Boundaries potentiate polycomb response element-mediated silencing. BMC Biol 19(1):113. https://doi.org/10.1186/s12915-021-01047-8
69. Erokhin M, Davydova A, Parshikov A, Studitsky VM, Georgiev P, Chetverina D (2013) Transcription through enhancers suppresses their activity in Drosophila. Epigenet Chromatin 6(1):31. https://doi.org/10.1186/1756-8935-6-31
70. Erokhin M, Elizar'ev P, Parshikov A, Schedl P, Georgiev P, Chetverina D (2015) Transcriptional read-through is not sufficient to induce an epigenetic switch in the silencing activity of Polycomb response elements. Proc Natl Acad Sci USA 112(48):14930–14935. https://doi.org/10.1073/pnas.1515276112
71. Mazina MY, Kovalenko EV, Vorobyeva NE (2021) The negative elongation factor NELF promotes induced transcriptional response of Drosophila ecdysone-dependent genes. Sci Rep 11(1):172. https://doi.org/10.1038/s41598-020-80650-1
72. Mazina MY, Ziganshin RH, Magnitov MD, Golovnin AK, Vorobyeva NE (2020) Proximity-dependent biotin labelling reveals CP190 as an EcR/Usp molecular partner. Sci Rep 10(1):4793. https://doi.org/10.1038/s41598-020-61514-0
73. Mazina MY, Kovalenko EV, Derevyanko PK, Nikolenko JV, Krasnov AN, Vorobyeva NE (2018) One signal stimulates different transcriptional activation mechanisms. Biochim Biophys Acta Gene Regul Mech 2:178–189. https://doi.org/10.1016/j.bbagrm.2018.01.016
74. Crosby MA, Miller C, Alon T, Watson KL, Verrijzer CP, Goldman-Levi R, Zak NB (1999) The trithorax group gene moira encodes a brahma-associated putative chromatin-remodeling factor in *Drosophila melanogaster*. Mol Cell Biol 19(2):1159–1170. https://doi.org/10.1128/MCB.19.2.1159
75. Mazina MY, Kocheryzhkina EV, Derevyanko PK, Vorobyeva NE (2016) The composition of Swi/Snf chromatin remodeling complex is stable during gene transcription. Tsitologiia 58(4):285–289
76. Mazina MY, Nikolenko YV, Krasnov AN, Vorobyeva NE (2016) SWI/SNF protein complexes participate in the initiation and elongation stages of Drosophila hsp70 gene transcription. Genetika 52(2):164–169
77. Vorobyeva NE, Mazina MU, Golovnin AK, Kopytova DV, Gurskiy DY, Nabirochkina EN, Georgieva SG, Georgiev PG, Krasnov AN (2013) Insulator protein Su(Hw) recruits SAGA and Brahma complexes and constitutes part of origin recognition complex-binding sites in the Drosophila genome. Nucl Acids Res 41(11):5717–5730. https://doi.org/10.1093/nar/gkt297
78. Lebedeva LA, Nabirochkina EN, Kurshakova MM, Robert F, Krasnov AN, Evgen'ev MB, Kadonaga JT, Georgieva SG, Tora L (2005) Occupancy of the Drosophila hsp70 promoter by a subset of basal transcription factors diminishes upon transcriptional

activation. Proc Natl Acad Sci USA 102(50):18087–18092. https://doi.org/10.1073/pnas.0509063102
79. Kovalenko EV, Mazina MY, Krasnov AN, Vorobyeva NE (2019) The Drosophila nuclear receptors EcR and ERR jointly regulate the expression of genes involved in carbohydrate metabolism. Insect Biochem Mol Biol 112:103184. https://doi.org/10.1016/j.ibmb.2019.103184
80. Mazina M, Vorob'eva NE, Krasnov AN (2013) Ability of Su(Hw) to create a platform for ORC binding does not depend on the type of surrounding chromatin. Tsitologiia 55(4):218–224
81. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9(4):357–359. https://doi.org/10.1038/nmeth.1923
82. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F, Manke T (2016) deepTools2: a next generation web server for deep-sequencing data analysis. Nucl Acids Res 44(W1):W160-165. https://doi.org/10.1093/nar/gkw257
83. Feng J, Liu T, Qin B, Zhang Y, Liu XS (2012) Identifying ChIP-seq enrichment using MACS. Nat Protoc 7(9):1728–1740. https://doi.org/10.1038/nprot.2012.101
84. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Gruning BA, Guerler A, Hillman-Jackson J, Hiltemann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D (2018) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucl Acids Res 46(W1):W537–W544. https://doi.org/10.1093/nar/gky379
85. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. Nucl Acids Res 37(Web Server issue):W202–W208. https://doi.org/10.1093/nar/gkp335
86. Herz HM, Mohan M, Garruss AS, Liang K, Takahashi YH, Mickey K, Voets O, Verrijzer CP, Shilatifard A (2012) Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. Genes Dev 26(23):2604–2620. https://doi.org/10.1101/gad.201327.112
87. Enderle D, Beisel C, Stadler MB, Gerstung M, Athri P, Paro R (2011) Polycomb preferentially targets stalled promoters of coding and noncoding transcripts. Genome Res 21(2):216–226. https://doi.org/10.1101/gr.114348.110
88. Wani AH, Boettiger AN, Schorderet P, Ergun A, Munger C, Sadreyev RI, Zhuang X, Kingston RE, Francis NJ (2016) Chromatin topology is coupled to Polycomb group protein subnuclear organization. Nat Commun 7:10291. https://doi.org/10.1038/ncomms10291
89. Pherson M, Misulovin Z, Gause M, Dorsett D (2019) Cohesin occupancy and composition at enhancers and promoters are linked to DNA replication origin proximity in Drosophila. Genome Res 29(4):602–612. https://doi.org/10.1101/gr.243832.118
90. Bergman CM, Carlson JW, Celniker SE (2005) Drosophila DNase I footprint database: a systematic genome annotation of transcription factor binding sites in the fruitfly, *Drosophila melanogaster*. Bioinformatics 21(8):1747–1749. https://doi.org/10.1093/bioinformatics/bti173
91. Biggin MD, Bickel S, Benson M, Pirrotta V, Tjian R (1988) Zeste encodes a sequence-specific transcription factor that activates the Ultrabithorax promoter in vitro. Cell 53(5):713–722. https://doi.org/10.1016/0092-8674(88)90089-x
92. Mohrmann L, Kal AJ, Verrijzer CP (2002) Characterization of the extended Myb-like DNA-binding domain of trithorax group protein Zeste. J Biol Chem 277(49):47385–47392. https://doi.org/10.1074/jbc.M202341200
93. Moses AM, Pollard DA, Nix DA, Iyer VN, Li XY, Biggin MD, Eisen MB (2006) Large-scale turnover of functional transcription factor binding sites in Drosophila. PLoS Comput Biol 2(10):e130. https://doi.org/10.1371/journal.pcbi.0020130
94. Timmerman C, Suppiah S, Gurudatta BV, Yang J, Banerjee C, Sandstrom DJ, Corces VG, Sanyal S (2013) The Drosophila transcription factor Adf-1 (nalyot) regulates dendrite growth by controlling FasII and Staufen expression downstream of CaMKII and neural activity. J Neurosci 33(29):11916–11931. https://doi.org/10.1523/JNEUROSCI.1760-13.2013
95. Lang M, Juan E (2010) Binding site number variation and high-affinity binding consensus of Myb-SANT-like transcription factor Adf-1 in Drosophilidae. Nucleic Acids Res 38(19):6404–6417. https://doi.org/10.1093/nar/gkq504
96. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M, Perez E, Uszkoreit J, Pfeuffer J, Sachsenberg T, Yilmaz S, Tiwary S, Cox J, Audain E, Walzer M, Jarnuczak AF, Ternent T, Brazma A, Vizcaino JA (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res 47(D1):D442–D450. https://doi.org/10.1093/nar/gky1106
97. Lo SM, Ahuja NK, Francis NJ (2009) Polycomb group protein suppressor 2 of zeste is a functional homolog of posterior sex combs. Mol Cell Biol 29(2):515–525. https://doi.org/10.1128/MCB.01044-08
98. Ha M, Kim VN (2014) Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol 15(8):509–524. https://doi.org/10.1038/nrm3838
99. Moshkovich N, Nisha P, Boyle PJ, Thompson BA, Dale RK, Lei EP (2011) RNAi-independent role for Argonaute2 in CTCF/CP190 chromatin insulator function. Genes Dev 25(16):1686–1701. https://doi.org/10.1101/gad.16651211
100. Cernilogar FM, Onorati MC, Kothe GO, Burroughs AM, Parsi KM, Breiling A, Lo Sardo F, Saxena A, Miyoshi K, Siomi H, Siomi MC, Carninci P, Gilmour DS, Corona DF, Orlando V (2011) Chromatin-associated RNA interference components contribute to transcriptional regulation in Drosophila. Nature 480(7377):391–395. https://doi.org/10.1038/nature10492
101. Nazer E, Dale RK, Chinen M, Radmanesh B, Lei EP (2018) Argonaute2 and LaminB modulate gene expression by controlling chromatin topology. PLoS Genet 14(3):e1007276. https://doi.org/10.1371/journal.pgen.1007276
102. Core L, Adelman K (2019) Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. Genes Dev 33(15–16):960–982. https://doi.org/10.1101/gad.325142.119
103. Dollinger R, Gilmour DS (2021) Regulation of promoter proximal pausing of RNA polymerase II in metazoans. J Mol Biol 433(14):166897. https://doi.org/10.1016/j.jmb.2021.166897
104. Nazer E, Dale RK, Palmer C, Lei EP (2018) Argonaute2 attenuates active transcription by limiting RNA Polymerase II elongation in *Drosophila melanogaster*. Sci Rep 8(1):15685. https://doi.org/10.1038/s41598-018-34115-1
105. Li J, Gilmour DS (2013) Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor. EMBO J 32(13):1829–1841. https://doi.org/10.1038/emboj.2013.111
106. King-Jones K, Korge G, Lehmann M (1999) The helix-loop-helix proteins dAP-4 and daughterless bind both in vitro and in vivo to SEBP3 sites required for transcriptional activation of the Drosophila gene Sgs-4. J Mol Biol 291(1):71–82. https://doi.org/10.1006/jmbi.1999.2963
107. Wong MM, Liu MF, Chiu SK (2015) Cropped, Drosophila transcription factor AP-4, controls tracheal terminal branching and cell growth. BMC Dev Biol 15:20. https://doi.org/10.1186/s12861-015-0069-6

108. Kal AJ, Mahmoudi T, Zak NB, Verrijzer CP (2000) The Drosophila brahma complex is an essential coactivator for the trithorax group protein zeste. Genes Dev 14(9):1058–1071
109. Shen X, Mizuguchi G, Hamiche A, Wu C (2000) A chromatin remodelling complex involved in transcription and DNA processing. Nature 406(6795):541–544. https://doi.org/10.1038/35020123
110. Alfieri C, Gambetta MC, Matos R, Glatt S, Sehr P, Fraterman S, Wilm M, Muller J, Muller CW (2013) Structural basis for targeting the chromatin repressor Sfmbt to Polycomb response elements. Genes Dev 27(21):2367–2379. https://doi.org/10.1101/gad.226621.113
111. Fereres S, Simon R, Mohd-Sarip A, Verrijzer CP, Busturia A (2014) dRYBP counteracts chromatin-dependent activation and repression of transcription. PLoS ONE 9(11):e113255. https://doi.org/10.1371/journal.pone.0113255
112. Sehadova H, Glaser FT, Gentile C, Simoni A, Giesecke A, Albert JT, Stanewsky R (2009) Temperature entrainment of Drosophila's circadian clock involves the gene nocte and signaling from peripheral sensory tissues to the brain. Neuron 64(2):251–266. https://doi.org/10.1016/j.neuron.2009.08.026
113. Amoyel M, Anderson AM, Bach EA (2014) JAK/STAT pathway dysregulation in tumors: a Drosophila perspective. Semin Cell Dev Biol 28:96–103. https://doi.org/10.1016/j.semcdb.2014.03.023
114. Grau DJ, Antao JM, Kingston RE (2010) Functional dissection of Polycomb repressive complex 1 reveals the importance of a charged domain. Cold Spring Harb Symp Quant Biol 75:61–70. https://doi.org/10.1101/sqb.2010.75.056
115. Chalkley GE, Moshkin YM, Langenberg K, Bezstarosti K, Blastyak A, Gyurkovics H, Demmers JA, Verrijzer CP (2008) The transcriptional coactivator SAYP is a trithorax group signature subunit of the PBAP chromatin remodeling complex. Mol Cell Biol 28(9):2920–2929. https://doi.org/10.1128/MCB.02217-07
116. Mohrmann L, Langenberg K, Krijgsveld J, Kal AJ, Heck AJ, Verrijzer CP (2004) Differential targeting of two distinct SWI/SNF-related Drosophila chromatin-remodeling complexes. Mol Cell Biol 24(8):3077–3088. https://doi.org/10.1128/MCB.24.8.3077-3088.2004
117. Yin JW, Wang G (2014) The mediator complex: a master coordinator of transcription and cell lineage development. Development 141(5):977–987. https://doi.org/10.1242/dev.098392
118. Kennison JA, Tamkun JW (1988) Dosage-dependent modifiers of polycomb and antennapedia mutations in Drosophila. Proc Natl Acad Sci USA 85(21):8136–8140. https://doi.org/10.1073/pnas.85.21.8136
119. de Ayala G, Alonso A, Gutierrez L, Fritsch C, Papp B, Beuchle D, Muller J (2007) A genetic screen identifies novel polycomb group genes in Drosophila. Genetics 176(4):2099–2108. https://doi.org/10.1534/genetics.107.075739
120. Chetverina D, Aoki T, Erokhin M, Georgiev P, Schedl P (2014) Making connections: insulators organize eukaryotic chromosomes into independent cis-regulatory networks. BioEssays 36(2):163–172. https://doi.org/10.1002/bies.201300125
121. Chetverina D, Fujioka M, Erokhin M, Georgiev P, Jaynes JB, Schedl P (2017) Boundaries of loop domains (insulators): determinants of chromosome form and function in multicellular eukaryotes. BioEssays. https://doi.org/10.1002/bies.201600233
122. Cubenas-Potts C, Corces VG (2015) Architectural proteins, transcription, and the three-dimensional organization of the genome. FEBS Lett 589(20 Pt A):2923–2930. https://doi.org/10.1016/j.febslet.2015.05.025
123. Mihaly J, Hogga I, Gausz J, Gyurkovics H, Karch F (1997) In situ dissection of the Fab-7 region of the bithorax complex into a chromatin domain boundary and a Polycomb-response element. Development 124(9):1809–1820
124. Kyrchanova O, Kurbidaeva A, Sabirov M, Postika N, Wolle D, Aoki T, Maksimenko O, Mogila V, Schedl P, Georgiev P (2018) The bithorax complex iab-7 Polycomb response element has a novel role in the functioning of the Fab-7 chromatin boundary. PLoS Genet 14(8):e1007442. https://doi.org/10.1371/journal.pgen.1007442
125. Gruzdeva N, Kyrchanova O, Parshikov A, Kullyev A, Georgiev P (2005) The Mcp element from the bithorax complex contains an insulator that is capable of pairwise interactions and can facilitate enhancer-promoter communication. Mol Cell Biol 25(9):3682–3689. https://doi.org/10.1128/MCB.25.9.3682-3689.2005
126. Vorobyeva NE, Erokhin M, Chetverina D, Krasnov AN, Mazina MY (2021) Su(Hw) primes 66D and 7F Drosophila chorion genes loci for amplification through chromatin decondensation. Sci Rep 11(1):16963. https://doi.org/10.1038/s41598-021-96488-0
127. Chen JD, Chan CS, Pirrotta V (1992) Conserved DNA binding and self-association domains of the Drosophila zeste protein. Mol Cell Biol 12(2):598–608. https://doi.org/10.1128/mcb.12.2.598-608.1992
128. Katsani KR, Hajibagheri MA, Verrijzer CP (1999) Co-operative DNA binding by GAGA transcription factor requires the conserved BTB/POZ domain and reorganizes promoter topology. EMBO J 18(3):698–708. https://doi.org/10.1093/emboj/18.3.698
129. Attardi LD, Tjian R (1993) Drosophila tissue-specific transcription factor NTF-1 contains a novel isoleucine-rich activation motif. Genes Dev 7(7B):1341–1353. https://doi.org/10.1101/gad.7.7b.1341
130. Espinas ML, Jimenez-Garcia E, Vaquero A, Canudas S, Bernues J, Azorin F (1999) The N-terminal POZ domain of GAGA mediates the formation of oligomers that bind DNA with high affinity and specificity. J Biol Chem 274(23):16461–16469. https://doi.org/10.1074/jbc.274.23.16461
131. Pereira A, Paro R (2017) Pho dynamically interacts with Spt5 to facilitate transcriptional switches at the hsp70 locus. Epigenet Chromatin 10(1):57. https://doi.org/10.1186/s13072-017-0166-9
132. Harvey R, Schuster E, Jennings BH (2013) Pleiohomeotic interacts with the core transcription elongation factor Spt5 to regulate gene expression in Drosophila. PLoS ONE 8(7):e70184. https://doi.org/10.1371/journal.pone.0070184
133. Papp B, Muller J (2006) Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. Genes Dev 20(15):2041–2054. https://doi.org/10.1101/gad.388706
134. Kyrchanova O, Maksimenko O, Stakhov V, Ivlieva T, Parshikov A, Studitsky VM, Georgiev P (2013) Effective blocking of the white enhancer requires cooperation between two main mechanisms suggested for the insulator function. PLoS Genet 9(7):e1003606. https://doi.org/10.1371/journal.pgen.1003606
135. Kang H, McElroy KA, Jung YL, Alekseyenko AA, Zee BM, Park PJ, Kuroda MI (2015) Sex comb on midleg (Scm) is a functional link between PcG-repressive complexes in Drosophila. Genes Dev 29(11):1136–1150. https://doi.org/10.1101/gad.260562.115

Authors and Affiliations

Darya Chetverina[1] · Nadezhda E. Vorobyeva[2] · Marina Yu. Mazina[3] · Lika V. Fab[4] · Dmitry Lomaev[1] · Alexandra Golovnina[1] · Vladic Mogila[5] · Pavel Georgiev[5] · Rustam H. Ziganshin[6] · Maksim Erokhin[4]

Nadezhda E. Vorobyeva
vorobyeva@genebiology.ru

Marina Yu. Mazina
magadovam@yandex.ru

Lika V. Fab
lika.fab7@gmail.com

Dmitry Lomaev
lomaevdv@gmail.com

Alexandra Golovnina
golovnina_aleksandra_2799@mail.ru

Vladic Mogila
v@mogila.org

Pavel Georgiev
georgiev_p@mail.ru

Rustam H. Ziganshin
rustam.ziganshin@gmail.com

[1] Group of Epigenetics, Institute of Gene Biology, Russian Academy of Sciences, 34/5 Vavilov Street, Moscow 119334, Russia

[2] Group of Dynamics of Transcriptional Complexes, Institute of Gene Biology, Russian Academy of Sciences, Moscow, Russia

[3] Group of Hormone-Dependent Transcriptional Regulation, Institute of Gene Biology, Russian Academy of Sciences, Moscow, Russia

[4] Group of Chromatin Biology, Institute of Gene Biology, Russian Academy of Sciences, 34/5 Vavilov Street, Moscow 119334, Russia

[5] Department of Control of Genetic Processes, Institute of Gene Biology, Russian Academy of Sciences, 34/5 Vavilov Street, Moscow 119334, Russia

[6] Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow 117997, Russia