Physical and functional interaction between SET1/COMPASS complex component CFP-1 and a Sin3S HDAC complex in *C. elegans*

Flore Beurton[1], Przemyslaw Stempor[2], Matthieu Caron[1], Alex Appert[2], Yan Dong[2], Ron A.-j. Chen[2], David Cluet[1], Yohann Couté[3], Marion Herbette[1], Ni Huang[2], Hélène Polveche[4], Martin Spichty[1], Cécile Bedet[1], Julie Ahringer[2,*] and Francesca Palladino[1,*]

[1]Laboratory of Biology and Modeling of the Cell, UMR5239 CNRS/Ecole Normale Supérieure de Lyon, INSERM U1210, UMS 3444 Biosciences Lyon Gerland, Université de Lyon, Lyon, France, [2]The Gurdon Institute and Department of Genetics, University of Cambridge, Cambridge, UK, [3]Grenoble Alpes, CEA, Inserm, BIG-BGE, 38000 Grenoble, France and [4]INSERM UMR 861, I-STEM, 28, Rue Henri Desbruères, 91100 Corbeil-Essonnes, France

ABSTRACT

The CFP1 CXXC zinc finger protein targets the SET1/COMPASS complex to non-methylated CpG rich promoters to implement tri-methylation of histone H3 Lys4 (H3K4me3). Although H3K4me3 is widely associated with gene expression, the effects of CFP1 loss vary, suggesting additional chromatin factors contribute to context dependent effects. Using a proteomics approach, we identified CFP1 associated proteins and an unexpected direct link between *Caenorhabditis elegans* CFP-1 and an Rpd3/Sin3 small (SIN3S) histone deacetylase complex. Supporting a functional connection, we find that mutants of COMPASS and SIN3 complex components genetically interact and have similar phenotypic defects including misregulation of common genes. CFP-1 directly binds SIN-3 through a region including the conserved PAH1 domain and recruits SIN-3 and the HDA-1/HDAC subunit to H3K4me3 enriched promoters. Our results reveal a novel role for CFP-1 in mediating interaction between SET1/COMPASS and a Sin3S HDAC complex at promoters.

INTRODUCTION

The CFP1/CXXC zinc finger protein targets the SET1/COMPASS complex to non-methylated CpG rich regions for trimethylation of histone H3 on Lys4 (H3K4me3) (1–5), a modification widely associated with active promoters (6–8). The roles of CFP1 and the SET1/COMPASS complex in gene regulation are unclear. In different systems, loss of individual subunits does not have widespread effects on transcription, with only small subsets of genes affected (1,2,9–14). The effects vary depending on context, consistent with potential interactions with other factors, and proposals that H3K4me3 may promote transcriptional memory and consistency (9,15,16).

In yeast, SET1 acts in a single complex known as COMPASS (complex of proteins associated with Set1) that is responsible for all H3K4 methylation (17,18). In mammals by contrast, six complexes have been isolated defined by the catalytic subunits SET1A, SET1B, MLL1, MLL2, MLL3 and MLL4 (reviewed in (19). The enzymatic activity of SET1/MLL family members is regulated by interactions with additional proteins, including Swd3/WDR5, Swd1/RbBP5, Bre2/ASH2 and Sdc1/hDPY30 that influence the state (mono-, di- or tri) of methylation deposited (20–22). In addition, unique subunits including CFP1 are associated with each complex and contribute to its specificity (3,23–25).

SET1/MLL complexes have non-redundant functions, as demonstrated by the distinct phenotypes and embryonic lethality caused by deletion of individual SET1/MLL genes (26–29). While SET1 proteins are responsible for global H3K4me3 at promoter regions in different organisms (30–34), MLL proteins deposit H3K4 methylation at specific genes or regulatory elements (35–37).

Caenorhabditis elegans contains a single homologue of SET1, named SET-2, one MLL-like protein, SET-16, and single homologs of WDR5, ASH2L, DPY30, RbBP5 and CFP1 (32,34,38,39), simplifying functional studies of SET1/MLL regulatory networks. Inactivation of SET-2,

---

WDR-5.1, DPY-30, RbBP5 and CFP-1 has shown that they all contribute to global H3K4 methylation in the germline and soma, and share common functions in somatic and germline development (32,34,38,40–43). To biochemically analyze the complex and identify associated proteins that may contribute to its functional outcome, we immunoprecipitated tagged CFP-1 and WDR-5.1, and identified copurifying proteins by mass spectrometry. In addition to identifying distinct SET-2/SET1 and SET-16/MLL complexes, we found that WDR-5.1 co-immunoprecipitates NSL histone acetyltransferase (HAT) complex subunits, consistent with its presence in multiple chromatin-associated complexes (44–46). Most importantly, we show that CFP-1 physically and functionally interacts with a conserved Rpd3/Sin3 small histone deacetylase complex, SIN3S. Mutants of SET-2/SET1 and SIN3S complex subunits share partly similar phenotypes, and CFP-1 is important for recruitment of both SIN-3 and the HDA-1/HDAC subunit to H3K4me3 enriched promoters. Our results reveal a novel role for CFP-1 in interacting with both SET-2/SET1 and SIN3S HDAC complexes to maintain the embryonic transcriptional program and influence both somatic and germline development.

MATERIALS AND METHODS

Strains and maintenance

Nematode strain maintenance was as described previously (47). The wild-type strain N2 (Bristol) was used as the reference strain. The strains used are as follows: PFR506 *qaIs22[HA::wdr-5.1;Cbunc-119(+)];wdr-5.1(ok1417)III*, JA1597 *pdpy-30::cfp-1::GFP*, PFR572 *qaIs22[HA::wdr-5.1;Cbunc-119(+)];wdr-5.1(ok1417)III;pdpy-30::cfp-1::GFP*, PFR625 *qaIs22[HA::wdr-5.1;Cbunc-119(+)];wdr-5.1(ok1417)III;pdpy-30::cfp-1::GFP;sin-3(tm1276)I*, PFR510 *set-2(bn129)lqC1dpy-19(e1259)glp-1(q339)[qIs26]III*, PFR624 *cfp-1(tm6369) IV/nT1[unc?(n754),let-?](IV;V)*, PFR391 *wdr-5.1(ok1417)III* out-crossed twice, PFR590 *sin-3(tm1276)* out-crossed twice, PFR630 *sin-3(tm1276)I/hT2[bli-4(e937)qIs48](I; III)*, PFR629 *set-2(bn129)III sin-3(tm1276)I*, PFR635 *set-2(bn129)III; cfp-1(tm6369)IV/nT1[unc?(n754),let?](IV;V)*, PFR636 *sin-3(tm1276)I;cfp-1(tm6369)IV/nT1[unc?(n754),let?](IV;V)*;dvIs70 [*hsp-16.2*p::GFP + *rol-6(su1006)*], PFR593 *athp-1(tm4223)III*, obtained from the National Bioresource Project (Japan) and outcrossed twice, PFR696 *athp-1;cfp-1* obtained by crossing *cfp-1(tm6369)* males with *athp-1(tm4223)* hermaphrodites.

ChIP experiments used PFR253 *set-2(bn129)* (outcrossed 10X) and TM6369 *cfp-1(tm6369)* (outcrossed 4X). To generate the PFR624 balanced strain, *cfp-1(tm6369)* animals were crossed with wildtype males to generate heterozygote males that were crossed again with [unc] animals segregating from AV112 (*mre-11(ok179),V/nT1,[unc?(n754),let-?](IV;V)*). [unc] animals were selected and screened by PCR for the presence of the *cfp-1(tm6369)* deletion. To generate the PFR630 balanced strain, *sin-3(tm1276)* animals were crossed with wildtype males to generate *sin-3(tm1276)* heterozygote males that were crossed to GFP(+) animals segregating

from *[F44E2.7(tm4302)/hT2]*. GFP(+) animals from this last cross were selected and screened by PCR for the presence of the *sin-3(tm1276)* deletion. Yeast two-hybrid strains were: EGY42 (MATa; *trp1, his3, ura3, leu2*); TB50 (MATα; *trp1, his3, ura3, leu2, rme1*). *cfp-1(tm6369)* is a deletion of 254 bp encompassing intron 4, exon 5 and intron 5 of *cfp-1*. It is predicted to produce a truncated CFP-1 protein of 374 aa lacking part of the conserved C-terminal domain. Primers used genotyping are listed in Supplementary Table S1.

Immunoprecipitation for proteomics

Immunoprecipitations were performed on frozen embryos prepared by hypochlorite treatment from strains grown at 20°C on enriched NGM. For all immunoprecipitations, wildtype embryos (N2) were treated in parallel to serve as negative control in the mass spectrometry analysis. For HA ::WDR-5.1 immunoprecipitations, embryos from PFR506 were flash-frozen immediately after hypochlorite treatment. For CFP-1::GFP immunoprecipitations, PFR572 late stage embryos were obtained by incubating the embryos collected by hypochlorite treatment for 4 h prior to flash freezing in liquid nitrogen. For each condition embryos were ground to powder, resuponded in IP buffer (50 mM HEPES/KOH pH 7,5; 300 mM KCl; 1 mM EDTA; 1 mM MgCl2; 0.2% Igepal-CA630 and 10% glycerol) containing complete protease inhibitors [Roche] and 1 mM PMSF, and sonicated. Protein extracts were recovered in supernantant following centrifugation at 20 000 g for 15 min at 4°C an 20°C and flash frozen in liquid nitrogen. Protein concentrations were estimated using the Bradford assay [Bio-Rad Protein Assay Dye]. For HA::WDR-5.1 immunoprecipitation, approximately 60 mg of total protein extract was incubated with protein G agarose beads [Sigma-Aldrich] in Bio-Spin Chromatography columns [Bio-Rad] for 30 min at 4°C on a rotator. Flow-through was collected and incubated with 240 μl slurry of anti-HA affinity matrix beads [Roche] in a fresh Bio-Spin Chromatography column for 90 min at 4°C on a rotator. The matrix was washed three times in IP buffer at 4°C and once in Benzo buffer (HEPES/KOH 50 mM pH 7,5; KCl 150 mM; EDTA 1 mM; MgCl$_2$ 1 mM; Igepal-CA630 0.2%; glycerol 10%). The matrix was then incubated in 400 μl of Benzo buffer containing 2500 units of benzonase [Sigma] for 1 h at 4°C and washed three times in IP buffer. Four successive elutions were performed at 37°C for 15 min each with HA peptide (250 μg/ml in 240 μl of IP buffer). The first three eluates were pooled and concentrated 20 times (final volume 35 μl) using a 10 kDa Amicon Ultra centrifugal device [Merck]. 1/70 and 1/700 of this eluate were resolved on a 4–12% NuPage Novex gel [Thermo Fischer] and the gel either stained with SilverQuest staining kit [Thermo Fischer] or analyzed by western blot with anti-HA antibody [Covance HA.11, clone 16B12]. 33 μl of the eluate was diluted with 11 μl of LDS4X buffer [Thermo Fischer] and analyzed by mass spectrometry. For CFP-1::GFP immunoprecipitation, ∼70 mg of total protein were incubated in IP buffer with 100 μl of GFP-TRAP MA beads slurry [Chromotek] for 3 h at 4°C on a rotator. Beads were collected with a magnet, washed three times in IP buffer and one time in Benzo buffer, and then treated with benzonase.

Eluates were recovered by incubation at 95°C for 10 min in 60 µl of LDS 1× buffer. 1/10 and 1/50 of this eluate were resolved on a 4–12% NuPage Novex gel [Thermo Fischer] and either stained with SilverQuest staining kit [Thermo Fischer] or analyzed by western blot with anti-GFP antibody [Sigma, 11814460001, clones 7.1 and 13.1] respectively, and 40 µl of the eluate was analyzed by mass spectrometry.

Mass spectrometry-based proteomic analyses

Proteins were stacked in the top of a SDS-PAGE gel (4-12% NuPAGE, Life Technologies) and stained with Coomassie blue R-250 before in-gel digestion using modified trypsin (Promega, sequencing grade) as previously described (48). Resulting peptides were analyzed by online nanoLC-MS/MS (UltiMate 3000 and LTQ-Orbitrap Velos Pro, Thermo Scientific). For this, peptides were sampled on a 300 µm × 5 mm PepMap C18 precolumn and separated on a 75 µm × 250 mm C18 column (PepMap, Thermo Scientific). MS and MS/MS data were acquired using Xcalibur (Thermo Scientific). Peptides and proteins were identified using Mascot (version 2.5.1) through concomitant searches against Uniprot (*C. elegans* taxonomy), classical contaminants database (homemade) and the corresponding reversed databases. The Proline software (http://proline. profiproteomics.fr) was used to filter the results (conservation of rank 1 peptides, peptide identification FDR < 1% as calculated on peptide-spectrum match scores by employing the reverse database strategy, minimum peptide score of 25, and minimum of 1 specific peptide per identified protein group) before performing a compilation, grouping and comparison of the protein groups from the different samples. In WDR-5.1 mass spectrometry, 242 proteins were found with a spectral count (SC) WDR-5.1 ≥ 3 and a SC control (without WDR-5::HA transgene) = 0 or SC WDR-5.1/SC control ≥5 for at least one replicate; in CFP-1 mass spectrometry 178 proteins were found with a SC CFP-1 ≥3 and a SC control (without CFP-1::GFP transgene) = 0.

Co-immunoprecipitation experiments

Co-immunoprecipitations with CFP-1::GFP were performed starting from 4 mg total protein embryonic extract from the strain containing the two transgenes CFP-1::GFP and HA::WDR-5.1. Samples were processed as in proteomic experiments. Co-immunoprecipitations with HA::WDR-5.1 were performed with the eluates sent to mass spectrometry analysis. Samples were processed as in proteomic experiments. Eluates were boiled in LDS sample buffer and analyzed on 4–12% NuPage Novex gels [Thermo Fischer] or Mini-PROTEAN TGX Stain-Free Precast gels [Bio-Rad] followed by western blotting. Antibodies used were: anti-GFP [Sigma, 11814460001, clones 7.1 and 13.1] (1/1000); anti-HA [Covance HA.11, clone 16B12] (1/2000); anti-HDA-1 [Novus Biologicals, 38660002] (1/2000); anti-DPY-30 [Novus Biologicals, 45110002] (1/5000); anti-ASH-2 (gift from B. Meyer) (1/4000); anti-MRG-1 antibody [Novus Biologicals, 35530002] (1/3000); anti-SIN-3 Q60131017(1/1000).

Plasmids construction for Y2H

Plasmids used for expression of BD and AD fusions were derived from pEG202 (Clontech; Genbank Accession Number U89960) and pJG4–5 plasmids (Clontech; Genbank Accession Number U89961, respectively (49). Constructions were generated by cloning the cDNA of the gene of interest in the *Xho*I restriction site of the pEG202 and pJG4–5 plasmids using the Gibson method (50). CFP-1 truncations were obtained by the same reaction using the *cfp-1* cDNA sequence as template. PCR reactions were carried out using pHusion polymerase and primers listed in Supplementary Table S1. All products were verified by sequencing. pSH18–34, bearing a β-galactosidase gene under the control of four overlapping LexA operators was used as reporter vector (51).

Interaction trap/two-hybrid system to identify interacting protein

Y2H assay is based on the LexA (BD)/B42 (AD) system (49). Cross-matings were performed in liquid phase (52). Competent haploid EGY42a cells were co-transformed with 1 µg of pSH18-34 (reporter vector) and 1 µg of BD construct. Competent TB50α cells were transformed with 1 µg of AD construct. Yeasts were selected for 3 days at 30°C on SD-UH (BD strains) and SD-W (AD strains) medium. Matings were performed overnight at 30°C in liquid YPAD (53). Cross-mating ensured that each hetero-interaction was tested twice (in both directions of the interaction matrix) and allowed the detection of homodimerisations. Diploids were amplified in selective liquid SD-UHW medium. For β-galactosidase assays, 50 µl of each diploids culture was inoculated (at OD$_{595\,nm}$ = 6) in 1 ml of pre-warmed (25°C) SGR-UHW medium supplemented with X-Gal (Thermo Scientific, #R0404) in Deepwell 96-well plates. Cultures were then incubated for 48 h at 25°C, centrifuged 5 min at 192g, resuspended in 300 µl, and transferred in flat bottom µClear Cellstar® plates (Greiner Bio one) for scanning and phenotype assessment.

Protein expression and GST pull-down assay

Full-length GST::CFP-1 and HIS$_6$::SIN-3 fragments (1–738, 699–1507) were amplified using primers listed in Supplementary Table S1 and subcloned into pGEX-6-P1 [Sigma GE28-9546-48] and pPROEX HTa (gift from L. Terradot, MMSB, Lyon), respectively, using the Gibson method (50). All proteins were expressed in BL21 Rosetta 2 [Merck-Millipore 71402]. Bacteria were grown to OD$_{600}$ 0.6 and protein expression induced with 1 mM IPTG at 16°C overnight. The pellet from 1 l of bacterial culture was resuspended in 10 ml lysis buffer (50 mM Tris pH 8,0 ; 300 mM NaCl; 0.1 mM EDTA; 0.1% Triton X-100 ; 0.05% NP-40; 1 mM MgCl$_2$; 5% glycerol) containing protease inhibitor [Roche 05056489001]. Samples were sonicated on ice and centrifuged at 20 000 g for 20 min. For pulldown assays 200 ul of GST::CFP-1 supernatant was mixed with 800 ul HIS$_6$::SIN-3[1–738] or HIS$_6$::SIN-3[699-1507], respectively, and incubated overnight at 4°C on a rotating wheel. Samples were submitted to GST purification on a Biosprint 15 automat from Qiagen. Samples were washed

three times with Lysis buffer and eluted with 50 ul of Lysis buffer containing 20 mM Glutathion. Eluted fractions were analyzed by western blot using a mouse anti-Histidine antibody [Sigma H1029] (1/3000) and Stain-Free gels (Bio-Rad).

Brood size and embryonic lethality assays

For each strain, 10 L4 worms were isolated to single plates in the presence of excess food at 20°C, and allowed to develop into egg-laying adults overnight. Adult animals were then transferred to fresh plates every 12 h until they ceased laying eggs. Plates were scored for number of viable progeny and dead embryos that failed to hatch 24 h after removal of the mother.

Fertility assay

Six independent lines were established from freshly thawed *sin-3(tm1276)* animals maintained as homozyotes, and homozygous *set-2(bn129)* and *cfp-1(tm6369)* animals obtained from balanced strains PFR 510 and PFR 624, respectively. For each line, six homozygous L4 stage animals were transferred to single plates with fresh *E. coli*, in the presence of excess food and cultivated at 25°C. From each generation, six worms were again picked to single plates until animals became sterile (fewer than 10 progeny/plate).

Characterization of nuclear divisions in intestinal nuclei

Adult animals were treated with hypochlorite solution to obtained L1 synchronized larva. L1 larva were transferred to 25°C for 48 hrs, until they developed into adults. Young adults were stained with DAPI staining and analyzed with a Zeiss 710 Confocal Microscope. Experiments were performed in three independent replicates and intestinal nuclei from a total of 150 worms for each strain were scored.

Comparison of gene expression changes in *cfp-1*, *set-2* and *sin-3* embryos

RNAs were extracted from wild-type, *cfp-1, set-2* and *sin-3* frozen early stage embryos prepared by hypochlorite treatment of young adults (>95% <200 cell stage). Two to three independent biological replicates were performed for each strain. RNAs were extracted with NucleoZol [Macherey-Nagel] according to manufacturer's instructions and treated with DNAse [Turbo-free DNAse, Ambion]. Integrity of RNA was assessed on Tape Station 4200 [Agilent]. RNA-seq librairies were generated at the GenomEast Platform [IGBMC, Strasbourg, France] using the directional mRNA-Seq SamplePrep [Illumina] and sequenced using the Illumina Hiseq 4000 technology. All RNA-seq data were mapped to the *C. elegans* reference genome (WS254) by RNA-STAR (Version 2.4.1d). Reads below a mapping score of 10 were filtered using SAMtools (Version 0.1.19). Of the 46 771 annotated genes, 20 183 were selected as protein coding genes and among them, 11 630 had sufficient read representation (baseMean > 10) for further analysis. The gene expression level in each sample was calculated by htseq-count (Version 0.7.2) and differential expression between the different strains was calculated with DESeq2 (54).

Gene expression data are available at GEO with the accession GSE110072.

Western blot analysis on histone marks

Embryos were obtained by hypochlorite treatment of adults grown on HB101 at 20°C and frozen in liquid nitrogen. Embryo pellets were resuspended in TNET buffer (50 mM Tris·HCl (pH 8), 300 mM NaCl, 1 mM EDTA, 0,5% Triton X-100 and protease inhibitors cocktail III [Merck]), lysed with zirconium beads [Lysing Matrix Y, MP Biomedicals #116960050] using a Precellys24 homogenizer and sonicated in a Bioruptor sonicator. Homogenates were centrifuged and supernatants aliquoted and frozen at −80°C. Total protein amount was quantified by the Bradford assay [Bio-Rad]. Serial dilutions of protein extracts were electrophoresed on 12% NuPage Novex gels for western blot analysis. Dilutions of wild type total protein extracts were analyzed to determine the upper limit of linearity of the following antibodies: anti-H3K4ac [Sigma, 07-539] (1/1000), anti-H3K9ac [Active Motif, 39137,39138] (1/2000), anti-H3K27ac [Active Motif, 39133] (1/2000), anti-H3K4me3 [Diagenode C15310003] (1/2000) and anti-H3 [Active Motif, 39163] (1/20 000). Each antibody was used on a separate blot loaded with the same extracts.

Heat shock assay

Synchronized L4 staged worms carrying a *phsp-16.2::gfp* reporter (dvIs70) (55) were shifted at 33°C for 30 min in a water bath and allowed to recover at 20°C for 60 min. before observation. Experiments were repeated 5–8 times, with similar results. Fluorescent expression analysis were carried out on a Zeiss AxioPlan 2 equipped with Nomarski optics coupled to a camera (CoolSNAP, Roper Scientific). Non-heat shocked controls were equally overexposed.

Chromatin immunoprecipitation

Wildtype, *cfp-1(tm6369)*, and *set-2(bn129)* mixed embryos were obtained by growing strains at 20°C in liquid culture using standard S-basal medium with HB101 bacteria. Strains were grown to the adult stage then bleached to obtain embryos, which were washed in M9, then frozen into 'popcorn' by dripping embryo slurry into liquid nitrogen. Chromatin immunoprecipitations and library preparations were conducted as in (56), using formaldehyde as a fixative for the H3K4me3 ChIPs (30 ug DNA, 2.5 ug antibody) and formaldehyde and EGS as fixatives for the SIN-3 (15 ug DNA, 2.5 ug antibody) and HDA-1 (30 ug DNA, 2.5 ug antibody) ChIPs. Approximately 10% *C. briggsae* chromatin extract was spiked into the *C. elegans* extract for the H3K4me3 ChIPs and 5% into the HDA-1 ChIPs. The HDA-1 antibody did not detect *C. briggsae* HDA-1 and so was not used for normalization. Two different antibodies to SIN-3 were raised through Strategic Diagnostics International by DNA immunization using aa427-576 (Q5986 and Q6013). Chromatin immunoprecipitations were conducted in duplicate with both SIN-3 antibodies in wild-type embryos; ChIP-seq patterns using these two SIN-3 antibodies were highly concordant (Supplementary Figure S1). Comparison of SIN-3 ChIP levels between wild-type and *cfp-1*

mutant embryos were done using SIN-3 antibody Q5986. HDA-1 ChIPs were done using Novus 38660002/Q2354 and H3K4me3 ChIPs used Abcam ab8580. The age distributions of mixed embryo collections were in the following proportions (% <300 cell / % over 300 cell, average of the two replicates): H3K4me3 ChIPs: WT N2, 51/49; *cfp-1*, 59/51, *set-2*, 54/46. SIN-3 and HDA-1 ChIPs: WT N2, 48/52; *cfp-1*, 49/51. RNA-seq was performed on matched wild-type and *cfp-1* mutant embryo collections and sequencing libraries constructed as in (56). Libraries were sequenced using an Illumina HiSeq1500 and aligned to ce11 (WBCel235) *C. elegans* genome assembly with STAR aligner using Ensembl v90 gene annotation for splice aware alignment. Reads were counted using HTSeq method implemented in R and differential expression was assessed using DESeq2 method (54).

SIN-3, HDA-1 and CFP-1::GFP ChIP-seq data processing

CFP-1::GFP (GEO GSE49870), SIN-3 and HDA-1 ChIP-seq reads were aligned to the ce11 assembly of the *C. elegans* genome using BWA v. 0.7.7 (57) with default settings (BWA-backtrack algorithm). The SAMtools v. 0.1.19 'view' utility was used to convert the alignments to BAM format. Normalized ChIP-seq coverage tracks was generated using the BEADS algorithm (58). ChIP-seq peaks were called for SIN-3, HDA-1 and CFP-1::GFP in wild-type embryos using MACS2 v. 2.1.1 (59) with a *q*-value cut-off of 0.05 and fragment size of 150bp against summed ChIP-seq input (GEO GSE87524). Peaks overlapping non-mappable (GEM-mappability < 25%; (60) or blacklisted regions (https://gist.githubusercontent. com/Przemol/ef62ac7ed41d3a84ad6c478132417770/ raw/56e98b99e6188c8fb3dfb806ff6f382fe91c27fb/ CombinedBlacklists.bed) were discarded. Peak summits were extended 150 bp upstream and downstream, creating 300 bp peak regions. Intersecting regions from the two replicates were kept and extended to 300 bp to obtain confident peak calls. SIN-3 peak calls are the intersection of peaks obtained using the Q5986 and Q6013 antibodies. To determine factor overlaps, the 300 bp peak call regions from CFP-1, SIN-3 and HDA-1 were intersected, also keeping regions with only one factor. Regions from this 'superset' were rescaled to 300bp and each region annotated for overlap with a CFP-1, SIN-3, HDA-1 or MRG-1 peak (MRG-1 peaks were obtained from (61); Supplementary Table S2 gives these regions and their annotations. We used SeqPlots (62) for k-means clustering of CFP-1::GFP and H3K4me3 signals in wild-type, *cfp-1* mutants, and *set-2* mutants to separate CFP-1::GFP peaks into strong and weak COMPASS sites, and to visualise CFP-1::GFP, HDA-1 SIN-3, MRG-1 (GEO GSE50333) and H3K4me3 ChIP-seq tracks as heatmaps. The IGV Genome Browser (63) was applied to visualise example regions. Strong and weak COMPASS peaks were assigned to promoters and genes based on overlap with promoter annotations in (64); for genes with no mapped promoter, peaks were annotated as promoters and assigned to genes if they were within 500 bp of a Wormbase gene start. ChIP-seq data generated in this study is available at GSE114715.

Spike-in normalization of H3K4me3 ChIP-seq

Sequencing reads from H3K4me3 ChIP and corresponding input samples were mapped to a concatenated reference genome sequence containing *C. elegans* ce11 and *C. briggsae* cb3 using BWA (62) and were then separated by species. Only reads that mapped uniquely (mapq> = 10) to non-blacklisted regions were kept. The spike-in ratios of *C. briggsae* to *C. elegans* chromatin present in the combined extract were calculated from the input sequence as *C. briggsae* read count divided by *C. elegans* read count. *C briggsae* H3K4me3 peaks were called from ChIP data using MACS2 (59) with default parameters. Scaling factors for each ChIP samples were calculated as corresponding spike-in ratio divided by *C. briggsae* H3K4me3 ChIP read count in peak regions in millions. These scaling factors were applied to *C. elegans* H3K4me3 ChIP raw coverage track. As a last step, ChIP background was removed from the scale coverage tracks by subtracting the mode and setting negative values to zero. The resulting tracks were used for visualization and analyzing H3K4me3 levels.

ChIP-seq signal quantifications

To compare SIN-3 and HDA-1 binding between wildtype and *cfp-1* mutant embryos, we quantified average BEADS normalized, *z*-scored signal on different peak sets. The signal was obtained using the *bigWigSummary* utility from Kent library (65) implemented in *rtracklayer* package in R. These signals were represented as overlaid violin plots (showing signal distribution) and Tukey box plots (showing estimation of statistical significance of difference between medians as notches) (66). The comparison of H3K4me3 levels in wt, *cfp-1*, and *set-2* mutants was done in the same way, using spike-in normalized signal tracks for quantification.

RESULTS

Co-immunoprecipitation of subunits of the *C. elegans* SET1/COMPASS complex

We used a proteomics approach to characterize *C. elegans* COMPASS-like complexes and search for associated proteins. In addition to the catalytic subunit, SET1/COMPASS complexes contain the core components ASH2, RbBP5, WDR5 and DPY30, and the unique subunits CFP1/CXXC and WDR82 (3,22,33,67–69). Using strains containing two previously described transgenes, CFP-1::GFP and HA::WDR-5.1 (38,70) (Supplementary Figure S2), we found that CFP-1::GFP coprecipitated HA::WDR-5.1, and that HA::WDR-5.1 coprecipitated CFP-1::GFP (Figure 1A, Supplementary Figure S3A). In addition, probing individual precipitates with anti-ASH-2 and anti-DPY-30 antibodies revealed the presence of both proteins (Figure 1). Because single or low copy tagged SET-2 protein could not be detected, we were unable to confirm its presence in the IP experiments.

The above experiments show that in embryos, CFP-1::GFP and HA::WDR-5.1 tagged proteins associated with each other *in vivo* and co-immunoprecipitate native ASH-2 and DPY-30, consistent with their incorporation into a SET1-related complex. To define SET1/MLL complexes
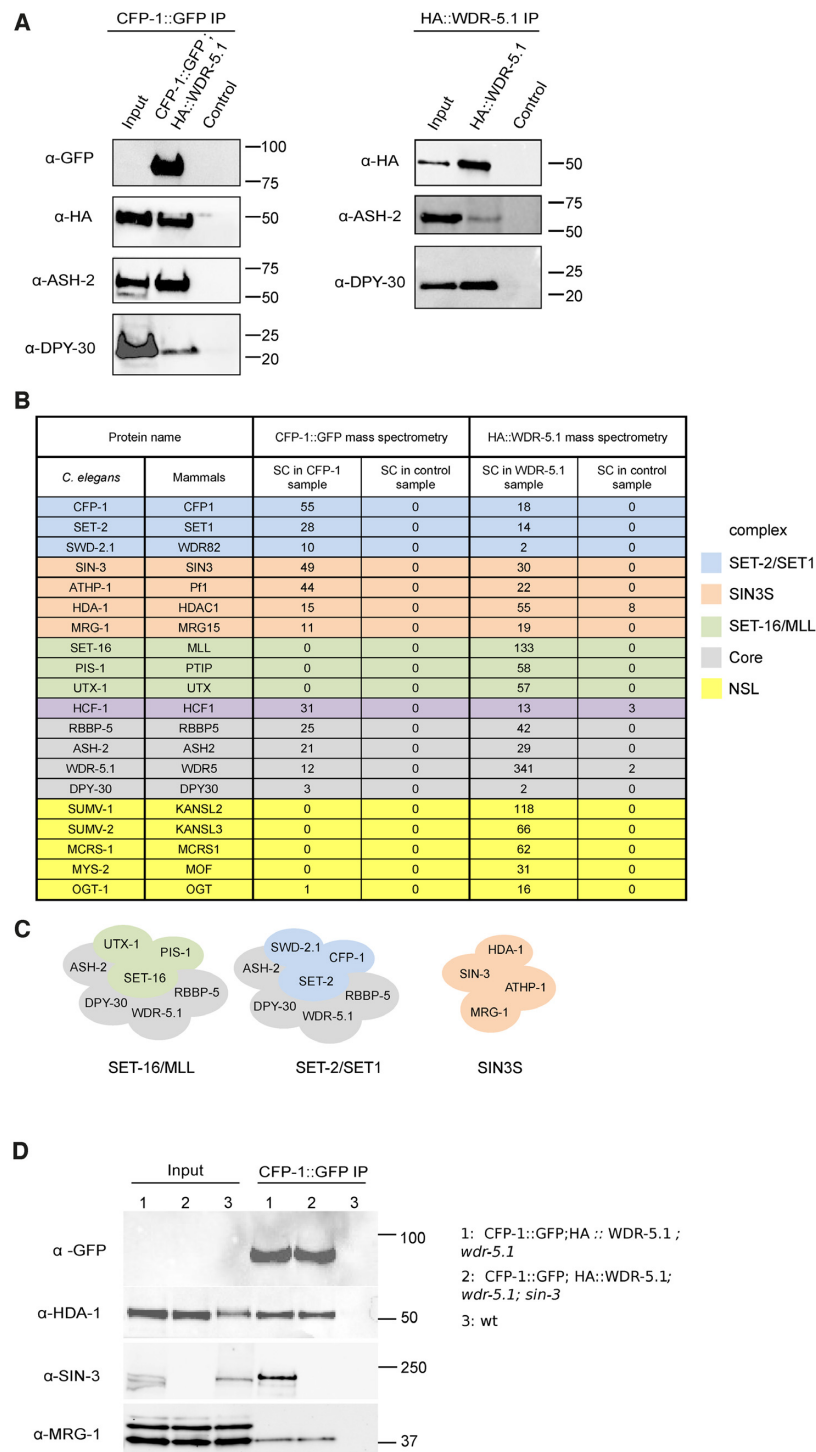
A

Protein name		CFP-1::GFP mass spectrometry		HA::WDR-5.1 mass spectrometry	
C. elegans	Mammals	SC in CFP-1 sample	SC in control sample	SC in WDR-5.1 sample	SC in control sample
CFP-1	CFP1	55	0	18	0
SET-2	SET1	28	0	14	0
SWD-2.1	WDR82	10	0	2	0
SIN-3	SIN3	49	0	30	0
ATHP-1	Pf1	44	0	22	0
HDA-1	HDAC1	15	0	55	8
MRG-1	MRG15	11	0	19	0
SET-16	MLL	0	0	133	0
PIS-1	PTIP	0	0	58	0
UTX-1	UTX	0	0	57	0
HCF-1	HCF1	31	0	13	3
RBBP-5	RBBP5	25	0	42	0
ASH-2	ASH2	21	0	29	0
WDR-5.1	WDR5	12	0	341	2
DPY-30	DPY30	3	0	2	0
SUMV-1	KANSL2	0	0	118	0
SUMV-2	KANSL3	0	0	66	0
MCRS-1	MCRS1	0	0	62	0
MYS-2	MOF	0	0	31	0
OGT-1	OGT	1	0	16	0

complex
- SET-2/SET1
- SIN3S
- SET-16/MLL
- Core
- NSL

Figure 1. The conserved SIN3 complex copurifies with CFP-1 and WDR-5.1. (**A**) Co-IP of CFP-1, WDR-5.1, ASH-2 and DPY-30. Western blot analysis of CFP-1::GFP (left panel) and HA::WDR-5.1 (right panel) purified complexes from embryos. For CFP-1::GFP, immunoprecipitations were performed on 4 mg total protein extract, and loadings were 1/200 of total protein extract for input and 1/2 of total elution volume. For HA::WDR-5.1, samples used for mass spectrometry analysis were loaded as follows: 1/8000 of total protein extract for input and 1/250 of elution for α-ASH-2; 1/60 000 of total protein extract for input and 1/200 of elution for α-DPY-30. Control samples were prepared from a wild-type strain without either transgene. (**B**) List of selected proteins identified by mass spectrometry of CFP-1::GFP or HA::WDR-5.1 immunoprecipitations, and their mammalian homologue. Subunits specific to the SET-2/SET1, SET-16/MLL, SIN3S and NSL complexes are highlighted in blue, green, orange and yellow, respectively. SET1/MLL core complex subunits are highlighted in grey. HCF-1 copurifies with both SET-2/SET1 and SET-16/MLL complexes. SC; Spectral Counts. (**C**) Cartoon representation of SET-2/SET1, SET-16/MLL and SIN3S complexes; subunits are highlighted as in (**B**). (**D**) Co-IP of CFP-1, HDA-1, SIN-3 and MRG-1. CFP-1::GFP immunoprecipitations were carried out on 5 mg total protein and analyzed by western blot using anti-HDA-1, SIN-3 and MRG-1 antibodies. Loadings were 1/5000 of total protein extract for input and 1/5 of total elution volume for anti-GFP and anti-HDA-1, 1/1000 of total protein extract for input and 1/5 of total elution volume for anti-SIN-3 and anti-MRG-1.

and identify additional associated proteins, we undertook mass spectrometry-based proteomic characterization of CFP-1::GFP and HA::WDR-5.1 immunoprecipitates. We reasoned that WDR-5.1, a core component of SET1/MLL and other chromatin complexes, should immunoprecipitate both SET1 and MLL-related complexes (21,71,72), while CFP-1 should specifically immunoprecipitate SET1, but not MLL-related complexes (2,3). Both tagged proteins were detected as unique bands in immunoprecipitates obtained using either anti-GFP or anti-HA antibodies (Supplementary Figure S3B), and as predominant bands by silver-staining (Supplementary Figure S3C). Tandem mass spectrometry (MS/MS) based proteomic analyses of the immunoprecipitates and comparison with eluates from negative controls identified both common and unique binding partners of CFP-1 and WDR-5 (Figure 1B, see Supplementary Table S3 for a full list). We found that CFP-1::GFP and HA::WDR-5.1 immunoprecipitates contained all common subunits of SET1/MLL complexes, including Swd1/RBBP-5, Bre2/ASH-2, and Sdc1/DPY-30. The orthologue of human host cell factor HCF1, a transcriptional regulator associated with COMPASS-like and other chromatin-associated complexes (73) was also identified in both immunoprecipitates. WDR-5.1 additionally coprecipitated specific components of an MLL-related complex including the MLL-like histone methyltransferases SET-16, the histone H3K27 demethylase UTX-1 and PIS-1 (74). Conversely, CFP-1 specifically immunoprecipitated SET-2/SET1 and SWD-2.1/WDR82, but not SET-16/MLL, consistent with it being a unique component of SET1, but not MLL complexes (10,69,75). These results define distinct SET-2/SET1 and SET-16/MLL complexes in *C. elegans* embryos (Figure 1C).

WDR-5.1 immunoprecipitates also contained subunits of the NSL histone acetyltransferase (HAT) complex, consistent with findings in other organisms (44,76–79). We identified the MOF homologue MYS-2, OGT-1, MCRS-1 and the NSL2 and NSL3 homologues SUMV-1 and SUMV-2, respectively. Homologues of two other NSL components found in other organisms, MBD-R2/PHF20 and NSL1/KANSL1, are not found in the *C. elegans* genome (80). Therefore, as in other species, *C. elegans* WDR-5.1 is found in the NSL complex as well as COMPASS/MLL complexes.

C. elegans homologs of the Sin3S complex copurify with CFP-1 and WDR-5.1

Four additional proteins, SIN-3, HDA-1, MRG-1 and ATHP-1, were reproducibly identified as top hits in both HA::WDR-5.1 and CFP-1::GFP immunoprecipitates (Figure 1B and Supplementary Table S3). These are homologs of subunits of the Rpd3/Sin3 small complex in yeast (Rpd3/Sin3S) and SHMP in mammalian cells (Figure 1B, see below) (81,82). In yeast and other organisms a second type of Rpd3/Sin3 complex is found (Rpd3L in yeast) defined by the presence of SAP30, SDS3, and other subunits (83–87). *C. elegans* does not harbor a SAP30 subunit, but has a single counterpart of SDS3, SUDS-3, which was absent from our proteomics analysis with CFP-1 and WDR-5. We will therefore refer to the complex of SIN-3, HDA-

1, MRG-1 and ATHP-1 as the *C. elegans* SIN3S complex (Figure 1C).

Functions ascribed to Rpd3/Sin3 complexes are varied and appear to be context dependent. Although typically referred to as corepressor complexes due to the presence of a histone deacetylase subunit, Rpd3/Sin3 complexes have been associated with both activation and repression of gene expression. In addition, the yeast Rpd3/Sin3S complex has been shown to repress cryptic transcription initiation in transcribed regions and to suppress antisense transcription initiation at promoters (81,88).

Sin3 proteins, which lack known DNA-binding motifs or enzymatic activity, are characterized by the presence of four paired amphipathic helices (PAH) with structural similarity to Myc family transcription factors (89), and a conserved HDAC-interacting domain (HID) (90). While mammals contain two Sin3 proteins (Sin3A and Sin3B) that share both overlapping and distinct functions (91–94), SIN-3 is the only *C. elegans* homologue. It contains a HID domain, and a single PAH most closely related to the highly conserved PAH1 in mammals (95) (Supplementary Figure S3D). *C. elegans* HDA-1 is one of three class I histone deacetylases (HDACs) in *C. elegans* and a component of several other chromatin complexes, as in other species (96,97). MRG-1, the *C. elegans* counterpart of the chromodomain (CD) protein Eaf3/MRG15, is also found in additional chromatin complexes (98–103), and ATHP-1 (AT Hook plus PHD finger transcription factor), a counterpart of Rco1/Pf1, contains two AT Hooks and a Forkhead-associated (FHA) domain that are not found in either Rco1 or Pf1 (Supplementary Figure S3D).

Western blot analysis on CFP-1::GFP immunoprecipitates using antibodies against endogenous MRG-1, HDA-1, and SIN-3 proteins confirmed the interactions between CFP-1 and SIN3 complex components detected by mass-spectrometry (Figure 1D and Supplementary Figure S3E). Interaction between CFP-1 and SIN-3 was also observed in young adults, showing that is not specific to embryos (Supplementary Figure S3F). We also confirmed that HDA-1 co-precipitates with WDR-5.1 (Supplementary Figure S3G). We further found that interaction of HDA-1 and MRG-1 with CFP-1 is not dependent on endogenous SIN-3, as both proteins are found in CFP-1 immunoprecipitates obtained from *sin-3* mutant extracts (Figure 1D). We conclude that CFP-1 physically interacts with a Sin3S complex, but may also interact with HDA-1 and MRG-1 in other contexts.

Subunits of the SET-2/SET1 and SIN3S complex physically interact

We used a yeast two-hybrid assay to assess potential physical interactions between components of the SIN3S and SET-2/SET1 complexes (49). A full-length cDNA of each SET-2/SET1 and SIN3S complex subunit was cloned into vectors to express DNA-binding (DB) and activation domain (AD) fusions. Western blot analysis confirmed expression of all cDNAs with the exception of *set-2* (Supplementary Figure S4A). Testing pairwise interactions of BD and AD fusions by cross-mating, we detected interaction between DPY-30 and ASH-2, and DPY-30 homodimerization

within the SET-2/SET1 complex, consistent with studies in other systems (20,45) (Figure 2A). In addition, we detected CFP-1 homodimerisation (Figure 2A). Within the SIN3S complex, we observed an interaction between MRG-1 and ATHP-1, and MRG-1 homodimerization (Figure 2A). Importantly, we found that CFP-1 interacted with the SIN3S complex components ATHP-1 and SIN-3.

The C-terminal domain of CFP-1 is necessary and sufficient for interaction with SIN-3

Mammalian CFP1 contains an N-terminal PHD domain that recognizes methylated H3K4, a Zn finger CXXC domain that binds to unmethylated CpG dinucleotides, a Set1 interaction domain (SID), a coiled-coiled leucine zipper (LZ) domain, and a cysteine-rich C-terminal domain (1,4,104–106) (Supplementary Figures S3D and S5). *C. elegans* CFP-1 contains all of these except for the PHD domain. To identify the domains that mediate interaction of CFP-1 with SIN-3 and ATHP-1, we expressed different regions of CFP-1 and tested their ability to interact with full length SIN-3 and ATHP-1 by Y2H as described above. Western blot analysis confirmed the expression of all CFP-1 constructs with the exception of DB 1–374 (Supplementary Figure S4B). We found that neither the N-terminal CXXC domain, nor the SID domain, were required for interaction with either SIN-3 or ATHP-1 (Figure 2B). The cysteine-rich C-terminal domain fragment interacted with SIN-3, and a larger fragment additionally containing the LZ domain was sufficient for interaction with ATHP-1. These results indicate that CFP-1 binds to SIN-3 through a region containing the cysteine-rich domain, and that interaction with ATHP-1 requires both this region and the LZ domain.

Using Y2H, we further showed that CFP-1 interacts with SIN-3 through its N-terminal region that contains the conserved PAH domain (Figure 2C), and with ATHP-1 through its C-terminal region (Figure 2D). Pull-down assays confirmed a direct interaction between the N-terminal fragment of SIN-3 and CFP-1 (Supplementary Figure S3H). Altogether, these results support the finding that CFP-1 physically interacts with the SIN3S complex.

Phenotypic similarity of SIN3S and SET-2/SET1 complex mutants

The physical interactions between CFP-1 and SIN3S complex components suggest that they may function in shared processes. To investigate this, we compared phenotypes of *set-2*, *cfp-1*, *sin-3* and *athp-1* mutants alone or in double mutant combinations, using null or strong loss of function alleles for all four genes (34,41,107). Similar to *set-2* mutants, we observed that *cfp-1, sin-3* and *athp-1* mutants also have reduced brood size at 20°C (32,34,41). *cfp-1* mutants showed extreme variability, with some animals showing a near-wild-type brood size, and others being completely sterile (Figure 3A). The brood size of *set-2; cfp-1* and *athp-1; cfp-1* double mutants is not reduced further compared to *cfp-1* single mutants, suggesting that SET-2 and ATHP-1 do not have CFP-1 independent fertility functions. However, *sin-3* fertility functions are independent or partially redundant with *set-2* and *cfp-1*, as brood size of *set-2; sin-3* is lower than that of the single mutants, and *cfp-1; sin-3*

double mutants showed a fully penetrant sterility that precluded scoring additional phenotypes (Figure 3A and data not shown). All single and double mutants also have a low level of embryonic lethality (Figure 3B).

set-2 mutants show transgenerational sterility at the stressful temperature of 25°C (32,38,41), and we found that *cfp-1, sin-3* and *athp-1* mutants also show this phenotype at 25°C. As expected, *set-2* mutants became sterile at generation F3–F4 (Figure 3C). We observed that *sin-3* and *athp-1* mutants become sterile at the F2–F3 generation, whereas the progeny of *cfp-1* mutants that were shifted to 25°C at the L4 stage were sterile (F1 generation).

We also observed that *cfp-1*, *set-2*, *sin-3* mutants have chromosome segregation defects in intestinal cells that become binucleate in the L1 stage (108). A similar trend was observed in *athp-1* mutants. Intestinal nuclei in adult animals were frequently connected by either thin or thick chromatin bridges in single mutants, and often completely failed to separate in *cfp-1* single, and *set-2;cfp-1*, *set-2;sin-3* and *athp-1;cfp-1* double mutants (Figure 3D). In summary, the similar phenotypes and genetic interactions suggest that SET-2/SET1 and SIN3 complexes are, at least in part, functionally linked in the germline and soma.

Similar steady-state gene expression changes in *set-2*, *cfp-1* and *sin-3* mutants

To ask whether SET-2/SET1 and SIN3 complexes have common roles in gene expression, we next performed RNA-sequencing (RNA-seq) on staged *cfp-1*, *set-2* and *sin-3* mutant embryos. Using DESeq2 (FDR < 0.05), we derived lists of differentially expressed genes in each mutant background, finding a similar number that were up- or down-regulated (Supplementary Table S4, Supplementary Figure S6). Consistent with the phenotypic similarities, we observed that gene expression changes detected in all three mutants show significant overlap (Figure 3E). Additionally, pairwise comparisons revealed gene expression changes shared only by *cfp-1* and *sin-3* mutants or only by *cfp-1* and *set-2* mutants, whereas *set-2* and *sin-3* do not show a specific association (Figure 3E). These shared patterns suggest that CFP-1 can act independently with COMPASS and SIN3S (Figure 3E).

Gene ontology (GO) term analysis showed enrichment for biological pathways related to translation, reproduction and development in all three mutant contexts (Supplementary Table 5). Downregulation of genes related to reproduction most likely reflects maternally inherited transcripts whose expression is altered in the germline of these mutants (41).

CFP-1 and SET-2/SET1 are needed for H3K4me3 at promoters

Previous studies showed that *cfp-1* or *set-2* inactivation results in greatly reduced global levels of H3K4me3 (32,34) (Supplementary Figure S7A). In addition, CFP-1 binding sites were shown to map to H3K4me3 marked promoters (70). To determine the roles of the two proteins on the pattern of H3K4me3 at CFP-1 sites, we compared H3K4me3 ChIP-seq signals in wildtype with those in *cfp-1* and *set-2*
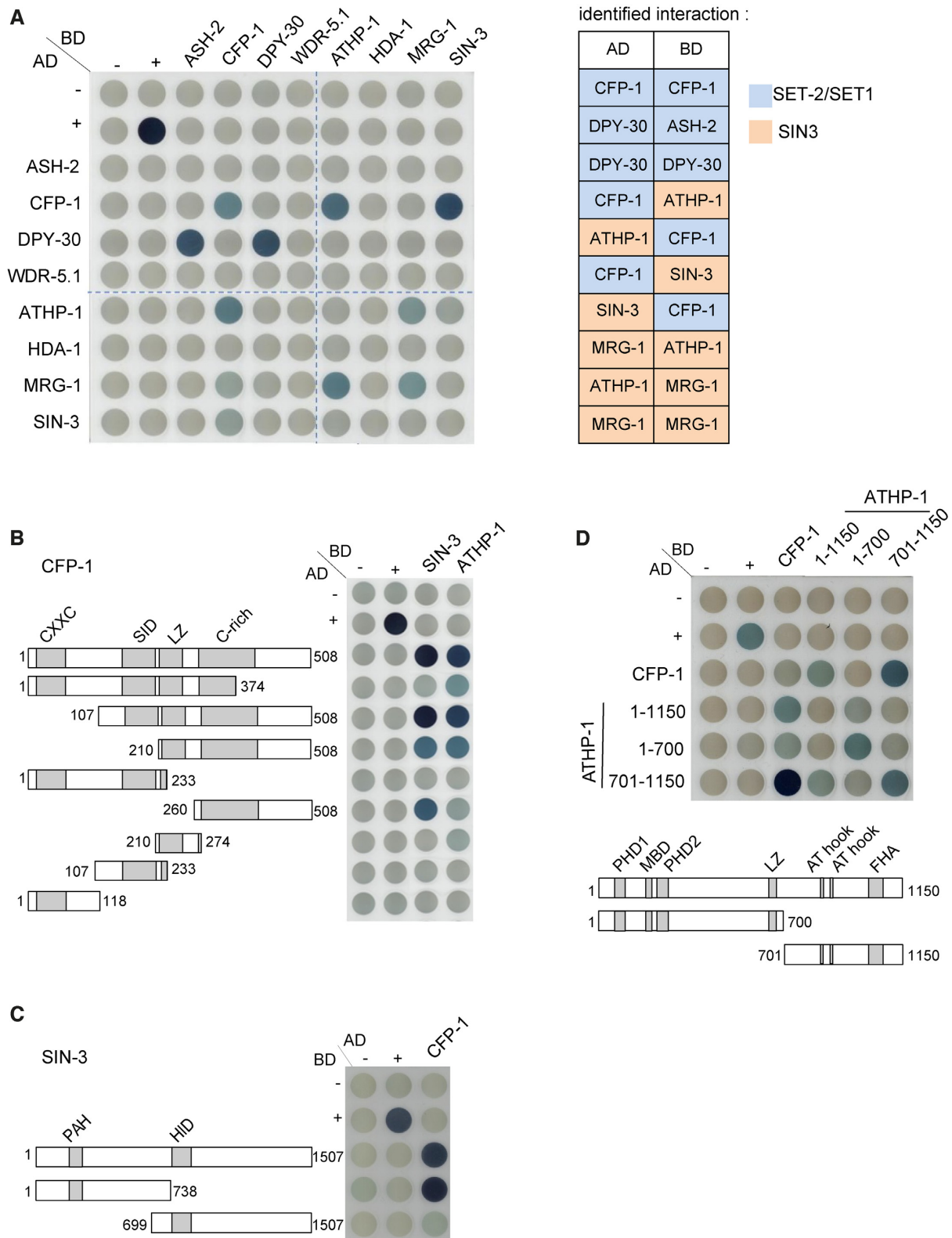
Figure 2. Interaction mapping between subunits of the SET-2/SET1 and SIN3S complexes by Y2H. (**A**) An interaction matrix was obtained by cross-mating of yeast haploid strains expressing different subunits as AD or BD fusion proteins. Positive control (+) is barnase/barstar interaction (133). Matings that lead to visually detectable staining in two independent experiments are reported in the tabular form on the right. Failure to detect the DB DPY-30/AD ASH-2 interaction is most likely due to DPY-30 homodimerization in the context of the DB domain fusion interfering with ASH-2 binding. (**B**) Interaction matrix of full length and truncated CFP-1 tested against full length SIN-3 and ATHP-1. CFP-1 truncations were constructed as BD and AD fusions, as indicated. (**C**, **D**) Interaction matrix of full length CFP-1 tested against N- and C-terminal fragments of SIN-3 (**C**) and ATHP-1 (**D**).
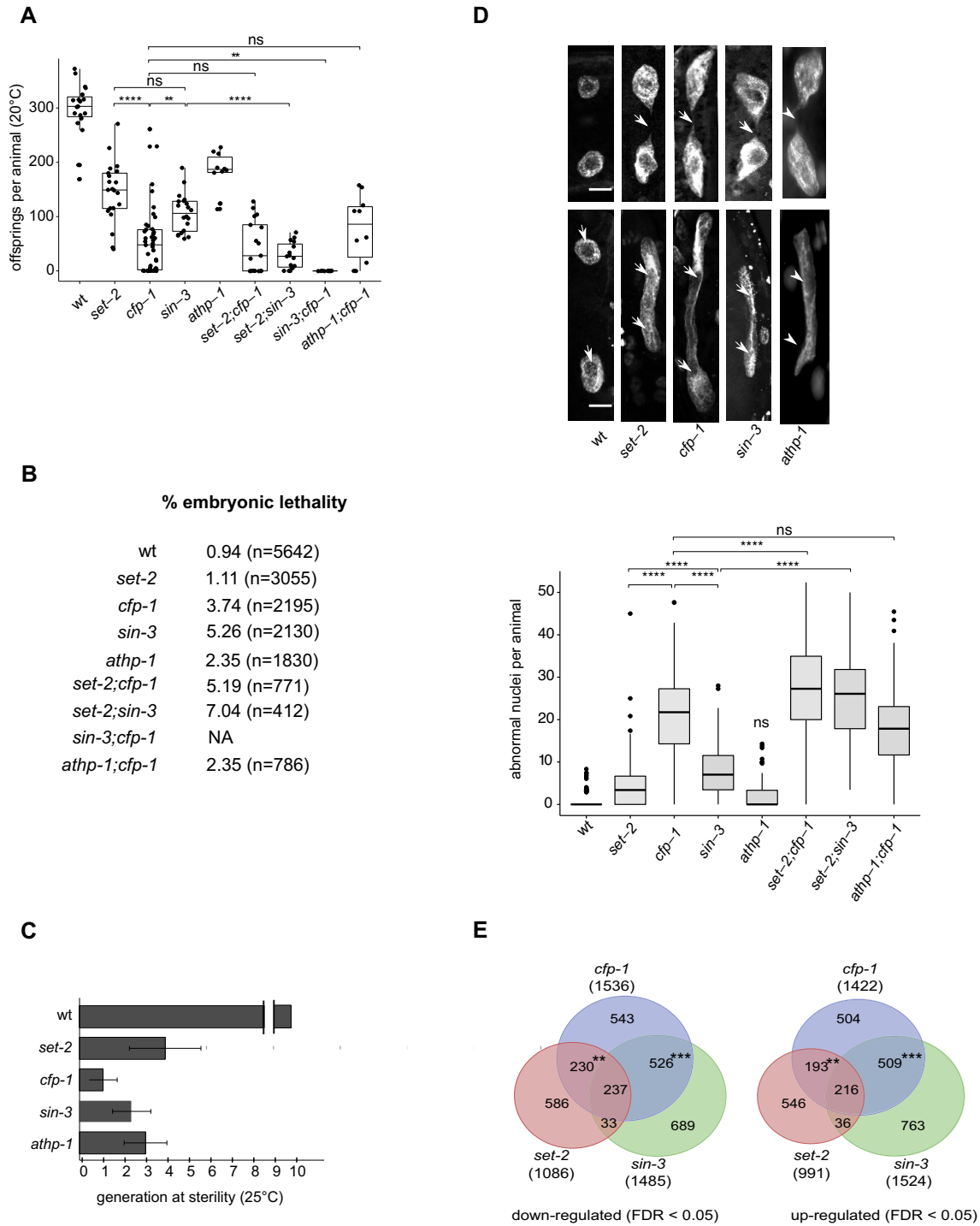
Figure 3. Loss of function of SET1/COMPASS and SIN3S complex components results in similar phenotypes and affects steady state gene expression. (**A**) Total number of progeny of single and double mutant animals of the given genotype. Multiple comparison was done using Wilcoxon (with Bonferroni correction) post hoc method following a significant Kruskal Wallis test; asterisks indicate a significant difference: $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$. (**B**) Embryonic lethality of single and double mutant strains; n = total number of embryos scored. (**C**) Fertility assays of *set-2, cfp-1, sin-3* and *athp-1* mutants grown at 25°C. Scoring was based on three to four biological replicates, with 6 independent lines each. Wildtype animals can be maintained for more that 40 generations without loss of fertility. (**D**) Confocal images of DAPI-stained intestinal nuclei from young adults. Examples of nuclear division abnormalities giving rise to thin chromatin bridges (arrows, top panel), or thick chromatin dense regions connecting two nuclei (arrowheads, bottom panel). Box plots show the total number of segregation defects (thin and thick chromatin bridges) per animal in single and double mutants of the given genotype ($n = 150$ worms for each strain). Multiple comparison was done using Wilcoxon (with Bonferroni correction) post hoc method following a significant Kruskal Wallis test as in (**A**). (**E**) Venn diagram showing the overlap between *cfp-1, set-2* and *sin-3* downregulated and upregulated genes. *P*-value for overlap between commonly misregulated genes in pairwise comparisons was calculated using Fisher's exact test in R: $**$*P*-value $< 10^{-10}$ and $***$*P*-value $< 10^{-100}$.

null mutant embryos, using a spike-in control for normalization. We observed strong reduction of H3K4me3 at CFP-1 sites in both *cfp-1* and *set-2* mutants (Figure 4A and Supplementary Figure S7B, see Methods). Using hierarchical clustering, we observed two classes of CFP-1 binding sites in wild-type embryos, both of which have reduced H3K4me3 in the two mutants. Sites with a high level of CFP-1 are strongly marked by H3K4me3, whereas sites with lower CFP-1 levels have low H3K4me3 marking (Figure 4A, B). We define the high H3K4me3 level sites as strong COMPASS targets, and the low level CFP-1 sites as weak COMPASS targets. The finding that the genomic distribution of H3K4me3 is similarly reduced in *cfp-1* and *set-2* mutants confirms that CFP-1 is needed for SET-2 activity at promoters.

Similar to findings in ES cells (1), we observed no clear relationship between gene expression changes in *cfp-1* mutants and promoter association of CFP-1. Both up- and down-regulated genes showed weak enrichment for CFP-1 binding, and only 2% of genes with CFP-1 peaks had significantly altered expression (Supplementary Figure S8A). The lack of a strong association between binding and gene expression change is consistent with evidence from various systems suggesting that additional factors, such as stress and age, influence the impact on transcription (11,109,110). For example, in yeast, SET1 acts primarily as a repressor of stress-induced transcription, with little effect of its loss under steady state conditions (9,11). In line with this, we found that *hsp-16.2::*GFP induction following heat shock was consistently stronger and more widespread in *cfp-1*, *sin-3* and *set-2* adult animals compared to wildtype, supporting a possible role in moderating the level of induced gene expression (Supplementary Figure S8B).

SIN3 complex components colocalize with CFP-1 at promoter regions

We next investigated how the distribution of SIN3 complex components relates to that of CFP-1. Using ChIP-seq analysis of SIN-3 in wild-type embryos, we observed that the pattern of SIN-3 binding was highly similar to that of CFP-1; 90% of SIN-3 peaks overlap a CFP-1 peak and 77% of these sites are found at promoters (Figures 5A–C, Supplementary Figure S9). In addition, as observed for CFP-1, SIN-3 levels are higher at strong COMPASS targets than at weak COMPASS targets (Figure 5B). We next determined the distribution of the SIN3 complex component HDA-1. We observed that HDA-1 was also present at most CFP-1 binding sites, with similar levels at strong and weak COMPASS targets (Figure 5A, B and Supplementary Figure S10). HDA-1 is additionally found at many sites that lack CFP-1 and SIN-3, presumably through its association with other proteins and complexes (Figure 5C) (97). Using previously published ChIP-seq data mapping MRG-1 in embryos (61), we observed weak enrichment at promoters and a broad distribution on the gene bodies of many actively transcribed genes (Supplementary Figure S9). In addition, 59% of sites harboring peaks of SIN-3, HDA-1 and CFP-1 ($n = 2707$) overlap an MRG-1 peak (Figure 5C). We also observed that SIN3 complex components SIN-3 and HDA-1 have a broader distribution than CFP-1 and are weakly enriched on gene bodies (Supplementary Figure S9). The finding that CFP-1 and SIN3 complex components extensively colocalize at promoter regions supports connected functions.

CFP-1 facilitates SIN-3 binding to H3K4me3 enriched promoter regions

The similarity in binding patterns together with our biochemical studies showing that CFP-1 physically associates with the SIN3S complex suggests a potential role in SIN3 chromatin recruitment. To investigate this possibility, we used ChIP-seq to map SIN-3 and HDA-1 binding in *cfp-1* mutant embryos. We observed that strong COMPASS targets had significantly reduced levels of both SIN-3 and HDA-1 in *cfp-1* mutants compared to wildtype (Figures 5A, B, D). In contrast weak COMPASS targets were largely unaffected (Figures 5A, B, D). HDA-1 sites that lack CFP-1 or SIN-3 binding and random genomic regions also showed no change in SIN-3 or HDA-1 levels in *cfp-1* mutants (Supplementary Figure S11). Together with the physical interaction results, we conclude that CFP-1 promotes recruitment of the SIN3 complex to strong COMPASS target sites.

DISCUSSION

In this study we identify a physical and functional interaction between CFP-1, the chromatin targeting subunit of the highly conserved SET1/COMPASS complex, and a SIN3S histone deacetylase complex similar to yeast Rpd3S and mouse SHMP containing SIN-3, HDA-1, MRG-1 and ATHP-1. We show that CFP-1 mediates interaction with the SIN3S complex through a direct interaction with SIN-3 and that it promotes recruitment of SIN-3 and HDA-1 at promoters. The interactions with SIN3S and other chromatin regulators identified by proteomics indicate that CFP-1 function extends beyond targeting the COMPASS complex to chromatin and support a role for CFP-1 in coordinating the activities of distinct chromatin complexes.

Our biochemical data also provide evidence for distinct SET-2/SET1 and SET-16/MLL related complexes in *C. elegans*, consistent with the presence of multiple H3K4 HMT complexes in metazoans (19). WDR-5.1 and CFP-1 both immunoprecipitated the core complex proteins RBBP-5, ASH-2, and DPY-30, as well as the SET1/COMPASS subunits SET-2/SET1 and SWD-2.1/WDR82 (111–113). However, WDR-5.1, but not CFP-1, immunoprecipitated unique subunits of the previously identified SET-16/MLL complex including the histone H3K27 demethylase UTX-1, and PIS-1 (74). WDR-5.1 also co-immunoprecipitated the NSL complex, consistent with its role as a central hub in several additional chromatin-associated complexes. Interestingly, in mammalian cells NSL has been shown to promote H3K4me2 activity by MLL1 (44), and we identified the single MLL1 homologue SET-16 with NSL subunits in our experiments, suggesting this activity may be conserved in *C. elegans*.

Y2H analyses showed that CFP-1 interacts with both SIN-3 and ATHP-1 subunits of the SIN-3 complex. We observed a direct interaction between CFP-1 and SIN-3 that is dependent on the C-terminus of CFP-1 containing the conserved cysteine-rich domain, and the N-terminal domain of
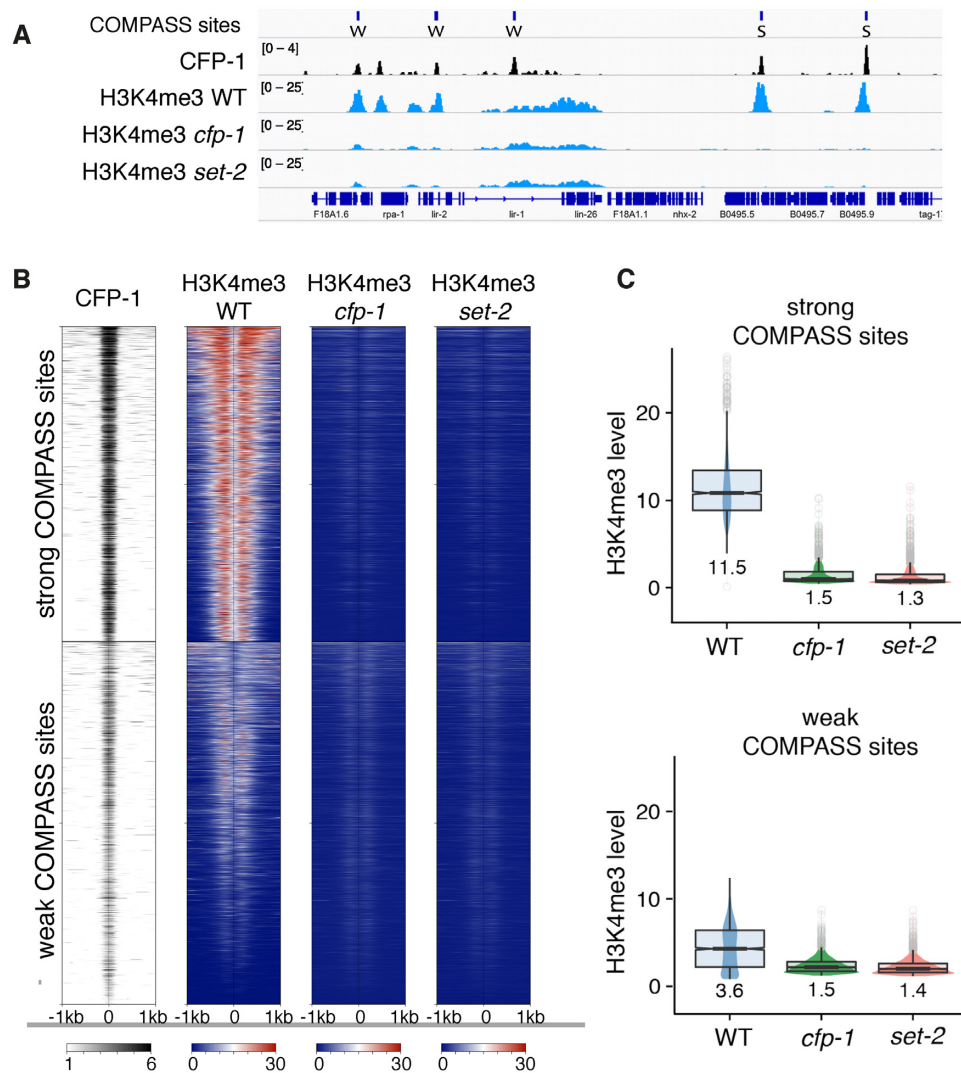
Figure 4. H3K4me3 at strong and weak COMPASS targets is dependent on CFP-1 and SET-2. (**A**) IGV browser view showing z-score BEADS normalized ChIP-seq signals of CFP-1::GFP, and H3K4me3 in wildtype, *cfp-1*, and *set-2* mutant embryos normalized to *C. briggsae* spike-in (see Methods). Top track shows locations of strong (S) and weak (W) COMPASS targets. (**B**) Heatmap of CFP-1::GFP, and H3K4me3 in wildtype, *cfp-1*, and *set-2* mutant embryos, using same tracks as in (A). (**C**) Quantification of H3K4me3 signal (*C. briggsae* normalized) in strong and weak COMPASS targets.

SIN-3 containing the highly conserved PAH domain. The PAH1 and PAH2 domains of mammalian SIN3 have been shown to facilitate SIN3 recruitment by transcription factors (95), however little is known about the function of the cysteine-rich C-terminal domain of mammalian CFP1. Although the C-terminal region of ATHP-1 that interacts with CFP-1 is not found in either yeast or mammalian proteins, it contains an FHA domain shown to mediate interaction between other transcription factors and the SIN3 complex in mammalian cells (114), raising the possibility that other proteins may fulfill a similar function in other species. Y2H analyses also confirmed physical interactions within SIN3 and SET-2/SET1 complexes and showed CFP-1 homodimerization, supporting studies suggesting dimerization of CFP1 within the SET1A/B complexes in human cells (28).

Supporting a functional link between COMPASS and SIN3S complexes, we observed similar germline and somatic phenotypes in *set-2, cfp-1, sin-3* and *athp-1* mutant animals, as well as shared gene expression changes between *set-2, cfp-1* and *sin-3*. We also found that CFP-1 regulates gene expression independently with SET-2 and SIN-3. These findings, together with the physical association and promoter co-occupancy of CFP-1 and SIN-3, suggest that CFP-1 may modulate gene expression by independently recruiting SET-2/SET1 and SIN3S complexes to promoters. Interestingly, the yeast CFP1 homologue Spp1 acts at meiotic recombination sites independently of *Set1*, and is found in at least two distinct Set1-independent complexes (115–117). COMPASS independent functions of Spp1 are also supported by studies showing that dissociation of Spp1 from SET1/COMPASS repurposed its function to promote transcriptional memory (118).

Previous genetic studies further support both unique and common functions for components of these complexes and other proteins isolated in our proteomics approach. For example, inactivation of the SIN3 complex subunits *sin-3* and
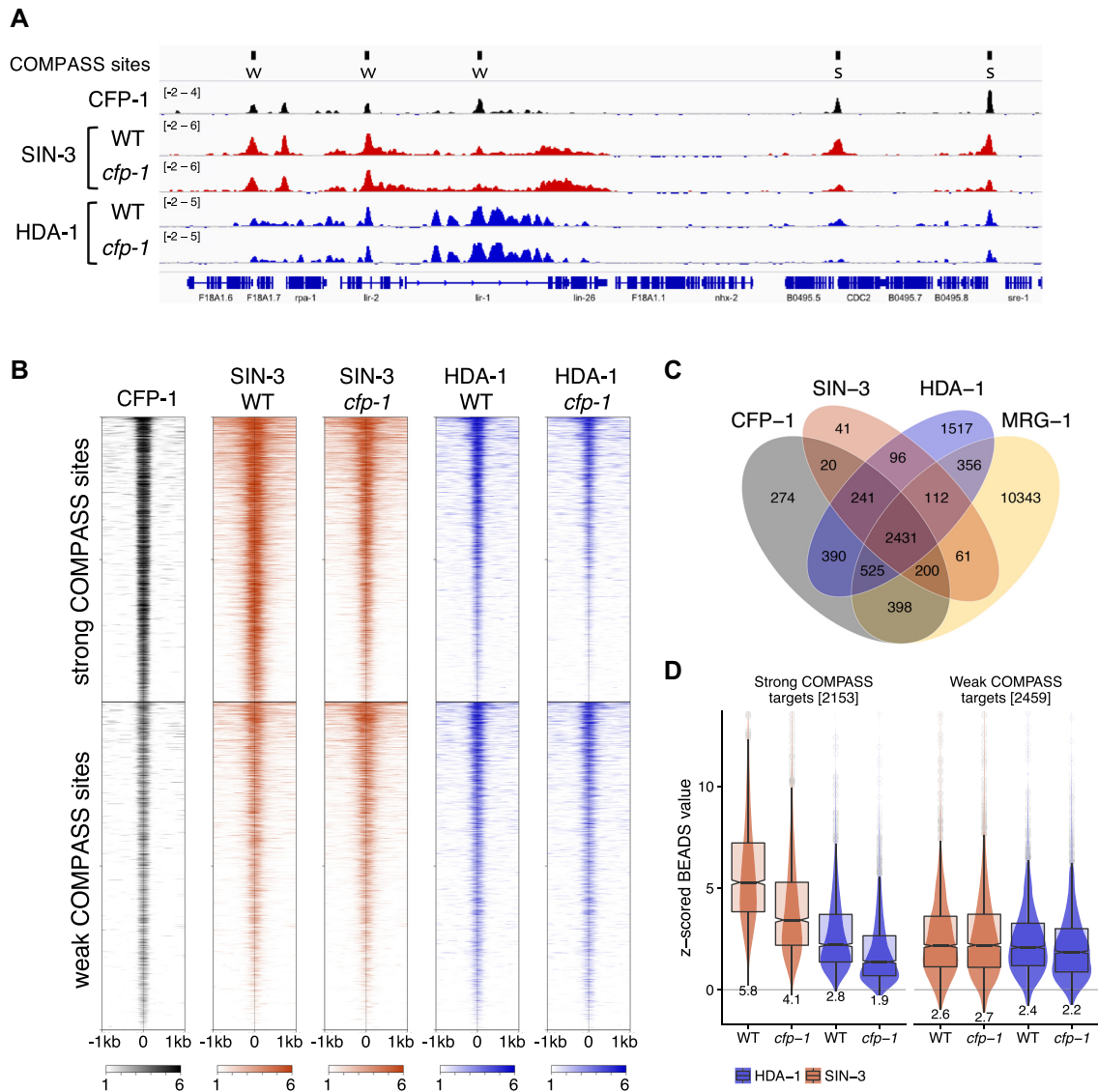
Figure 5. SIN-3 and HDA-1 require CFP-1 for recruitment to strong COMPASS targets. (**A**) IGV browser view showing *z*-scored BEADS normalized ChIP-seq signals from mixed embryos of indicated strains. Top track shows locations of strong (S) and weak (W) COMPASS targets. (**B**) Heatmap of *z*-scored BEADS normalized ChIP-seq signals from mixed embryos over strong and weak COMPASS targets in indicated strains. (**C**) Venn diagram showing overlap of CFP-1, SIN-3, HDA-1 and MRG-1 ChIP-seq peaks. (**D**) Quantification of normalized z-scored SIN-3 and HDA-1 signal in wildtype and *cfp-1* mutants at strong and weak COMPASS targets.

mrg-1, the NSL complex subunits *sumv-1* and *sumv-2*, and the SET-2/SET1 complex subunits *cfp-1, wdr-5.1* and *dpy-30,* but not *set-2,* can all suppress the synthetic multivulval (SynMuv) phenotype resulting from mutations in repressive chromatin factors (119,120). A subset of these genes, including *sin-3, mrg-1, wdr-5.1* and *dpy-30,* but neither *set-2* nor *cfp-1,* also suppress the larval lethality resulting from inactivation of *lin-35*/Rb in a sensitized background (121). We found that CFP-1 promotes binding of SIN-3 and HDA-1 at strong COMPASS dependent promoters, consistent with direct recruitment dependent on the activity of the SET-2/SET1 complex. SIN-3 binding at weak COMPASS regions was not affected. Interestingly, in mammalian cells COMPASS and MLL complex subunits have been shown

to cofractionate with SIN3, further supporting a connection between the complexes (122,123).

A prevailing view is that the regulatory functions of both SET1/COMPASS and Rpd3/Sin3 complexes are context dependent, but these functions are not well understood (10,75). For example, knock-out of Sin3 in different systems results in both gene activation and repression (124–127), and we observed no clear relationship between gene expression changes and SIN-3 binding. Similarly, loss of CFP1 or SET1 in a wide range of different systems causes surprisingly few gene expression changes relative to the number of genes marked by H3K4me3, with no clear relationship between expression and marking under steady state or induced conditions (1,10,11,109,128,129). We also ob-

served no reproducible change in bulk acetylation in *sin-3*, *cfp-1* or *set-2* mutants (Supplementary Figure S7A). SIN-3 and CFP-1 could alter gene expression through transient changes in acetylation and methylation dynamics that cannot be detected in the context of dividing embryos (130). Alternatively, or in addition, properties such as local nucleosome density and dynamics could also be affected (131). Our finding that expression from a heat-shock inducible promoter increased in the absence of CFP-1, SIN-3 or SET-2 is consistent with proposed roles in response to external stimuli in yeast (11). Interestingly, recent data suggests that AMPK signaling may regulate SET-2/SET1 complex activity in response to stress (132). Future work on defined loci will be needed to understand these regulatory functions. Because of the high degree of conservation between mammalian and *C. elegans* SET1/COMPASS and Rpd3/Sin3 complexes, our finding that they functionally interact through a direct physical interaction with CFP-1 contributes towards understanding the complexity of interactions between chromatin associated proteins with distinct activities.

DATA AVAILABILITY

Data generated in this study are available from GEO: gene expression data, GSE110072; ChIP-seq data, GSE114715. Other ChIP-seq datasets used are CFP-1::GFP (GSE49870; Ref. (70)) and MRG-1 (GSE50333; Ref. from (61))

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Brown,D.A., Di Cerbo,V., Feldmann,A., Ahn,J., Ito,S., Blackledge,N.P., Nakayama,M., McClellan,M., Dimitrova,E., Turberfield,A.H. *et al.* (2017) The SET1 complex selects actively transcribed target genes via multivalent interaction with CpG Island Chromatin. *Cell Rep.*, **20**, 2313–2327.
2. Clouaire,T., Webb,S., Skene,P., Illingworth,R., Kerr,A., Andrews,R., Lee,J.-H., Skalnik,D. and Bird,A. (2012) Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes Dev.*, **26**, 1714–1728.
3. Lee,J.H. and Skalnik,D.G. (2005) CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *J. Biol. Chem.*, **280**, 41725–41731.
4. Mahadevan,J. and Skalnik,D.G. (2016) Efficient differentiation of murine embryonic stem cells requires the binding of CXXC finger protein 1 to DNA or methylated histone H3-Lys4. *Gene*, **594**, 1–9.
5. Thomson,J.P., Skene,P.J., Selfridge,J., Clouaire,T., Guy,J., Webb,S., Kerr,A.R.W., Deaton,A., Andrews,R., James,K.D. *et al.* (2010) CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature*, **464**, 1082–1086.
6. Bernstein,B.E., Kamal,M., Lindblad-Toh,K., Bekiranov,S., Bailey,D.K., Huebert,D.J., McMahon,S., Karlsson,E.K., Kulbokas,E.J. 3rd, Gingeras,T.R. *et al.* (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*, **120**, 169–181.
7. Heintzman,N.D., Stuart,R.K., Hon,G., Fu,Y., Ching,C.W., Hawkins,R.D., Barrera,L.O., Van Calcar,S., Qu,C., Ching,K.A. *et al.* (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.*, **39**, 311–318.
8. Schneider,R., Bannister,A.J., Myers,F.A., Thorne,A.W., Crane-Robinson,C. and Kouzarides,T. (2004) Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat. Cell Biol.*, **6**, 73–77.
9. Howe,F.S., Fischl,H., Murray,S.C. and Mellor,J. (2017) Is H3K4me3 instructive for transcription activation? *BioEssays News Rev. Mol. Cell. Dev. Biol.*, **39**, 1–12.
10. Clouaire,T., Webb,S. and Bird,A. (2014) Cfp1 is required for gene expression-dependent H3K4 trimethylation and H3K9 acetylation in embryonic stem cells. *Genome Biol.*, **15**, 451.
11. Weiner,A., Chen,H.V., Liu,C.L., Rahat,A., Klien,A., Soares,L., Gudipati,M., Pfeffner,J., Regev,A., Buratowski,S. *et al.* (2012) Systematic dissection of roles for chromatin regulators in a yeast stress response. *PLoS Biol.*, **10**, e1001369.
12. Lenstra,T.L., Benschop,J.J., Kim,T., Schulze,J.M., Brabers,N.A., Margaritis,T., van de Pasch,L.A., van Heesch,S.A., Brok,M.O., Groot Koerkamp,M.J. *et al.* (2011) The specificity and topology of chromatin interaction pathways in yeast. *Mol. Cell*, **42**, 536–549.
13. Margaritis,T., Oreal,V., Brabers,N., Maestroni,L., Vitaliano-Prunier,A., Benschop,J.J., van Hooff,S., van Leenen,D., Dargemont,C., Geli,V. *et al.* (2012) Two distinct repressive mechanisms for histone 3 lysine 4 methylation through promoting 3′-end antisense transcription. *PLos Genet.*, **8**, e1002952.
14. Hörmanseder,E., Simeone,A., Allen,G.E., Bradshaw,C.R., Figlmüller,M., Gurdon,J. and Jullien,J. (2017) H3K4 methylation-dependent memory of somatic cell identity inhibits reprogramming and development of nuclear transfer embryos. *Cell Stem Cell*, **21**, 135–143.
15. Fabrizio,P., Garvis,S. and Palladino,F. (2019) Histone methylation and memory of environmental stress. *Cells*, **8**, 339.
16. Lee,B.B., Choi,A., Kim,J.H., Jun,Y., Woo,H., Ha,S.D., Yoon,C.Y., Hwang,J.-T., Steinmetz,L., Buratowski,S. *et al.* (2018) Rpd3L HDAC links H3K4me3 to transcriptional repression memory. *Nucleic Acids Res.*, **46**, 8261–8274.
17. Miller,T., Krogan,N.J., Dover,J., Erdjument-Bromage,H., Tempst,P., Johnston,M., Greenblatt,J.F. and Shilatifard,A. (2001) COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 12902–12907.
18. Roguev,A., Schaft,D., Shevchenko,A., Pijnappel,W.W., Wilm,M., Aasland,R. and Stewart,A.F. (2001) The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *EMBO J.*, **20**, 7137–7148.
19. Shilatifard,A. (2012) The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annu. Rev. Biochem.*, **81**, 65–95.

20. Dehe,P.M., Dichtl,B., Schaft,D., Roguev,A., Pamblanco,M., Lebrun,R., Rodriguez-Gil,A., Mkandawire,M., Landsberg,K., Shevchenko,A. *et al.* (2006) Protein interactions within the Set1 complex and their roles in the regulation of histone 3 lysine 4 methylation. *J. Biol. Chem.*, **281**, 35404–35412.

21. Dou,Y., Milne,T.A., Ruthenburg,A.J., Lee,S., Lee,J.W., Verdine,G.L., Allis,C.D. and Roeder,R.G. (2006) Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat. Struct. Mol. Biol.*, **13**, 713–719.

22. Steward,M.M., Lee,J.S., O'Donovan,A., Wyatt,M., Bernstein,B.E. and Shilatifard,A. (2006) Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. *Nat. Struct. Mol. Biol.*, **13**, 852–854.

23. Lee,S., Lee,D.K., Dou,Y., Lee,J., Lee,B., Kwak,E., Kong,Y.Y., Lee,S.K., Roeder,R.G. and Lee,J.W. (2006) Coactivator as a target gene specificity determinant for histone H3 lysine 4 methyltransferases. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 15392–15397.

24. Narayanan,A., Ruyechan,W.T. and Kristie,T.M. (2007) The coactivator host cell factor-1 mediates Set1 and MLL1 H3K4 trimethylation at herpesvirus immediate early promoters for initiation of infection. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 10835–10840.

25. Tyagi,S., Chabes,A.L., Wysocka,J. and Herr,W. (2007) E2F activation of S phase promoters via association with HCF-1 and the MLL family of histone H3K4 methyltransferases. *Mol. Cell*, **27**, 107–119.

26. Bledau,A.S., Schmidt,K., Neumann,K., Hill,U., Ciotta,G., Gupta,A., Torres,D.C., Fu,J., Kranz,A., Stewart,A.F. *et al.* (2014) The H3K4 methyltransferase Setd1a is first required at the epiblast stage, whereas Setd1b becomes essential after gastrulation. *Development*, **141**, 1022–1035.

27. Glaser,S., Schaft,J., Lubitz,S., Vintersten,K., van der Hoeven,F., Tufteland,K.R., Aasland,R., Anastassiadis,K., Ang,S.L. and Stewart,A.F. (2006) Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Development*, **133**, 1423–1432.

28. Lee,J., Saha,P.K., Yang,Q.H., Lee,S., Park,J.Y., Suh,Y., Lee,S.K., Chan,L., Roeder,R.G. and Lee,J.W. (2008) Targeted inactivation of MLL3 histone H3-Lys-4 methyltransferase activity in the mouse reveals vital roles for MLL3 in adipogenesis. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 19229–19234.

29. Yu,B.D., Hess,J.L., Horning,S.E., Brown,G.A. and Korsmeyer,S.J. (1995) Altered Hox expression and segmental identity in Mll-mutant mice. *Nature*, **378**, 505–508.

30. Ardehali,M.B., Mei,A., Zobeck,K.L., Caron,M., Lis,J.T. and Kusch,T. (2011) *Drosophila* Set1 is the major histone H3 lysine 4 trimethyltransferase with role in transcription: dSet1 is the major H3K4 trimethyltransferase. *EMBO J.*, **30**, 2817–2828.

31. Hallson,G., Hollebakken,R.E., Li,T., Syrzycka,M., Kim,I., Cotsworth,S., Fitzpatrick,K.A., Sinclair,D.A.R. and Honda,B.M. (2012) dSet1 is the main H3K4 di- and tri-methyltransferase throughout Drosophila development. *Genetics*, **190**, 91–100.

32. Li,T. and Kelly,W.G. (2011) A role for Set1/MLL-related components in epigenetic regulation of the Caenorhabditis elegans germ line. *PLos Genet.*, **7**, e1001349.

33. Wu,M., Wang,P.F., Lee,J.S., Martin-Brown,S., Florens,L., Washburn,M. and Shilatifard,A. (2008) Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. *Mol. Cell. Biol.*, **28**, 7337–7344.

34. Xiao,Y., Bedet,C., Robert,V.J., Simonet,T., Dunkelbarger,S., Rakotomalala,C., Soete,G., Korswagen,H.C., Strome,S. and Palladino,F. (2011) Caenorhabditis elegans chromatin-associated proteins SET-2 and ASH-2 are differentially required for histone H3 Lys 4 methylation in embryos and adult germ cells. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 8305–8310.

35. Denissov,S., Hofemeister,H., Marks,H., Kranz,A., Ciotta,G., Singh,S., Anastassiadis,K., Stunnenberg,H.G. and Stewart,A.F. (2014) Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. *Dev. Camb. Engl.*, **141**, 526–537.

36. Hu,D., Gao,X., Morgan,M.A., Herz,H.M., Smith,E.R. and Shilatifard,A. (2013) The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. *Mol. Cell. Biol.*, **33**, 4745–4754.

37. Wang,P., Lin,C., Smith,E.R., Guo,H., Sanderson,B.W., Wu,M., Gogol,M., Alexander,T., Seidel,C., Wiedemann,L.M. *et al.* (2009) Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. *Mol. Cell. Biol.*, **29**, 6074–6085.

38. Simonet,T., Dulermo,R., Schott,S. and Palladino,F. (2007) Antagonistic functions of SET-2/SET1 and HPL/HP1 proteins in C. elegans development. *Dev. Biol.*, **312**, 367–383.

39. Fisher,K., Southall,S.M., Wilson,J.R. and Poulin,G.B. (2010) Methylation and demethylation activities of a C. elegans MLL-like complex attenuate RAS signalling. *Dev. Biol.*, **341**, 142–153.

40. Greer,E.L., Maures,T.J., Hauswirth,A.G., Green,E.M., Leeman,D.S., Maro,G.S., Han,S., Banko,M.R., Gozani,O. and Brunet,A. (2010) Members of the H3K4 trimethylation complex regulate lifespan in a germline-dependent manner in C. elegans. *Nature*, **466**, 383–387.

41. Robert,V.J., Mercier,M.G., Bedet,C., Janczarski,S., Merlet,J., Garvis,S., Ciosk,R. and Palladino,F. (2014) The SET-2/SET1 histone H3K4 methyltransferase maintains pluripotency in the Caenorhabditis elegans germline. *Cell Rep.*, **9**, 443–450.

42. Xu,L. and Strome,S. (2001) Depletion of a novel SET-domain protein enhances the sterility of mes-3 and mes-4 mutants of Caenorhabditis elegans. *Genetics*, **159**, 1019–1029.

43. Han,S., Schroeder,E.A., Silva-García,C.G., Hebestreit,K., Mair,W.B. and Brunet,A. (2017) Mono-unsaturated fatty acids link H3K4me3 modifiers to C. elegans lifespan. *Nature*, **544**, 185–190.

44. Cai,Y., Jin,J., Swanson,S.K., Cole,M.D., Choi,S.H., Florens,L., Washburn,M.P., Conaway,J.W. and Conaway,R.C. (2010) Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the male-specific lethal (MSL) complex. *J. Biol. Chem.*, **285**, 4268–4272.

45. Cho,Y.-W., Hong,T., Hong,S., Guo,H., Yu,H., Kim,D., Guszczynski,T., Dressler,G.R., Copeland,T.D., Kalkum,M. *et al.* (2007) PTIP Associates with MLL3- and MLL4-containing Histone H3 Lysine 4 Methyltransferase Complex. *J. Biol. Chem.*, **282**, 20395–20406.

46. Suganuma,T., Gutiérrez,J.L., Li,B., Florens,L., Swanson,S.K., Washburn,M.P., Abmayr,S.M. and Workman,J.L. (2008) ATAC is a double histone acetyltransferase complex that stimulates nucleosome sliding. *Nat. Struct. Mol. Biol.*, **15**, 364–372.

47. Brenner,S. (1974) The genetics of caenorhabditis elegans. *Genetics*, **77**, 71–94.

48. Casabona,M.G., Vandenbrouck,Y., Attree,I. and Couté,Y. (2013) Proteomic characterization of Pseudomonas aeruginosa PAO1 inner membrane. *Proteomics*, **13**, 2419–2423.

49. Golemis,E.A. and Brent,R. (1992) Fused protein domains inhibit DNA binding by LexA. *Mol. Cell. Biol.*, **12**, 3006–3014.

50. Gibson,D.G. (2011) Enzymatic assembly of overlapping DNA Fragments. In: *Methods in Enzymology*. Elsevier, Vol. **498**, pp. 349–361.

51. Estojak,J., Brent,R. and Golemis,E.A. (1995) Correlation of two-hybrid affinity data with in vitro measurements. *Mol. Cell. Biol.*, **15**, 5820–5829.

52. Ito,T., Chiba,T., Ozawa,R., Yoshida,M., Hattori,M. and Sakaki,Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 4569–4574.

53. Kolonin,M.G., Zhong,J. and Finley,R.L. (2000) Interaction mating methods in two-hybrid systems. *Methods Enzymol.*, **328**, 26–46.

54. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.

55. Link,C.D., Cypser,J.R., Johnson,C.J. and Johnson,T.E. (1999) Direct observation of stress response in Caenorhabditis elegans using a reporter transgene. *Cell Stress Chaperones*, **4**, 235–242.

56. McMurchy,A.N., Stempor,P., Gaarenstroom,T., Wysolmerski,B., Dong,Y., Aussianikava,D., Appert,A., Huang,N., Kolasinska-Zwierz,P., Sapetschnig,A. *et al.* (2017) A team of heterochromatin factors collaborates with small RNA pathways to combat repetitive elements and germline stress. *eLife*, **6**, e21666.

57. Li,H. and Durbin,R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.*, **26**, 589–595.

58. Cheung,M.-S., Down,T.A., Latorre,I. and Ahringer,J. (2011) Systematic bias in high-throughput sequencing data and its correction by BEADS. *Nucleic Acids Res.*, **39**, e103.

59. Feng,J., Liu,T., Qin,B., Zhang,Y. and Liu,X.S. (2012) Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.*, **7**, 1728–1740.

60. Derrien,T., Estellé,J., Marco Sola,S., Knowles,D.G., Raineri,E., Guigó,R. and Ribeca,P. (2012) Fast computation and applications of genome mappability. *PLoS One*, **7**, e30377.

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

62. Stempor,P. and Ahringer,J. (2016) SeqPlots - Interactive software for exploratory data analyses, pattern discovery and visualization in genomics. *Wellcome Open Res.*, **1**, 14.

63. Thorvaldsdóttir,H., Robinson,J.T. and Mesirov,J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.*, **14**, 178–192.

64. Janes,J., Dong,Y., Schoof,M., Serizay,J., Appert,A., Cerrato,C., Woodbury,C., Chen,R., Gemma,C., Huang,N. *et al.* (2018) Chromatin accessibility dynamics across development and ageing. *eLife*, **7**, e37344.

65. Kent,W.J., Zweig,A.S., Barber,G., Hinrichs,A.S. and Karolchik,D. (2010) BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinforma. Oxf. Engl.*, **26**, 2204–2207.

66. Tukey,J.W. (1977) Some thoughts on clinical trials, especially problems of multiplicity. *Science*, **198**, 679–684.

67. Cosgrove,M.S and Patel,A. (2010). Mixed lineage leukemia: a structure-function perspective of the MLL1 protein. *FEBS J.*, **277**, 1832–1842.

68. Takahashi,Y. -h., Westfield,G.H., Oleskie,A.N., Trievel,R.C., Shilatifard,A. and Skiniotis,G. (2011) Structural analysis of the core COMPASS family of histone H3K4 methylases from yeast to human. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 20526–20531.

69. Lee,J.H. and Skalnik,D.G. (2008) Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A Histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes. *Mol. Cell. Biol.*, **28**, 609–618.

70. Chen,R.A.-J., Stempor,P., Down,T.A., Zeiser,E., Feuer,S.K. and Ahringer,J. (2014) Extreme HOT regions are CpG-dense promoters in C. elegans and humans. *Genome Res.*, **24**, 1138–1146.

71. van Nuland,R., Smits,A.H., Pallaki,P., Jansen,P.W., Vermeulen,M. and Timmers,H.T. (2013) Quantitative dissection and stoichiometry determination of the human SET1/MLL histone methyltransferase complexes. *Mol. Cell. Biol.*, **33**, 2067–2077.

72. Patel,A., Dharmarajan,V., Vought,V.E. and Cosgrove,M.S. (2009) On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex. *J. Biol. Chem.*, **284**, 24242–24256.

73. Zargar,Z.U. and Tyagi,S. (2012) Role of Host Cell Factor-1 in cell cycle regulation. *Transcription*, **3**, 187–192.

74. Vandamme,J., Lettier,G., Sidoli,S., Di Schiavi,E., Norregaard Jensen,O. and Salcini,A.E. (2012) The C. elegans H3K27 demethylase UTX-1 is essential for normal development, independent of its enzymatic activity. *PLos Genet.*, **8**, e1002647.

75. Lee,J.H., Tate,C.M., You,J.S. and Skalnik,D.G. (2007) Identification and Characterization of the Human Set1B Histone H3-Lys4 Methyltransferase Complex. *J. Biol. Chem.*, **282**, 13419–13428.

76. Dias,J., Van Nguyen,N., Georgiev,P., Gaub,A., Brettschneider,J., Cusack,S., Kadlec,J. and Akhtar,A. (2014) Structural analysis of the KANSL1/WDR5/KANSL2 complex reveals that WDR5 is required for efficient assembly and chromatin targeting of the NSL complex. *Genes Dev.*, **28**, 929–942.

77. Milne,T.A., Dou,Y., Martin,M.E., Brock,H.W., Roeder,R.G. and Hess,J.L. (2005) MLL associates specifically with a subset of transcriptionally active target genes. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 14765–14770.

78. Raja,S.J., Charapitsa,I., Conrad,T., Vaquerizas,J.M., Gebhardt,P., Holz,H., Kadlec,J., Fraterman,S., Luscombe,N.M. and Akhtar,A. (2010) The nonspecific lethal complex is a transcriptional regulator in Drosophila. *Mol. Cell*, **38**, 827–841.

79. Zhao,X., Su,J., Wang,F., Liu,D., Ding,J., Yang,Y., Conaway,J.W., Conaway,R.C., Cao,L., Wu,D. *et al.* (2013) Crosstalk between NSL histone acetyltransferase and MLL/SET complexes: NSL complex functions in promoting histone H3K4 Di-methylation activity by MLL/SET complexes. *PLoS Genet.*, **9**, e1003940.

80. Hoe,M. and Nicholas,H.R. (2014) Evidence of a MOF histone acetyltransferase-containing NSL complex in *C. elegans*. *Worm*, **3**, e982967.

81. Carrozza,M.J., Li,B., Florens,L., Suganuma,T., Swanson,S.K., Lee,K.K., Shia,W.-J., Anderson,S., Yates,J., Washburn,M.P. *et al.* (2005) Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell*, **123**, 581–592.

82. Jelinic,P., Pellegrino,J. and David,G. (2011) A novel mammalian complex containing Sin3B mitigates histone acetylation and RNA polymerase II progression within transcribed loci. *Mol. Cell. Biol.*, **31**, 54–62.

83. Alland,L., David,G., Shen-Li,H., Potes,J., Muhle,R., Lee,H.-C., Hou,H., Chen,K. and DePinho,R.A. (2002) Identification of mammalian Sds3 as an integral component of the Sin3/histone deacetylase corepressor complex. *Mol. Cell. Biol.*, **22**, 2743–2750.

84. Fleischer,T.C., Yun,U.J. and Ayer,D.E. (2003) Identification and characterization of three new components of the mSin3A corepressor complex. *Mol. Cell. Biol.*, **23**, 3456–3467.

85. Lechner,T., Carrozza,M.J., Yu,Y., Grant,P.A., Eberharter,A., Vannier,D., Brosch,G., Stillman,D.J., Shore,D. and Workman,J.L. (2000) Sds3 (suppressor of defective silencing 3) is an integral component of the yeast Sin3[middle dot]Rpd3 histone deacetylase complex and is required for histone deacetylase activity. *J. Biol. Chem.*, **275**, 40961–40966.

86. Sardiu,M.E., Smith,K.T., Groppe,B.D., Gilmore,J.M., Saraf,A., Egidy,R., Peak,A., Seidel,C.W., Florens,L., Workman,J.L. *et al.* (2014) Suberoylanilide hydroxamic acid (SAHA)-induced dynamics of a human histone deacetylase protein interaction network. *Mol. Cell. Proteomics MCP*, **13**, 3114–3125.

87. Spain,M.M., Caruso,J.A., Swaminathan,A. and Pile,L.A. (2010) *Drosophila* SIN3 isoforms interact with distinct proteins and have unique biological functions. *J. Biol. Chem.*, **285**, 27457–27467.

88. Churchman,L.S. and Weissman,J.S. (2011) Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature*, **469**, 368–373.

89. Kadamb,R., Mittal,S., Bansal,N., Batra,H. and Saluja,D. (2013) Sin3: insight into its transcription regulatory functions. *Eur. J. Cell Biol.*, **92**, 237–246.

90. Laherty,C.D., Yang,W.M., Sun,J.M., Davie,J.R., Seto,E. and Eisenman,R.N. (1997) Histone deacetylases associated with the mSin3 corepressor mediate mad transcriptional repression. *Cell*, **89**, 349–356.

91. Ayer,D.E., Lawrence,Q.A. and Eisenman,R.N. (1995) Mad-Max transcriptional repression is mediated by ternary complex formation with mammalian homologs of yeast repressor Sin3. *Cell*, **80**, 767–776.

92. Cowley,S.M., Iritani,B.M., Mendrysa,S.M., Xu,T., Cheng,P.F., Yada,J., Liggitt,H.D. and Eisenman,R.N. (2005) The mSin3A chromatin-modifying complex is essential for embryogenesis and T-cell development. *Mol. Cell. Biol.*, **25**, 6990–7004.

93. Dannenberg,J.-H., David,G., Zhong,S., van der Torre,J. and Wong,W.H. Depinho, R.A. (2005) mSin3A corepressor regulates diverse transcriptional networks governing normal and neoplastic growth and survival. *Genes Dev.*, **19**, 1581–1595.

94. David,G., Grandinetti,K.B., Finnerty,P.M., Simpson,N., Chu,G.C. and Depinho,R.A. (2008) Specific requirement of the chromatin modifier mSin3B in cell cycle exit and cellular differentiation. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 4168–4172.

95. Sahu,S.C., Swanson,K.A., Kang,R.S., Huang,K., Brubaker,K., Ratcliff,K. and Radhakrishnan,I. (2008) Conserved themes in target recognition by the PAH1 and PAH2 domains of the Sin3 transcriptional corepressor. *J. Mol. Biol.*, **375**, 1444–1456.

96. Kelly,R.D.W. and Cowley,S.M. (2013) The physiological roles of histone deacetylase (HDAC) 1 and 2: complex co-stars with multiple leading parts. *Biochem. Soc. Trans.*, **41**, 741–749.

97. Passannante,M., Marti,C.-O., Pfefferli,C., Moroni,P.S., Kaeser-Pebernard,S., Puoti,A., Hunziker,P., Wicky,C. and Müller,F. (2010) Different Mi-2 complexes for various developmental functions in Caenorhabditis elegans. *PLoS One*, **5**, e13681.

98. Hajduskova,M., Baytek,G., Kolundzic,E., Gosdschan,A., Kazmierczak,M., Ofenbauer,A., Beato Del Rosal,M.L., Herzog,S.,

Ul Fatima,N., Mertins,P. *et al.* (2018) MRG-1/MRG15 is a barrier for germ cell to neuron reprogramming in caenorhabditis elegans. *Genetics*, **211**, 121–139.

99. Bleuyard,J.-Y., Fournier,M., Nakato,R., Couturier,A.M., Katou,Y., Ralf,C., Hester,S.S., Dominguez,D., Rhodes,D., Humphrey,T.C. *et al.* (2017) MRG15-mediated tethering of PALB2 to unperturbed chromatin protects active genes from genotoxic stress. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 7671–7676.

100. Chen,M., Tominaga,K. and Pereira-Smith,O.M. (2010) Emerging role of the MORF/MRG gene family in various biological processes, including aging. *Ann. N. Y. Acad. Sci.*, **1197**, 134–141.

101. Huang,C., Yang,F., Zhang,Z., Zhang,J., Cai,G., Li,L., Zheng,Y., Chen,S., Xi,R. and Zhu,B. (2017) Mrg15 stimulates Ash1 H3K36 methyltransferase activity and facilitates Ash1 Trithorax group protein function in Drosophila. *Nat. Commun.*, **8**, 1649.

102. Iwamori,N., Tominaga,K., Sato,T., Riehle,K., Iwamori,T., Ohkawa,Y., Coarfa,C., Ono,E. and Matzuk,M.M. (2016) MRG15 is required for pre-mRNA splicing and spermatogenesis. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, E5408–E5415.

103. Smith,H.F., Roberts,M.A., Nguyen,H.Q., Peterson,M., Hartl,T.A., Wang,X.-J., Klebba,J.E., Rogers,G.C. and Bosco,G. (2013) Maintenance of interphase chromosome compaction and homolog pairing in Drosophila is regulated by the condensin cap-h2 and its partner Mrg15. *Genetics*, **195**, 127–146.

104. Butler,J.S., Lee,J.-H. and Skalnik,D.G. (2008) CFP1 interacts with DNMT1 independently of association with the Setd1 Histone H3K4 methyltransferase complexes. *DNA Cell Biol.*, **27**, 533–543.

105. Tate,C.M., Lee,J.-H. and Skalnik,D.G. (2009) CXXC finger protein 1 contains redundant functional domains that support embryonic stem cell cytosine methylation, histone methylation, and differentiation. *Mol. Cell. Biol.*, **29**, 3817–3831.

106. Voo,K.S., Carlone,D.L., Jacobsen,B.M., Flodin,A. and Skalnik,D.G. (2000) Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. *Mol. Cell. Biol.*, **20**, 2108–2121.

107. Choy,S.W., Wong,Y.M., Ho,S.H. and Chow,K.L. (2007) C. elegans SIN-3 and its associated HDAC corepressor complex act as mediators of male sensory ray development. *Biochem. Biophys. Res. Commun.*, **358**, 802–807.

108. Hedgecock,E.M. and White,J.G. (1985) Polyploid tissues in the nematode Caenorhabditis elegans. *Dev. Biol.*, **107**, 128–133.

109. Pu,M., Wang,M., Wang,W., Velayudhan,S.S. and Lee,S.S. (2018) Unique patterns of trimethylation of histone H3 lysine 4 are prone to changes during aging in Caenorhabditis elegans somatic cells. *PLoS Genet.*, **14**, e1007466.

110. Cruz,C., Della Rosa,M., Krueger,C., Gao,Q., Horkai,D., King,M., Field,L. and Houseley,J. (2018) Tri-methylation of histone H3 lysine 4 facilitates gene expression in ageing cells. *eLife*, **7**, e34081.

111. Dou,Y., Milne,T.A., Tackett,A.J., Smith,E.R., Fukuda,A., Wysocka,J., Allis,C.D., Chait,B.T., Hess,J.L. and Roeder,R.G. (2005) Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell*, **121**, 873–885.

112. Patel,A., Vought,V.E., Dharmarajan,V. and Cosgrove,M.S. (2011) A novel non-set domain multi-subunit methyltransferase required for sequential nucleosomal histone H3 methylation by the mixed lineage leukemia protein-1 (MLL1) core complex. *J. Biol. Chem.*, **286**, 3359–3569.

113. Shinsky,S.A., Monteith,K.E., Viggiano,S. and Cosgrove,M.S. (2015) Biochemical Reconstitution and Phylogenetic Comparison of Human SET1 Family Core Complexes Involved in Histone Methylation. *J. Biol. Chem.*, **290**, 6361–6375.

114. Shi,X., Seldin,D.C. and Garry,D.J. (2012) Foxk1 recruits the Sds3 complex and represses gene expression in myogenic progenitors. *Biochem. J.*, **446**, 349–357.

115. Karányi,Z., Halász,L., Acquaviva,L., Jónás,D., Hetey,S., Boros-Oláh,B., Peng,F., Chen,D., Klein,F., Géli,V. *et al.* (2018) Nuclear dynamics of the Set1C subunit Spp1 prepares meiotic recombination sites for break formation. *J. Cell Biol.*, **217**, 3398–3415.

116. Acquaviva,L., Székvölgyi,L., Dichtl,B., Dichtl,B.S., de La Roche Saint André,C., Nicolas,A. and Géli,V. (2013) The COMPASS subunit Spp1 links histone methylation to initiation of meiotic recombination. *Science*, **339**, 215–218.

117. Adam,C., Guérois,R., Citarella,A., Verardi,L., Adolphe,F., Béneut,C., Sommermeyer,V., Ramus,C., Govin,J., Couté,Y. *et al.* (2018) The PHD finger protein Spp1 has distinct functions in the Set1 and the meiotic DSB formation complexes. *PLoS Genet.*, **14**, e1007223.

118. D'Urso,A., Takahashi,Y.-H., Xiong,B., Marone,J., Coukos,R., Randise-Hinchliff,C., Wang,J.-P., Shilatifard,A. and Brickner,J.H. (2016) Set1/COMPASS and Mediator are repurposed to promote epigenetic transcriptional memory. *eLife*, **5**, e16691.

119. Cui,M., Kim,E.B. and Han,M. (2006) Diverse chromatin remodeling genes antagonize the Rb-involved SynMuv pathways in C. elegans. *PLos Genet.*, **2**, e74.

120. Yücel,D., Hoe,M., Llamosas,E., Kant,S., Jamieson,C., Young,P.A., Crossley,M. and Nicholas,H.R. (2014) SUMV-1 antagonizes the activity of synthetic multivulva genes in Caenorhabditis elegans. *Dev. Biol.*, **392**, 266–282.

121. Fay,D.S. and Yochem,J. (2007) The SynMuv genes of Caenorhabditis elegans in vulval development and beyond. *Dev. Biol.*, **306**, 1–9.

122. Wysocka,J., Myers,M.P., Laherty,C.D., Eisenman,R.N. and Herr,W. (2003) Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1. *Genes Dev.*, **17**, 896–911.

123. Nakamura,T., Mori,T., Tada,S., Krajewski,W., Rozovskaia,T., Wassell,R., Dubois,G., Mazo,A., Croce,C.M. and Canaani,E. (2002) ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. *Mol. Cell*, **10**, 1119–1128.

124. Gajan,A., Barnes,V.L., Liu,M., Saha,N. and Pile,L.A. (2016) The histone demethylase dKDM5/LID interacts with the SIN3 histone deacetylase complex and shares functional similarities with SIN3. *Epigenetics Chromatin*, **9**, 4.

125. Saha,N., Liu,M., Gajan,A. and Pile,L.A. (2016) Genome-wide studies reveal novel and distinct biological pathways regulated by SIN3 isoforms. *BMC Genomics*, **17**, 111.

126. Saunders,A., Huang,X., Fidalgo,M., Reimer,M.H., Faiola,F., Ding,J., Sánchez-Priego,C., Guallar,D., Sáenz,C., Li,D. *et al.* (2017) The SIN3A/HDAC Corepressor complex functionally cooperates with NANOG to promote pluripotency. *Cell Rep.*, **18**, 1713–1726.

127. Yao,C., Carraro,G., Konda,B., Guan,X., Mizuno,T., Chiba,N., Kostelny,M., Kurkciyan,A., David,G., McQualter,J.L. *et al.* (2017) Sin3a regulates epithelial progenitor cell fate during lung development. *Dev. Camb. Engl.*, **144**, 2618–2628.

128. Ramakrishnan,S., Pokhrel,S., Palani,S., Pflueger,C., Parnell,T.J., Cairns,B.R., Bhaskara,S. and Chandrasekharan,M.B. (2016) Counteracting H3K4 methylation modulators Set1 and Jhd2 co-regulate chromatin dynamics and gene transcription. *Nat. Commun.*, **7**, 11949.

129. Murray,S.C., Lorenz,P., Howe,F.S., Wouters,M., Brown,T., Xi,S., Fischl,H., Khushaim,W., Rayappu,J.R., Angel,A. *et al.* (2019) H3K4me3 is neither instructive for, nor informed by, transcription. BioRxiv doi: https://doi.org/10.1101/709014, 19 July 2019, preprint: not peer reviewed.

130. Weiner,A., Hsieh,T.-H.S., Appleboim,A., Chen,H.V., Rahat,A., Amit,I., Rando,O.J. and Friedman,N. (2015) High-resolution chromatin dynamics during a yeast stress response. *Mol. Cell*, **58**, 371–386.

131. Chen,X.-F., Kuryan,B., Kitada,T., Tran,N., Li,J.-Y., Kurdistani,S., Grunstein,M., Li,B. and Carey,M. (2012) The Rpd3 core complex is a chromatin stabilization module. *Curr. Biol.*, **22**, 56–63.

132. Demoinet,E., Li,S. and Roy,R. (2017) AMPK blocks starvation-inducible transgenerational defects in Caenorhabditis elegans. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E2689–E2698.

133. Schreiber,G. and Fersht,A.R. (1995) Energetics of protein-protein interactions: analysis of the barnase-barstar interface by single mutations and double mutant cycles. *J. Mol. Biol.*, **248**, 478–486.