Molecular Cell, Volume *54*

Supplemental Information

Identification of Regulators of the Three-Dimensional Polycomb Organization by a Microscopy-Based Genome-wide RNAi Screen

Inma Gonzalez, Julio Mateos-Langerak, Aubin Thomas, Thierry Cheutin, and Giacomo Cavalli

Figure S1

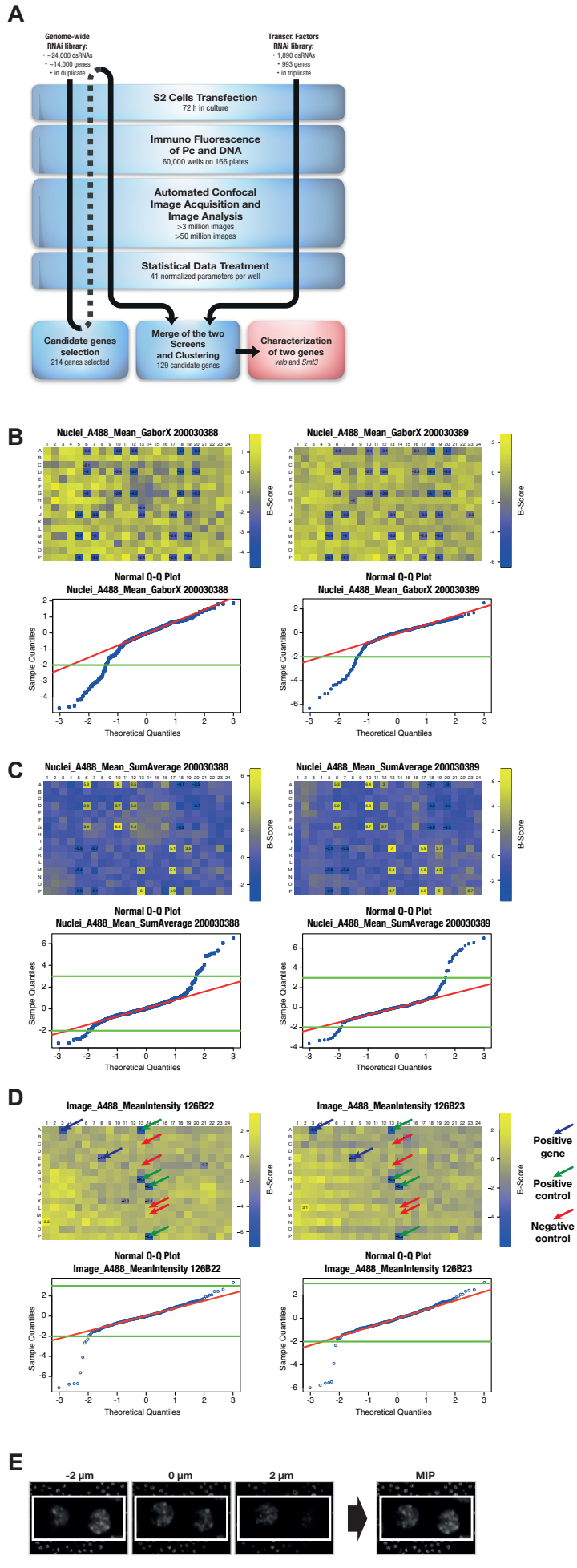
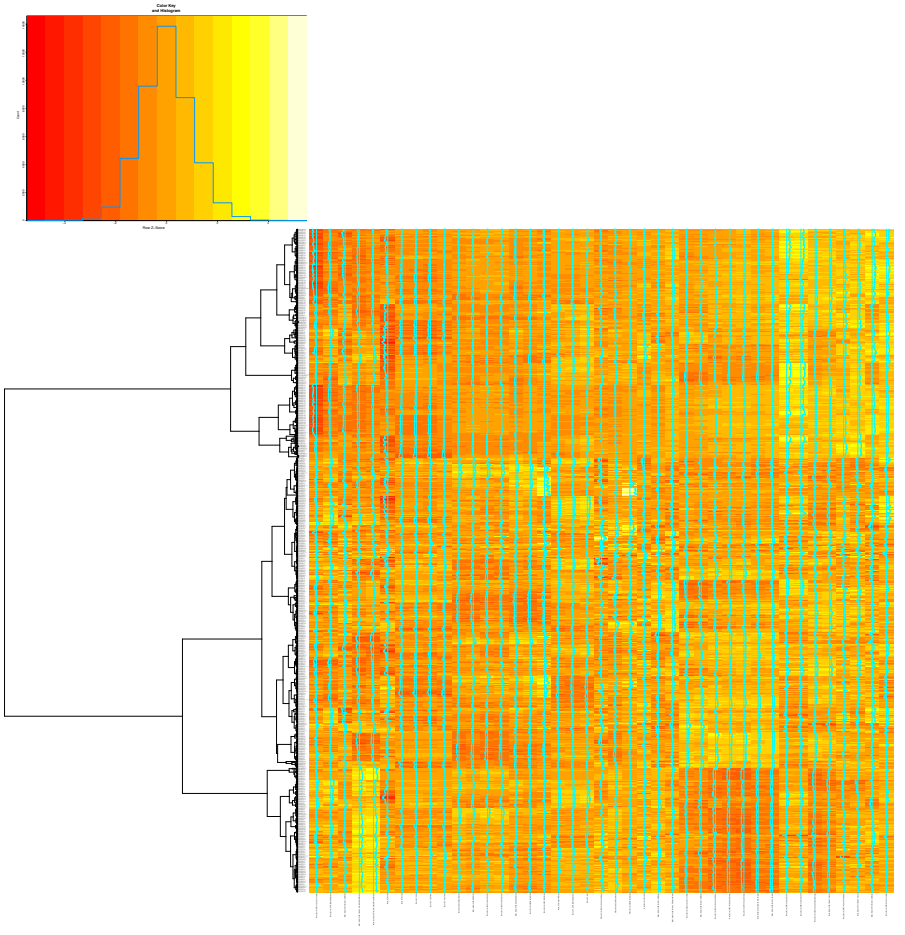

A
Genome-wide RNAi library:
• ~24,000 dsRNAs
• ~14,000 genes
• in duplicate
Transcr. Factors RNAi library:
• 1,890 dsRNAs
• 993 genes
• in triplicate
S2 Cells Transfection
72 h in culture
Immuno Fluorescence of Pc and DNA
60,000 wells on 166 plates
Automated Confocal Image Acquisition and Image Analysis
>3 million images
>50 million images
Statistical Data Treatment
41 normalized parameters per well
Candidate genes selection
214 genes selected
Merge of the two Screens and Clustering
129 candidate genes
Characterization of two genes
velo and Smt3
B
Nuclei_A488_Mean_GaborX 200030388
Nuclei_A488_Mean_GaborX 200030389
B-Score
Normal Q-Q Plot
Nuclei_A488_Mean_GaborX 200030388
Nuclei_A488_Mean_GaborX 200030389
Sample Quantiles
Theoretical Quantiles
C
Nuclei_A488_Mean_SumAverage 200030388
Nuclei_A488_Mean_SumAverage 200030389
B-Score
Normal Q-Q Plot
Nuclei_A488_Mean_SumAverage 200030388
Nuclei_A488_Mean_SumAverage 200030389
Sample Quantiles
Theoretical Quantiles
D
Image_A488_MeanIntensity 126B22
Image_A488_MeanIntensity 126B23
B-Score
Positive gene
Positive control
Negative control
Normal Q-Q Plot
Image_A488_MeanIntensity 126B22
Normal Q-Q Plot
Image_A488_MeanIntensity 126B23
Sample Quantiles
Theoretical Quantiles
E
-2 µm
0 µm
2 µm
MIP

Figure S2

A


Color Key
and Histogram
Count
Row Z-Score

B

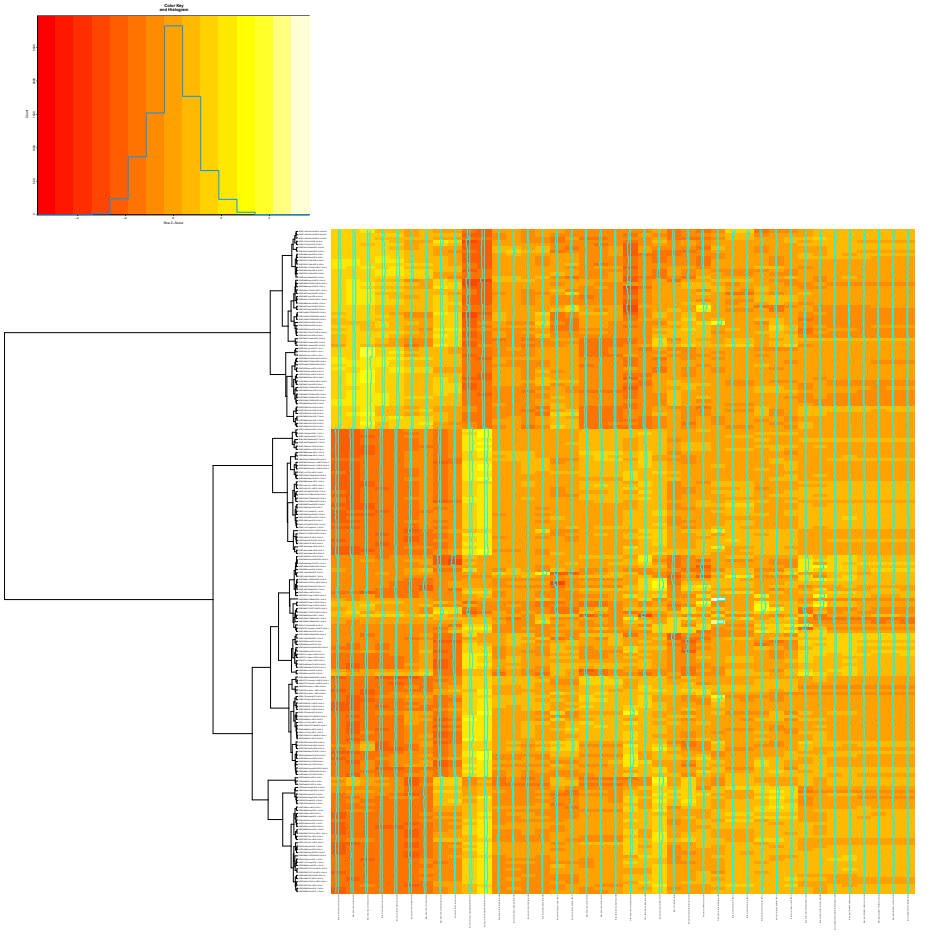
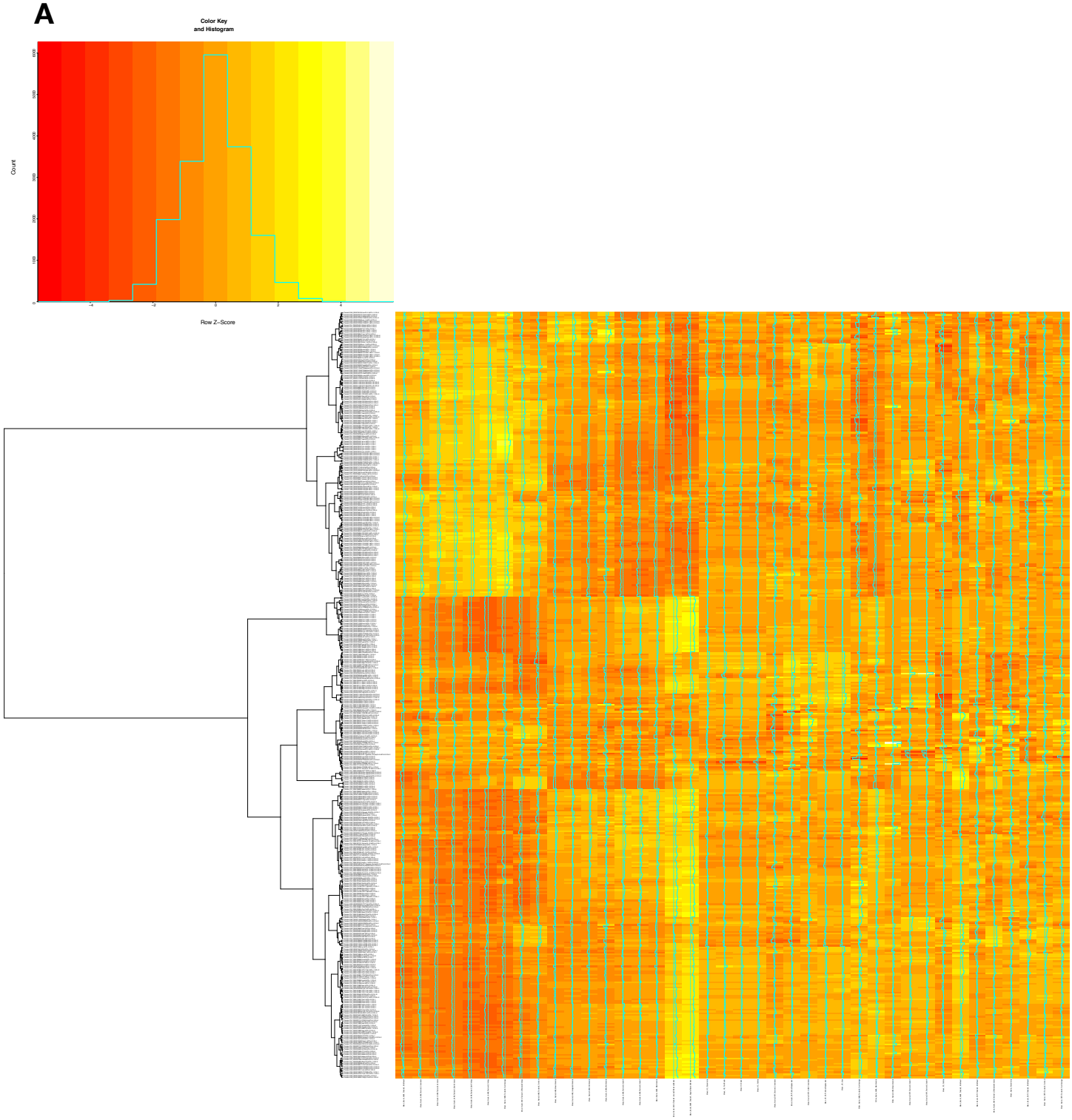

Color Key
and Histogram
Count
Row Z-Score

Figure S3

A


Color Key
and Histogram
Count
Row Z-Score

B

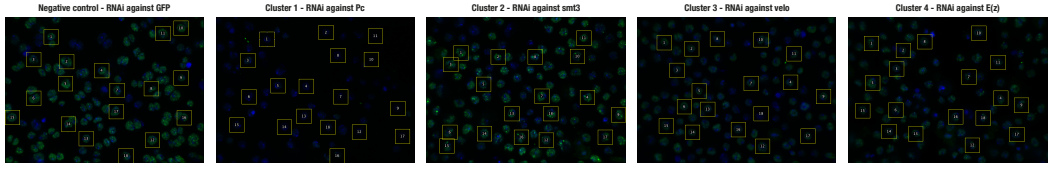

Negative control - RNAi against GFP
Cluster 1 - RNAi against Pc
Cluster 2 - RNAi against smt3
Cluster 3 - RNAi against velo
Cluster 4 - RNAi against E(z)

Figure S4

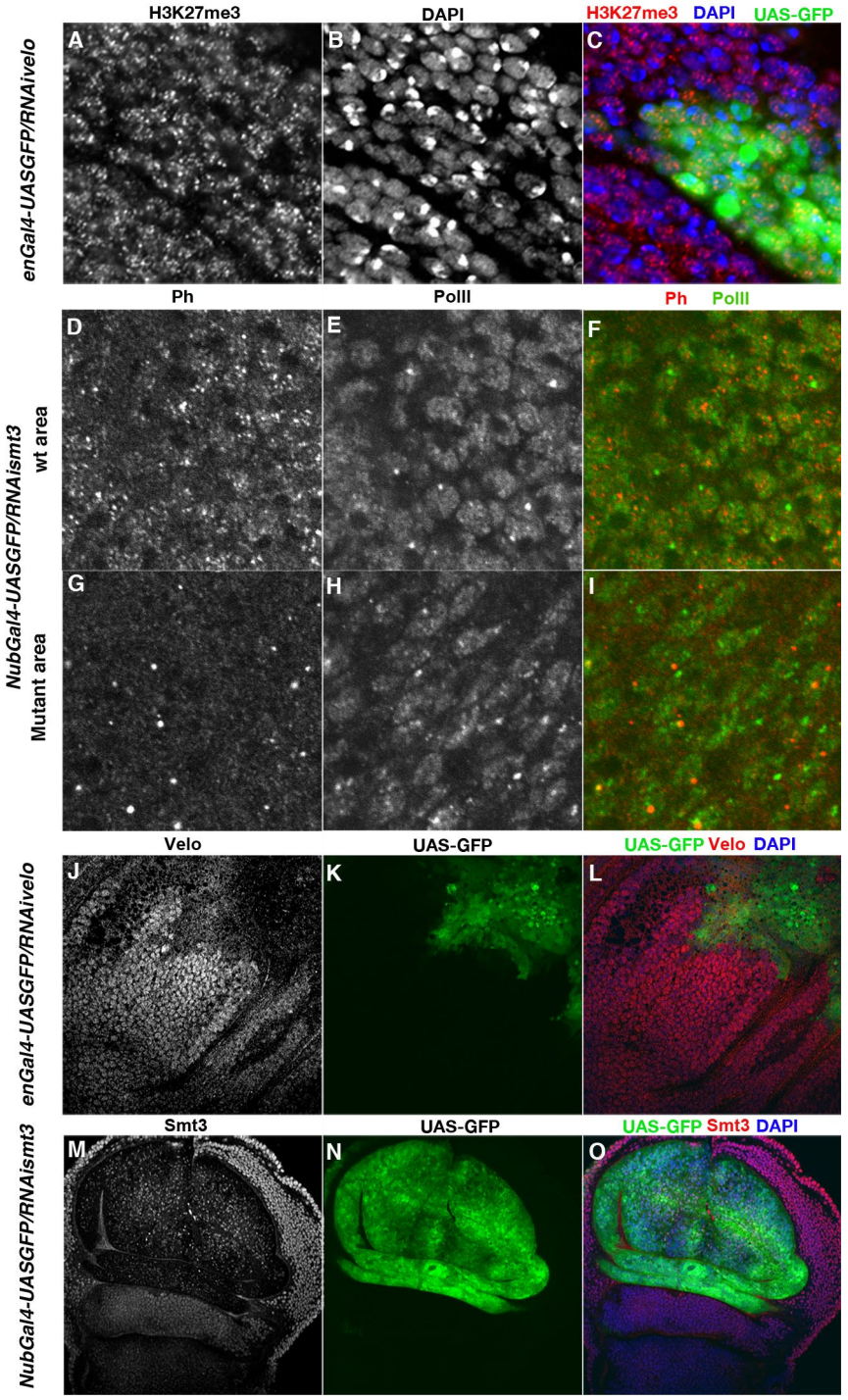

H3K27me3
DAPI
H3K27me3 DAPI UAS-GFP
enGal4-UASGFP/RNAivelo
A
B
C
Ph
PolII
Ph PolII
NubGal4-UASGFP/RNAismt3
wt area
Mutant area
D
E
F
G
H
I
Velo
UAS-GFP
UAS-GFP Velo DAPI
enGal4-UASGFP/RNAivelo
J
K
L
Smt3
UAS-GFP
UAS-GFP Smt3 DAPI
NubGal4-UASGFP/RNAismt3
M
N
O

Figure S5

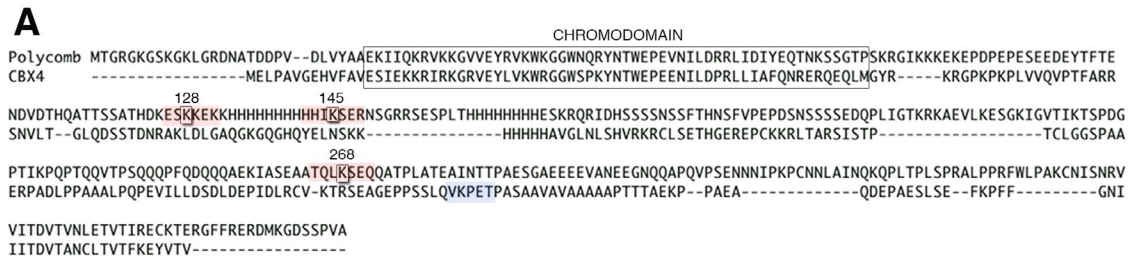
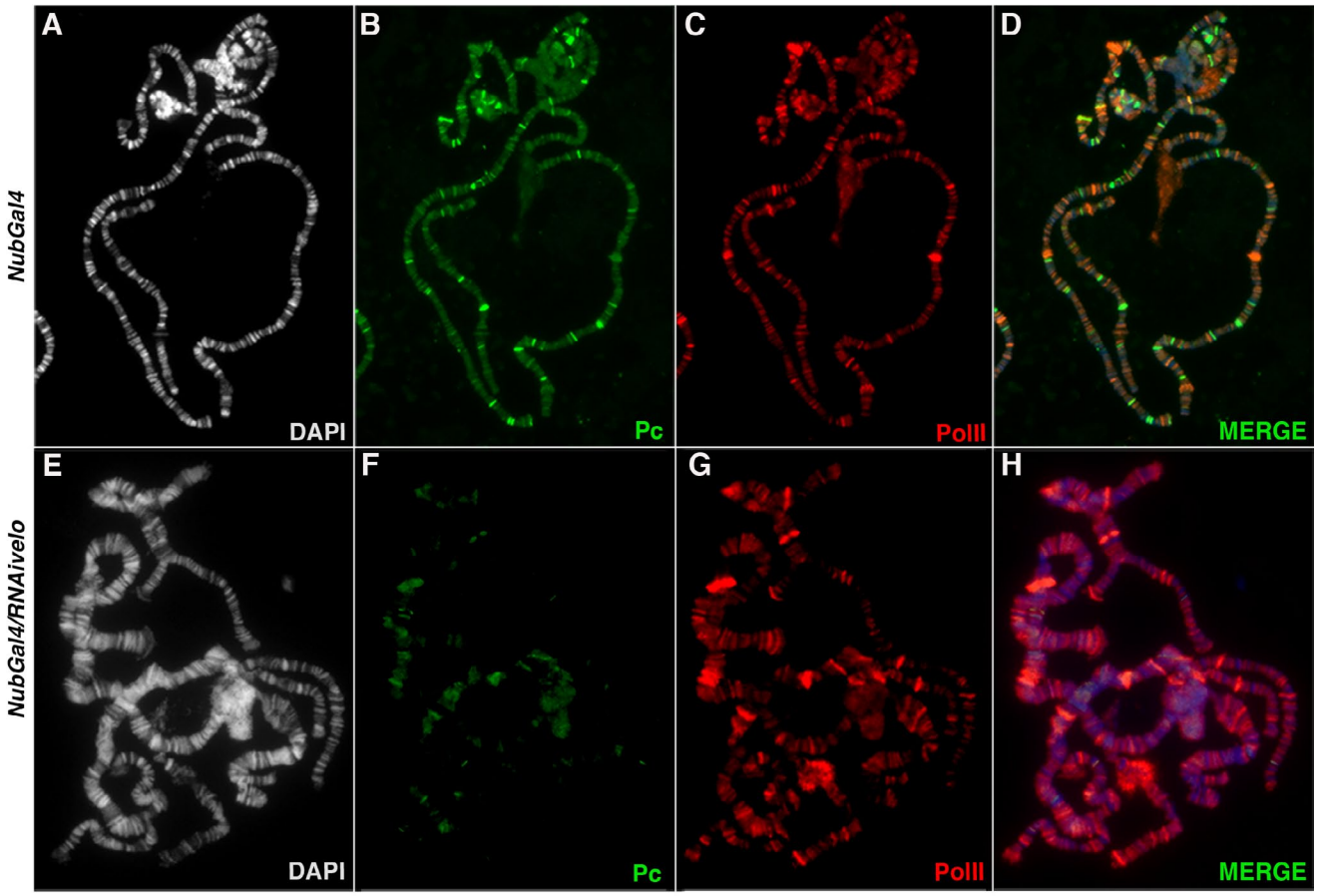
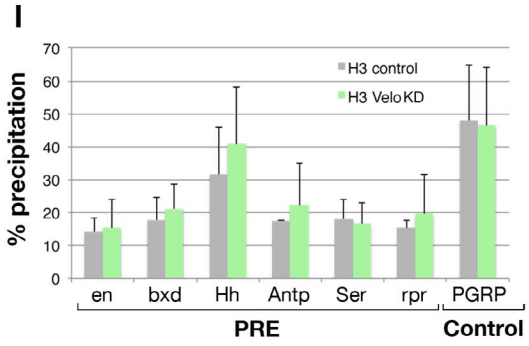
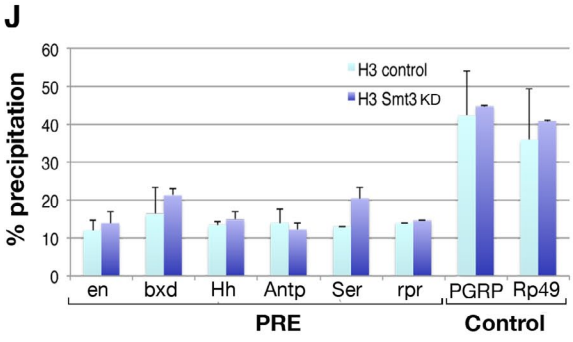
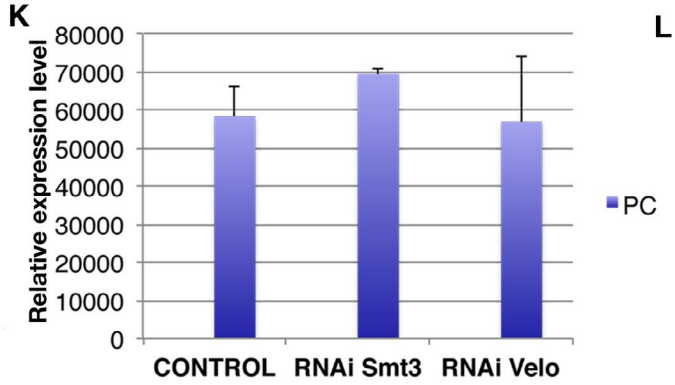

A
CHROMODOMAIN
Polycomb MTGRGKGSKGKLGRDNATDDPV--DLVYAAEKIIQKRVKKGVVEYRVKWKGGWNQRYNTWEPEVNILDRRLIDIYEQTNKSSGTPSKRGIKKKEKEPDPEPESEEDEYTFTE
CBX4 ----------MELPAVGEHVFAVESIEKKRIRKGRVEYLVKWRGGWSPKYNTWEPEENILDPRLLIAFQNRERQEQLMGYR-----KRGPKPKPLVVQVPTFARR
128
145
NDVDTHQATTSSATHDKESKKEKKHHHHHHHHHHIKSERNSGRRSESPLTHHHHHHHHESKRQRIDHSSSSNSSFTHNSFVPEPDSNSSSSEDQPLIGTKRKAEVLKESGKIGVTIKTSPDG
SNVLT--GLQDSSTDNRAKLDLGAQGKGQGHQYELNSKK----------HHHHHAVGLNLSHVRKRCLSETHGEREPCKKRLTARSISTP----------TCLGGSPAA
268
PTIKPQPTQQVTPSQQQPFQDQQQAEKIASEAATQLKSEQQATPLATEAINTTPAESGAEEEEVANEEGNQQAPQVPSENNNIPKPCNNLAINQKQPLTPLSPRALPPRFWLPAKCNISNRV
ERPADLPPAAALPQPEVILLDSDLDEPIDLRCV-KTRSEAGEPPSSLQVKPETPASAAVAVAAAAAPTTTAEKP--PAEA----------QDEPAESLSE--FKPFF---------GNI
VITDVTVNLETVTIRECKTERGFFRERDMKGDSSPVA
IITDVTANCLTVTFKEYVTV----------
B
Total extract
Cytoplasm
Nuclei
Soluble part
Chromatin
WT dsvelo WT dsvelo WT dsvelo WT dsvelo WT dsvelo
aVelo
aVelo
aH3
short exposure
long exposure
C
SUMO, not PC 2549
H3K4me3 3999
RNA-PolII 6240
797
177
1205
930
2095
237
2978
D
Integrated Intensity
Maximum Intensity
Std. Dev. of the Intensity
Pc
Pc–3RK
Pc–SUMO
p = 0.68
p = 0.36
p = 0.68
p = 9.4 e-5
p < 2 e-16
p = 9.4 e-5
p = 0.00221
p = 2.6 e-12
p = 0.00033
VELO
PH
VELO PH
E
F
G
SUMO
PH
SUMO PH
H
I
J

Figure S6


A
B
C
D
NubGal4
DAPI
Pc
PolII
MERGE
E
F
G
H
NubGal4/RNAivelo
DAPI
Pc
PolII
MERGE


I
% precipitation
H3 control
H3 Velo KD
en
bxd
Hh
Antp
Ser
rpr
PGRP
PRE
Control


J
% precipitation
H3 control
H3 Smt3 KD
en
bxd
Hh
Antp
Ser
rpr
PGRP
Rp49
PRE
Control


K
Relative expression level
PC
CONTROL
RNAi Smt3
RNAi Velo

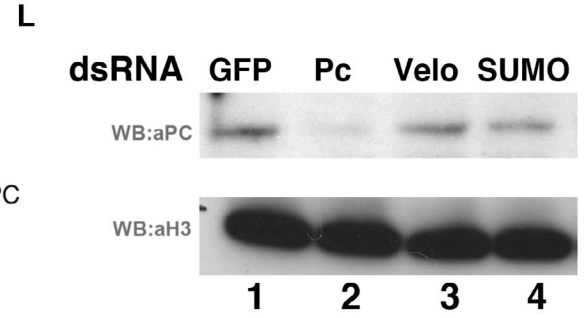

L
dsRNA
GFP
Pc
Velo
SUMO
WB:aPC
WB:aH3
1
2
3
4

Figure S1. Schematic representation of the different phases of the screening procedure, example plates and Maximum Intensity Projection. Related to Figure 1

(A) Schematic representation of the different phases of the screening procedure. Each blue box represents a stage of the workflow. Two lines represent the flow of the data through the different stages: on the left, a line for the procedure followed in the genome-wide screen and on the right for the Transcription Factors screen. In the case of the genome-wide screen, the candidate genes selected after the first passage were confirmed in a secondary screen using a similar procedure. The lists of candidate genes from both screens are merged and classified together. Two genes are characterized in detail. (B) and (C) upper half. Graphical representations of two replicas of the control plates showing a color code for the B-scores of two example parameters: (B) the mean GaborX of all nuclei for Pc-IF and (C), the mean Sum Average of all nuclei for Pc-IF. Most wells are not transfected and show a B-score close to 0. Wells with cells transfected with dsRNA targeting the corresponding genes show how Pc staining parameters can be affected in different ways: in A, all control wells show a more or less strong negative B-score value, while in B some of them show a strong negative B-score and others a strong positive B-score. This exemplifies how the phenotypes can be characterized in an unbiased way by combining multiple parameters. (B) and (C), lower half. Q-Q plots of the same parameters. Q-Q plots show one blue dot per well positioned at the intersection between the Quantile position (x axis) it should have in an ideal normal distribution and its experimental Quantile position (y axis). The vast majority of the wells (excluding control wells) is assumed to be negative and is used to calculate the normal distribution of the wells in the sample, marked with a red line. Any well significantly deviating from the red

line is a candidate. Green lines represent significance thresholds. (D) Upper half: graphical representations of two replicas of a typical screen plate showing a colour code for the B-score normalized mean intensity of the Pc-IF of each well. Green and red arrows show the position of the positive and negative controls respectively. Blue arrows point to two positive wells for this particular parameter. (D) Lower half, Q-Q plots for the same plates and parameter. (E). For every position in the well, we acquired three optical sections separated by 2 μm. Nuclear shape was homogenized by Maximum Intensity Projection.

Figure S2. Hierarchical clustering of positive genes from individual screens. Related to Figure 1

(A) and (B). Hierarchical clustering results for the candidate wells in the genome-wide screen (A) and the transcription factors screen (B). On the left, a dendrogram represents, in 2D, the n-dimensional distance between the different wells. The B-score for every well and parameter is color coded as shown in the scale and by the position of the cyan line within each colored square. Labels show: DRSC code for the dsRNA, gene targeted, Hits in Public Screens (HiPS) for how many times the dsRNA has been found as positive in a screen and X19 denoting how many other genes have homology of at least 19 nucleotides as a measure for possible off-targets. Some details of A are shown on main figure 1.

Figure S3. Hierarchical clustering of merged positive genes. Related to Figures 1 and 2

(A). Hierarchical clustering results for the candidate wells after merging the positive genes in the transcription factors screen and the genes from the genome-wide screen,

confirmed as positives in the secondary screen. On the left, a dendrogram represents, in 2D, the n-dimensional distance between the different wells. The B-score for every well and parameter is color coded as shown in the scale and by the position of the cyan line within each colored square. Labels show: DRSC code for the dsRNA, gene targeted, Hits in Public Screens (HiPS) for how many times the dsRNA has been found as positive in a screen and X19 denoting how many other genes have homology of at least 19 nucleotides as a measure for possible off-targets. (B). Sample full field of view images of the phenotype produced by one gene KD for each cluster. In blue, DAPI and in green Pc-IF. The squares mark the position of the nuclei that have been cut out and amplified in main figure 2.

Figure S4. Distribution of H3K27me3 and RNA Polymerase II in wing imaginal disc after knock down of *velo* and *smt3*. Related to Figure 3

(A) H3K27me3 immunostaining in wild type region of imaginal discs *enGal4-UASGFP/RNAivelo* flies (B) DNA labeling with DAPI. (C) Merge including GFP. The GFP marks the expression domain of the en GAL4 line. (D) PH immunostaining in wild type region of the wing imaginal disc. (E) PolII immunostaining in wild type region of the wing imaginal disc. (F) Merge of D and E. (G) PH immunostaining in the *smt3* knock down area of the wing imaginal disc. (H) PolII immunostaining in the *smt3* knock down area of the wing imaginal disc. (I) Merge of G and H. (J) Velo immunostaining in the *enGal4-UASGFP/UAS-RNAivelo* wing imaginal disc. (K) GFP marking the expression domain of the en GAL4 line. (L) Merge of J, K and DAPI. (M) SUMO immunostaining in the *NubGal4-UASGFP/UAS-RNAismt3* wing imaginal disc. (K) GFP marking the expression domain of the NubGAL4 line. (L) Merge of M, K and DAPI.

Figure S5. Sequence alignment of Drosophila Pc and human CBX4/Pc2 and nuclear distribution of velo. Related to Figure 4

(A) The consensus SUMOylation motif is labeled in red in Pc and in blue in CBX4. The square shows the lysine residue to be SUMOylated. (B) Cell fractionation experiment. Total Protein extracts corresponding to the cytoplasmatic fraction, the nuclear fraction (divided into soluble fraction and chromatin fraction) were analyzed by Western Blot with Velo (long and short exposures are shown) and H3 antibodies**.** . (C) Venn diagrams showing overlap between bound regions of indicated proteins. (D and G) Subcellular localization of Velo and SUMO in wing imaginal disc. Immunostaining analysis of Imaginal discs using Velo and PH antibodies. Velo forms nuclear foci, a subset of them (red circles) are co-localizing with Ph. (D) Quantitative analysis of PC-GFP redistribution. Integrated intensity of the GFP signal in individual nuclei shows that there is no significant difference in expression among the different constructs. In contrast, Maximum Intensity analysis, showing the maximum measured intensity normalized by the integrated intensity, demonstrates that the PC-3RK has stronger peak intensity, whereas the PC-SUMO fusion stains weaker. Similarly, Standard Deviation of the Intensities in the nuclei are also normalized for the expression levels and also show significant differences among the different constructs. n PC-SUMO = 1612, n PC = 974 and n PC-3RK = 964. p values are calculated using a pairwise t-test with pooled SD and multiple testing corrected. Values in the y-axis are arbitrary intensity values where 0 is absolute no photons and 1 equals the maximum intensity in a 16bit image (65536 grey levels).

Figure S6. Effect of Velo KD and smt3 KD on chromatin associated proteins.

Related to Figure 5

(A-H) Immunostaining analysis of polytene chromosomes. (A-D) Inmunostaining of control *UAS-Dcr2;nubGal4* chromosomes (A) labeled with DAPI, (B) with anti-PC and (C) with anti-PolII. (D) Merge of A, B ad C. (E-H) Immunostaining of *UAS-Dcr2;nubGal4-UASGFP/UAS-RNAivelo* chromosomes (E) labeled with DAPI, (F) with anti-PC and (G) with anti-PolII. (H) Merge of E, F and G. (I) qChIP analysis using H3 antibodies of *MS1096Gal4/UAS-RNAiGFP* wing imaginal discs (control, grey bars) and *MS1096Gal4/UAS-RNAivelo* wing imaginal discs (velo KD, green bars). ChIP signal levels are represented as percentage of input and error bars represent standard deviation. (J) qChIP analysis using H3 antibodies of *MS1096Gal4/UAS-RNAiGFP* wing imaginal discs (control, light blue bars) and MS1096Gal4/UAS-RNAismt3 wing imaginal discs (smt3 KD, dark blue bars). ChIP signal levels are represented as percentage of input and error bars represent standard deviation. (K) RT-qPCR experiments showing expression levels of Polycomb RNA in wild type wing *MS1096Gal4* (control) and wings where *velo* (RNAi velo) or *smt3* (RNAi smt3) are knocked down respectively using the *MS1096Gal4* driver. RNA levels for each gene were normalized to those of the housekeeping gene Rp49. Error bars represent standard deviation. (L) Western blot for Polycomb of S2 cells transfected with dsRNA targeting: *GFP* (lane 1), *Pc* (lane 2), *velo* (lane 3), or *smt3* (lane 4).

Table S1. Table containing the ensemble of genes obtained in the different phases of the screen.

(A) List of candidate genes found as positives in the primary Genome-wide Screen. (B) List of of dsRNAs identifying candidate genes found as positives in the primary Genome-wide Screen. (C) List of candidate genes found as positives in the primary Transcription Factors Screen. (D) List of of dsRNAs identifying candidate genes found as positives in the primary Transcription Factors Screen. (E) List candidate genes confirmed as positives in the secondary Genome-wide Screen. (F) List of dsRNAs identifying candidate genes confirmed as positives in the secondary Genome-wide Screen. (G) Merged list of candidate genes found as positives in the secondary Genome-wide and in the Transcription factors screens. Duplicates arising from KD of the same genes by different dsRNAs have been removed from the list of the "Final Screen Positives" sheet. (H) Merged list of dsRNAs identifying candidate genes found as positives in the secondary Genome-wide and in the Transcription factors screens. Strengths of the phenotypes are found in the corresponding tables. (I) Clusters distribution. For each gene, dsRNA and replica (represented by plate/row/col) the cluster number is indicated.

Supplemental Experimental Procedures

Cell Culture, dsRNA synthesis and transfections

Drosophila S2 cells were cultured at 23 °C in Schneider's insect medium (Sigma) supplemented with 10% heat-inactivated fetal bovine serum tested for Insect cell culture (Intermountain Scientific and Sigma) and 1:100 Penicillin-Streptomycin (Gibco).

For the dsRNA synthesis we followed the protocols described by the *Drosophila* RNAi

Screening Center (http://www.flyrnai.org/DRSC-PRS.html). We produced dsRNA by *in vitro* transcription using the MEGAscript T7 kit (Ambion). PCR amplicons were generated from genomic DNA with the primers designed by the DRSC where the T7 promoter was added:

velodsRNA: DRSC08841

SUMOdsRNA: DRSC03611

PcdsRNA: DRSC24966

dsRNA was purified using NucAway Spin Columns (Ambion). For RNAi experiments, cells were transfected by bathing in 6-well plates. 1.5 $x10^6$ cells per well were resuspended in 1 mL of serum free medium and incubated with 10 μg dsRNA at RT for 30 min, before addition of 3 mL complete media with 10% FBS to each well. The cells were harvested 3 days later.

Screen protocol

We initially screened two RNAi libraries from the Drosophila RNAi Screening Center (DRSC. Harvard Medical School, Boston, USA. http://www.flyrnai.org). We screened the DRSC Genome-wide RNAi Library version 2.0 in duplicate and the DRSC TRXN library in triplicate. These RNA libraries are pre-plated into high optical quality bottom CellCarrier 384-well plates (Perkin Elmer, Waltham, USA). Apart from the mentioned libraries, we custom prepared two Control Plates where known PcG genes were knocked down. These two plates contained multiple wells of dsRNA targeting:

0 esc (DRSC28970). 6 wells

1 Pc (DRSC11873). 6 wells

2 Pc (DRSC24966). 6 wells

3 ph-d (DRSC18819). 6 wells

4 ph-d (DRSC23191). 6 wells

5 ph-p (DRSC18820). 6 wells

6 pho (DRSC17219). 6 wells

7 Sce (DRSC23442). 6 wells

8 Su(z)12 (DRSC26081). 6 wells

9 TopoII (DRSC03459). 6 wells

The remaining 324 wells were not transfected and served as negative controls.

Every procedure of the screen was performed on batches of 4 or 6 384-well plates. To transfect the dsRNA, for each well, 10^4 cells in 10µL in serum free media were seeded into wells containing 5µL of ~0.05µg/µL dsRNA in water (~0.25µg dsRNA per well). 45 min after seeding the cells, 30µL of 15% FBS media was added to each well.

For every plate, four wells contained dsRNA targeting the Green Fluorescent Protein (GFP). We used these wells as negative controls. Additionally, each plate had 4 wells containing dsRNA targeting the gene *th* and 4 wells containing dsRNA targeting the gene *rho*. These knockdowns produce cell death and large nuclear sizes respectively, which is easily identifiable under a phase contrast microscope. These wells were used to qualitatively asses the efficiency of the transfections before the IF procedure. In each plate, we custom added dsRNA targeting *Pc* in 4 wells. These wells were later used to assess the efficiency of the experiment qualitatively during image acquisition and quantitatively in later phases.

Cells were fixed 72 hours after dsRNA transfection, Pc was fluorescently immunostained (IF) and DNA was fluorescently labeled with DAPI. 10µL of 16% PFA (Thermo

Scientific, Hudson, USA) was added to each well followed by centrifugation at 1000 rpm for 1 min. 10 min after adding the PFA we proceeded to wash 3 times for 5 minutes with PBS. By adding directly the PFA to the culture medium and centrifuging the cells, we minimized cell loss during subsequent procedures.

During all the procedure, we used a WellMate (Thermo Scientific) to add the reagents and a VP186L manual manifold (V&P Scientific Inc. San Diego, USA) coupled to a vacuum pump to remove the reagents. To minimize and localize the loss of cells, we tilted the plates and gently removed the reagents always from the same corner. The IF procedure is as follows: (1) incubation for 5 min with 10µL of 0.5% (v/v) Triton X-100 in PBS, (2) three washes of 5 min with 30µL of PBS, (3) block 1 h with 30µL of 20% (v/v) Western Blocking Reagent (Roche Applied Science, Penzberg, Germany) in PBS, (4) incubate 2 h with 20µL of rabbit-anti-Pc (Grimaud et al., 2006) at a 1:1000 dilution in the same blocking solution, (5) wash 3 times 5 min with 30µL of 0.1% (w/v) Tween-20 in PBS, (6) incubate 1h with 20 µL goat-anti-rabbit labeled with AlexaFluor 488 (Invitrogen, Carlsbad, USA) diluted 1:500 in blocking solution, (7) wash 3 times 5 min with 30µL of 0.1% (w/v) Tween-20 in PBS, (8) stain DNA for 5 min with 30µL of 0.5 mg/mL DAPI in PBS, (9) wash twice 5 min with 30µL of PBS and (10) added PBS as imaging medium. All steps were carried out at room temperature. After IF, plates were sealed to prevent evaporation and we immediately proceeded to imaging.

Image acquisition

Screened plates were imaged in an Opera microscope (Perkin Elmer). This microscope is a spinning disk confocal microscope equipped with a 60x/NA1.2 and providing a lateral resolution of 0.25µm. Sequential images were acquired using the corresponding settings

to capture DAPI and Pc-IF signals and a binning of 2x2. Exposure parameters were adjusted to reach the same approximate maximum intensities in all experiments. For each well, 8 or 12 positions were selected (8 in the genome wide screen and 12 in the other screens) avoiding the area where reagents were aspirated during the IF procedure. At each position, 3 different z optical slices were imaged, at 3, 5 and 7 μm from the coverslip z position marked by the microscope's autofocus system.

All other images were acquired in either a Zeiss CLSM780 or a Leica SP8 confocal microscope. In both microscopes we used high magnification oil immersion optics. For the automated image acquisition of S2 cells we used the Leica SP8 in combination with MatrixScreener (Leica). Cuticles in dark field were acquired using a Zeiss Axioimager equipped with a 10x objective.

Image analysis

The images corresponding to the three z optical slices were maximum intensity projected along the z-axis into a single image using ImageJ (U. S. National Institutes of Health, Bethesda, Maryland, USA, http://imagej.nih.gov/ij/). This produced a more homogeneous nuclear size while keeping a good contrast (Figure S2D) and conveniently reduced the dataset by a factor of three. We also used ImageJ to convert the FLEX proprietary file format into a convenient tiff standard.

The rest of the image analysis procedure was performed using CellProfiler (Carpenter et al., 2006) (Broad Institute Imaging Platform, MIT/Harvard, Cambridge, USA, http://www.cellprofiler.com). In short, nuclei, heterochromatin and euchromatin were segmented using the DAPI signal. These will be referred to as "objects". For the images as a whole and for every single object parameters were measured. These

parameters describe the intensities, intensity distributions (granularity and texture) of the images and objects, and the shapes of the objects in both channels (DAPI and IF of Pc) as well as intensity correlations between them. For a detailed list of all the parameters descriptions and the algorithms used in their calculation, we refer to the user's manual of CellProfiler (version 1) (http://www.cellprofiler.org). For visual inspection of the results, colored images were exported where the outlines of detected objects were superimposed. We visually confirmed the accurate selection of nuclei in random samples of images.

Images were analyzed at the Orchestra computer cluster of Harvard Medical School and the resulting measurements were exported as CSV tables.

For GFP redistribution quantification, GFP and DAPI images were maximum intensity projected and used to segment the transfected cells using similar procedures as for the genome wide screen. Integrated intensities were measured on average intensity projections and Maximum and Standard Deviation intensities were measured in maximum intensity projections.

Statistical procedures

Data analysis was performed using the R software package (R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org). Data tables were summarized into three values per parameter and per well: mean, median and standard deviations for all parameters assigned to images in the well and for all the parameters assigned to objects in the well.

To be able to compare all parameters across plates and independently of the units, we normalized all the data using the B-score (Brideau et al., 2003). Once normalized, we produced, for every plate and parameter, a graphical representation of the two or three

replicates where a color code represents the B-score (Figures S2). In the same sheet, a second plot shows the wells that are automatically selected as weak, moderate and strong hits. The criteria to consider a well as a hit for a particular parameter are as follows:

- Weak: All replicas show for that parameter a B-score above 1.4 or below -1.4.
- Moderate: For that parameter, one replica shows a B-score above 2.5 and the others above 1.2 or one replica shows a B-score below -2.5 and the others below -1.2
- Strong: All replicas show for that parameter a B-score above 2.5 or below -2.5

Parameters were separated into two groups, the ones measured on the DAPI signal and the ones measured on the IF of Pc. Wells were graded according to their highest grading in the DAPI parameters and their highest grading in the IF-Pc parameters. Wells scoring in the DAPI parameters higher than in the IF-Pc were not considered further. Following these procedures, all positive controls (Pc KDs) were detected as strong hits and none of the negative controls (GFP KDs) were detected as hits.

We observed the graphical representations with all the parameters on the Control Plates (Figure S2) and extracted a list of 41 parameters that were significantly changed by our positive controls (known PcG genes KDs) and used this list of parameters to cluster the genes. Clustering (Figures 1, S3 and S4) was performed using the gplots package in R. Euclidean n dimensional distance was computed in combination with ward clustering. To cluster the positive genes into a discrete number of groups (Figure 2) we used k-means clustering. We tested different arbitrary numbers for the k-means clustering and found that the data could be robustly classified into four clusters. We defined four groups in 25 different random start configurations and a maximum of 500 iterations.

Secondary screen images were analyzed similarly and data were normalized using the Normalized Percentage Inhibition method. This method does not assume the majority of the samples to be negative, but places the value of a well measurement in a scale delimited between the negative and positive control means, that are assigned the values of 0 and 1 respectively.

Statistical significance in boxplots (Figure 2) was calculated using Welch's t-test. P-values were corrected for multiplicity of comparisons using the Bonferroni method and a number of comparisons equal to 53000, corresponding to all the dsRNAs screened, including replicas.

Fly strains, handling and genetics

The binary GAL4/UAS system (Brand and Perrimon, 1994) was used at different temperatures, 25ºC and 29ºC, with the following GAL4 drivers: *en-Gal4*, *nub-GAL4* and *nanos-GAL4*, available at the Bloomington *Drosophila* Stock Centre. Note that the enGal4 line could not be used for KD of *smt3* because the imaginal discs are very small, since SUMOylation is required for proliferation of imaginal discs during development (Kanakousaki and Gibson, 2012). RNAi stocks are coming either from the Vienna *Drosophila* RNAi Centre (*velo* KK112690) or from the TRiP stock collection (*smt3* P-JF02869).

Wings of adult flies and cuticles preparation

To analyze the wings of adult flies, flies stored in a mixture of ethanol:glycerol (3:1) were washed in water at 60ºC for 5 minutes, cut and incubated with KOH 10% at 60ºC

for 10 minutes. KOH was removed and cuticles washed with water, dehydrated in ethanol and mounted in Euparal.

For cuticle preparations, 24-hour-old embryos were mounted in Hoyer's and visualized with dark field microscopy.

Plasmid Constructions

The constructs were made using the Drosophila Gateway Vector Collection (Drosophila Genomics Resource Center at Indiana University.): destination vectors pAGW, pAFW, pAHW, pAFHW (p-Promoter-tag-protein). The promoter used was A= Actin5C. The proteins were tagged with: G= EGFP, F= 3xFLAG, H= 3xHA and FH=3xFLAG-3xHA. cDNA clones were obtained from Drosophila Genomics Resource Center (https://dgrc.cgb.indiana.edu/). The PCR products were amplified from cDNA of *velo* (LD44253-B) with the primers (5'-3')

CG10107 S	caccATGAACGATGAAGATTTCTCCCGTGGCG
CG10107 AS + stop	ctaCGGCTCTATCTTCCGGATCTTCAGT

from cDNA of Pc (RE66837) with the primers

Pc FL S	caccATGACTGGTCGAGGCAAGGGGA
Pc FL AS + stop	ttaTCAAGCTACTGGCGACGAATCG

from genomic DNA the gene *Ubc9* with the primers

Ubc9 FL S	caccATGTCCGGCATTGCTATTACACG
Ubc9 FL AS	ctaCTCAGTGGCCGCCATGGCG

and from genomic DNA the gene *Smt3*

SUMO S	caccATGTCTGACGAAAAGAAGGGAGGT
SUMO AS	ttaTGGAGCGCCACCAGT

All PCR products were cloned into pENTR/D vectors (Invitrogen), confirmed by sequencing, and transferred to the appropriate vectors by recombination using LR Clonase (Invitrogen).

Generation of Polycomb mutants and Velo mutant in the catalytic domain

Site-directed mutagenesis was carried out with the QuikChange Site-Directed Mutagenesis Kit (Agilent technology) using the following primers to generate the point mutations of PC in the Lys 129, 145 and 268:

a386g S	CCACGATAAGGAGTCGAGGAAGGAGAAGAAGCACC
a386g AS	GGTGCTTCTTCTCCTTCCTCGACTCCTTATCGTGG
a434g S	CCATCATCACCACATCAGGTCCGAACGCAACAGTG
a434g AS	CACTGTTGCGTTCGGACCTGATGTGGTGATGATGG
a803g S	CTGCAACGCAGCTGAGATCTGAGCAGCAGGC
a803g AS	GCCTGCTGCTCAGATCTCAGCTGCGTTGCAG

The following primers were used to generate a point mutation in velo disrupting the catalytic domain. These primers introduce a change from codon TGC to AGC at amino acid 502 of the short *velo* transcript resulting in a cysteine to serine change:

t1504a S	GAACAACTTCACCGATAGCGGCCTGTATCTGCTGC
t1504a AS	GGCAGCAGATACAGGCCGCTATCGGTGAAGTTGTTC

Immunostaining analysis of imaginal discs and polytene chromosomes

Imaginal discs were dissected and stained as previously described (Gonzalez and Busturia, 2009). The primary antibodies used were rabbit anti-Pc (1:500) (Grimaud et al., 2006), rabbit anti-Ph (1:200) (Schuettengruber et al., 2009), mouse anti-Ubx (1:20) (White and Wilcox, 1984), mouse anti-ABD-B (1:20) (Celniker et al., 1989), rabbit anti-H3K27 (Millipore 07-449), mouse anti-FLAG (M2, Sigma), mouse anti-HA (HA-7, Sigma) and mouse anti-PolII (Millipore 05-623), rabbit anti-Velo (1/100) and rabbit anti-Sumo (1/200). Images were generated using a LSM510 META (Zeiss) confocal microscope and subsequently analyzed using Photoshop (Adobe Systems Inc.).

Drosophila polytene chromosomes preparation and immunostaining was performed as previously described (Lavrov et al., 2004). The primary antibodies used were: goat anti-Ph (1/500), rabbit anti-Pc (1/100), rabbit anti-Velo (1/200) and rabbit anti-SUMO (1/100). Chromosomes were counterstained with DAPI (Sigma). Images of labeled chromosomes were acquired with a Leica microscope equipped with a digital camera and analyzed using Photoshop.

Antibody production

The Gateway cloning technology (Invitrogen) was used for generating epitope-tagged fusion proteins. The anti-Velo antibody was raised against the 95 amino acids at the C-terminus of Velo amplified from the cDNA (LD44235) with the primers:

Velo Cter S caccATGGTGAACGAGGAGCCAGCGTTGGCAGGGAAAACC

Velo Cter AS CTACGGCTCTATCTTCCGGATCTTCAGT

The anti-SUMO antibody was produced against full length SUMO protein (SUMO S and SUMO A). The PCR products were cloned into pENTR/D vectors. For generating the His-tagged Velo and SUMO fusion protein, the Clonase II reaction from the p-entry clones and pDEST17 (Invitrogen) was performed following manufacturer's instructions. Fusion proteins were used as antigens for injection into rabbits, which showed a low pre-immunization background against fly protein extracts (data not shown). The specificity of the antibodies was tested by Western blot analysis and immunostaining experiments of control (WT) and *velo* or *smt3* knock down (FigureS5J-O).

***In vitro* SUMOylation assay**

SUMO-3 Conjugation Kit (BostonBiochem) was used for the *in vitro* SUMOylation assay. The reaction was carried out at 37°C for 1 hr with the mixture including E1 (44 ng), UbcH9 (0.2 μg), SUMO-1 (2.64 μg), ATP (1 mM), and GST-Pc (1 μg) as substrate purified from bacteria. All four SUMOs have high homology with *Drosophila* Smt3. However SUMO3 has the highest identity (71%) to Smt3.

***In vivo* SUMOylation/deSUMOylation assay**

Drosophila S2 cells were cultured to a density of $2x10^6$ cells/mL in 75 cm^2 flasks in a final volume of 5 mL. For SUMOylation assays of endogenous Pc, 10 µg Flag-HA-SUMO were transfected using Effectene Transfection Reagent (Qiagen). For SUMOylation assays of Flag-HA-Pc a total of 10 µg of DNA were transfected using Effectene Transfection Reagent including 2.5 µg Flag-tagged SUMO, 2.5 µg Flag-HA-Pc, 2.5 µg Flag-HA-Ubc9 and 2.5 µg of the empty vector. For deSUMOylation assays of Flag-HA-Pc a total of 10 µg of DNA were transfected using Effectene Transfection Reagent including 2.5 µg Flag-tagged SUMO, 2.5 µg Flag-HA-Pc, 2.5 µg Flag-HA-Ubc9 and 2.5 µg of Flag-HA-Velo.

One day after transfection cells were treated following the protocol described by Bruderer et al. (Bruderer et al., 2011). Cells were harvested by centrifugation (500g, 4 ºC, 15 min), washed 4 times with ice-cold PBS containing 200 mM IAA, then incubated on ice in buffer (10 mM HEPES, pH 7.9, 1.5 mM MgCl2, 10 mM KCl2, 0.07% NP-40, complete protease inhibitor cocktail tablets (Roche) and 200 mM iodoacetamide) for 15 min. Cells were disrupted by dounce homogenisation on ice. Nuclei were collected by centrifugation (2000 G, 4 ºC, 5 min). The nuclei pellet was resuspended in 300µl of Buffer1 (50 mM Tris at pH 7.5, 150 mM NaCl, 1% NP-40, 0.5% deoxycholate, 10 mM iodoacetamide and protease inhibitor tablets) and sonicated to shear the DNA (two times during 15 sec). 50 µL were taken to use as input. The remaining 250 µL of nuclei fraction was pre-cleared with protein A for 1h at 4 °C, and incubated overnight with antibodies and protein A

beads. Then protein A beads were washed 5X with Buffer 1 at 4 °C for 5 min each time, resuspended in 25 µL of Samples Buffer 2X and denatured for 5 min at 95 °C.

Cell fractionation

Nuclear and cytoplasmatic fractions were separated as described above. The nuclear fraction was ultracentrifuged at 100,000 rpm for 1 h to separate the chromatin fraction and the soluble fraction. The chromatin pellet was resuspended in 250 µL of Buffer 1 and aliquots of each fraction were analyzed by Western Blot analysis.

Cell cycle analysis by FACS

10^6 S2 cells were trypsinized, resuspended in serum containing medium and centrifuged at 1100 rpm for 5 min. Cells were washed with 3-4 mL of PBS, centrifuged at 1100 rpm for 5 min and resuspended in 1 mL PBS. 2.5 mL of ethanol 100 % were added drop-wise while vortexing and cells were placed at -20 °C overnight. Cells were centrifuged at 2000 rpm at 4 °C for 10 min. Cells were washed with 3-4 mL PBS, resuspended by pipetting, centrifuged at 2000 rpm for 10 min. Pellet was resuspended in 1 mL of PBS, RNAse A at 50 µg/mL were added and incubated 30-60 min at 37 °C. Cells were incubated in the dark for 10-15 min with 20-50 µg/mL Propidium iodide (PI) prior to flow cytometry analysis. PI was detected with FL3-H detector on FACS Calibur. Aggregates were eliminated by FL3-A vs FL3-W plot.

RT-qPCR

Third-instar larval imaginal wing discs were dissected in Schneider's *Drosophila* Medium (Sigma) and 30–40 discs were taken for RNA isolation using TRIzol reagent (Invitrogen). 300–400 ng of total RNA were used for the RT reaction. RT was performed using the Superscript III First Strand Synthesis Kit (Invitrogen) following manufacturer's instructions and using oligo-dT primers. cDNA quantifications were performed by real-time PCR, using a Roche Light Cycler and the Light Cycler FastStart DNA Master SYBR green I kit (Roche). Expression levels were normalized to Rp49. Primer sequences used were:

PC_ex2_S1	AGCAGGAGTTTGTAAATTCCCCG
PC_ex2_AS1	TGCCTTTATGGGTTCAGAGCGG

Primers used for Quantitative qChIP experiments of wing imaginal discs

The precipitated DNA was analyzed by qPCR using following primers for several known PREs. Binding levels were normalized to the Input.

Fab7 boundary:

Fab7 sense	AGGAAGAGAGCGGAAAGTGCA
Fab7 antisense	CGGTCGCTCTTAGCCAATACTCTT

bxdPRE:

bxd_mull_S1	AAGGCGAAAGAGAGCAC

bxd_mull_AS1	CGTTTTAAGTGCGACTGAG
enPRE:	
en2 S	GGCTTGTTAGGCAGCAATATGAC
en2 AS	TGAACAGTGCCGCTATATGACC
HhPRE:	
hhprom_S1	TATGCTGCATCATCTGGTTGTC
hhprom_AS1	CACTATCGCCTCGAGTTCATTC
AntpPRE:	
AntpPRE1 S	TGGCCGAGTTTATATCGAAGCG
AntpPRE1 AS	CGGCCAACTTGTGTTGTTGTTC
SerPRE:	
S1 sense Ser	GCTCATCACCGCTCGAAGCAC
S1 antisense Ser	TCTCTCGCGTGACTTGCTCCG
rprPRE:	
rpr prom_AS1	AAAAACACGCTTGGCAACAG
rpr prom_S1	GCTATTTATACCTGGTTCTCTCACG
PGRP:	
PGRPnasc_AS1	CTTACTCAAAACCGAAGAGATCG

PGRPnasc_S1 CCTGGTGAATGATAGCTTACTCTG

Chromatin immunoprecipitation

ChIP on wing imaginal discs was done as described in (Schuettengruber et al., 2009). 50 wing imaginal discs were used per IP experiment. For ChIP on S2 cells, cells were cultured to 90% confluence in 75-cm flask, and ChIP was done essentially as described in (Schuettengruber et al., 2009) with the following modifications: 10x106 cells were used per IP and cross linked chromatin was resuspended in lysis buffer [1% SDS, 10 mM EDTA 50 mM Tris, pH8.1 plus Roche Protease Inhibitor Mixture] and sonicated using a Bioruptor (Diagenode) for 15 times for 30 s each, with 30 s of pause between each to obtain sheared chromatin with an average length of about 100–600 bp. The chromatin was diluted 1:10 with dilution buffer [0.01% SDS, 1.1% Triton X-100, 1,2 mM EDTA 16.7 mM Tris, pH8.1, 167 mM NaCl plus Roche Protease Inhibitor Mixture] for immunoprecipitation. The washes of the beads were done with RIPA, High Salt buffer and LiCL buffer respectively. Libraries for paired-end sequencing libraries for sequencing were prepared from 10 ng of precipitated DNA using the TruSeq ChIP Sample Preparation Kit of Illumina with size selection for products 350 bp +/- 50. The libraries were sequenced on a HiSeq 2000 machine (Illumina), following the manufacturer's protocol.

ChIPseq data analysis

Sequencing overview

The data have been sequenced in duplicate using HiSeq2000. Sequencing was using multiplexing techniques. Duplicates were sequenced in different runs. Base-calling, demultiplexing, adapter removal, pre-quality controls and alignment on *Drosophila melanogaster* reference genome have been performed using Casava 1.8 from Illumina with standard parameter settings. After alignment, each dataset contained at least 28 million of reads for Pc, 30 million for Sumo, 19 million for Velo and 31 million for each input replicate. We then estimated the genome coverage, which is the number of nucleotides on the genome that are covered by the tags that are sequenced. Every 100,000 reads the part of the genome that is covered is computed. By plotting the curve of the part of the genome that is covered by the reads as a function of the number of reads sequenced, a plateau (which is defined as a gain lower than 3bp per extra read) is obtained with 8.7 million of reads for Pc, 8.4 million of read for Sumo and 7.8 million of reads for Velo. This indicates that the genome coverage was complete in all experiments. The diversities (number of distinct genomic positions across the reads) go from 42% to 67% for Pc, 40% to 77% for Sumo and 23% to 34% for Velo, which is a standard value range in ChIP-Seq experiments. The correlations between replicates were then computed using the Pearson correlation test. Pearson correlation coefficients were calculated based on reads mapping to chromosome 3R. The chromosome was divided into 500 bp bins. Each read was assigned into an individual bin if its 5' starting coordinate was included within the start and end interval of the bin. All reads within each bin were summed and Pearson correlation was calculated between pairs of binned datasets. The scores are 0.82

between PC duplicates, 0.83 between Sumo duplicates, 0.76 between Velo duplicates and 0.73% between the input duplicates.

Peak calling

All the analysis were performed using MACS (Zhang et al., 2008), version 1.4.2, with a genome size of 120Mb and a pvalue fixed to 10^{-5} on cumulated duplicates. Of the peaks given by MACS, only those with fold enrichment above 2 and coverage above 20x were kept. The first replicate of PC has a mean coverage of 13.9x on the genome, whereas the mean coverage on the enriched regions is 63.7. The first Sumo replicate has a mean coverage of 12.7x on the genome, to be compared with a 40x mean coverage in enriched regions. Finally, the first Velo replicate has a mean coverage of 7.6 on the genome, whereas the mean coverage in enriched regions is 25.5.

Assignment of genes to enriched regions

The method is based on the one used by ModEncode, where regulatory elements in *Drosophila* were mapped relative to putative target genes (Negre et al., 2011). A gene is considered as marked if an enriched region is located between 1kb upstream of one of the TSSs of the gene, and a distance downstream equal to the minimum between the length of the longest gene transcript and 2kb. The annotations used to assign genes to enriched peaks were computed using dedicated scripts and the annotated genome version 5.52 from Flybase.

Comparison between PC, SUMO and Velo bound regions

To produce the Venn diagram, the enriched regions of PC, SUMO and Velo have been grouped into clustered regions. A clustered region is a genomic region where overlaps between regions enriched by two or three of the factors have been detected. Each clustered region is thus defined as starting at the most proximal coordinate and ending at most distal coordinate of the corresponding regions of overlap. Therefore, the final regions retained for PC, SUMO and Velo contain the clustered regions as well as individual regions for each of the factors in the cases where no overlap exists.

FRAP experiments

FRAP experiments were performed on a Carl Zeiss CLSM780 microscope equipped with GaAsP detectors. Nuclei of wing imaginal disc expressing PC-GFP were half bleached 4s after the beginning of 2D time-lapse experiments during 5s. 2D time-series were recorded for 2 min at a rate of 1 frame every 2s. Measurements of average intensity of PC-GFP were done using the ZEN software package (Zeiss).

Supplemental References

Brand, A.H., and Perrimon, N. (1994). Raf acts downstream of the EGF receptor to determine dorsoventral polarity during Drosophila oogenesis. Genes & development *8*, 629-639.

Brideau, C., Gunter, B., Pikounis, B., and Liaw, A. (2003). Improved statistical methods for hit selection in high-throughput screening. J Biomol Screen *8*, 634-647.

Bruderer, R., Tatham, M.H., Plechanovova, A., Matic, I., Garg, A.K., and Hay, R.T. (2011). Purification and identification of endogenous polySUMO conjugates. EMBO Rep *12*, 142-148.

Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J.*, et al.* (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol *7*, R100.

Celniker, S.E., Keelan, D.J., and Lewis, E.B. (1989). The molecular genetics of the bithorax complex of Drosophila: characterization of the products of the Abdominal-B domain. Genes & development *3*, 1424-1436.

Gonzalez, I., and Busturia, A. (2009). High levels of dRYBP induce apoptosis in Drosophila imaginal cells through the activation of reaper and the requirement of trithorax, dredd and dFADD. Cell Res *19*, 747-757.

Grimaud, C., Bantignies, F., Pal-Bhadra, M., Ghana, P., Bhadra, U., and Cavalli, G. (2006). RNAi components are required for nuclear clustering of Polycomb group response elements. Cell *124*, 957-971.

Kanakousaki, K., and Gibson, M.C. (2012). A differential requirement for SUMOylation in proliferating and non-proliferating cells during Drosophila development. Development *139*, 2751-2762.

Lavrov, S., Dejardin, J., and Cavalli, G. (2004). Combined immunostaining and FISH analysis of polytene chromosomes. Methods Mol Biol *247*, 289-303.

Negre, N., Brown, C.D., Ma, L., Bristow, C.A., Miller, S.W., Wagner, U., Kheradpour, P., Eaton, M.L., Loriaux, P., Sealfon, R.*, et al.* (2011). A cis-regulatory map of the Drosophila genome. Nature *471*, 527-531.

Schuettengruber, B., Ganapathi, M., Leblanc, B., Portoso, M., Jaschek, R., Tolhuis, B., van Lohuizen, M., Tanay, A., and Cavalli, G. (2009). Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. PLoS Biol *7*, e13.

White, R.A., and Wilcox, M. (1984). Protein products of the bithorax complex in Drosophila. Cell *39*, 163-171.

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