Supplementary Information:

Sex comb on midleg (Scm) is a functional link between PcG repressive complexes in *Drosophila*

Hyuckjoon Kang[1,2], Kyle A. McElroy[1-3], Youngsook Lucy Jung[1,4], Artyom A. Alekseyenko[1,2], Barry M. Zee[1,2], Peter J. Park[1,4], and Mitzi I. Kuroda[1,2*]

[1]Division of Genetics, Brigham and Women's Hospital, Boston, MA 02115

[2]Department of Genetics, Harvard Medical School, Boston, MA 02115

[3]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138

[4]Center for Biomedical Informatics, Harvard Medical School, Boston, MA 02115

[*]Corresponding author:

Email: mkuroda@genetics.med.harvard.edu; Phone: 617-525-4520; Fax: 617-525-4522

Supplemental Table Legends

Table S1. Comparison of repressor complex components co-purified by Pc, E(z) and Scm BioTAP-XL pull-down. The proteins characterized as belonging to PcG and other repressor complexes are listed along with the total peptide counts recovered from Pc-C-BioTAP, BioTAP-N-E(z) and BioTAP-N-Scm pull-downs in embryos and BioTAP-N-Scm pull-down in S2 cells. Proteins are grouped and color-coded by repressive complexes. The component CoRest is a shared component between the CtBP and LINT complexes. Protein length, in amino acids, is indicated.

Table S2. Proteins identified after BioTAP-XL mass spectrometry of Pc and E(z) pull-downs in embryos. All proteins recovered from Pc-C-BioTAP and BioTAP-N-E(z) pull-downs, as well as input and mock, in embryos are listed along with UnitProtKB entry name and the unique & total peptide counts.

Table S3. Proteins identified after BioTAP-XL mass spectrometry of Scm pull-downs in embryos and S2 cell lines. All proteins recovered from BioTAP-N-Scm pull-down and input in S2 cells and embryos are listed along with UnitProtKB entry name and the unique & total peptide counts. Two embryonic BioTAP-N-Scm pull-down mass spec results are from technical replicates, and two input mass spec results are from biological replicates.

Kang_Table S1.

Complex	Components	Length	Peptide counts (BioTAP-XL pull-down)				
			Embryo				S2 cell
			Pc	E(z)	Scm		Scm
PRC1	Psc	1601	110	7	44	43	59
	Su(z)2	1368	89	0	38	33	99
	Sce	435	67	2	22	16	46
	ph-p	1589	39	2	42	20	56
	Pc	390	27	6	11	11	12
	Scm	877	23	24	93	35	167
	ph-d	1359	17	0	3	0	5
PRC2	Su(z)12	900	4	113	35	25	80
	esc	425	0	64	12	8	30
	Nurf55	430	4	47	16	7	13
	Pcl	1043	0	45	27	27	55
	E(z)	760	0	44	11	8	17
	Jarid2	2351	2	44	0	0	4
	jing	1582	0	31	0	0	0
	escl	462	0	2	0	0	6
PhoRC	Pho	520	2	6	16	17	18
	Sfmbt	1220	13	1	25	19	27
CtBP	G9a	1637	2	0	42	38	85
	Rpd3	521	17	5	12	20	27
	Su(var)3-3	890	0	0	3	6	20
	CtBP	386	18	9	7	9	13
	CG9932	2171	0	0	6	14	25
	peb	1894	2	0	16	18	0
	CoRest	590	4	0	14	9	39
LINT	L(3)mbt	1477	0	0	12	17	23
	Lint-1	602	0	0	4	8	9

Supplemental Figure Legends

Figure S1. Validation of Pc-, E(z)- and Scm- BioTAP tagged proteins expressed in transgenic flies. (A) A BioTAP tag containing 2xProtein A moieties and a Biotinylation accepting sequence was added to the C-terminus of Pc in its genomic context, and a BioTAP tag containing 1xProtein A moiety and the Biotinylation accepting sequence was added to the N-terminus of the E(z) cDNA and Scm in its genomic context. Pc and Scm are expressed under the control of their native promoters and the expression of E(z) is controlled by the α-tubulin 1 promoter. **(B)** Western blot analysis of nuclear extracts from embryos expressing BioTAP fusion proteins. Tagged proteins were detected by Peroxidase Anti-peroxidase (PAP) antibody against the Protein A epitope. **(C)** Polytene chromosome immunostaining from larvae of BioTAP transgenic flies. The BioTAP fusion proteins were detected by PAP antibody (red) in a wild type background, and DNA was counterstained with Hoechst (blue). **(D)** Viability rescue tests of trans-heterozygotic mutants by BioTAP tagged fusion proteins indicate that the BioTAP-tagged constructs are functional.

Figure S2. BioTAP-XL purification strategy. Nuclear extracts from *Drosophila* embryos or cell lines expressing the BioTAP-tagged bait were cross-linked using formaldehyde and sonicated. Tandem affinity purification was performed to minimize non-specific contaminants. The first purification using rabbit IgG−agarose beads was eluted under denaturing conditions (6M Urea and 0.2% SDS) and the subsequent purification was performed using streptavidin–agarose beads. After stringent washing, peptides released by on-bead trypsin digestion were identified by LC-MS/MS and DNA samples were analyzed by high-throughput sequencing.

Figure S3. Correlations between Scm, E(z), Pc and H3K27me3 in embryos. (**A)** Smoothed density scatter plots between BioTAP-N-Scm and BioTAP-N-E(z) (left), Pc-C-BioTAP (middle), or H3K27me3 (modENCODE, ChIP-seq; right). The listed r value is the genome-wide Pearson correlation coefficient using 1 kb bins. Scm is highly correlated with E(z), Pc and H3K27me3 in embryos. Red lines indicate Loess regression lines. (**B)** Heat map of pairwise genome-wide Pearson correlations between Scm, E(z), Pc and H3K27me3 in embryos.

Figure S4. Genome-wide BioTAP-N-Scm profile in S2 cells. (**A)** Genome browser view of a representative region of Chr3R showing that the BioTAP-N-Scm ChIP-sequencing profile significantly overlaps with Pc-C-BioTAP and H3K27me3 in S2 (Scm and Pc from the current study; H3K27me3 ChIP-chip profile in modENCODE). (**B)** Correlations between Scm and other factors in S2 cells. Pearson correlation coefficients between Scm and other factors profiled from the modENCODE project show that Scm is most positively correlated with H3K27me3, Pc and E(z) in S2 cells. Please note that the datasets from modENCODE for other factors in S2 cells were generated by ChIP-chip.

Figure S5. Comparison of binding profiles between BioTAP-fusion Pc & Scm and PcG proteins from modENCODE in S2 cells. Genome browser shots of PcG proteins and H3K27me3 at the BX-C (Bithorax complex) **(A)** and at a second representative PcG site, the *wingless* (*wg*) PRE **(B)** show that BioTAP-fusion Pc and Scm have sharp peaks at PREs and a broad binding pattern beyond PREs, overlapping with H3K27me3 from modENCODE. The positions of genes and their transcript structures are depicted below PcG binding profiles.

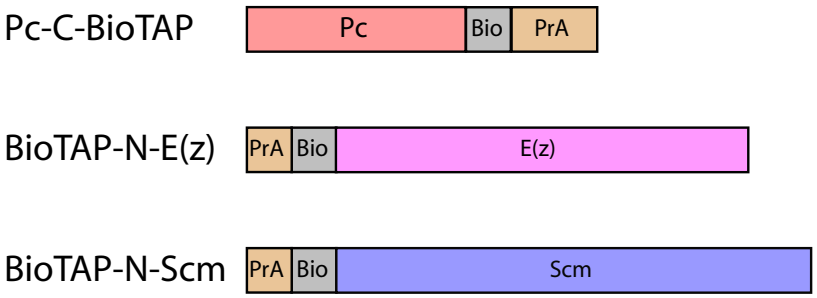
Figure S6. Recruitment of Scm and Pc are independent of each other and of PRC2. The P{GawB}c729 salivary gland-specific GAL4 line was crossed with Oregon R, UAS-*Pc* shRNA, UAS-*E(z)* shRNA, or UAS-*Scm* shRNA TRiP lines. Polytene chromosomes from the resulting third instar larvae were immunostained with anti-Scm **(A, C, E,** and **G)** or anti-Pc **(B, D, F,** and **H)** (red) and DNA was counterstained with Hoechst (blue). Compared to wild-type **(A** and **B)**, induction of *Pc* or *Scm* RNAi leads to significant decrease of each cognate protein on polytene chromosomes, respectively **(D** and **G)**. However, *Pc* and *E(z)* knockdowns do not result in significant depletion of Scm binding on polytene chromosomes **(C** and **E)**, although increased variability of the immunostaining after RNAi was observed. The independence of binding is similar for Pc. Pc still strongly binds at several sites after induction of *E(z)* and *Scm* RNAi. **(F** and **H)**.

Figure S7. The effects of RNAi knockdown of Scm, Pc, and E(z) on H3K27me3 distribution on polytene chromosomes. The P{GawB}c729 salivary gland-specific GAL4 line was crossed with **(A)** Oregon R, **(B)** UAS-*E(z)* shRNA, **(C)** UAS-*Pc* shRNA, or **(D)** UAS-*Scm* shRNA TRiP lines. Polytene chromosomes from the resulting third instar larvae were immunostained with anti-H3K27me3 (green). DNA was counter-stained with Hoechst (blue). **(E)** H3K27me3 binding pattern on wild-type chromosome 3R shows strong labeling bands of H3K27me3 between cytogenetic map 89D (Bithorax complex region) and 100A. **(F)** The strong band signals shown in (E) are decreased and weak bands are concomitantly strengthened in *Scm* RNAi. Dashed lines connect the same cytogenetic map sites between wild-type and *Scm* RNAi.

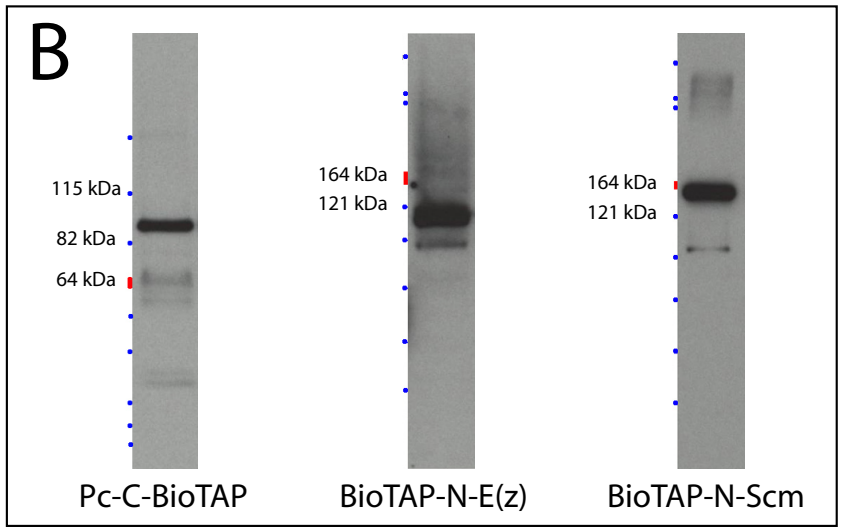
Figure S8. Induction of *Scm* RNAi in the posterior compartment of the wing imaginal disc does not affect H3K27me3 levels. After induction of *Scm* RNAi **(A-F)** or *E(z)* RNAi **(G-I)** driven by *engrailed* (*en*)-GAL4, wing imaginal discs were labeled with Hoechst **(A,D,** and **G)** and antibodies against Scm **(C)** or H3K27me3 **(F** and **I)**. Expression of GFP marks the posterior compartment where *Scm* knockdown **(B** and **E)** or *E(z)* knockdown **(H)** is induced. *Scm* RNAi results in a decrease in Scm levels **(C)** but no detectable change in H3K27me3 levels **(F)**, whereas *E(z)* RNAi compromises the morphology of wing discs and causes the depletion of H3K27me3 **(I)**.

Figure S9. The effects of overexpressed Scm-SAM domain on PcG proteins and H3K27me3 distributions on polytene chromosomes. The P{GawB}c729 GAL4 line was crossed with Oregon R or the UAS-SAM transgenic line, and polytene chromosomes were immunostained with anti-Pc **(A** and **B)** (red), anti-H3K27me3 **(C** and **D)** (green), and anti-E(z) **(E** and **F)** (green). DNA was counterstained with Hoechst. Overlays of the antibody staining and DNA are shown in **A-D**. For clarity, the anti-E(z) immunostaining is presented alone, with the corresponding Hoechst stain inlaid over the image. Left panels **(A, C,** and **E)** show wild-type distributions of target proteins or modifications detected by the indicated antibodies. **(B)** Pc binding is still observed after overexpression of the Scm-SAM domain. **(D)** Overexpression of SAM results in redistribution of H3K27me3 such as strong labeling of the chromocenter and changes in the distribution of signal of strong and weak bands, much like the effect of *Scm* knockdown. **(F)** Significant depletion of E(z) bands is caused by the overexpressed SAM domain, though two E(z) bands on average are consistently observed, in contrast with *Scm* knockdown where none remain (Fig 4D).

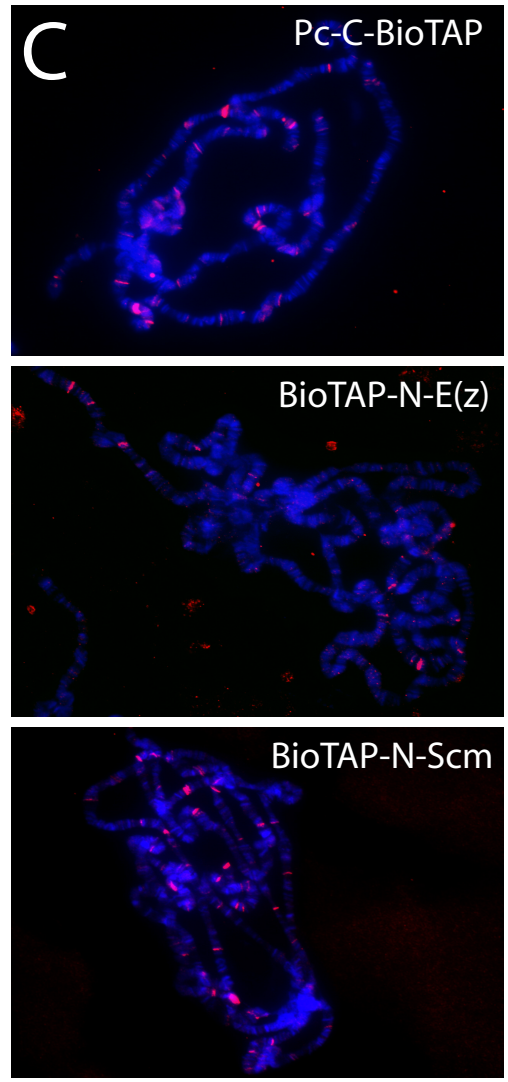
A



B



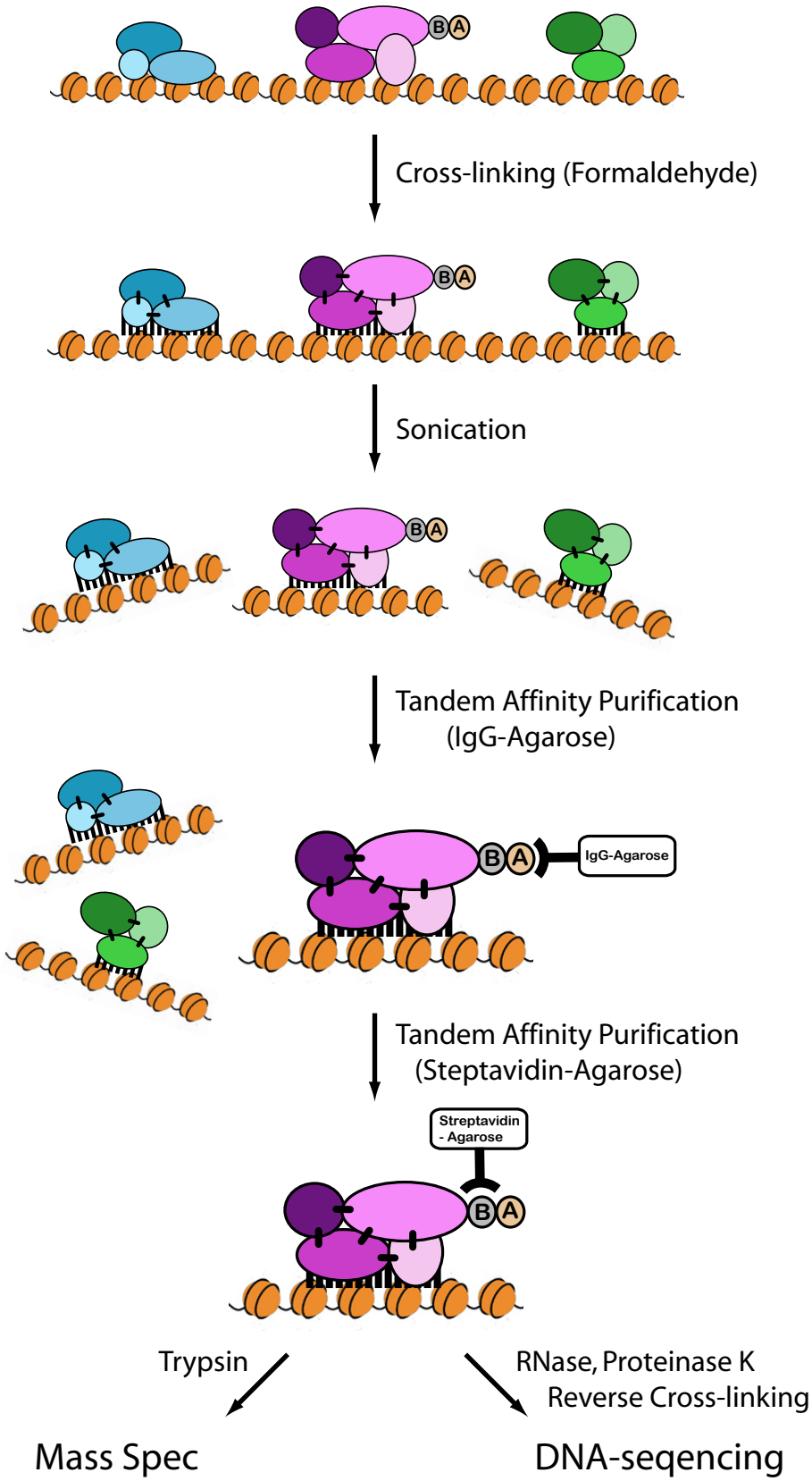
C



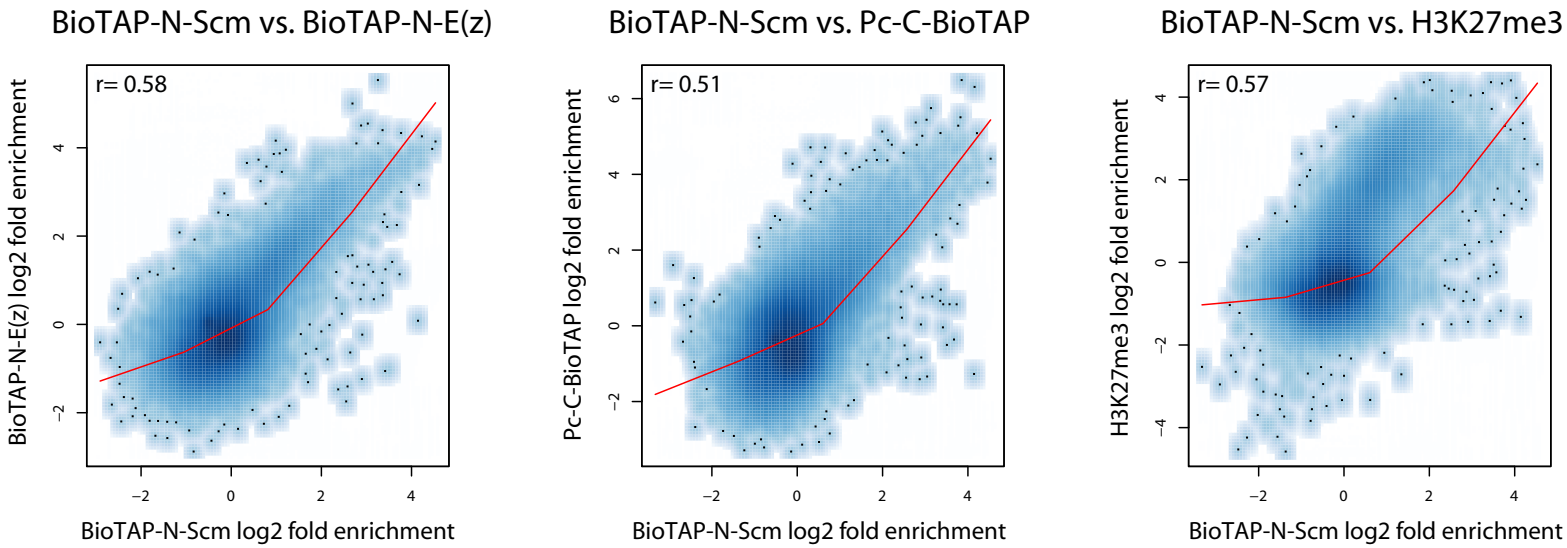
D

Experimental cross		Genotype	Expected mendelian proportion of F1 with this genotype	Observed proportion of viable F1 with this genotype
Pc	w ; P{w$^+$,Pc-C-BioTAP}/+ ; Pc^3/TM2 X w ; +/+ ; Pc^1/TM6B	w ; P{w$^+$,Pc-C-BioTAP}/+ ; Pc^3/Pc^1	0.143	0.119
	w ; P{w$^+$,Pc-C-BioTAP}/+ ; Pc^3/TM2 X w ; +/+ ; Pc^2/TM6B	w ; P{w$^+$,Pc-C-BioTAP}/+ ; Pc^3/Pc^2	0.143	0.181
E(z)	w ; P{w$^+$,BioTAP-N-E(z)}/+ ; E(z)^{63}/TM3 X w ; +/+ ; Df(3L)lxd15/TM3	w ; P{w$^+$,BioTAP-N-E(z)}/+ ; E(z)^{63}/Df(3L)lxd15	0.2	0.09
	w ; P{w$^+$,BioTAP-N-E(z)}/+ ; E(z)^{12}/TM3 X w ; +/+ ; E(z)^{61}/TM3	w ; P{w$^+$,BioTAP-N-E(z)}/+ ; E(z)^{12}/E(z)^{61}	0.2	0.29
Scm	w ; P{w$^+$,BioTAP-N-Scm}/+ ; Scm^{D1}/TM3 X w ; +/+ ; Df(3R)BSC468/TM6C	w ; P{w$^+$,BioTAP-N-Scm}/+ ; Scm^{D1}/Df(3R)BSC468	0.143	0.232

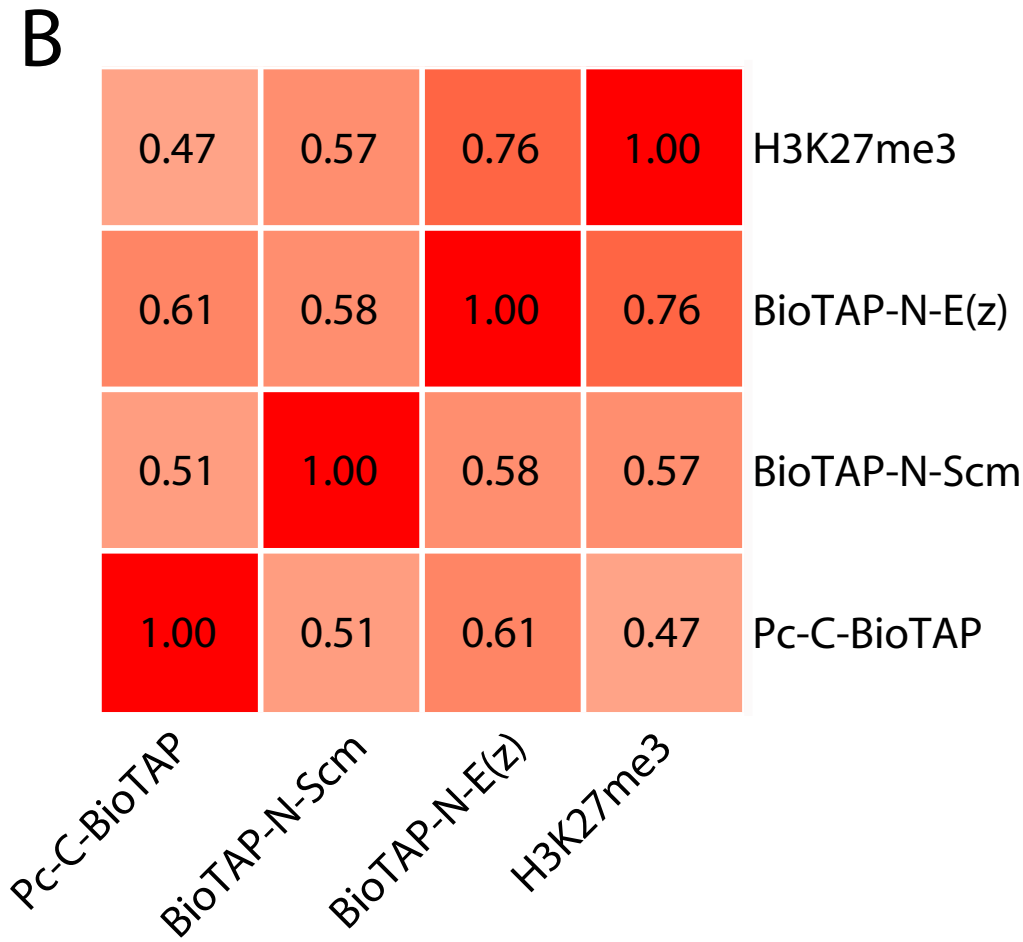
Kang_FigS2.


B
A
Cross-linking (Formaldehyde)
Sonication
Tandem Affinity Purification
(IgG-Agarose)
IgG-Agarose
Tandem Affinity Purification
(Steptavidin-Agarose)
Streptavidin
- Agarose
Trypsin
RNase, Proteinase K
Reverse Cross-linking
Mass Spec
DNA-seqencing

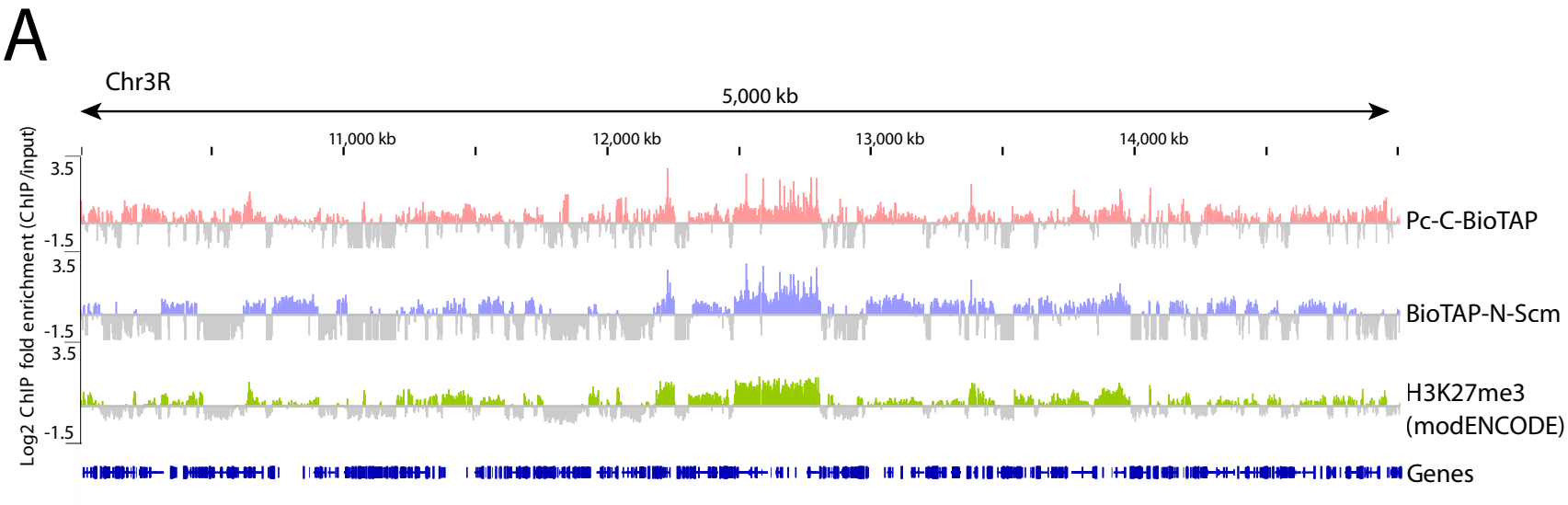
A

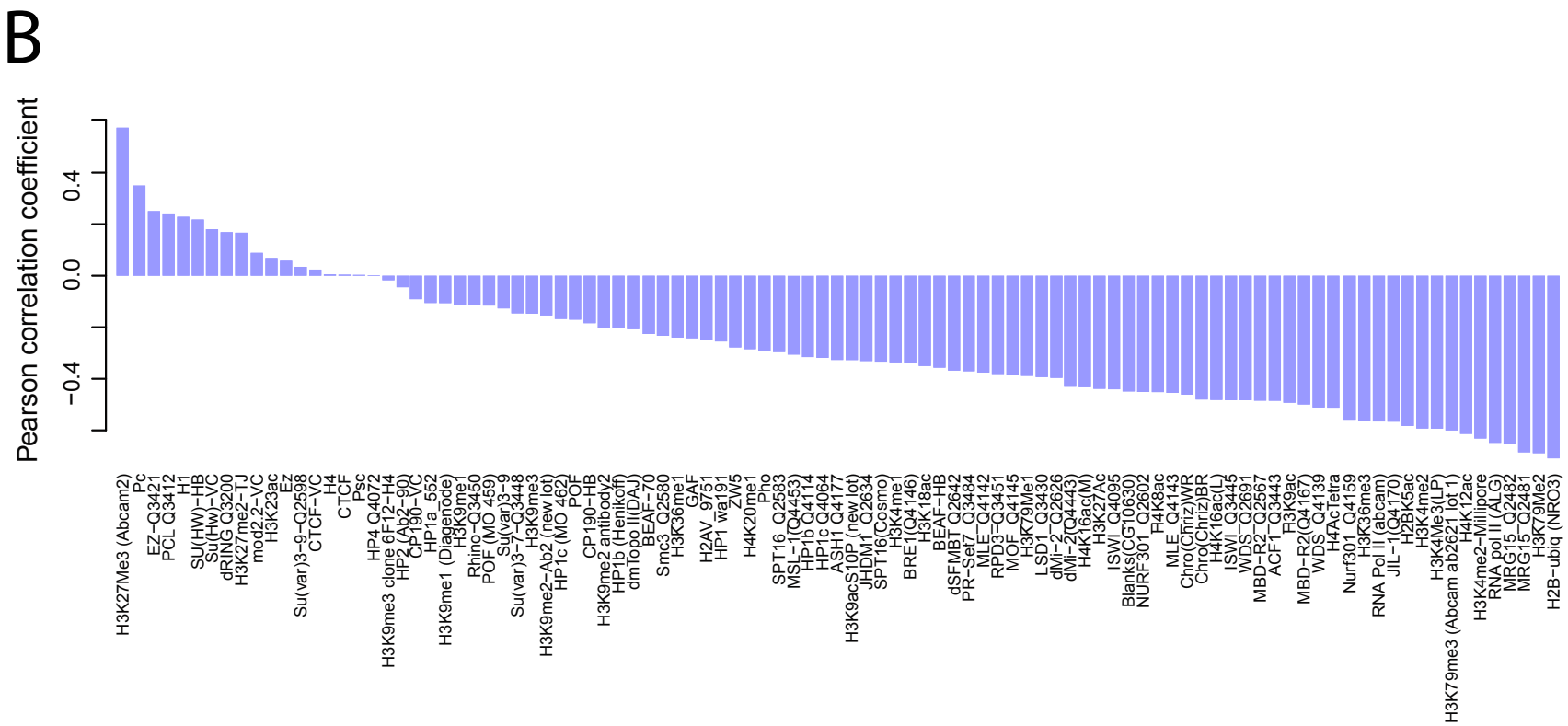

BioTAP-N-Scm vs. BioTAP-N-E(z)
r= 0.58
BioTAP-N-E(z) log2 fold enrichment
BioTAP-N-Scm log2 fold enrichment
BioTAP-N-Scm vs. Pc-C-BioTAP
r= 0.51
Pc-C-BioTAP log2 fold enrichment
BioTAP-N-Scm log2 fold enrichment
BioTAP-N-Scm vs. H3K27me3
r= 0.57
H3K27me3 log2 fold enrichment
BioTAP-N-Scm log2 fold enrichment

B

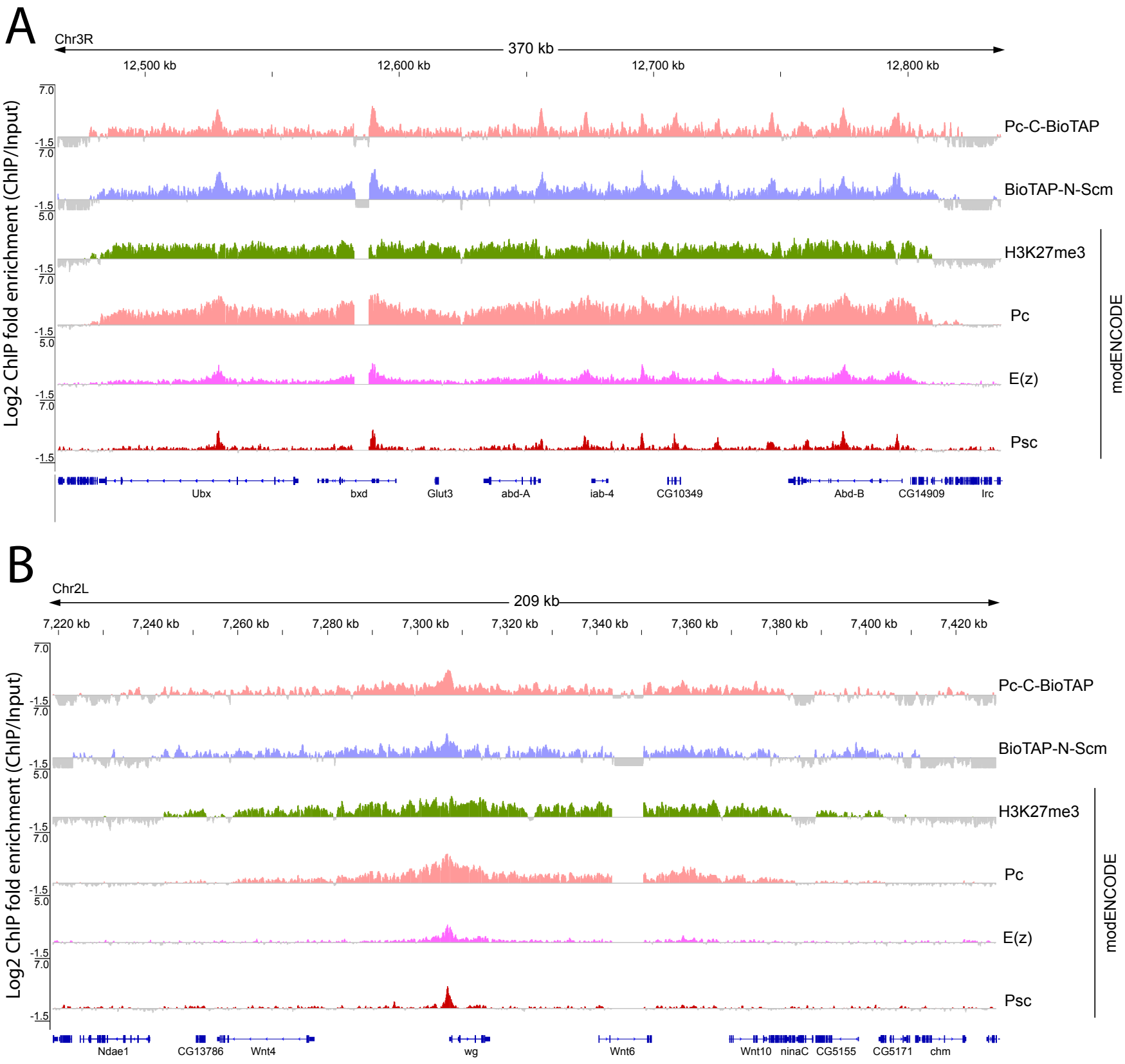

0.47
0.57
0.76
1.00
H3K27me3
0.61
0.58
1.00
0.76
BioTAP-N-E(z)
0.51
1.00
0.58
0.57
BioTAP-N-Scm
1.00
0.51
0.61
0.47
Pc-C-BioTAP
Pc-C-BioTAP
BioTAP-N-Scm
BioTAP-N-E(z)
H3K27me3

Kang_FigS4.

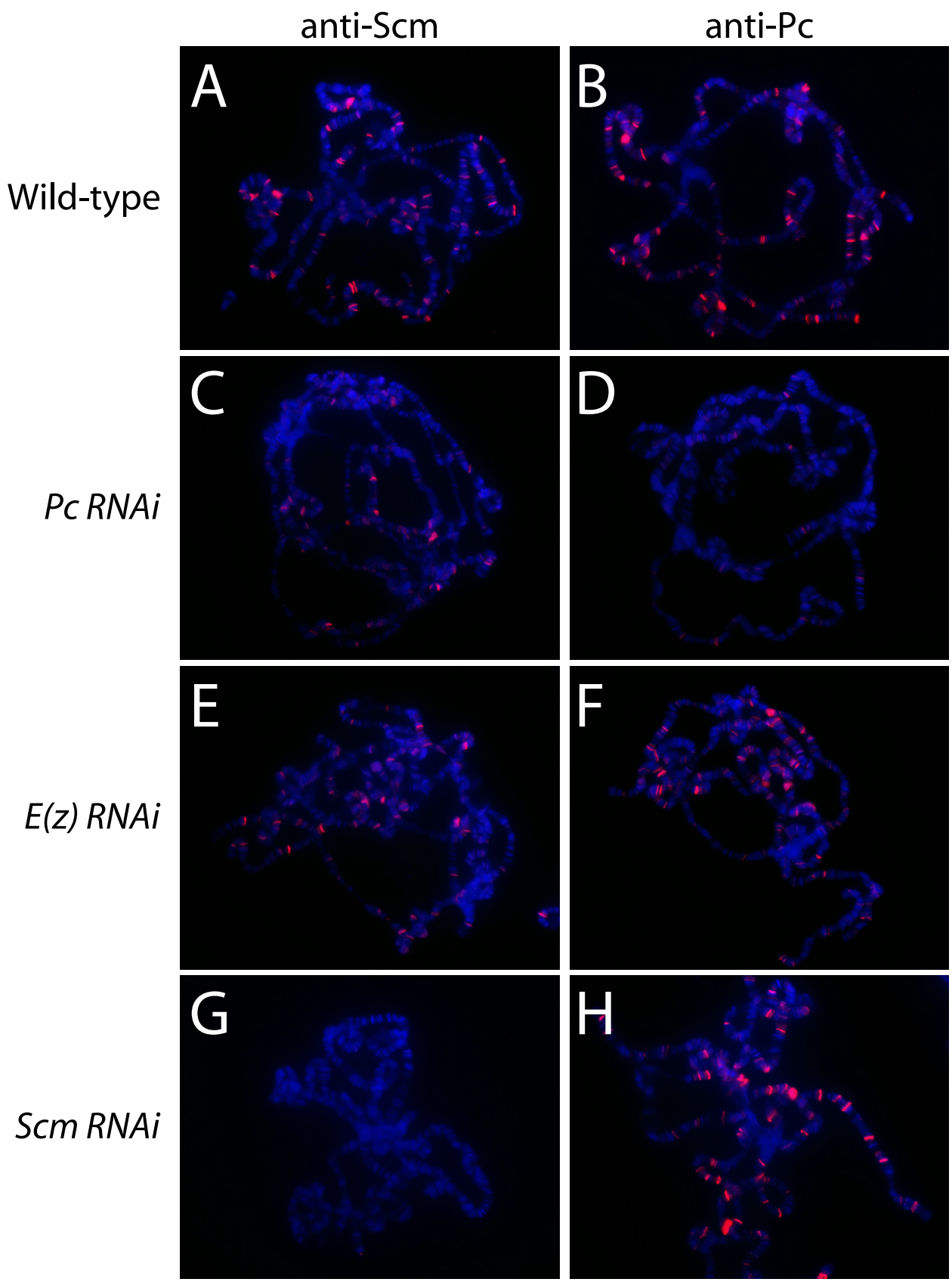

A
Chr3R
5,000 kb
11,000 kb
12,000 kb
13,000 kb
14,000 kb
Log2 ChIP fold enrichment (ChIP/input)
3.5
-1.5
Pc-C-BioTAP
BioTAP-N-Scm
H3K27me3 (modENCODE)
Genes

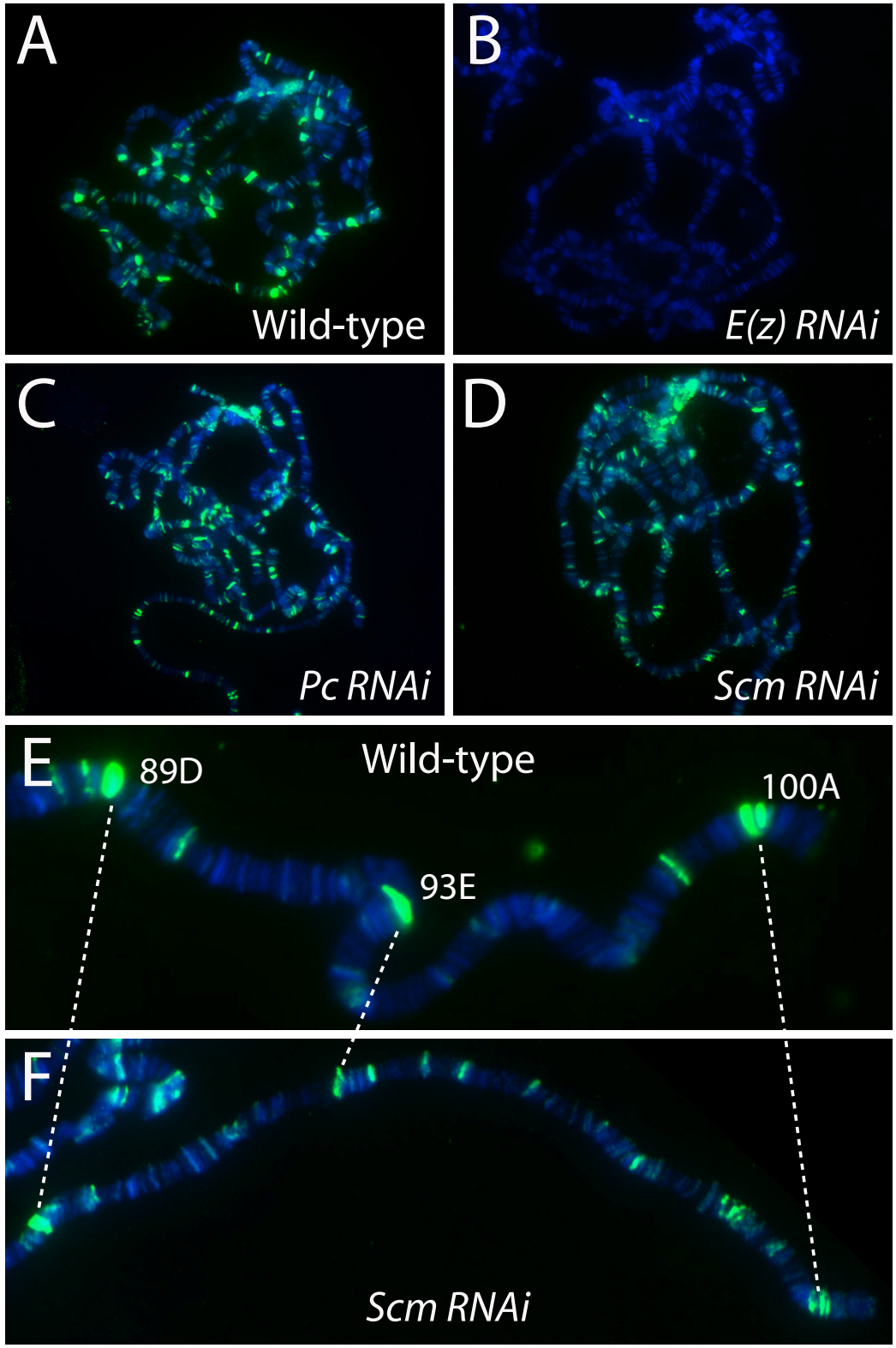

B
Pearson correlation coefficient
0.4
0.0
−0.4
H3K27Me3 (Abcam2)
Pc
EZ−Q3421
PCL Q3412
H1
SU(HW)−HB
Su(Hw)−VC
dRING Q3200
H3K27me2−TJ
mod2.2−VC
H3K23ac
Ez
Su(var)3−9−Q2598
CTCF−VC
H4
CTCF
Psc
HP4 Q4072
H3K9me3 clone 6F12−H4
HP2 (A02−90)
CP190−VC
HP1a_552
H3K9me1 (Diagenode)
H3K9me1
Rhino−Q3450
POF (MO 459)
Su(var)3−9
Su(var)3−7−Q3448
H3K9me3
H3K9me2−Ab2 (new lot)
HP1c (MO 462)
POF
CP190−HB
H3K9me2 antibody2
HP1b (Henikoff)
dmTopo II(DAJ)
BEAF−70
Smc3 Q2580
H3K36me1
GAF
H2AV 9751
HP1 wa191
ZW5
H4K20me1
Pho
SPT16 Q2583
MSL−1(Q4453)
HP1b Q4114
HP1c Q4064
ASH1 Q4177
H3K9acS10P (new lot)
JHDM1 Q2634
SPT16(Cosmo)
H3K4me1
BRE1(Q4146)
H3K18ac
BEAF−HB
dSFMBT Q2642
PR−Set7 Q3484
MLE Q4142
RPD3−Q3451
MOF Q4145
H3K79Me1
LSD1 Q3430
dMi−2 Q2626
dMi−2(Q4443)
H4K16ac(M)
H3K27Ac
ISWI Q4095
Blanks(CG10630)
NURF301 Q2602
H4K8ac
MLE Q4143
Chro(Chriz)WR
Chro(Chriz)BR
H4K16ac(L)
ISWI Q3445
WDS−Q2691
MBD−R2−Q2567
ACF1−Q3443
R3K9ac
MBD−R2(Q4167)
WDS Q4139
H4AcTetra
Nurf301 Q4159
H3K36me3
RNA Pol II (abcam)
JIL−1(Q4170)
H2BK5ac
H3K4me2
H3K4Me3(LP)
H3K79me3 (Abcam ab2621 lot 1)
H4K12ac
H3K4me2−Millipore
RNA pol II (ALG)
MRG15−Q2481
MRG15 Q2482
H3K79Me2
H2B−ubiq (NRO3)

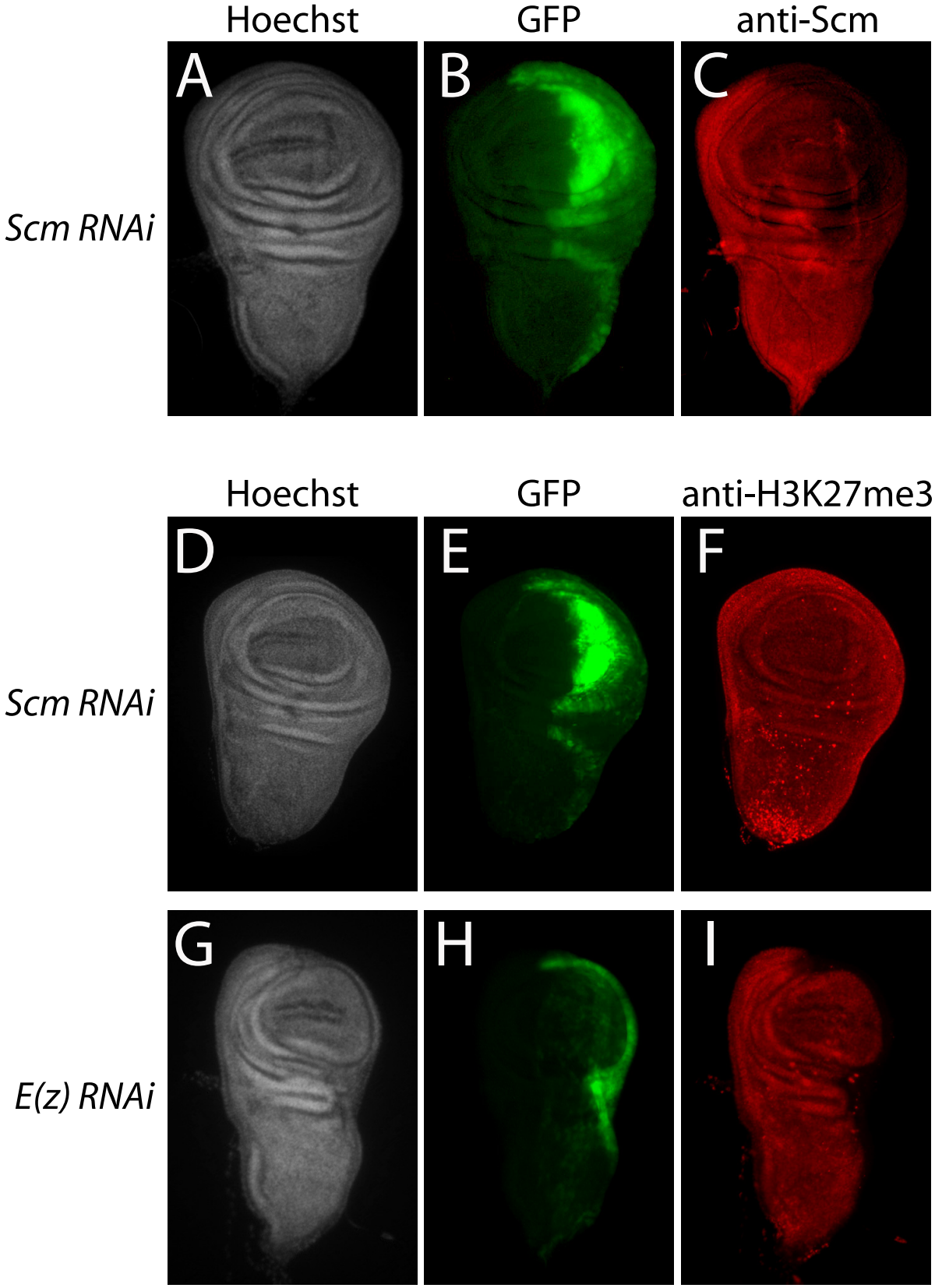
Kang_FigS5.

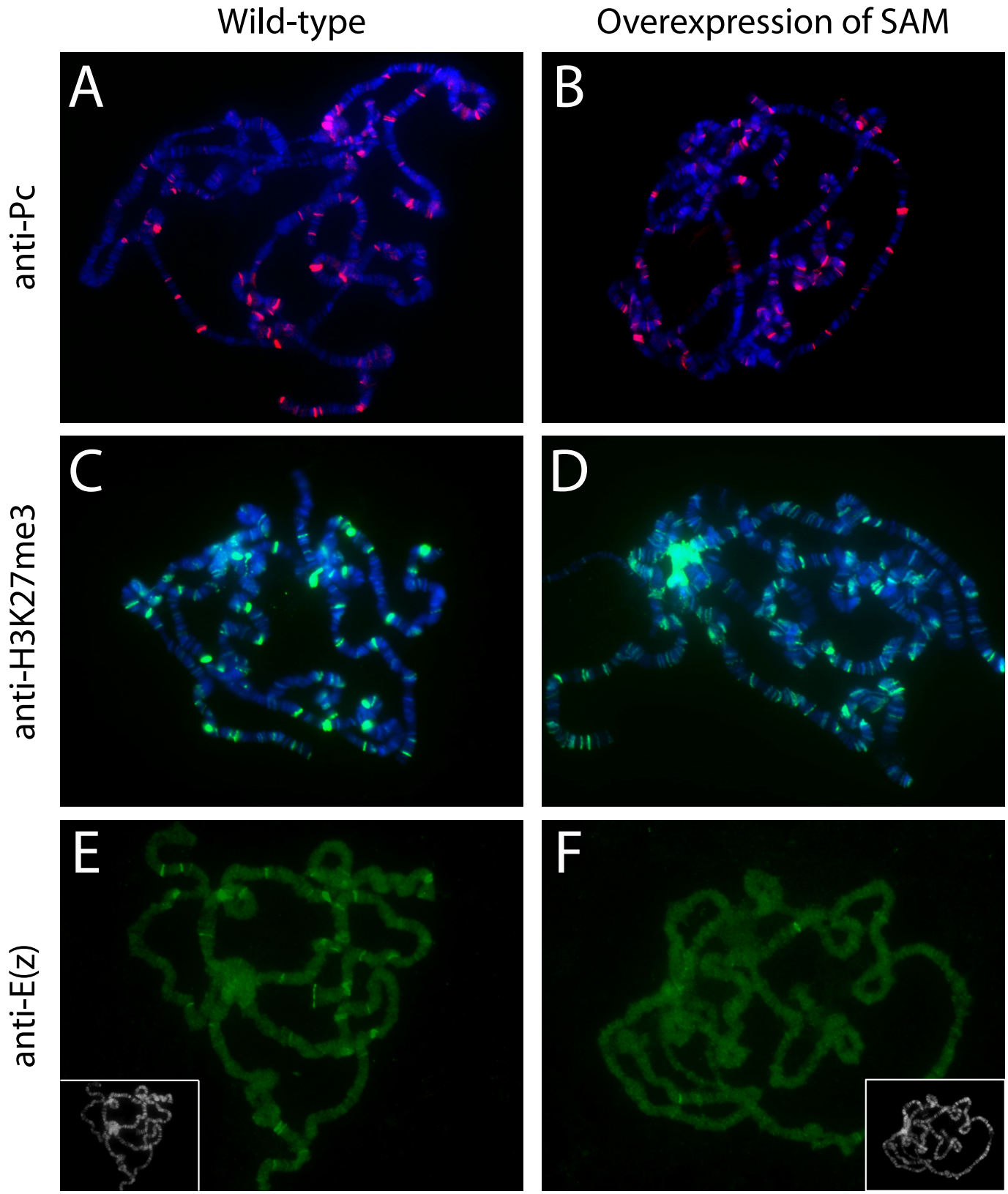

A
Chr3R
370 kb
12,500 kb
12,600 kb
12,700 kb
12,800 kb
Log2 ChIP fold enrichment (ChIP/Input)
Pc-C-BioTAP
BioTAP-N-Scm
H3K27me3
Pc
E(z)
Psc
modENCODE
Ubx
bxd
Glut3
abd-A
iab-4
CG10349
Abd-B
CG14909
lrc
B
Chr2L
209 kb
7,220 kb
7,240 kb
7,260 kb
7,280 kb
7,300 kb
7,320 kb
7,340 kb
7,360 kb
7,380 kb
7,400 kb
7,420 kb
Log2 ChIP fold enrichment (ChIP/Input)
Pc-C-BioTAP
BioTAP-N-Scm
H3K27me3
Pc
E(z)
Psc
modENCODE
Ndae1
CG13786
Wnt4
wg
Wnt6
Wnt10
ninaC
CG5155
CG5171
chm

Kang_FigS6.




A
Wild-type
B
E(z) RNAi
C
Pc RNAi
D
Scm RNAi
E
Wild-type
89D
93E
100A
F
Scm RNAi


Hoechst
GFP
anti-Scm
A
B
C
Scm RNAi
Hoechst
GFP
anti-H3K27me3
D
E
F
Scm RNAi
G
H
I
E(z) RNAi


Wild-type
Overexpression of SAM
anti-Pc
anti-H3K27me3
anti-E(z)
A
B
C
D
E
F