Spring 2019 – Epigenetics and Systems Biology
Discussion Session (Epigenetics)
Michael K. Skinner – Biol 476/576
Week 5 (February 7)

Epigenetics (History / Molecular Processes/ Genomics)

Primary Papers

1. Singer J, et al. (1977) J Biol Chem. 10;252(15):5509-13. (Classic)
2. Holliday and Pugh (1975) Science 24;187(4173):226-32
3. Wen, et al. (2012) BMC Genomics 13:566
4. Schuettengruber, et al. (2017) Cell 171:34

Discussion

Student 10 – Ref #1 & 2 above

- How did BRdU effect DNA methylation?
- What new assay for DNA methylation was developed?
- What observations were used to document DNA methylation?

Student 11 – Ref #3 above

- How does euchromatin correlate to DNA methylation?
- What is CTCF and its function?
- What integration in epigenetics is observed?

Student 12 – Ref #4 above

- What are chromatin remodeling proteins?
- What is the Compass family and functions?
- What is the function of the epigenetic modifications?

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 252, No. 15, Issue of August 10, pp. 5509-5513, 1977
Printed in U.S.A.

5-Methylcytosine Content of Rat Hepatoma DNA Substituted with Bromodeoxyuridine*

(Received for publication, February 7, 1977)

JUDITH SINGER,‡ ROBERT H. STELLWAGEN,§ JOAN ROBERTS-EMS,‡ AND ARTHUR D. RIGGS‡

From the ‡Division of Biology, City of Hope National Medical Center, Duarte, California 91010 and the §Department of Biochemistry, School of Medicine, University of Southern California, Los Angeles, California 90033

The aim of these experiments was to test whether incorporation of bromodeoxyuridine into DNA affects DNA methylation. Rat hepatoma (HTC) cells in culture were labeled for two generations with [^{14}C]bromodeoxyuridine and [^{3}H]thymidine to yield DNA which was 2.1, 20.6, 52.6, and 95.0% bromodeoxyuridine-substituted in the newly made strands. The DNA then was fractionated into highly repetitive, moderately repetitive, and single copy sequences. As determined by a comparison of ^{14}C and ^{3}H counts per min, the percentage of substitution with bromodeoxyuridine was found to be the same in each repetition class. The 5-methylcytosine content of each fraction was determined using high pressure liquid chromatography. It was found that bromodeoxyuridine, even at a level of substitution into newly made DNA of 95%, has no effect on the 5-methylcytosine content of DNA. At all levels of bromodeoxyuridine substitution, highly repetitive DNA has slightly more 5-methylcytosine (3.0% of total cytosine) than does single copy DNA or moderately repetitive DNA (2.3%). The 5-methylcytosine content of whole HTC DNA is the same as that of rat liver DNA (2.4%).

The function of DNA methylation in eukaryotic cells is not known, but it has often been suggested that methylation may affect gene expression and differentiation. Riggs (1), Holliday and Pugh (2), and Sager and Kitchin (3) have published recent reviews and models for differentiation based on DNA methylation by sequence-specific enzymes. Bromodeoxyuridine is known to block differentiation (4–8), as well as the expression of some genes of differentiated cells (9–12), but its molecular mechanism of action has not been determined. A unifying hypothesis is that BrdUrd[1] affects gene expression by changing the level or pattern of DNA methylation. Therefore, we determined the effect of BrdUrd on DNA methylation and report these results here. We have also measured the extent of incorporation of BrdUrd into repetitive and single copy DNA, since it has been reported previously that in rat and chick embryo cells (13, 14) low concentrations of BrdUrd are selectively incorporated into repetitive and "intermediate" DNA.

* This work was supported by a Cancer Center research grant to the University of Southern California from the National Cancer Institute and by a grant to A. D. R. from the National Institutes of Health.

[1] The abbreviations used are: BrdUrd, bromodeoxyuridine; dThd, thymidine; C_0t, concentration × time (mol·s·liter^{-1}).

The DNA we used was that of HTC cells, a line derived from a rat hepatoma (15), whose response to BrdUrd has been well characterized (11, 12). Among the findings have been that the rate of synthesis of tyrosine aminotransferase (EC 2.6.1.5) is reduced by BrdUrd, and that incorporation of BrdUrd into DNA is required for this effect (11).

We find that in rat hepatoma cells BrdUrd is not selectively incorporated into repetitive DNA and does not affect the 5-methylcytosine content of DNA. We also describe an assay for the quantitative detection of 5-methylcytosine in DNA which is quite sensitive (~10 ng detectable), and does not require the use of radioactive label.

EXPERIMENTAL PROCEDURES

Growth and Labeling of Cells—HTC cells were grown for two generations (46 h) in modified Swim's 77 medium (16) containing 0.1 mM hypoxanthine, 1 μM aminopterin, and either 10 μM thymidine or various mixtures of [*methyl*-^{3}H]dThd and [2-^{14}C]BrdUrd (final concentration of dThd + BrdUrd, 10 μM). Cells were harvested in late log phase (10^6 cells/ml) and washed twice with 0.14 M NaCl, 0.01 M potassium phosphate buffer, pH 7.6. Specific activities of radioactive labels were 5 μCi/μmol except for Sample B (see "Results"), in which case the specific activity of [^{14}C]BrdUrd was 50 μCi/μmol.

DNA Purification—Washed cells were suspended in 1% sodium dodecyl sulfate, 0.1 M EDTA, 0.01 M Tris/HCl, pH 8.0 (approximately 1 ml/5 × 10^6 cells). The solution was incubated for 1 h at 37° with RNase A (final concentration 20 μg/ml), then with pronase (previously autodigested for 1 h at 37°) at a final concentration of 500 μg/ml for 1 to 2 h at 55°. The extract was allowed to cool, then sodium perchlorate (5 M) was added to a final concentration of 1 M. The solution was extracted several times with chloroform:isoamyl alcohol (24:1) until a clear interface was obtained. The crude DNA was dialyzed against standard saline citrate, then precipitated with 2 volumes of 95% ethanol, and resuspended in standard saline citrate (2 ml).

The DNA was further treated with RNase A (25 μg/ml for ½ h at 37°) and with autodigested pronase (100 μg/ml for ½ h at 37°) and extracted again with chloroform/isoamyl alcohol until a clear interface was obtained. After ethanol precipitation, the DNA was resuspended in 0.1 N NaOH and incubated at 37° overnight to degrade RNA completely. The solution then was neutralized with 0.1 N HCl and dialyzed against 0.06 M sodium acetate, pH 6.8.

Shearing DNA—Twenty milliliters of glycerol was added to 10 ml of DNA (about 100 μg/ml) in 0.06 M sodium acetate, and the mixture was sheared using a Virtis 60 homogenizer at 50,000 rpm for 30 min (17). DNA was precipitated by making the solution 1.5 to 2 times the standard saline citrate, then adding 2 volumes of 95% ethanol and storing overnight at −20°. After centrifugation, DNA was resuspended in 0.06 M sodium acetate, passed through a Chelex column (Bio-Rad) to remove heavy metal ions, dialyzed against 0.01 M ammonium acetate, and lyophylized.

Sizing DNA – All sheared DNA was sized on isokinetic sucrose gradients by previously published methods (18, 19).

Reassociation – Sheared DNA was dissolved in sodium phosphate buffer at a concentration and ionic strength (0.03, 0.12, or 0.48 M sodium phosphate buffer) so that, where possible, C_0t values would be attained within 30 min to 2 h (17). The DNA was boiled for 5 min in sealed capillaries or Pasteur pipettes, then incubated at 54, 60, or 70° in 0.03, 0.12, or 0.48 M sodium phosphate buffer, respectively. The incubations were ended by quick-freezing the samples.

Analysis of DNA on Hydroxylapatite – Hydroxylapatite was defined and boiled in 0.03 M sodium phosphate buffer before use. To prevent sample degradation, sodium phosphate buffer was passed through Chelex, and both the columns and the glass beads used as column supports were rinsed in 1 mM EDTA.

Following procedures described by Britten *et al.* (17), each sample was diluted to 200 μg/ml, then applied to a hydroxylapatite column (100 μg of DNA, 0.5 ml of hydroxylapatite) in 0.03 M sodium phosphate buffer at 54° or 0.12 M sodium phosphate buffer at 60°. After elution of single-stranded DNA from the column in 0.12 M sodium phosphate buffer at 60°, the temperature was raised to 95° to elute double-stranded DNA. We found no significant difference in our results whether the sample was applied in 0.03 M sodium phosphate buffer or in 0.12 M sodium phosphate buffer.

Hydrolysis of DNA and Assay for 5-Methylcytosine Content – DNA was suspended in 200 μl of 88% formic acid, sealed in Pasteur pipettes, and hydrolyzed at 180° for 20 min (20). The hydrolysate was evaporated to dryness under a stream of N_2 gas, then resuspended in 10 to 50 μl of 0.1 N HCl.

The bases were analyzed by high pressure liquid chromatography. About 15 μl of sample was applied to a Partisil SCX K218 column (Reeve Angel) via an injection port (Altex 905-23), and eluted in 0.045 M ammonium phosphate buffer, adjusted to a pH of 2.3 with 2 M phosphoric acid. The column was run at ambient temperature at a pressure of 550 p.s.i., by use of a Milton Roy Mini-Pump (model GK) (flow rate ~60 ml/h). The bases were detected by their absorbance at 280 nm measured with a UA5 absorbance monitor (ISCO) with a 19-μl flow cell. (The 280 nm filter was chosen to maximize the molar absorbance of 5-methylcytosine.) The lamps used with this absorbance monitor (General Electric G4S11) had half-life of approximately 1 month; for each new lamp, a new calibration curve was necessary to quantitate the bases.

RESULTS

Effect of Bromodeoxyuridine on Enzyme Specific Activity – As summarized in Table I, BrdUrd affects the specific activities of many products of HTC cells. The extent of substitution of BrdUrd for dThd in HTC cell DNA can be varied in a predictable way by varying the relative concentrations of the two nucleosides in the medium when the *de novo* synthesis of deoxythymidylate has been inhibited (21). Under these conditions, there is no effect of any level of substitution on the rate of cell growth for two generations. At longer times, substitution levels above 50% markedly reduce the growth rate (21). The lower levels of substitution are most selective for specific enzymes like tyrosine aminotransferase. Our aim was to see whether these same levels of substitution of BrdUrd would bring about a change in the methylation of DNA.

Isolation and Fractionation of DNA – HTC cells were grown in the presence of [^{14}C]BrdUrd and [^{3}H]dThd so as to yield DNA substituted 2.1, 20.6, 52.6, and 95.0% in the newly made strands (Samples B, C, D, and E, respectively, Table II). Since it is widely believed that repetitive sequences may have a regulatory function (22), we fractionated the DNA into repetitive and single copy sequences before analyzing it for 5-methylcytosine content.

First we obtained a reassociation profile of HTC DNA (Fig.

TABLE II

Percentage of bromodeoxyuridine substitution in HTC DNA

Rat hepatoma cells were labeled with [^{14}C]BrdUrd and [^{3}H]dThd to yield DNA substituted with BrdUrd in the newly made strands as indicated in the first column. Each sample was fractionated into highly repetitive, moderately repetitive, and single copy DNA (see text and Fig. 1). The ^{3}H and ^{14}C cpm were determined for each DNA sample and used to calculate the ratio of BrdUrd/[BrdUrd + dThd].

Sample	$\frac{[^{14}C]BrdUrd}{[^{14}C]BrdUrd + [^{3}H]dThd} \times 100$			
	Total DNA	Component 1, highly repetitive (double-stranded C_0t 10)	Component 2, moderately repetitive (single-stranded C_0t 10; double-stranded C_0t 1000)	Component 3, single copy (single-stranded C_0t 1000)
B	2.1	2.0[a]	1.9[a]	2.1
C	20.6	20.2	20.9	20.5
D	52.6	50.6	51.1	49.7
E	95.0	97.1	96.9	98.1

[a] Sample B was incubated to C_0t 2.

TABLE I

Effects of BrdUrd on HTC cell products

Cellular product	Change in specific activity with BrdUrd	Substitution producing one-half maximal effect	Time needed to observe effect	Reference
		%	h	
Tyrosine aminotransferase	Decrease	~25	<24	11, 12, 21
Alcohol dehydrogenase	Decrease	25–50	>24	12, 21
Glucose-6-P dehydrogenase	Decrease[a]		>48	12
Lactate dehydrogenase	Decrease[a]		>48	12
Malate dehydrogenase	Increase	~75	>48	21
Acid phosphatase	Increase	~75	>48	21
Intracisternal A-type particles	Increase	<50	>48	[b]
Extracellular C-type particles	No change			[b]
Alanine aminotransferase	No change[a]			12
Ornithine decarboxylase	No change			21
Glucocorticoid-inducible cell-surface factor	No change[a]			11
Glucocorticoid receptor	No change[a]			11
Protein kinase	No change[a]			[c]
cAMP phosphodiesterase	No change[a]			[c]

[a] Examined only at a substitution level of ~50%.

[b] H. W. Weber, J. O'Brien, A. Geddes, and R. H. Stellwagen, unpublished observations.

[c] R. H. Stellwagen, unpublished observations.

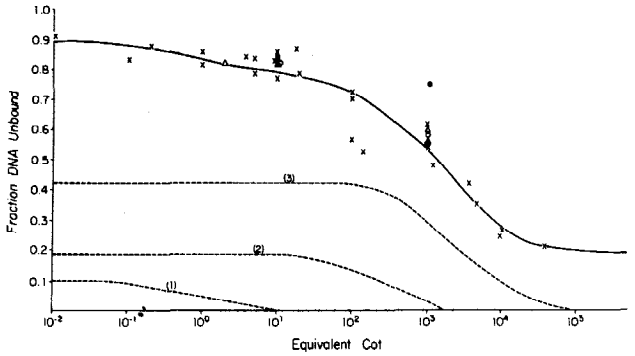


FIG. 1. C_0t curve of rat hepatoma DNA. DNA was purified, sheared to 400 nucleotides, incubated to the C_0t values indicated, then fractionated on hydroxylapatite columns as described under "Experimental Procedures." The experimental points were fit by computer to a theoretical curve with three components, having $C_0t_{1/2}$ values 0.9, 230, and 2900 mol·s·liter^{-1}, for Components 1, 2, and 3, respectively. The $C_0t_{1/2}$ values for each component were taken from a 106-point reassociation curve of rat DNA.[2] ×, unsubstituted rat hepatoma DNA; △, ▲, ○, and ●, Samples B, C, D, and E, respectively, of BrdUrd-substituted DNA (Table II).

1). Our data were found to be consistent with an extensive study of rat DNA.[2] We used our data, the second order rate constants obtained by W. Pearson, and a computer program based on least squares analysis (23) to generate the theoretical three-component curve seen in Fig. 1. From Fig. 1, the $C_0t_{1/2}$ values for the three components, as well as their relative amounts, can easily be determined (Component 1, $C_0t_{1/2}$ = 0.9, 9.9%; Component 2, $C_0t_{1/2}$ = 230, 18.3%; Component 3, $C_0t_{1/2}$ = 2900, 42%). Component 1, then, represents DNA sequences present an average of about 3000 times/genome, while Component 2 represents sequences present 10 times/genome, and Component 3 represents single copy DNA.

Then we fractionated the unsubstituted and BrdUrd-substituted DNA on the basis of the reassociation curve. The DNA was incubated to a C_0t of 10; then single-stranded DNA was separated from double-stranded DNA on hydroxylapatite. Fig. 1 indicates that fractionation at a C_0t of 10 should yield nearly complete separation of Component 1 (highly repetitive) from the others; control hybridization experiments with ^{14}C-labeled repetitive Component 1 DNA driven by excess whole HTC DNA confirmed that this was the case.

We reassociated the remaining single-stranded DNA to a C_0t of 1000, and again used hydroxylapatite to obtain single-stranded and double-stranded fractions. Fig. 1 shows that at a C_0t value of 1000, DNA still single-stranded should be about 90% Component 3, while the double-stranded fraction should be approximately 70% Component 2 and 30% Component 3. For simplicity, despite the overlap between the two fractions, we will refer to DNA single-stranded at a C_0t of 1000 as "single copy," and to DNA double-stranded at a C_0t of 1000 as "moderately repetitive."

Incorporation of Bromodeoxyuridine into DNA – To determine whether BrdUrd was incorporated preferentially into repetitive DNA, as has been reported for rat embryo and chick embryo DNA (13, 14), the ratio of [^{14}C]BrdUrd to [^{14}C]BrdUrd + [^{3}H]dThd was determined for total DNA and for DNA fractionated as just described. Table II shows that the incorporation of BrdUrd is the same for highly repetitive, moderately repetitive, and single copy DNA, even in the case where only 2.1% of the newly made DNA is BrdUrd-substituted.

[2] W. Pearson and J. B. Bonner, manuscript submitted.

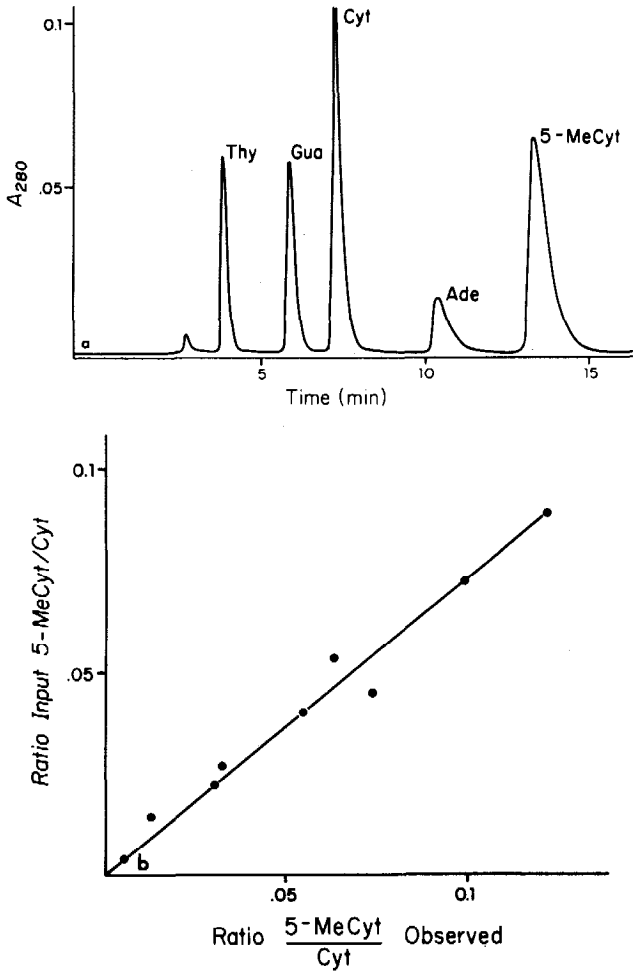


FIG. 2. *a*, elution profile of a base mixture. Ten microliters of a solution containing 1 μg each of thymine, guanine, cytosine, 5-methylcytosine, and adenine were applied to a Partisil SCX column (0.21 × 30 cm) and eluted with 0.045 M ammonium phosphate buffer, pH 2.3, as described under "Experimental Procedures." *b*, standard curve of 5-methylcytosine and cytosine. Solutions containing 2 μg of cytosine and increasing amounts of 5-methylcytosine were applied to a Partisil SCX cation exchange column and eluted as in *a*. Areas under the peaks were compared by weight.

Assay for 5-Methylcytosine – To detect small amounts of 5-methylcytosine, we separated the bases by high pressure liquid chromatography, monitoring elution by absorbance at 280 nm where 5-methylcytosine absorbs maximally. Fig. 2*a* shows the separation of a standard base mixture. Fig. 2*b* shows that measuring peak areas gives good quantitation for 5-methylcytosine; 10 ng of 5-methylcytosine, the smallest amount measured in Fig. 2*b*, represents the approximate limit of sensitivity. Thus only about 5 to 10 μg of DNA is needed for each assay.

5-Methylcytosine in Bromodeoxyuridine-substituted DNA – To determine the effect of BrdUrd on DNA methylation, the samples were hydrolyzed in formic acid and analyzed for 5-methylcytosine by the above high pressure liquid chromatography system. An high pressure liquid chromatography elution profile for DNA 20% substituted by BrdUrd in the newly made strands is shown in Fig. 3 and the results for all samples are listed in Table III. The data clearly show that bromodeoxyuridine has no discernible effect on the 5-methylcytosine content of whole rat hepatoma DNA or on its single copy, moderately repetitive, or highly repetitive components.

TABLE III

5-Methylcytosine content of BrdUrd-substituted HTC DNA

Rat hepatoma cells were labeled with bromodeoxyuridine as described in Table I. DNA was purified, sheared to 400 nucleotides, fractionated by rate of reassociation, then analyzed for 5-methylcytosine content. Samples measured in triplicate are shown with standard deviations.

Sample	$\frac{\text{BrdUrd}}{\text{BrdUrd + dThd}} \times 100$	$\frac{\text{5-Methylcytosine}}{\text{Cytosine + 5-Methylcytosine}} \times 100$			
		Whole DNA	Highly repetitive (double-stranded C_0t 10)	Moderately repetitive (single-stranded C_0t 10; double-stranded C_0t 1000)	Single copy (single-stranded C_0t 1000)
A	0.	2.40 ± 0.09	3.23 ± 0.25	2.32 ± 0.12	2.20 ± 0.23
B	2.1	2.57 ± 0.08	2.9[a]	1.86[a]	2.16 ± 0.03
C	21.6	2.49 ± 0.15	3.24 ± 0.13	2.32 ± 0.08	2.37 ± 0.11
D	52.6	2.09 ± 0.17	2.84 ± 0.04	2.46 ± 0.19	2.29 ± 0.18
E	95.0	2	2.96 ± 0.08	2.05 ± 0.13	2.21 ± 0.14

[a] Sample B was incubated to a C_0t of 2.

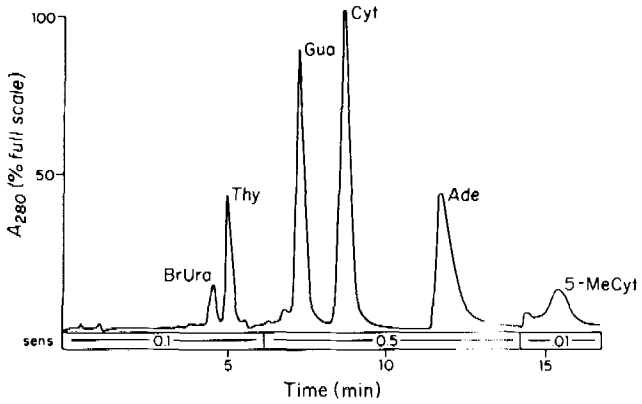


FIG. 3. Elution profile of DNA. Repetitive Component 2 DNA from Sample C (20% substituted in newly made strands with BrdUrd, Table III) was hydrolyzed, applied to a Partisil SCX column, and eluted as outlined under "Experimental Procedures." Changes in the sensitivity of the ultraviolet monitor are indicated.

However, there is a slight but statistically significant difference ($p < 0.001$) between highly repetitive and single copy sequences, the level of methylation of single copy DNA (2.3%) being approximately 0.75 that of repetitive DNA (3.0%). Since rat DNA is 40% GC (24), one can calculate that single copy DNA has an average of 1 5-methylcytosine/110 base pairs, while repetitive DNA has 1 5-methylcytosine/80 base pairs.

Methylation of Rat Liver DNA versus HTC DNA – Since HTC cells are derived from cancerous liver (15), it was of interest to compare the methylation of DNA from rat hepatoma with that of normal liver DNA. We found that for normal rat liver, 5-methylcytosine is 2.28 ± 0.32% of total cytosine, not significantly different from 2.40 ± 0.09%, the value for rat HTC cells (Table III). Therefore, major changes in the extent of DNA methylation are not necessary for growth typical of cancer cells.

DISCUSSION

In *Escherichia coli*, both the *lac* repressor and restriction enzymes are now known to be sensitive to subtle changes in the major groove of the DNA helix (25–28). We have argued that the introduction of a 5-methyl group in the major groove by formation of 5-methylcytosine does affect the binding of regulatory proteins and that DNA methylation may have a regulatory function (1). BrdUrd substitution in DNA is known to affect the binding of *lac* repressor (25) and histones (29). Therefore, we hypothesized that BrdUrd substitution might affect the activity of DNA methylating enzymes and thereby affect differentiation. Holliday and Pugh (2) also reasoned similarly. We decided to test the hypothesis that BrdUrd affects methylation by using a cell line (HTC) whose growth is not affected by BrdUrd for at least two generations and whose responses to BrdUrd are well documented (Table I).

Our results argue against the idea tested. We find that even high levels of BrdUrd substitution do not affect the overall 5-methylcytosine content of rat hepatoma DNA (Table III). By fractionating the DNA into repetitive and nonrepetitive samples, we hoped to detect changes in methylation which otherwise might have been obscured. However, as Table III also shows, no change in the 5-methylcytosine content of repetitive or single copy DNA is evident after BrdUrd substitution. Therefore, it now seems unlikely that BrdUrd affects differentiation by changing DNA methylation. However, it should be noted that, because of the high background level of 5-methylcytosine in mammalian DNA, we cannot rule out the possibility that BrdUrd is affecting the 5-methylcytosine content of DNA to a small extent. A doubling, for example, of the 5-methylcytosine content of 5% of the repetitive sequences would be undetectable.

For mouse (30, 31) and chicken[3] repetitive DNA is more highly methylated than single copy DNA. The same trend is seen for rat DNA, but the differences are small (3.0 *versus* 2.4%) (Table III). The meaning of the difference between the 5-methylcytosine content of single copy *versus* repetitive DNA (2.4 *versus* 3.0%) is difficult to assess. Immunofluorescent studies have indicated that for some species the centromeric regions of chromosomes are highly methylated (32), so a functional significance is possible. Another possibility is simply a "jackpot effect" caused by chance repetition of some methylated sequences.

We have found no selective incorporation of BrdUrd into repetitive sequences, even at a ratio of BrdUrd to dThd where BrdUrd replaces only 2% of the thymine (Table II). This finding conflicts with previous reports of selective incorporation of BrdUrd into the repetitive and intermediate DNA of rat and chick embryo cells (13, 14). The discrepancy may reflect the fact that HTC cells are not an embryonic system or might be due to different experimental procedures.

There is one report of a doubling in the 5-methylcytosine content of BHK cells after transformation with polyoma virus (33). Our finding that the DNA of HTC cells is not differently methylated from that of rat liver cells is evidence against the notion that abnormal methylation of DNA is necessarily corre-

[3] J. Singer and J. Roberts-Ems, unpublished data.

lated with cancerous growth. Our results are consistent with those of Gantt *et al.* (34), who showed that the 5-methylcytosine content of a mouse non-neoplastic cell line was the same as that of a neoplastic line derived from it.

The assay we used to detect 5-methylcytosine involves base separation by high pressure liquid chromatography and detection by ultraviolet absorption. This method offers several advantages (*a*) speed (only 15 min/sample), (*b*) since 10 ng of 5-methylcytosine are detectable, much less material is required than for paper chromatography, and (*c*) the problems of pool size inherent in most radioactive labeling are avoided (for a discussion see Ref. 35). The sensitivity achieved is in the same range as has been currently attained using mass spectroscopy (36).

Acknowledgments—We thank Dr. William Pearson for the gift of rat liver DNA and for the use of his computer program. We also thank the people in Eric Davidson's laboratory for their kind help.

REFERENCES

1. Riggs, A. D. (1975) *Cytogenet. Cell Genet.* **14**, 9–25
2. Holliday, R., and Pugh, J. E. (1975) *Science* **187**, 226–232
3. Sager, R., and Kitchin, R. (1975) *Science* **189**, 426–433
4. Bischoff, R., and Holtzer, H. (1970) *J. Cell Biol.* **44**, 134–150
5. Weintraub, H., Campbell, G. L., and Holtzer, H. (1972) *J. Mol. Biol.* **70**, 337–350
6. Walther, B. T., Pictet, R. L., David, J. D., and Rutter, W. J. (1974) *J. Biol. Chem.* **249**, 1953–1964
7. Lasher, R., and Cahn, R. D. (1969) *Dev. Biol.* **19**, 415–435
8. Coleman, A. W., Coleman, J. R., Kankel, D., and Werner, I. (1970) *Exp. Cell Res.* **59**, 319–328
9. Holthausen, H. S., Chacko, S., Davidson, E. A., and Holtzer, H. (1969) *Proc. Natl. Acad. Sci. U. S. A.* **63**, 864–870
10. Silagi, S., and Bruce, S. A. (1970) *Proc. Natl. Acad. Sci. U. S. A.* **66**, 72–78
11. Stellwagen, R. H., and Tomkins, G. M. (1971) *J. Mol. Biol.* **56**, 167–182
12. Stellwagen, R. H., and Tomkins, G. M. (1971) *Proc. Natl. Acad. Sci. U. S. A.* **68**, 1147–1150
13. Schwartz, S. A., and Kirsten, W. H. (1974) *Proc. Natl. Acad. Sci. U. S. A.* **71**, 3570–3574
14. Strom, C. M., and Dorfman, A. (1976) *Proc. Natl. Acad. Sci. U. S. A.* **73**, 1019–1023
15. Thompson, E. B., Tomkins, G. M., and Curran, J. F. (1966) *Proc. Natl. Acad. Sci. U. S. A.* **56**, 296–303
16. Stellwagen, R. H. (1974) *Biochim. Biophys. Acta* **338**, 428–439
17. Britten, R., Graham, D. E., and Neufeld, B. E. (1974) *Methods Enzymol.* **29E**, 363–418
18. Davidson, E. H., Hough, B. R., Klein, W. H., and Britten, R. J. (1975) *Cell* **4**, 217–238
19. Noll, H. (1967) *Nature* **215**, 360–363
20. Wyatt, G. R. (1951) *Biochem. J.* **48**, 584–590
21. O'Brien, J. C., and Stellwagen, R. H. (1977) *Exp. Cell Res.* in press
22. Britten, R. J., and Davidson, E. H. (1969) *Science* **165**, 349–357
23. Pearson, W., Davidson, E. H., and Britten, R. J. (1977) *Nucleic Acid Res.* in press
24. Sober, H. A., ed (1970) *CRC Handbook of Biochemistry* 2nd Ed, p. H97, Chemical Rubber Co. Press, Cleveland
25. Lin, S.-Y., and Riggs, A. D. (1972) *Proc. Natl. Acad. Sci. U. S. A.* **69**, 2574–2576
26. Lin, S.-Y., and Riggs, A. D. (1971) *Biochem. Biophys. Res. Commun.* **45**, 1542–1547
27. Lin, S.-Y., and Riggs, A. D. (1976) *Biochim. Biophys. Acta* **432**, 185–191
28. Yuan, R., and Meselson, M. (1970) *Proc. Natl. Acad. Sci. U. S. A.* **65**, 357–362
29. Lin, S.-Y., Lin, D., and Riggs, A. D. (1976) *Nucleic Acids Res.* **3**, 2183–2191
30. Salomon, R., Kaye, A. M., and Herzberg, M. (1969) *J. Mol. Biol.* **43**, 481–592
31. Harbers, K., Harbers, B., and Spencer, J. H. (1975) *Biochem. Biophys. Res. Commun.* **66**, 738–746
32. Miller, O. J., Schnedl, W., Allen, J., and Erlanger, B. F. (1974) *Nature* **251**, 636–637
33. Rubery, E. D., and Newton, A. A. (1973) *Biochim. Biophys. Acta* **324**, 24–36
34. Gantt, R., De Oca, F. M., and Evans, V. J. (1973) *In Vitro* **8**, 288–294
35. Kappler, J. W. (1971) *J. Cell Physiol.* **78**, 33–36
36. Deutsch, J., Razin, A., and Sedat, J. (1976) *Anal. Biochem.* **72**, 586–592

5-Methylcytosine content of rat hepatoma DNA substituted with bromodeoxyuridine.

J Singer, R H Stellwagen, J Roberts-Ems and A D Riggs

J. Biol. Chem. 1977, 252:5509-5513.

DNA Modification Mechanisms and Gene Activity during Development

Developmental clocks may depend on the enzymic modification of specific bases in repeated DNA sequences.

R. Holliday and J. E. Pugh

It is generally accepted that the differentiated state of a given type of cell is associated with the activity of a particular set of genes, together with the total inactivity of those sets associated with the differentiation of other cell types. It is also clear that the differentiated state of dividing or nondividing cells is often extremely stable. In this article we suggest mechanisms that may account for this stability and that also attempt to explain the ordered switching on or off of genes during development.

The phenotype of the organism depends on the genotype, and the genetic contribution from both parents is in almost all cases equal. Since the ultimate control of development resides in the genetic material, the actual program must be written in base sequences in the DNA. It is also clear that cytoplasmic components can have a powerful or overriding influence on genomic activity in particular cells, yet these cytoplasmic components are, of course, usually derived from the activity of genes at some earlier stage in development. A continual interaction between cytoplasmic enzymes and DNA sequences is an essential part of the model to be presented.

Modification Enzymes

In bacteria, enzymes exist which modify DNA by methylating adenine in the 6-position (*1*). These enzymes are extremely specific in their action; they modify bases at particular positions in short defined sequences of DNA, which, at least in some instances, form a palindrome. (A palindrome in DNA is an inverted duplication, with twofold rotational symmetry. The $3' \rightarrow 5'$ base sequence is the same on each strand.) These modifications prevent the DNA being degraded by restriction enzymes, which are equally specific in their action. In higher organisms, bases are also modified: 5-methylcytosine is a common component of DNA (*2*), and 6-methyladenine has been identified in simple eukaryotes (*3*). It is not yet known whether these modifications occur at specific sites. In the case of transfer RNA (tRNA) of both bacteria and higher organisms, a number of bases are modified at specific sites (*4*).

The methylation of adenine in DNA is not heritable in the usual sense, but a bacterium with a modification enzyme could, in principle, have a very different phenotype to one without if the presence or absence of methylation affected transcription. Hawthorne (*5*) and Scarano (*6*) have suggested that certain other base modifications could lead to heritable changes in base sequences and that these could control the activity of adjacent structural genes. We explore these possibilities further and suggest that such changes could operate developmental clocks which turn genes on or off after a specific number of cell divisions. In addition, we propose that the same ordered control of the transcription of genes could be achieved by the methylation of bases, without changes in sequence. The modification mechanisms are as follows.

1) To explain the instability of the mating type loci in certain strains of yeast, Hawthorne (*5*) has suggested that an operator region could exist in two alternative states. One state has A · T (adenine · thymine) base pairs at particular sites in the controlling region, and the other has G · C (guanine · cytosine) base pairs at the same sites. The transition from A · T to G · C or G · C to A · T requires cell division, and it occurs as follows. The modification of adenine at particular sites could occur by the removal of an amino group at the 6-position. This gives rise to inosine (I), and it is known that inosine base pairs with cytosine rather than thymine (*7*). Therefore one round of replication after modification will produce A · T and I · C, and at the following replication a G · C base pair is formed. The reverse transition occurs by the action of a second modifying enzyme, which removes an amino group from the 6-position of cytosine to give uracil (U). After two rounds of replication, the original A · T base pair is restored. These transitions are illustrated in Fig. 1a. The change in base sequence that occurs is irreversible if only one of the two modifying enzymes is present. Since there is now genetic evidence that mismatched bases in DNA are repaired to give normal base pairing (*8*), Hawthorne suggests that the modification occurs in the short stretches of single-stranded DNA in the replication fork. There would therefore be no opportunity for the repair of mismatched bases such as I · T or G · U.

2) Another possibility, which has been discussed by Scarano (*6*), in connection with the problem of differentiation, depends on the methylation of cytosine at the 5-position, followed by deamination at the 6-position to give thymine. In this way a G · C pair would be changed to an A · T pair after replication (Fig. 1b). The amination of thymine to 5-methylcytosine, which pairs with G, will give the reverse change.

3) To maintain methylated bases in DNA, a modification enzyme must always be present. To segregate methylated from nonmethylated cells, two enzymes are necessary. One model is illustrated in Fig. 2. The first enzyme, E1, methylates one strand within a stretch of palindromic DNA and the other strand just outside this stretch. This does not provide a substrate for the second enzyme, E2, until replication occurs, and then only one of the daughter molecules is methylated by E2. This enzyme resembles bacterial modification enzymes in adding a methyl group to the other half of the palindrome, but it differs in not acting on nonmethylated DNA. [The same discrimination between half-methylated and nonmethylated DNA is shown by restriction enzymes of bacteria (*1*).] Therefore, once both strands of the se-

R. Holliday is head of the Genetics Division, National Institute for Medical Research, Mill Hill, London NW7 1AA, England, and J. E. Pugh is a Medical Research Council Scholar in the same division.

quence are methylated, all subsequent progeny are modified, provided that E2 remains present. The other product of the first division segregates further methylated cells as long as E1 is present, but in its absence only nonmethylated progeny will be formed. The methylated state could be extremely stable, as the methyl groups would be diluted out only if the modification enzyme is lost through mutation in its structural gene. However, an essential part of our model is the switching on or off of the genes for modification enzymes during development or differentiation.

In bacteria and their viruses, specific mechanisms for the control of gene activity at the level of transcription are well known, and it has been shown that operator regions have palindromic features (*9*). It is generally believed that similar control mechanisms must exist in higher organisms. There are several simple ways in which changes in base sequence or methylation could determine whether or not a particular gene is transcribed, some of which have already been discussed by Venner and Reinert (*10*). One possibility is that the sequence where modification occurs is also an operator sequence to which a repressor binds. In one state of the DNA the repressor binds to the operator and the contiguous structural gene is inactive. In the other state the repressor does not bind to the operator and transcription occurs. Alternatively, modification could occur in the promoter sequence, that is, the short region of DNA to which the transcribing RNA polymerase first binds; in one modified state the gene would be transcribed and in the other it would not. [It is known in several instances that promoter regions contain short palindromes, since they can be attacked by restriction enzymes (*11*).] Binding sites for RNA polymerase will be common to many or all structural genes, yet the modification enzyme is specifically inactivating or activating particular genes. We must therefore postulate that the specificity of binding is provided by a defined sequence adjacent to the promoter, but that modification actually occurs in the promoter region. A third possibility, even simpler, is that base changes would simply introduce (or remove) a polypeptide chain terminating sequence within a structural gene.

In the subsequent discussion we often refer to enzymes which modify DNA as controlling enzymes and to their substrates as controlling sequences.

Somatic Segregation of Gene Activities

The modifications outlined in Figs. 1 and 2 can result in the formation of an unaltered cell, together with one in which a particular gene is activated or inactivated. This situation is like that of a stem line cell which continually divides to form cells with new functions. The stem line cell is unstable, but the daughter cells which are modified are quite stable. However, although the switching on of a single gene may commit the cell to differentiation, it is unlikely to be sufficient to bring about all the changes required for differentiation. One obvious possibility is that the first activated gene codes for another modifying enzyme that is active at several other sites in the genome, which have the same controlling sequence. This may, for instance, shut off genes whose activity is necessary for cell division and turn on others which synthesize those proteins that give the cell its particular properties. It is easy to see how somatic segregation could lead to the triggering of sequential regulatory events, or the type of cascade regulation discussed by Britten and Davidson (*12*) and Pontecorvo (*13*). A further possibility is where a gene is switched on transiently. A controlling enzyme may modify the controlling sequence adjacent to its own structural gene. The enzyme is first switched on by the action of some other controlling enzyme, but as soon as it is synthesized it overrides the action of the first enzyme by reversing the modification. In this way a controlling enzyme would only be present for one or two cell divisions, but during this time it could, of course, affect the activity of other genes.

Certain complications could arise when one considers the possible segre-

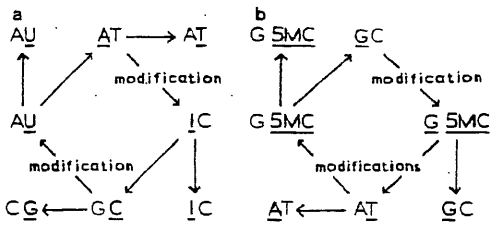


Fig. 1. (a) The transition from an A · T to a G · C base pair and the reverse transition, based on the scheme of Hawthorne (*5*). The modification of adenine occurs by the removal of an amino group at the 6-position to give inosine (I), which pairs with cytosine (C). The modification of cytosine occurs by the removal of an amino group at the 6-position to give uracil (U), which pairs with adenine. The modifications occur at the time of replication. Bases in conserved strands are underlined. (b) The transition from a G · C to an A · T base pair and the reverse transition, based on the scheme of Scarano (*6*). Cytosine is methylated at the 5-position (5 MC) and then deaminated at the 6-position to give thymine. Only one replication is necessary for the change from G · C to A · T. The reverse change involves the addition of an amino group to thymine at the 6-position to give 5-methylcytosine, which pairs with G at the subsequent replication.

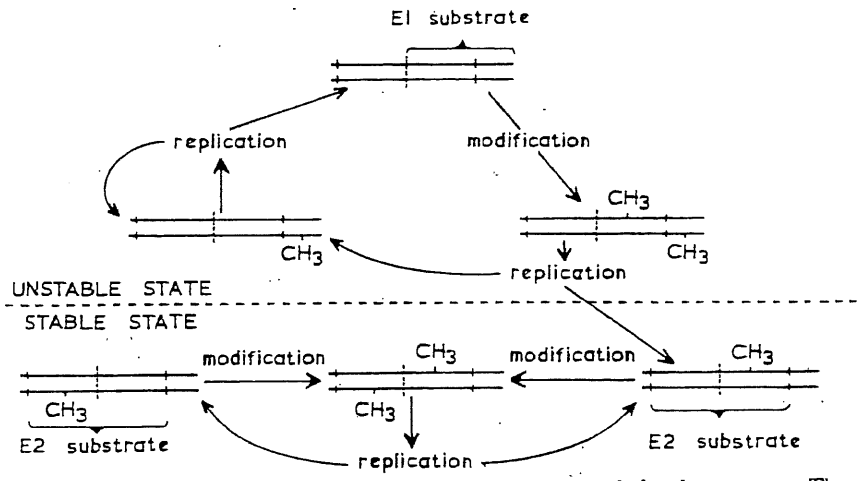


Fig. 2. The segregation of methylated DNA from an unmethylated precursor. The first modification enzyme, E1, methylates one-half of a palindromic sequence and an adjacent sequence in the complementary strand. Replication provides one substrate for the second enzyme, E2, which methylates the other half of the palindrome and all subsequent progeny molecules. In the presence of E1 and E2, unmodified or partially modified cells continually give rise to stable modified ones. If E1 is inactivated or disappears, stable modified and unmodified cells are formed.

gation events that can occur in the division of a diploid cell. Each controlling sequence occurs on each homolog, and therefore modification at both sites will result in the production of daughters with one, two, or no modifications in the ratio 1 : 2 : 1. In many instances this may not matter, as a single activated gene may set in train all the required changes. Another possibility which we favor is that controlling sequences are modified and activate genes on only one of the two homologous chromosomes. In differentiated antibody forming cells, only one of the two structural genes in a heterozygote is transcribed (*14*), and in female mammals only one of the two X chromosomes is active (*15*). The mechanisms discussed suggest how modification could occur in only one of two homologous chromosomes in a cell. Suppose the controlling sequence is a substrate for both a repressor and a modification enzyme, E1, but the affinity of the repressor is very much greater. There is therefore a low probability of modification, and a very much lower probability that both controlling sequences on both chromosomes will be modified in one cell generation. Once modification has occurred, it prevents repressor binding and allows transcription of the adjacent structural genes. The products of these inactivate E1 and switch off its synthesis. Provided that the initial modification is in both arms of a palindromic sequence, a maintenance enzyme, E2, keeps one chromosome methylated in all subsequent generations, as in the lower half of Fig. 2, whereas its homolog is unmethylated and remains so. We do not propose that this simple model will alone account for whole X chromosome inactivation.

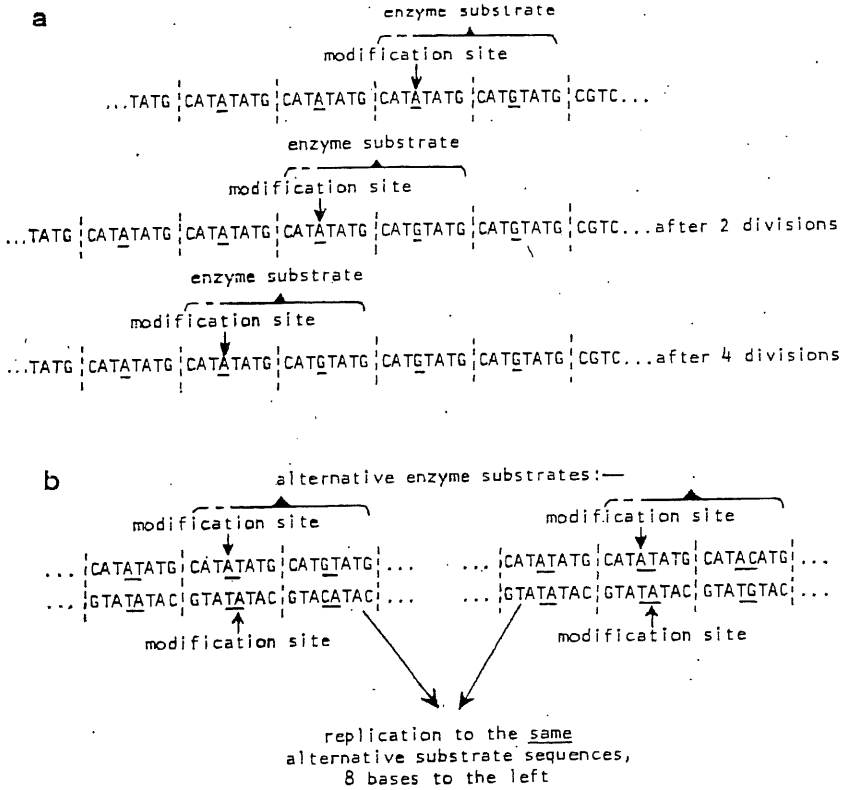


Fig. 3. (a) A mechanism for counting cell divisions based on the A · T → G · C transition in Fig. 1a. The modification enzyme recognizes the first sequence of eight bases, because it contains G at the 5-position, together with the whole or some part of the sequence to its left. The A at the 5-position of this second sequence is changed to G to give a new recognition sequence eight bases to the left. When all the sequences have been successively modified, a structural gene at the extreme left of the repeated controlling sequences (not shown here) is activated. (b) The modification of both strands of a controlling sequence. The controlling enzyme recognizes a sequence which has a G · C pair at either the 4- or 5-position. (If each controlling sequence forms a short palindrome, the recognition sequences can be structurally identical, as shown here.) In both cases it modifies A at the 4- and 5-position of the sequence to the left. Both these modified sequences then become recognition sequences after replication. The modifications therefore move progressively from right to left and count cell divisions as in (a).

Segregation of gene activities is not the only important event, as we have to consider also the mechanisms whereby all the progeny from a particular cell are altered or differentiated in the same way at a particular time in development. The application of the model to this situation is developed below.

Developmental Clocks

It can be readily seen how in principle the modification mechanism could enable a cell to count the number of divisions it has gone through during a particular stage in development. Consider the hypothetical repeating sequences shown in Fig. 3a. At the right end there is a sequence to which the modification enzyme binds. This sequence is first modified by an A → I → G transition. When this has occurred, the site of action for the enzyme has now moved eight bases to the left. This process will be repeated as many times as there are repeats of the sequence. At the end of the precisely determined number of divisions, the operator or promoter site is altered in the way that has been mentioned and the developmental switch comes into operation.

Such a developmental clock will not operate precisely if the bases modified are on one strand. In this case modified and unmodified strands segregate, and the subsequent progeny of a single cell will have modified a varying number of control sequences. This difficulty is avoided if both strands are modified (Fig. 3b). There is a binding sequence for the controlling enzyme which can exist in two forms, differing in at least two base pairs. It is adjacent to a very similar sequence which will be modified in both strands by the enzyme. These sequences when modified become the same as the binding sequences. Therefore the modifications move progressively from one end of the region of repetitive DNA to the other, and the divisions are counted.

Methylating enzymes may also count cell divisions. Of several possible mechanisms we describe one (Fig. 4): The clock is started by E1, which acts on a specific substrate at one end of the repeated sequences. It methylates one strand, and this is an essential signal for the second enzyme, E2, which inserts further methyl groups on both strands within the next sequence. This enzyme cannot act on DNA methylated on both strands in one sequence, but it does so

after replication again provides substrates with only one methylated strand. In this way an additional segment of the clock is modified at every division until the end is reached, whether or not all the sequences behind the growing points are methylated.

In both these types of clock, all the offspring from a progenitor cell will reach the same stage of development after they have gone through some specified number of divisions. The clock may, of course, trigger one or several segregation events that lead to specific differences in cell types within the clone. Separate clocks could run sequentially at the same time, overlapping each other in time within one cell lineage.

Britten and Davidson (*12*) have pointed out that the existence of multiple repeats of DNA sequences in the genome suggests that common regulatory sequences may be adjacent to many different structural genes. The developmental clocks that we have described would suggest an additional function for repetitive DNA which is not transcribed. These sequences would be tandem repeats of palindromes. Evidence for the existence of many such sequences in the DNA of higher organisms has been obtained (*16*).

Development of the Chick Limb Bud

The recent experiments on the early development of the chick wing (*17*) provide a convincing example of a developmental clock. The tip of the limb bud, which is called the progress zone, contains dividing cells, and the products of division form in strict sequence the various structures of the limb from its base to the extremity. If the progress zone from a limb in which the basic structures are nearly fully formed is transplanted to a very young limb from which the progress zone has been removed, then none of the structures are produced. On the other hand, if a young progress zone replaces one on the end of a wing which has already laid down all essential structures, then another wing is formed at the end of the first. In this case, the sequence of bone rudiments would be humerus, radius or ulna, hand, humerus, radius or ulna, hand. These results show that there is a temporal order in the laying down of successive structures, and this order might very well be related to the number of cell divisions that have elapsed in the cells of the progress zone.

A Clock for Aging?

The life-span of an organism is under genetic control, and it has frequently been asserted that there must be a developmental program for aging. More specifically, it has been suggested that the aging program might be related to division potential of cells, because diploid cells in culture have a clearly defined life-span which is dependent on the number of population doublings rather than chronological time (*18*). Current interest in mechanisms of aging has centered around error theories, for which some evidence has been published. If, instead, the life-span of these cells is programed, we think that a clock of the type outlined in Figs. 3 and 4 might provide the necessary specificity in doubling potential before senescence and cell death occurred. When the clock runs out, there are many possible deleterious or lethal events that might be triggered. For instance, the enzymes for chromosome replication or any other essential cellular function may be switched off; alternatively, there may be a general reduction in the accuracy of information transfer between macromolecules.

There is no doubt that programed death of certain tissues or groups of cells is a normal component of embryogenesis and development (*19*). This program could be based on the clock mechanisms we suggest; furthermore if restriction enzymes (specific deoxyribonucleases) (*1*) occur in higher organisms, substrates for these might be created by the loss or gain of modification enzymes in particular cells, and this would be followed by the degradation of the DNA and death of these cells.

The Developmental Program

The combination of developmental clocks and precise segregation mechanisms which together determine which genes will be activated provides the essential requirement for an ordered genetic program for development. One can describe the determined changes as being part of a developmental tree, where, at precise times during development, cells branch out into different

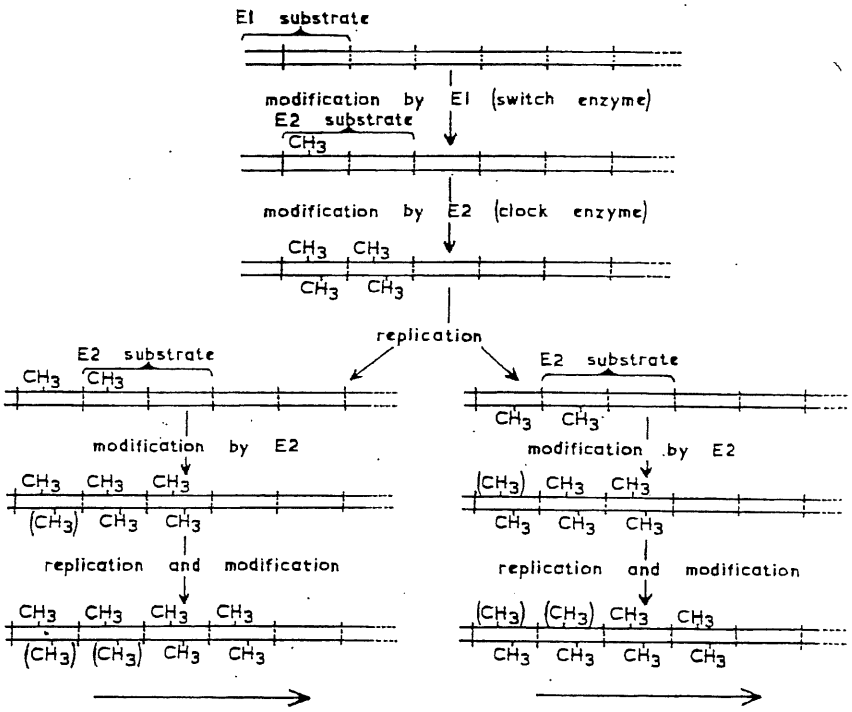


Fig. 4. A mechanism for counting cell divisions based on the methylation of palindromic controlling sequences. The first enzyme, E1, switches on the clock by recognizing a starter sequence, at the extreme left, which is adjacent to the first of the repeated sequences of the clock. One strand of this sequence is methylated by E1, and this provides a substrate for E2, which inserts three more methyl groups in the first two controlling sequences. E2 does not act further once both strands are modified. However, after replication new substrates of E2 are formed, allowing the next sequence to be methylated. (All the sequences behind the "growing point" may become modified, but this does not affect the clock mechanism.)

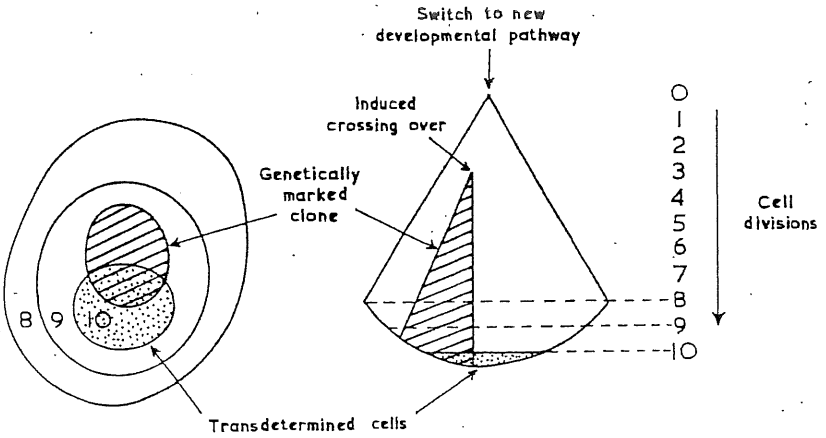


Fig. 5. An explanation for transdetermination based on an event in a single cell. The diagram on the right represents a clone derived from a cell in which a developmental clock has been triggered. It takes ten divisions for the clock to induce transdetermination in a group of cells. The induction of mitotic crossing-over by irradiation occurs after the clock has started and it produces a genetically marked clone (or patch of cells, as shown on the left) which can overlap the transdetermined region (*33*).

sublines that later themselves become subdivided into more diverse classes. At each stage the cells become more and more tied to a specific pathway of development and more and more distinct from cells derived from branches of the tree. Moreover, once a differentiated state has been reached, the model explains why it is so stable. Base changes are obviously stable, and methylated bases are maintained provided that the necessary enzyme is present. They would be lost only as a result of mutation in the structural gene for the modification enzyme.

Although the model suggests that development is clonal, it should be noticed that specific events can occur in groups of cells rather than individual ones. For instance, after fertilization a developmental clock or clocks may be set so that after n divisions one or more segregation events are triggered. At this time the 2^n cells that have been formed may segregate into two or more types of cell. Embryonic cells with specific cell surfaces can recognize each other (*20*); these cell types may therefore aggregate together into groups. This is possibly what happens when different embryos are fused to form mosaic allophenic mice (*21*).

The model described is more likely to provide an explanation for the ordered development of embryos of the mosaic type than it is for those of the regulative type. In the former, exemplified by *Drosophila* (*22*), the parts of the embryo are rigidly determined to develop into particular larval or adult structures. Removal of groups of cells of the determined embryo results in loss of specific differentiated structures. In the latter, removal of parts of the developing organism, or the inhibition of cell division, may simply result in the formation of a smaller complete organism at some later stage in development (*23*).

We do not wish to underestimate the importance of a cell's environment in the determination of its subsequent fate during development. It is, for instance, widely believed that the pattern of development is determined by fields or gradients set up by organizing or signaling cells or groups of cells (*24*). We would simply point out that the origin of such a situation must initially depend on programed differences between cells in the developing organism, and some of these differences could come about in the way suggested. Moreover, some of the switches or clocks we have discussed could be triggered by hormonal or other influences, or alternatively they may determine how a cell will respond to such stimuli. It may be significant that some cells can accept positional information only if they are first appropriately conditioned by cell division (*25*).

Totipotent Nuclei

In extensive experiments by Gurdon and his associates (*26*), nuclei from differentiated cells were injected into anucleate eggs. For instance, nuclei from tadpole intestinal epithelial cells supported embryonic development to a stage where many types of differentiated cell were present, and in some instances adult animals were formed. In other experiments, nuclei derived from adult skin tissue were successfully transplanted into anucleate eggs. These results show that nuclei in differentiated cells can be reprogramed by egg cytoplasm: they are totipotent because they can subsequently give rise to all other types of cell. There are, however, types of differentiated cell such as neurons, which contain nuclei that do not support normal development after transplantation, and in these cases the changes in genetic activity that occur during differentiation may be irreversible.

The modification mechanisms described are all reversible; therefore it is possible that a battery of specific enzymes exists in the egg cytoplasm which recognize controlling sequences and reverse many—although not necessarily all—modifications. The specificity of these enzymes may be somewhat less than those which originally introduced the modifications during development. For instance, all the modified bases in the repeated sequences of a developmental clock could be erased at a single step. It is not at all unlikely that there is a special mechanism for reprograming in the egg cytoplasm, since apart from the transplantation experiments just mentioned, the oocyte and the spermatozoan are highly specialized products of meiotic division, the nuclei of which themselves have to be reprogramed. Nevertheless, we find it hard to believe that reprograming could actually involve base changes in the DNA, and for this reason we tend to prefer the version of our model that depends on methylation of particular bases. A general demethylating enzyme is a possibility (provided that it was removed before the embryo started development), but we prefer the following alternative.

If the egg cytoplasm contains no maintenance enzyme, then methyl groups will simply be diluted out during the early cleavage divisions. At each nuclear division the number of chromosomes containing methyl groups will be reduced by half. After x divisions the probability of any one chromosome remaining modified is $2n/2^x$, where n is the haploid chromosome number. Where n is between 10 and 30, it would need between 11 and 13 divisions to reduce the number of cells containing at least one modified chromosome to 1 percent. We suggest that an initial clock, immune from the diluting out process,

would set in train the whole process of development after approximately this number of nuclear divisions has occurred.

Effect of Bromodeoxyuridine on Differentiation

Perhaps the strongest evidence that DNA is directly implicated in differentiation comes from numerous studies with the thymidine analog bromodeoxyuridine (BrdU). It has frequently been shown that low concentrations of BrdU which are nontoxic to cells specifically inhibit differentiation or development. There is no effect if excess thymidine is added at the same time as the analog, but in several instances once the BrdU is incorporated into DNA, the block in differentiation cannot be reversed by adding excess thymidine. Only a few of the many examples of the specific action of BrdU can be mentioned [for a full review, see (*27*)]. Myogenic cells can be cultivated in vitro for several days. After this time, DNA synthesis ceases, the cells fuse to form multinucleate tubules and synthesize the contractile proteins actin and myosin. Bromodeoxyuridine does not prevent the myoblasts from proliferating, but its presence, even for one cell division, completely inhibits their differentiation (*28*). Mesoderm of the chick limb bud differentiates into cartilage in cell culture, and this differentiation is irreversibly blocked by the substitution of approximately 2 percent of the thymine in DNA with the bromo analog. After treatment with BrdU is terminated, the analog rapidly disappears from the dividing cells, but even so differentiation does not then occur (*29*). In other instances, the analog is diluted out by replication and differentiation follows. Finally, it has been shown that BrdU blocks the development of embryos if applied at an early cleavage stage (*30*).

We propose that these effects are brought about by the substitution of a bromine atom for the methyl group on the 5-position of thymine, and this prevents the normal modification of controlling sequences during development. This could occur either by preventing the loss or gain of a methyl group of a particular pyrimidine base (for instance, in the change from cytosine to thymine previously mentioned) or, more generally, by altering the action of modification enzymes on controlling sequences containing BrdU-substituted DNA.

Determination and Transdetermination in Drosophila

The stability of differentiated cells has already been mentioned, but we now turn to the remarkable studies of Hadorn and his associates (*22, 31*), who have demonstrated that the determined state of undifferentiated larval cells can be very stable. The adult structures of *Drosophila* are formed during metamorphosis from imaginal discs in the larva. Imaginal disc tissue can be grown in the abdomen of adult flies and continuously propagated by transfer of pieces of tissue to fresh adults. Disc tissue reimplanted in larvae differentiates during metamorphosis to produce a particular adult structure, such as part of a wing, leg, or antenna. This is triggered by the hormone ecdysone, which activates the developmental program and allows the further events required for differentiation to proceed. A particular line of disc cells is determined to produce a specific adult structure. This is inherited from cell to cell, as disc fragments have been subcultured for more than 70 transfer generations over a period of several years without any change in determined state. We suggest that this stability is due to the inheritance of appropriate modifications in their DNA. Sometimes disc tissue that is determined to develop in one direction spontaneously changes to another determined state. This transdetermination never occurs in the absence of proliferation; indeed, its frequency is related to the number of cell divisions which have occurred. If during growth the modifications are occasionally lost, then these cells may move into a determined state distinct from the first one. It is a characteristic of transdetermination that specific changes occur more frequently than others and that successive changes follow particular pathways.

Kauffman (*32*) has presented a detailed model for determination based on the setting of a number of bistable states, or developmental switches. The various pathways for transdetermination are explicable if the setting of the switches alters with given frequencies, one state changing to the other more frequently than the reverse change. His analysis is quite consistent with our model if the switches are modified or unmodified states of particular controlling sequences and one change, for instance the failure to methylate DNA, is more frequent than the reverse.

By use of mitotic crossing-over to mark particular groups of cells, it is possible to show that a patch of tissue in which transdetermination has occurred can occasionally overlap one which has arisen as a result of mitotic crossing-over in a single cell. It is therefore impossible for each patch to be an individual clone, and it has consequently been argued that transdetermination occurs in groups of cells (*33*). However, this argument no longer holds if transdetermination depends on two events: first, the reversion in one cell to a predetermined state, then a given amount of proliferation to a new determined state. If only a proportion of the cells in the clone have undergone sufficient divisions to reach the new determined state (and such cells are known to aggregate together), then it is quite possible for the patch from mitotic crossing-over to be included within this larger clone, only part of which has undergone transdetermination (Fig. 5).

Homoeotic mutants are those that produce developmental defects analogous to transdetermination. For instance, the mutation *aristapedia* in *Drosophila* results in the development of a leg structure in place of part of an antenna (*34*). Such mutants may have a defect either in a controlling enzyme, which fails to recognize a particular controlling sequence, or alternatively they might have an altered controlling sequence which is not recognized by the appropriate controlling enzyme. As a result, cells are channeled into an alternative developmental pathway. It has been shown that a homoeotic mutant can mimic transdetermination in that the developmental abnormality originates in a group of cells rather than in one. But in this case the cells are part of a larger clone, the whole organism, with a particular genetic defect. In a similar way a patch of transdetermined cells could originate from a larger clone derived from a cell with altered DNA.

Conclusions

We are aware that no direct evidence exists for specific modification enzymes in eukaryotes, let alone that such enzymes might exercise control of gene activities. Nevertheless, in view of our almost complete ignorance of the mechanism for the unfolding of the genetic program during development, it seems justifiable to suggest speculative hypotheses that may lead to meaningful experi-

mental approaches, particularly when these hypotheses are based on some of the known features of modification systems in bacteria. It is significant that Sager (*35*) has argued, from a quite different viewpoint, that restriction and modification mechanisms may exist in higher organisms.

A direct search for specific modification enzymes and modified bases in specific sequences will be difficult, as the number of controlling sequences of any one type in the genome may be only one or a few. Methylases have been identified in sea urchin embryos (*36*), and there is evidence that the distribution of methyl groups in DNA is not random. It may be significant that the doublet CpG is the most highly methylated (*6, 36*), but occurs much less frequently than expected from the overall base composition of eukaryotic DNA (*37*). A search for the transition of cytosine to thymine by methylation and deamination has not so far been successful (*38*).

Although further study of methylases and the pattern of methylation of certain families of reiterated DNA in different tissues or at different stages of development might well be profitable, we feel that it is unlikely that biochemical studies alone will provide direct evidence for our model. The use of developmental mutants is probably essential, since by comparison with wild-type organisms it may be possible to identify the nature of their biochemical defects. We would predict two general classes of mutant: those with altered controlling sequences, which may be dominant (as in the case of operator constitutive mutations); and those with altered controlling enzymes, which would usually be recessive and obtainable in temperature sensitive form. Analysis of developmental pathways can be assisted by the use of homoeotic mutants, and in this connection we agree with McClintock (*39*), who has emphasized that, if the ordered processes of development are deranged, then genes which usually become active at very specific times may instead be activated spasmodically or in random fashion during development. Her studies with maize [for a review, see (*40*)] have led to the discovery of unstable states and controlling elements. The latter not only control the stability and level of expression of nearby genes, but also transpose from one chromosomal location to another. The possibility of transposition of genetic elements has also been discussed in connection with the problem of immunoglobulin synthesis from genes coding for constant and variable regions (*41*). Many of the properties of such systems as McClintock's could, we believe, be explained on the basis of repeated sequences of controlling DNA, which could dissociate from and reassociate with several chromosomal sites by means of genetic recombination. What may now be needed is an examination of these genetic elements in a higher organism in which both biochemical and genetic studies can be undertaken.

References and Notes

1. W. Arber, in *The Bacteriophage Lambda*, A. D. Hershey, Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1971), p. 83; M. Meselson, R. Yuan, J. Heywood, *Annu. Rev. Biochem.* **41**, 447 (1972); J. D. Smith, W. Arber, U. Kühnlein, *J. Mol. Biol.* **63**, 1 (1972).
2. J. Doskocil and F. Sorm, *Biochim. Biophys. Acta* **55**, 953 (1962); G. R. Wyatt, *Biochem. J.* **48**, 584 (1951).
3. M. A. Gorovsky, S. Hattman, G. L. Pleger, *J. Cell Biol.* **56**, 697 (1973).
4. D. Soll, *Science* **173**, 293 (1971).
5. D. C. Hawthorne, personal communication and in preparation.
6. E. Scarano, *Adv. Cytopharmacol.* **1**, 13 (1971).
7. D. R. Davies and A. Rich, *J. Am. Chem. Soc.* **80**, 1003 (1958).
8. R. Holliday, *Genetics*, in press.
9. W. Gilbert, N. Maizels, A. Maxam, *Cold Spring Harbor Symp. Quant. Biol.* **38**, 845 (1973); T. Maniatis, B. G. Barrell, M. Ptashne, J. Donelson, *Nature (Lond.)* **250**, 394 (1974); B. Lewin, *Cell* **2**, 1 (1974).
10. H. Venner and H. Reinert, *Z. Allg. Mikrobiol.* **13**, 613 (1973).
11. B. Allet, R. J. Roberts, R. F. Gesteland, R. Solem, *Nature (Lond.)* **249**, 217 (1974); R. Maurer, T. Maniatis, M. Ptashne, *ibid.*, p. 221.
12. R. J. Britten and E. H. Davidson, *Science* **165**, 349 (1969); E. H. Davidson and R. J. Britten, *Q. Rev. Biol.* **48**, 565 (1973).
13. G. Pontecorvo, *Proc. R. Soc. Lond. Ser. B* **158**, 1 (1963).
14. B. Pernis, *Cold Spring Harbor Symp. Quant. Biol.* **33**, 333 (1967).
15. M. F. Lyon, *Biol. Rev. (Camb.)* **47**, 1 (1972).
16. D. A. Wilson and C. A. Thomas, *J. Mol. Biol.* **84**, 115 (1974); T. R. Cech, A. Rosenfeld, J. E. Hearst, *ibid.* **81**, 299 (1973).
17. D. Summerbell, J. Lewis, L. Wolpert, *Nature (Lond.)* **244**, 492 (1973).
18. L. Hayflick, *Exp. Cell Res.* **37**, 614 (1965); J. Ponten, *Inst. Natl. Santé Rech. Med.* **27**, 53 (1973).
19. A. Glucksmann, *Biol. Rev. (Camb.)* **26**, 59 (1951); J. W. Saunders, Jr., *Science* **154**, 604 (1966).
20. S. Roth, *Dev. Biol.* **18**, 602 (1968); R. Nöthiger, *Wilhelm Roux' Arch. Entwicklungsmech. Org.* **155**, 269 (1964).
21. B. Mintz, *Symp. Soc. Exp. Biol.* **25**, 345 (1971); M. N. Nesbitt and S. M. Gartler, *Annu. Rev. Genet.* **5**, 143 (1971).
22. H. Ursprung and R. Nöthiger, *Biology of Imaginal Discs* (Springer Verlag, Berlin, 1972).
23. J. S. Huxley and G. R. de Beer, *The Elements of Experimental Embryology* (Cambridge Univ. Press, Cambridge, 1934); J. Cooke, *Nature (Lond.)* **243**, 55 (1973).
24. L. Wolpert, *J. Theor. Biol.* **25**, 1 (1969); F. H. C. Crick, *Symp. Soc. Exp. Biol.* **25**, 429 (1971).
25. P. A. Lawrence, F. H. C. Crick, H. Munro, *J. Cell Sci.* **11**, 815 (1972).
26. J. B. Gurdon, *J. Embryol. Exp. Morphol.* **10**, 622 (1962); ——— and R. A. Laskey, *ibid.* **24**, 227 (1970); J. B. Gurdon and V. Vehlinger, *Nature (Lond.)* **210**, 1240 (1966).
27. H. Holtzer, H. Weintraub, R. Mayne, B. Mochan, *Curr. Top. Dev. Biol.* **7**, 229 (1972).
28. R. Bischoff and H. Holtzer, *J. Cell Biol.* **48**, 523 (1971).
29. D. Levitt and A. Dorfman, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1253 (1972).
30. M. Gontcharoff and D. Mazia, *Exp. Cell Res.* **46**, 315 (1967); R. Tenser and J. Brachet, *Differentiation* **1**, 51 (1973).
31. E. Hadorn, *Brookhaven Symp. Biol.* **18**, 148 (1965); H. Wildermuth, *Sci. Prog.* **58**, 329 (1970); W. Gehring, *J. Embryol. Exp. Morphol.* **79**, 731 (1973).
32. S. A. Kauffman, *Science* **181**, 310 (1973).
33. W. Gehring, *Dev. Biol.* **16**, 438 (1967).
34. E. Balkaschina, *Wilhelm Roux' Arch. Entwicklungsmech. Org.* **115**, 448 (1929).
35. R. Sager and R. Kitchin, personal communication and in preparation; R. Sager and Z. Ramanis, *Theor. Appl. Genet.* **43**, 101 (1973).
36. P. Grippo, M. Iaccarino, E. Parisi, E. Scarano, *J. Mol. Biol.* **36**, 195 (1968).
37. J. H. Subak-Sharpe, *Br. Med. Bull.* **23**, 161 (1967); J. M. Morrison, H. M. Keir, H. Subak-Sharpe, L. V. Crawford, *J. Gen. Virol.* **1**, 101 (1967).
38. T. W. Sneider, *J. Mol. Biol.* **79**, 731 (1973).
39. B. McClintock, *Cold Spring Harbor Symp. Quant. Biol.* **16**, 13 (1951); *ibid.* **21**, 197 (1956); *Brookhaven Symp. Biol.* **8**, 58 (1955); *ibid.* **18**, 164 (1965).
40. J. R. S. Fincham and G. R. K. Sastry, *Annu. Rev. Genet.*, in press.
41. W. J. Dreyer and J. C. Bennett, *Proc. Natl. Acad. Sci. U.S.A.* **54**, 864 (1965); E. S. Lennox and M. Cohn, *Annu. Rev. Biochem.* **36**, 365 (1967); J. A. Gally and G. M. Edelman, *Annu. Rev. Genet.* **6**, 1 (1972).

42. We thank Drs. B. Alberts, J. Cooke, F. H. C. Crick, R. M. Gaze, J. B. Gurdon, B. Lewin, A. McLaren, Zh. A. Medvedev, and L. E. Orgel for their helpful comments.

BMC Genomics

RESEARCH ARTICLE Open Access

Euchromatin islands in large heterochromatin domains are enriched for CTCF binding and differentially DNA-methylated regions

Bo Wen[1,4†], Hao Wu[2†], Yuin-Han Loh[3†], Eirikur Briem[1], George Q Daley[3] and Andrew P Feinberg[1*]

Abstract

Background: The organization of higher order chromatin is an emerging epigenetic mechanism for understanding development and disease. We and others have previously observed dynamic changes during differentiation and oncogenesis in large heterochromatin domains such as Large Organized Chromatin K (lysine) modifications (LOCKs), of histone H3 lysine-9 dimethylation (H3K9me2) or other repressive histone posttranslational modifications. The microstructure of these regions has not previously been explored.

Results: We analyzed the genome-wide distribution of H3K9me2 in two human pluripotent stem cell lines and three differentiated cells lines. We identified > 2,500 small regions with very low H3K9me2 signals in the body of LOCKs, which were termed as euchromatin islands (EIs). EIs are 6.5-fold enriched for DNase I Hypersensitive Sites and 8-fold enriched for the binding of CTCF, the major organizer of higher-order chromatin. Furthermore, EIs are 2–6 fold enriched for differentially DNA-methylated regions associated with tissue types (T-DMRs), reprogramming (R-DMRs) and cancer (C-DMRs). Gene ontology (GO) analysis suggests that EI-associated genes are functionally related to organ system development, cell adhesion and cell differentiation.

Conclusions: We identify the existence of EIs as a finer layer of epigenomic architecture within large heterochromatin domains. Their enrichment for CTCF sites and DNAse hypersensitive sites, as well as association with DMRs, suggest that EIs play an important role in normal epigenomic architecture and its disruption in disease.

Keywords: Epigenetics, H3K9me2, Euchromatin islands, CTCF, DNA methylation

Background

Epigenetics involves information retained during cell division other than DNA sequence per se, and both DNA methylation and post-translational modifications of histones are fundamental in understanding normal development and disease [1-3]. Genome-scale localization of histone modifications had been extensively mapped in mammalian genomes [4-10]. While most of these studies focused on local regulatory elements such as promoters and enhancers, global organization of the chromatin has not been well understood.

Recent evidence indicates that repressive histone modifications form large scale domains in both mouse and human genomes. We had previously identified large blocks of H3 lysine 9 dimethylation (H3K9me2), termed Large Organized Chromatin K9-modifications (LOCKs), which affect more than 40% of the mouse genome in liver cells [11]. LOCKs significantly overlap with lamina-associated domains (LADs) [12] and are associated with domain-wide gene silencing in a tissue-specific manner. Importantly, both coverage and domain size of LOCKs increase upon differentiation of mouse embryonic stem cells (ESCs) [11]. On the other hand, genome-scale reduction of LOCKs was seen in epithelial-to-mesenchymal transition (EMT) induced by TGF-β treatment of mouse hepatocytes, a process in which cells gain stem cell-like and malignant-type traits [13]. Similarly, large blocks of other repressive marks (H3K9me3 and H3K27me3) are also found to expand in human lung fibroblasts compared

* Correspondence: afeinberg@jhu.edu
† Equal contributors
[1]Center for Epigenetics and Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA
Full list of author information is available at the end of the article

with human ESCs [14], and those blocks/LOCKs expand in breast cancer cells relative to normal epithelial cells [15]. Furthermore, large H3K9me3 and H4K20me3 blocks specifically coat olfactory receptor (OR) gene clusters in mouse olfactory epithelium but not in liver [16]. Taken together, these data demonstrate that large heterochromatin domains are highly dynamic in differentiation and tumorigenesis.

DNA methylation has been tightly linked to development and disease [1]. We previously reported that differentially methylated regions (DMRs) related to tissue specificity (T-DMRs), colon cancer (C-DMRs) and reprogramming (R-DMRs) have largely common targets in the genome and are strongly associated with local regulation of adjacent genes [17,18]. Whole genome bisulfite sequencing had found partial methylated domains (PMDs) which are highly methylated in human ESCs but partially methylated in fibroblasts [19]. Similar large hypomethylation blocks relative to normal cells have been identified in colon cancer [20] and breast cancer cells [15], and loss of methylation in these regions is accompanied by acquisition of large domains of H3K9me3 and H3K27me3 [15].

Surprisingly, the relationship of H3K9me2 LOCKs/blocks to DMRs has not been previously assessed. In the course of this investigation, we identified a new chromatin unit we term "euchromatin island" which may serve as a fulcrum between DNA methylation and chromatin in development.

Results

We analyzed whole genome distribution of H3K9me2 by ChIP-chip using a highly specific monoclonal antibody in two human pluripotent stem cell (PSC) lines (human ESC H1, human iPSC ADA-38) and three primary differentiated cell lines: human astrocytes (HA), human aortic endothelial cells (HAEC) and human pulmonary fibroblasts (HPF). For differentiated cells, we used early passages of primary cells instead of immortalized cell lines to avoid potentially aberrant epigenetic changes due to long time cell culture and immortalization of the cells [21]. These differentiated lines represent three germ layers: ectoderm (HA), mesoderm (HAEC) and endoderm (HPF).

We normalized the ChIP-chip data as described [11] to calculate the log2 ratios of ChIP/Input comparable among cell types. By using the 90^{th} quantile as a cutoff to define large domains, the genome coverage of LOCKs was found to increase from 17.5-24% in PSC lines, to 39.3-44.8% in differentiated cells, and the average sizes of LOCK expanded from 142–171 kb in PSC lines, to 233–315 kb in the differentiated. The trends were the same when we used different cutoffs to define LOCKs (Additional file 1: Table S1), consistent with our previous findings that LOCKs increase after mouse ESC differentiation [11]. For example, in the *WSCD2* gene locus, only some small H3K9me2 peaks can be seen in the PSCs, but the H3K9me2 enriched regions expanded to ~350 kb long and cover the whole gene body and its flanking regions in the differentiated cells (Figure 1A).

To validate the ChIP-chip data, we performed quantitative PCR (qPCR) on 23 loci using independently prepared ChIP and input DNA samples from four cell types. For all the cases, the quantitative differences of H3K9me2 enrichments within and among samples detected by ChIP-chip were well validated by qPCR (Additional file 2: Figure S1). Overall, the ChIP/Input log2 ratios of microarray (ChIP-chip) and qPCR were strongly correlated (R^2 = 0.87, Figure 1B), indicating that the ChIP-chip data are of high quality.

To reveal the relationship between dynamics of H3K9me2 and DNA methylation on a large scale, we compared genome-wide distributions of LOCKs (this study), PMDs in fibroblasts [14], and DNA hypomethylation blocks in colon cancer [20]. LOCKs in fibroblasts (HPF) largely overlap PMDs (Additional file 3: Figure S2A), and overall 61.5% regions of LOCKs in HPF coincide with PMDs ($p < 0.001$, based on 1,000 permutations), and H3K9me2 signals in the regions of PMDs are higher than non-PMD regions (Additional file 3: Figure S2B). Furthermore, more than 80% LOCK regions in HPF were contained within DNA hypomethylation blocks found in colon cancer tissues (Additional file 3: Figure S2). Thus, our data support a strong correlation between LOCKs and DNA hypomethylation blocks in human cells.

On closer examination of the microstructure of the LOCKs, we noticed that many small H3K9me2-depleted regions are located in the body of LOCKs. These regions are a few kb in length, and away from the LOCK boundaries. We found that these regions are abundant in the genome, and they appear to be associated with open chromatin (see below). Thus, we termed these regions Euchromatin Islands (EIs). As an example, an EI was found near the transcription start sites (TSSs) of the cadherin 11 gene (*CDH11*, Figure 2A), of which epigenetic disruption was associated with metastasis of human cancers [22]. Other examples of EIs include within the gene body of *PDILT*, a testis-specific gene; and downstream of the glycoprotein 2 (*GP2*) gene (Figure 2B).

Then we developed a statistical algorithm to identify EIs genome-wide (see Methods). We identified 758 to 2,465 EIs across cell types, with average sizes from 4.4 to 5.9 kb (Table 1 and Additional file 4: Table S2). These EIs form strong dips relative to adjacent LOCK regions as demonstrated by average H3K9me2 densities (Additional file 5: Figure S3). We have performed replicates on one array of the "Mouse ChIP-chip 2.1M

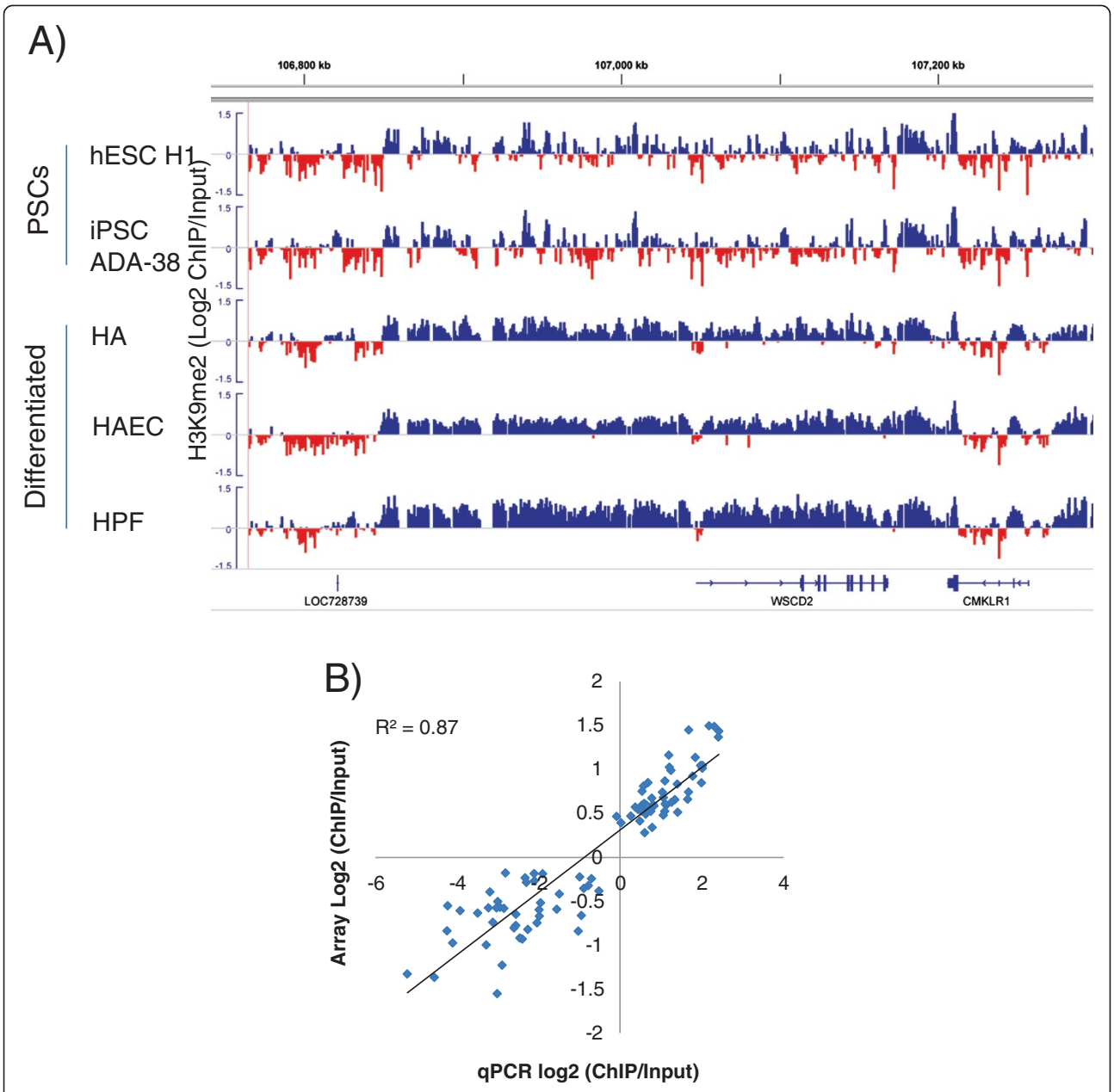


Figure 1 Global pattern of H3K9me2 in human pluripotent and differentiated cells. A) A representative region that shows LOCKs in differentiated cells (HA, HAEC and HPF) but not PSCs (H1 and ADA-38). Shown are H3K9me2 signals of ChIP-chip experiments in a ~500 kb long region containing *WSCD2* gene. Positive and negative log2 (ChIP/Input) rations are shown in blue and red, respectively. **B**) qPCR validation of ChIP-chip data. By comparing four cell lines in 23 regions (Figure S1), enrichments (log2 ChIP/Input) measured by ChIP-chip (Y axis) and ChIP-qPCR (X axis) are highly correlated (R^2 = 0.87).

Whole-Genome Tiling sets", which covers 10% of the genome. The EIs detected from the replicate experiments have high concordance with the ones from whole genome arrays (Additional file 6: Figure S4). Percentage of EIs detected from whole genome arrays that can also be detected from replicate arrays are 76.3% for H1, 74.1% for ADA-38, 63.3% for HA, 84.4% for HAEC and 71.4% for HPF. To exclude the possibility that EIs resulted from lack of histones in these regions, we plotted nucleosome density around EIs, and no depletion of nucleosomes was observed in EIs (Additional file 7: Figure S5), indicating that the observation of EIs is not due to nucleosome positioning.

Among the five cell lines, 4.6% to 12.7% of EIs coincided with transcriptional start sites (TSSs), which associated with 60 to 409 genes across cell types. Compared to random, the enrichment at TSS ranged from 2.7 (in ADA-38) to 7.8 (HA), with randomization p-values

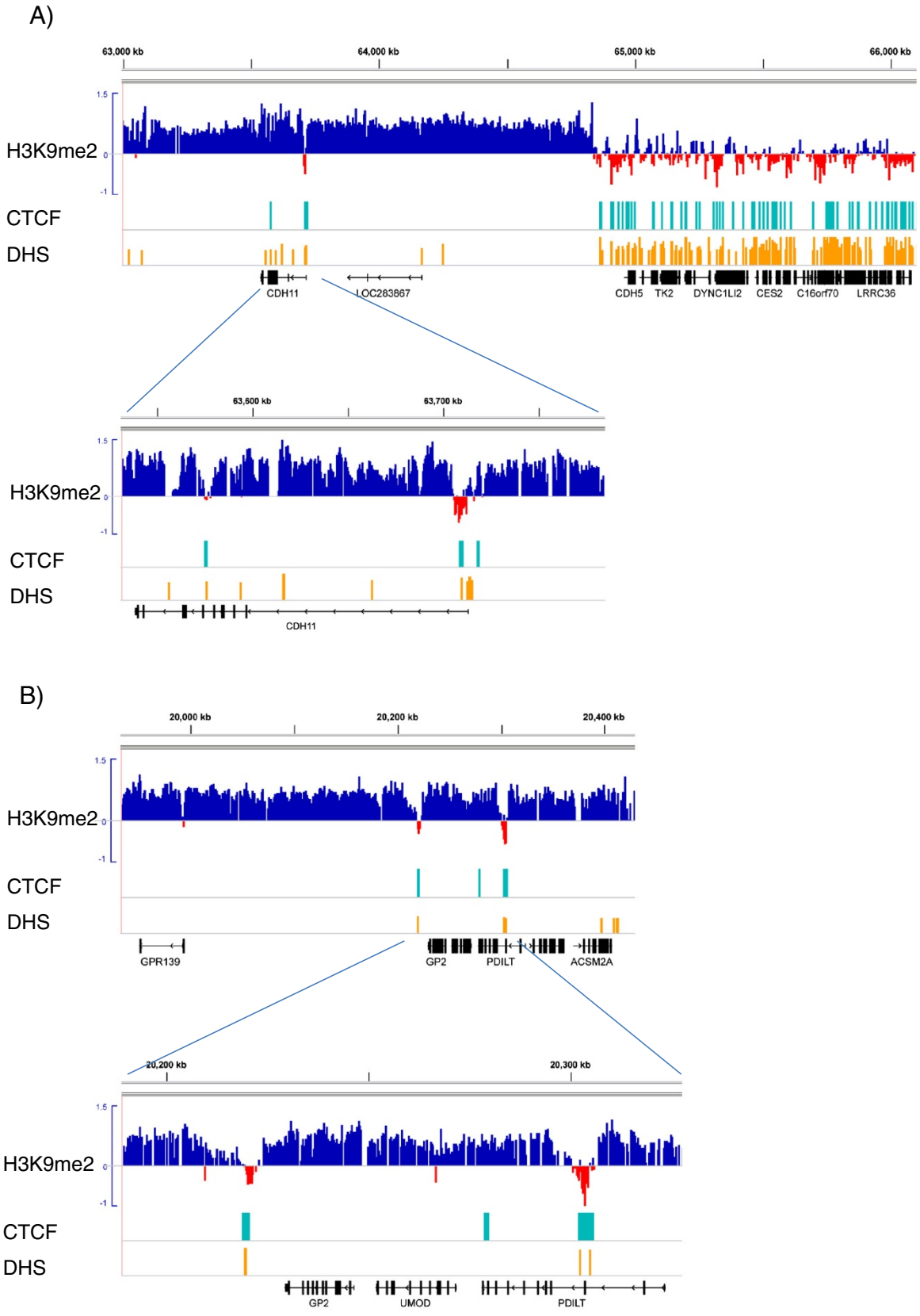


Figure 2 (See legend on next page.)

(See figure on previous page.)
Figure 2 Euchromatin islands (EIs) in LOCKs overlap CTCF interacting regions and DNase hypersensitive sites (DHSs). H3K9me2 log ratios of HAEC are shown on the top track. CTCF binding regions and DHSs of HUVEC are denoted as light blue and orange bars, respectively. EIs are small regions with strong negative signals within the body of LOCKs. **A**) Shown is a 3 Mb long region (top) containing CHD11 genes (zoomed-in view on the bottom), a member of the cadherin gene family. CFCF interacting regions and DHSs are highly depleted in the H3K9me2 blocks (LOCKS), but overlap the EI located near the TSS of *CDH11* gene. **B**) Additional examples of EIs near GP2 and PDILT genes.

$< 10^{-3}$ for all cell lines (Table 1). We further investigated the spatial relationship between EI and CpG islands (CGI). We found that 4.7% to 17% of EIs overlapped with CGIs, with enrichment ranging from 0.6 to 2.1. The randomization test suggested that EIs significantly overlapped with CGI in differentiated cells, but not in ES and iPS cells (Table 1).

To probe the chromatin features of EIs, we compared locations of EIs in H1, HAEC and HPF with public datasets of comparable cell lines [10,23]. Interestingly, EIs highly coincide with regions interacting with CCCTC-binding factor (CTCF), the major organizer of higher-order chromatin in mammalian genomes (Figure 2). Overall, up to 61.3% of EIs overlap with CTCF binding regions, which are 8.2-fold enriched compared with the random pattern ($P < 10^{-3}$, Table 2). Furthermore, up to 49% of EIs overlap with DNase hypersensitive sites (DHSs), the hallmark of open chromatin, which is 6.5-fold enrichment compared with the random ($P < 10^{-3}$, Figure 2 and Table 2). We further explored the overlaps of EIs with other histone modifications, and found that EIs highly overlaps with H3K4me3 (Enrichment up to 5.3) and H3K9ac (Enrichment up to 3.3), but less enrich for H3K27me3 (Enrichments from 1.7 to 2.2) and H3K36me3 (Enrichments from 0.7 to 2.1). The enrichments are similar among the three cell types. In addition, we investigated the enrichment by comparing EIs with random pattern within LOCK regions, and got similar results and even stronger enrichments for CTCF (up to 13.9 fold, Table 2).

We then asked whether there is any association between EIs and DMRs. For this purpose, we compared genomic locations of EIs with DMRs identified by CHARM array [17,18]. We found that EIs are highly enriched for DMRs distinguishing tissue types (T-DMRs). For example, EIs near TSSs of nitric oxide synthase 1 (*NOS1*), xylosyltransferase I (*XYLT1*) and heparan sulfate (glucosamine) 3-O-sulfotransferase 1 (*HS3ST1*) all overlap T-DMRs (Figure 3). Overall, a large fraction of EIs (39-62% across the five cell types) overlap with T-DMRs, with enrichment from 2.1 to 2.9 folds relative to random patterns (Table 3; $p < 10^{-3}$).

We further tested the relationship between EIs and DMRs associated with reprogramming (R-DMRs). Similar to T-DMRs, R-DMRs were more methylated in iPSC cells compared to fibroblasts (Hyper R-DMRs) were also significantly enriched in EIs of all the cell types, whose enrichments ranging from 2.2 to 3.2 fold. However, R-DMRs less methylated in iPSC cells (Hypo R-DMRs) were highly enriched in EIs of differentiated cells (5.1 to 6.2 folds enrichment, P values all $< 10^{-3}$), but much less enriched in PSCs (1.2 to 3.1 folds of enrichment). Importantly, EIs in HPF, the same cell type of parental cells in reprogramming, are strongly enriched for hypomethylated R-DMRs (enrichment = 6.1, $P < 10^{-3}$), whereas those in iPSCs did not significantly overlap with hypomethylated R-DMRs (enrichment = 1.2, P = 0.33), indicating a coordinated hypomethylation in these EIs during reprogramming.

We then compared EI locations with colon cancer-associated DMRs (C-DMRs) and observed an opposite trend to that of R-DMRs. EIs in 4 out of 5 cell lines were significantly enriched for C-DMRs more methylated in colon cancers (hypermethylated C-DMRs), and the enrichment ranged from 3.8 to 5.4 fold (Table 3). In contrast, all five cells types were not significantly enriched for C-DMRs less methylated in cancers (hypomethylated C-DMRs). These results were further confirmed by comparing EIs with an independent list of C-DMRs discovered by whole genome bisulfite sequencing [20]. EIs of all five cell lines significantly overlapped hypermethylated C-DMRs (enrichment from 4.1 to 7.6 fold, P values all $< 10^{-3}$), whereas none of them were significantly enriched for hypomethylated C-DMRs (Table 3). Interestingly, almost all EIs (98%) that associated with hypermethylated C-DMRs also overlap CGIs. These data suggest that EIs in normal cells may become hypermethylated in cancers.

To explore the biological role of EIs, we compared expression levels of genes associated with EIs, of genes with LOCKs but not EIs, and of genes not overlapping LOCKs (Figure 4A). It is clear that expressions of genes overlapping EIs are significantly higher than those of within LOCKs but not EIs (*t*-test, p < 2–16). To further test whether EI associated genes are regulated by other histone marks, we investigated the relationship between H3K36me3/H3K27me3 and genes with EIs, with LOCKs and without LOCKs (Figure 4B). In either category (with or without K36me3/K27me3), genes at LOCK regions always have the lowest expression and genes at non-LOCK regions have the highest. However, genes with EIs have expressions in the middle, and positively (negatively) associated with H3K36me3 (H3K27me3),

Table 1 Overlap of EIs with CpG islands (CGIs) and transcription start sites (TSSs)

Cell line	Number of EIs	Average Size of EIs (bp)	% of EIs overlap CGIs (En.[b], P[c])	% of EIs overlap TSSs (En.[b], P[c])	# of genes associated with EIs[a]
H1	1,060	4880	9.7 (1.3, 0.06)	7.9 (5.0, $<10^{-3}$)	119
ADA-38	758	4401	4.7 (0.6, 0.99)	4.6 (2.7, $<10^{-3}$)	60
HA	2,254	5029	14.2 (1.8, $<10^{-3}$)	12.4 (7.8, $<10^{-3}$)	338
HAEC	2,359	5867	13.6 (1.6, $<10^{-3}$)	11.3 (5.4, $<10^{-3}$)	373
HPF	2,465	5477	17 (2.1, $<10^{-3}$)	12.7 (6.5, $<10^{-3}$)	409

[a] 1 kb up and downstream of TSS overlapping EIs; [b] Enrichment compared to random patterns; [c] p values calculated by 1,000 permutations.

indicating that EI related genes could be regulated by these two marks.

Then we conducted Gene Ontology (GO) analysis with genes whose TSSs are associated with EIs. EI-associated genes in differentiated cells were strongly associated with 1) biological processes such as system development, cell adhesion and cell differentiation; 2) cellular compartments of plasma membrane and synapse, and 3) molecular function of ion binding and channel activity (Table 4).

Finally, to test whether EIs are associated with specific cellular functions, we compared the location of EIs among the three differentiated cell lines. Based on our current strategy to define EIs, ~50% of them are cell type specific (Figure 5A). It should be noted that the detection of EI is based on the definition of LOCKs as well as the amount of reduction of H3K9me2 levels within LOCK bodies. Some tissue specific EIs may be due to differential LOCKs or different amount of H3K9me2 reductions among cell types. Due to these reasons the number of tissue specific EIs is likely an over-estimate. New technology with higher resolution and dynamic range, such as ChIP-seq, will help achieve better accuracy and specificity in tissue comparisons. Nevertheless, we found that some tissue specific EIs are biologically meaningful. For example, an EI is located near the TSS of Down syndrome cell adhesion molecule gene (*DSCAM*) in astrocytes (HA) but not the other two cell types (Figure 5B). It was shown that Dscam diversity is essential for neuronal circuit assembly [24], and genetic variations of this gene were associated with Down syndrome and congenital heart disease (DSCHD) [25] and

Table 2 Overlaps (%) of EIs with chromatin marks[a]

	EIs overlap with	Observed	Random within LOCKs	Fold enriched[c]	P value[b]	Random at WG	Fold enriched[c]	P value[b]
HAEC	CTCF	61.3	4.4	13.9	$<10^{-3}$	7.5	8.2	$<10^{-3}$
	DHSs	48.9	10.1	4.8	$<10^{-3}$	7.5	6.5	$<10^{-3}$
	H3K4me3	23.3	6.8	3.4	$<10^{-3}$	5.6	4.2	$<10^{-3}$
	H3K27me3	46.0	34.1	1.4	$<10^{-3}$	27.8	1.7	$<10^{-3}$
	H3K36me3	20.0	14.1	1.4	$<10^{-3}$	9.7	2.1	$<10^{-3}$
	H3K9ac	17.7	6.0	2.9	$<10^{-3}$	5.3	3.3	$<10^{-3}$
H1	CTCF	46.9	9.0	5.2	<10-3	11.8	4.0	$<10^{-3}$
	DHSs	36.0	7.1	5.1	<10-3	18.1	2.0	$<10^{-3}$
	H3K4me3	13.2	1.8	7.4	<10-3	1.4	9.8	$<10^{-3}$
	H3K27me3	21.6	10.8	2.0	<10-3	9.8	2.2	$<10^{-3}$
	H3K36me3	11.0	16.1	0.7	1	15.2	0.7	1
	H3K9ac	28.9	11.8	2.4	<10-3	11.5	2.5	$<10^{-3}$
HPF	CTCF	56.9	7.9	7.2	$<10^{-3}$	9.6	5.9	$<10^{-3}$
	DHSs	71.0	27.7	2.6	$<10^{-3}$	19.2	3.7	$<10^{-3}$
	H3K4me3	30.8	8.1	3.8	$<10^{-3}$	13.0	2.4	$<10^{-3}$
	H3K27me3	53.1	34.2	1.6	$<10^{-3}$	29.0	1.8	$<10^{-3}$
	H3K36me3	22.6	16.5	1.4	$<10^{-3}$	24.9	0.9	0.85
	H3K9ac	13.2	2.5	5.2	$<10^{-3}$	6.2	2.1	$<10^{-3}$

[a] EIs of HAEC, H1 and HPF were compared with chromatin marks of HUVEC, H1 and normal lung fibroblasts, respectively (ref.10).
[b]P values were calculated by 1000 permutations.
[c]Enrichment is calculated as the ratio of observed to random.

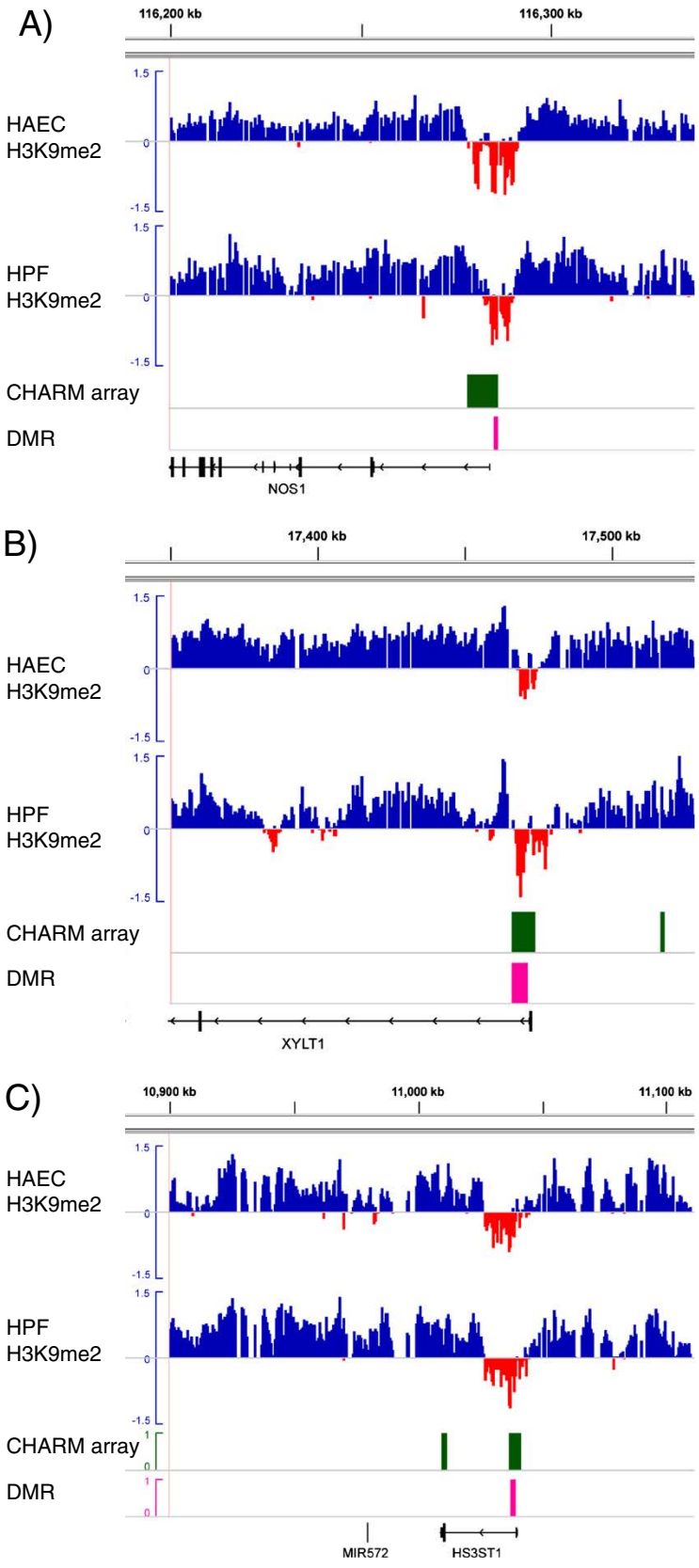


Figure 3 (See legend on next page.)

(See figure on previous page.)
Figure 3 EIs are enriched for differential methylation regions (DMRs). The H3K9me2 signals of HAEC and HPF are compared with regions of T-DMRs (pink bars). Regions of CHARM array are denoted by green bars. EIs (red dips) clearly overlap T-DMRs near the TSSs of *NOS1* (**A**), *XYLT1* (**B**) and *HS3HT1* (**C**).

bipolar disorder [26]. Furthermore, an EI is found on the 5' end of myocardin (*MYOCD*) gene in HA and HPF but HAEC (Figure 5C). Myocardin is a coactivator of serum response factor which specifically expressed in cardiac and smooth muscle cells [27], and promoter variation of this gene was proposed as a biomarker of cardiac hypertrophy [28]. These data suggest that EIs may be important in regulating specific cellular functions.

Discussion

In summary, by examining the genome-wide distribution of H3K9me2 in human PSCs and differentiated cells, we found a novel microstructure within heterochromatin domains of thousands of small euchromatin islands (EIs) located within large H3K9me2 blocks (LOCKs). EIs are strongly associated with open chromatin regions (DHSs), active chromatin marks (H3K4me3 and H3K9ac) and higher-order chromatin organizers (CTCF). Furthermore, EIs are highly enriched for DMRs associated with tissue specificity (T-DMRs), reprogramming (R-DMRs) and cancers (C-DMRs). This association is particularly strong for hypomethylated R-DMRs and hypermethylated C-DMRs. Genes associated with EIs are enriched for annotations of system development, cellular differentiation and cell adhesion. These results suggested that EIs may coordinate higher order chromatin and mediate co-regulation of DNA methylation in reprogramming and tumorigenesis. However, further experimental work is needed to address the functional relevance of EIs and their strong association with CTCF and DMRs.

In this study, we compared H3K9me2 profiles with publicly available epigenomic data generated from similar cell types. This strategy may lead to biased estimation of the enrichments of EIs with other epigenetic marks, because patterns of EIs may be different between the two samples. Comparison of exactly matched cell lines and cultures could assess the association between them more accurately.

Note that a previous literature used the term "euchromatic islands" in a completely different context, simply to describe chromatin regions with H3K4me3 and CpG islands, essentially describing promoter regions of active genes [4,29]. As that term was rarely used previously, and to convey a completely different meaning, we do not think there will be confusion with our newly defined (and differently spelled) "euchromatin islands" or EIs, namely H3K9me2 depleted regions/islands within an ocean of heterochromatin (LOCKs), enriched for regulatory elements such as enhancers (DHSs) and insulators (CTCF). Thus, EIs are novel units of the genomic "tool box" which may be important in epigenetic regulation as suggested by their strong association with DMRs (Table 3).

Higher-order organization of the genome remains a highly active area to be explored. Recent evidence indicates the presence of spatial compartments of active and repressive chromatin domains as general principles of genome organization in mammalian cells [30,31], and CTCF mediates intra- and inter-chromosomal interactions by tethering chromatin regions binding CTCF [32]. It would be interesting to explore the possibility that euchromatin islands act as "anchors" for the interactions among heterochromatin domains or between heterochromatic and enchromatic regions. Moreover, the relationships between EIs and heterochromatin formation, and the biophysical features of EIs are interesting questions for future investigation.

Although evidence provided in this and other studies have indicated that large heterochromatin domains

Table 3 Percentage of EIs overlap with DMRs

Cell line	Overlaps with DMRs (%)																				
	T-DMR			R-DMR						C-DMR						Whole genome C-DMR					
				Hyper			Hypo			Hyper			Hypo			Hyper			Hypo		
	%	En.	p	%	En.	p	%	En.	p	%	En.	p	%	En.	p	%	En.	p	%	En.	p
H1	53.4	2.8	$<10^{-3}$	19.4	3.1	$<10^{-3}$	5.8	3.1	0.02	11.7	3.8	$<10^{-3}$	4.9	1.4	0.24	5.6	5.2	$<10^{-3}$	0.9	1	0.31
ADA-38	39.1	2.1	$<10^{-3}$	17.4	3.2	$<10^{-3}$	2.2	1.2	0.33	6.5	2.2	0.05	6.5	2	0.08	4.1	4.1	$<10^{-3}$	0.4	0.5	0.84
HA	60.4	2.9	$<10^{-3}$	19.3	3.1	$<10^{-3}$	11	5.5	$<10^{-3}$	16.7	5.4	$<10^{-3}$	4.5	1.3	0.25	8.6	7.3	$<10^{-3}$	1	1.1	0.34
HPF	59.5	2.8	$<10^{-3}$	17.5	2.8	$<10^{-3}$	12.8	6.1	$<10^{-3}$	17.5	5.1	$<10^{-3}$	5.7	1.5	0.09	9.4	7.6	$<10^{-3}$	1.3	1.4	0.04
HAEC	62.3	2.9	$<10^{-3}$	14.3	2.2	$<10^{-3}$	11.7	5.6	$<10^{-3}$	15.7	4.5	$<10^{-3}$	4.6	1.2	0.29	7	5.3	$<10^{-3}$	1.1	1.1	0.27

Notes: En. = Enrichment. Statistical significances were tested by 1000 permutations.

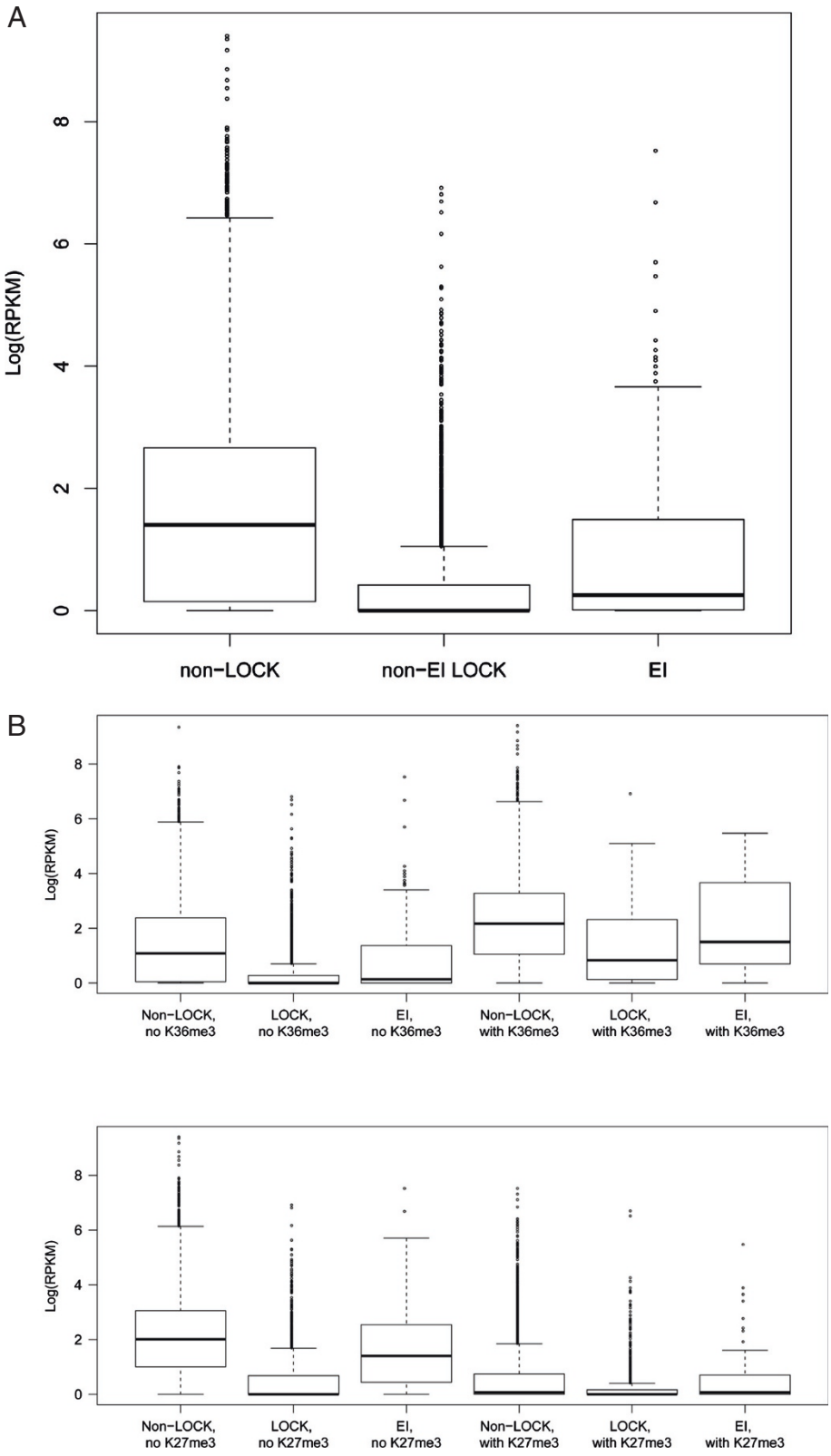


Figure 4 Expression of genes associated with EIs. We compared expression levels for genes with TSS at different regions. Expression values are RPKM (read per kb per million reads) for lung fibroblast IMR90 [7]. **A**) boxplot of expression level of genes with TSS 1) overlapping EIs; 2) overlapping LOCKs but not EIs; and 3) not overlapping EIs. **B**) Relationship between H3K36me3/H3K27me3 and expression of EI associated genes.

Table 4 Top 10 GO terms of genes associated with EIs

GO terms	% of genes	Fold en.	FDR
Biological processes			
nervous system development	12.6	2.2	9E-11
developmental process	25.5	1.6	4E-09
system development	20.5	1.7	4E-09
anatomical structure development	21.5	1.6	1E-08
multicellular organismal development	23.4	1.6	2E-08
cell adhesion	8.7	2.4	8E-08
biological adhesion	8.7	2.4	8E-08
cell differentiation	15.2	1.8	4E-07
cell development	8.0	2.4	4E-07
multicellular organismal process	31.1	1.4	4E-07
Cellular compartment			
plasma membrane part	21.1	1.9	4E-14
plasma membrane	30.3	1.6	1E-12
membrane part	44.2	1.3	1E-09
integral to plasma membrane	12.6	2.1	3E-09
intrinsic to plasma membrane	12.7	2.1	5E-09
membrane	47.0	1.3	2E-08
intrinsic to membrane	37.3	1.4	2E-07
integral to membrane	35.9	1.4	1E-06
synapse	5.3	3.0	2E-06
extracellular region	16.7	1.7	6E-06
Molecular function			
calcium ion binding	9.7	2.1	4E-06
gated channel activity	4.5	2.9	9E-05
molecular transducer activity	17.5	1.5	6E-04
signal transducer activity	17.5	1.5	6E-04
substrate specific channel activity	5.1	2.5	6E-04
ion channel activity	4.9	2.5	8E-04
channel activity	5.1	2.4	1E-03
passive transmembrane transporter activity	5.1	2.4	1E-03
cation channel activity	3.9	2.8	2E-03
ligand-gated channel activity	2.3	3.6	1E-02

are highly dynamic in stem cell differentiation and tumorigenesis [11-15], Lienert et al. indicated that genome coverage of H3K9me2 domains do not increase globally during neuronal differentiation of mouse ES cells [33]. First of all, lineage specificity of differentiated cells may explain the conflicts. As reported in our earlier work [11] the amount of LOCKs detected from brain and ES cells are comparable (9.8% vs. 4%), whereas the amount is very high in liver (45.6%). The Lienert study used in vitro differentiated neurons as differentiated cells which is more similar to brain. Furthermore, the inconsistence may be due to sensitivities of different statistical methods for finding large domains, heterogeneity of stem cells, and so on. Notably, extensive deduction of LOCKs during EMT suggested that quantitative differences of these large domains may be functionally important [13]. Nevertheless, further studies on homogenous stem cell populations may be helpful to address these debates. Whatever it holds, functionally investigations of these large domains should provide important insight toward how higher-order chromatin affects normal development and disease.

Conclusions

In conclusion, we have explored the microstructure of LOCKS and indentified thousands of euchromatin islands (EIs), which may be served as a finer layer of epigenomic architecture within large heterochromatin domains. The strong association of EIs with CTCF sites, DNAse hypersensitivies sites, and DMRs suggests that EIs play an important role in normal epigenomic architecture and its disruption in disease.

Methods

Cell culture

Human H1 ESCs and ADA-38 iPSCs were cultured as described [34]. Primary Human Pulmonary Fibroblasts (HPF), Human Aortic Endothelial Cells (HAEC) and Human Astrocytes (HA) were purchased from ScienCell Research Laboratories (San Diego, CA), and cultured as recommended by ScienCell.

ChIP-chip

ChIP-chip experiments were performed as described [11], using a commercial monoclonal antibody (Abcam, ab1220), which specifically recognizes H3K9me2 but not other modifications [35]. The passage numbers for cells used for ChIP analysis were P46 for H1, P59 for ADA-38 and P2 for primary cells from ScienCell. We first mapped whole genome distribution of H3K9me2 using "Mouse ChIP-chip 2.1M Economy Whole-Genome Tiling arrays (4 arrays per set) from NimbleGen", with 203 bp of median probe spacing. Then we repeated the microarray experiments on one of the Mouse ChIP-chip 2.1 M Whole-Genome Tiling sets, whose median probe spacing is 100 bp. The replicate array covers 10% of the genome, including part of chromosome 6 (111,920, 005-170,893,515), whole chromosome 7 and part of chromosome 8 (521–74,730,105). For the replicate experiments, cell cultures, ChIP sample preparation, labeling and hybridization were performed independently.

ChiP-chip data analysis

Data were first normalized by partial quantile normalization, then LOCKs were detected based on the smoothing values of normalized log2 ratios of data

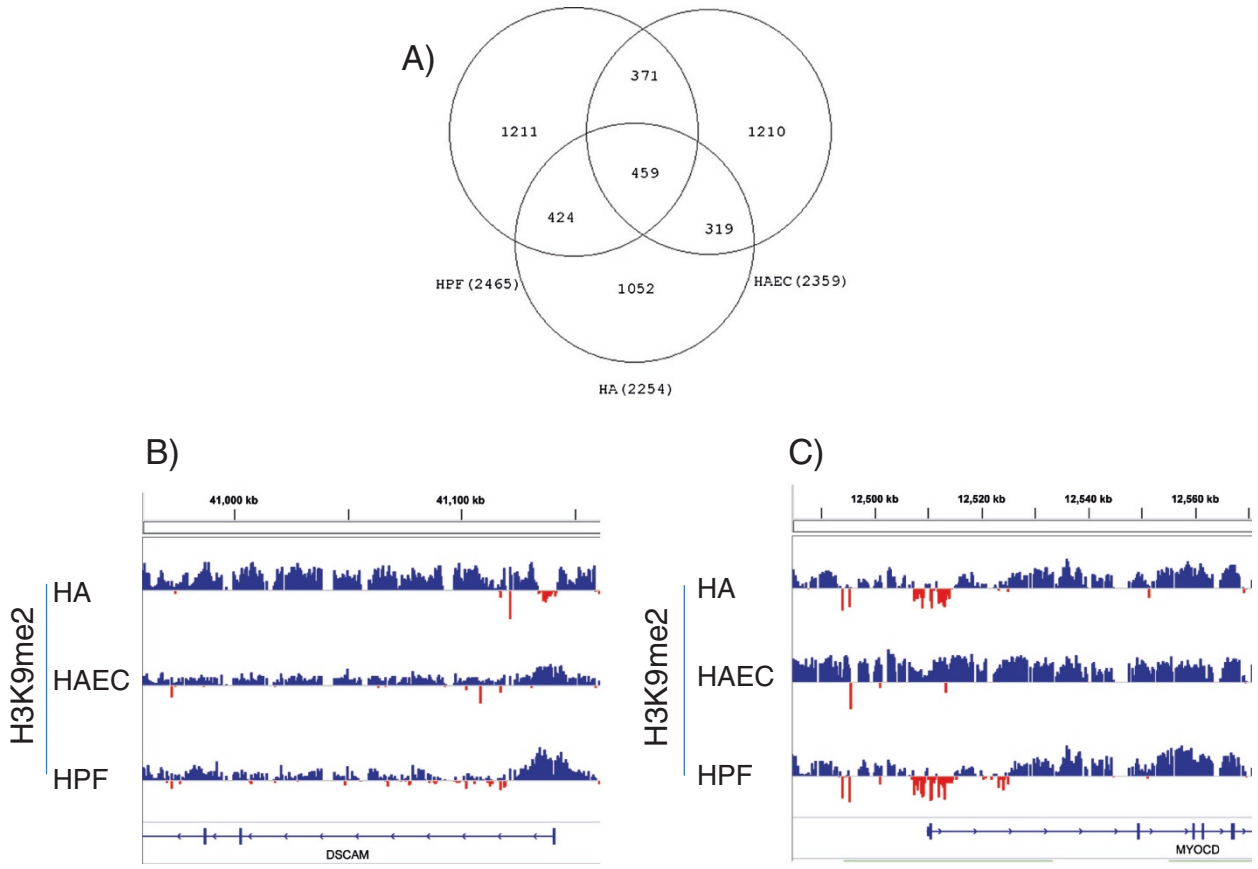


Figure 5 Cell type specific EIs. A) The Venn diagram shows overlaps of EIs among three differentiated cell lines (HA, HAEC and HPF). **B**-**C**) Shown are examples of EIs which differ among cell types. The annotation of tracks is the same as in Figures 2 and 3. An EI is found near the 5′ end of the myocardin gene (*MYOCD*) in HA and HPF but not in HAEC. An EI is found in the TSS of the Down syndrome cell adhesion molecule gene (*DSCAM*) specifically in HA but not in other two cell types (**C**).

between ChIP and input channels [11]. The euchromatin islands (EIs) are defined as short regions within LOCK body that have low H3K9me2 methylation levels. To detect such regions we designed the following smoothing based approach. The log2 ratios for probes within LOCKs were first smoothed using 5000 bp window. The relatively short smoothing window is used to capture the signal variations in small regions. Genomic regions with smoothed value less than 1% of all the smoothed values were defined as EIs. It is required that EIs are at least 1000 bps long and contain at least 10 probes. EIs less than 1000 bps apart will be merged into one. It is also required that the EIs are at least 20000 bps away from the LOCK boundaries. This is because the log2 ratios are smaller at LOCK boundaries. Such requirement prevents mistakenly taking LOCK boundaries as EIs. A flow diagram showing the algorithm for detecting EIs was provided in Additional file 8: Figure S6. Microarray data have been submitted to GEO database (accession numbers: GSE37335).

To compute the enrichment of EI overlapping other genomic features (CTCF, DHS, etc.), we first calculated the percent of EIs overlapping the feature. Then a set of genomic regions was randomly sampled. The number and lengths of the random regions match the EI list. The random regions were then compared with the feature to obtain a percentage of overlapping. Such process was repeated 1000 times. The percentages obtained from the process form the null distribution for percentage of overlapping. The p-values and enrichments were computed based on the null distribution. The p-values were then corrected for multiple testing using Bonferroni correction. Publicly available datasets used for analysis were listed in Additional file 9: Table S3.

Quantitative PCR (qPCR)

Experiments of qPCR were conducted as described [11]. Primer sequences are provided in Additional file 10: Table S4.

GO analysis

GO analysis was performed using DAVID tools as described [36], using the list of genes overlapping EIs of the three differentiated cell lines (HA, HAEC and HPF).

Additional files

Additional file 1: Table S1. Description: Genome coverage and average size of LOCKs in human PSCs and differentiated cells.

Additional file 2: Figure S1(A-D). Description: qRCR validation of H3K9me2 ChIP-chip data on 23 loci. Upper panels show log2 (ChIP/Input) ratios of microarrays and green bars denote regions selected for qPCR validation; lower panels present qPCR enrichments of ChIP over input in the selected regions.

Additional file 3: Figure S2. Description: LOCKs overlap partial methylation domains (PMDs). (A) One representative region (on chromosome 17) where LOCKs and PMDs overlap, green and orange bars show locations of LOCK (green) and PMD (orange), and hypomethylation blocks (purple), respectively; (B) H3K9me2 density in and out of PMDs. X-axis is the probe log2 ratios between ChIP and control samples. Y-axis is the the probability density.

Additional file 4: Table S2. Description: Coordinates of EIs (HG18).

Additional file 5: Figure S3. Description: Average H3K9me2 densities in EIs and their adjacent regions.

Additional file 6: Figure S4. Description: H3K9me2 ChIP-chip experiments in whole genome (WG) and replicate (rep) arrays.

Additional file 7: Figure S5. Description: Nucleosome density in EIs and adjacent regions. We compared common EIs of HA, HAEC and HPF with nucleosome maps of GM12878 (Supplementary Table S3), to overcome potential lineage specificity among those cell types.

Additional file 8: Figure S6. Description: Flow diagram of EI detection.

Additional file 9: Table S3. Description: Public datasets used for analysis.

Additional file 10: Table S4. Description: qPCR primer sequences.

Abbreviations

EI: Euchromatin island; DHS: DNase hypersensitive site; DMR: Differential methylation region; H3K9me2: H3 lysine 9 dimethylation; LOCK: Large organized chromatin k9-modification; ESC: Embryonic stem cell; PMD: Partial methylated domains; PSC: Pluripotent stem cell; HA: Human astrocytes; HAEC: Human aortic endothelial cell; HPF: Human pulmonary fibroblast; TSS: Transcription start site; CTCF: CCCTC-binding factor.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

BW and APF conceived the project and designed the study. BW performed cell culture of primary cells, ChIP and qPCR; HW performed data analysis; YL and GQD generated PSC lines; BE conducted microarray analysis; BW, HW and APF prepared the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by National Institute of Health (NIH) grant R37 CA54358.

Author details

[1]Center for Epigenetics and Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [2]Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA, USA. [3]Division of Pediatric Hematology/Oncology, Children's Hospital Boston and Howard Hughes Medical Institute, Boston, MA, USA. [4]Current address: Institutes of Biomedical Sciences, Shanghai Medical College, Fudan University, Shanghai, China.

Received: 28 September 2012 Accepted: 19 October 2012
Published: 26 October 2012

References

1. Feinberg AP: **Phenotypic plasticity and the epigenetics of human disease.** *Nature* 2007, **447**(7143):433–440.
2. Berger SL: **The complex language of chromatin regulation during transcription.** *Nature* 2007, **447**(7143):407–412.
3. Kouzarides T: **Chromatin modifications and their function.** *Cell* 2007, **128**(4):693–705.
4. Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B: **A high-resolution map of active promoters in the human genome.** *Nature* 2005, **436**(7052):876–880.
5. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K: **High-resolution profiling of histone methylations in the human genome.** *Cell* 2007, **129**(4):823–837.
6. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, *et al*: **Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome.** *Nat Genet* 2007, **39**(3):311–318.
7. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, *et al*: **Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.** *Nature* 2007, **448**(7153):553–560.
8. Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ, *et al*: **Combinatorial patterns of histone acetylations and methylations in the human genome.** *Nat Genet* 2008, **40**(7):897–903.
9. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, *et al*: **Histone modifications at human enhancers reflect global cell-type-specific gene expression.** *Nature* 2009, **459**(7243):108–112.
10. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, *et al*: **Mapping and analysis of chromatin state dynamics in nine human cell types.** *Nature* 2011, **473**(7345):43–49.
11. Wen B, Wu H, Shinkai Y, Irizarry RA, Feinberg AP: **Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells.** *Nat Genet* 2009, **41**(2):246–250.
12. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W, *et al*: **Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions.** *Nature* 2008, **453**(7197):948–951.
13. McDonald OG, Wu H, Timp W, Doi A, Feinberg AP: **Genome-scale epigenetic reprogramming during epithelial-to-mesenchymal transition.** *Nat Struct Mol Biol* 2011, **18**(8):867–874.
14. Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, Pelizzola M, Edsall LE, Kuan S, Luu Y, Klugman S, *et al*: **Distinct epigenomic landscapes of pluripotent and lineage-committed human cells.** *Cell Stem Cell* 2010, **6**(5):479–491.
15. Hon GC, Hawkins RD, Caballero OL, Lo C, Lister R, Pelizzola M, Valsesia A, Ye Z, Kuan S, Edsall LE, *et al*: **Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer.** *Genome Res* 2012, **22**(2):246–258.
16. Magklara A, Yen A, Colquitt BM, Clowney EJ, Allen W, Markenscoff-Papadimitriou E, Evans ZA, Kheradpour P, Mountoufaris G, Carey C, *et al*: **An epigenetic signature for monoallelic olfactory receptor expression.** *Cell* 2011, **145**(4):555–570.
17. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, *et al*: **The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores.** *Nat Genet* 2009, **41**(2):178–186.
18. Doi A, Park IH, Wen B, Murakami P, Aryee MJ, Irizarry R, Herb B, Ladd-Acosta C, Rho J, Loewer S, *et al*: **Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts.** *Nat Genet* 2009, **41**(12):1350–1353.
19. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, *et al*: **Human DNA methylomes at base**

resolution show widespread epigenomic differences. *Nature* 2009, **462**(7271):315–322.
20. Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, Wen B, Wu H, Liu Y, Diep D, *et al*: **Increased methylation variation in epigenetic domains across cancer types.** *Nat Genet* 2011, **43**(8):768–775.
21. Paz MF, Fraga MF, Avila S, Guo M, Pollan M, Herman JG, Esteller M: **A systematic profile of DNA methylation in human cancer cell lines.** *Cancer Res* 2003, **63**(5):1114–1121.
22. Carmona FJ, Villanueva A, Vidal A, Munoz C, Puertas S, Penin RM, Goma M, Lujambio A, Piulats JM, Mesia R, *et al*: **Epigenetic Disruption of Cadherin-11 in Human Cancer Metastasis.** *J Pathol* 2012.
23. ENCODE: **A user's guide to the encyclopedia of DNA elements (ENCODE).** *PLoS Biol* 2011, **9**(4):e1001046.
24. Hattori D, Demir E, Kim HW, Viragh E, Zipursky SL, Dickson BJ: **Dscam diversity is essential for neuronal wiring and self-recognition.** *Nature* 2007, **449**(7159):223–227.
25. Barlow GM, Chen XN, Shi ZY, Lyons GE, Kurnit DM, Celle L, Spinner NB, Zackai E, Pettenati MJ, Van Riper AJ, *et al*: **Down syndrome congenital heart disease: a narrowed region and a candidate gene.** *Genet Med* 2001, **3**(2):91–101.
26. Amano K, Yamada K, Iwayama Y, Detera-Wadleigh SD, Hattori E, Toyota T, Tokunaga K, Yoshikawa T, Yamakawa K: **Association study between the Down syndrome cell adhesion molecule (DSCAM) gene and bipolar disorder.** *Psychiatr Genet* 2008, **18**(1):1–10.
27. Wang D, Chang PS, Wang Z, Sutherland L, Richardson JA, Small E, Krieg PA, Olson EN: **Activation of cardiac gene expression by myocardin, a transcriptional cofactor for serum response factor.** *Cell* 2001, **105**(7):851–862.
28. Kontaraki JE, Parthenakis FI, Patrianakos AP, Karalis IK, Vardas PE: **Myocardin gene regulatory variants as surrogate markers of cardiac hypertrophy - study in a genetically homogeneous population.** *Clin Genet* 2008, **73**(1):71–78.
29. Guccione E, Martinato F, Finocchiaro G, Luzi L, Tizzoni L, Dall' Olio V, Zardo G, Nervi C, Bernard L, Amati B: **Myc-binding-site recognition in the human genome is determined by chromatin context.** *Nat Cell Biol* 2006, **8**(7):764–770.
30. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, *et al*: **Comprehensive mapping of long-range interactions reveals folding principles of the human genome.** *Science* 2009, **326**(5950):289–293.
31. Kalhor R, Tjong H, Jayathilaka N, Alber F, Chen L: **Genome architectures revealed by tethered chromosome conformation capture and population-based modeling.** *Nat Biotechnol* 2012, **30**(1):90–98.
32. Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, Lee CW, Ye C, Ping JL, Mulawadi F, *et al*: **CTCF-mediated functional chromatin interactome in pluripotent cells.** *Nat Genet* 2011, **43**(7):630–638.
33. Lienert F, Mohn F, Tiwari VK, Baubec T, Roloff TC, Gaidatzis D, Stadler MB, Schubeler D: **Genomic prevalence of heterochromatic H3K9me2 and transcription do not discriminate pluripotent from terminally differentiated cells.** *PLoS Genet* 2011, **7**(6):e1002090.
34. Park IH, Zhao R, West JA, Yabuuchi A, Huo H, Ince TA, Lerou PH, Lensch MW, Daley GQ: **Reprogramming of human somatic cells to pluripotency with defined factors.** *Nature* 2008, **451**(7175):141–146.
35. Egelhofer TA, Minoda A, Klugman S, Lee K, Kolasinska-Zwierz P, Alekseyenko AA, Cheung MS, Day DS, Gadel S, Gorchakov AA, *et al*: **An assessment of histone-modification antibody quality.** *Nat Struct Mol Biol* 2011, **18**(1):91–93.
36. da Huang W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.** *Nat Protoc* 2009, **4**(1):44–57.

doi:10.1186/1471-2164-13-566
Cite this article as: Wen *et al.*: Euchromatin islands in large heterochromatin domains are enriched for CTCF binding and differentially DNA-methylated regions. *BMC Genomics* 2012 **13**:566.

Genome Regulation by Polycomb and Trithorax: 70 Years and Counting

Bernd Schuettengruber,[1,*] Henri-Marc Bourbon,[2] Luciano Di Croce,[3,4,5,*] and Giacomo Cavalli[1,*]
[1]Institute of Human Genetics, UMR9002 CNRS, University of Montpellier, 141 Rue de la Cardonille, 34396 Montpellier Cedex 5, France
[2]Centre de Biologie du Développement, UMR5547 CNRS/UPS, Centre de Biologie Intégrative, Université de Toulouse, 31000 Toulouse, France
[3]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, 08003 Barcelona, Spain
[4]Universitat Pompeu Fabra (UPF), 08002 Barcelona, Spain
[5]ICREA, Pg. Lluis Companys 23, 08010 Barcelona, Spain
*Correspondence: bernd.schuttengruber@igh.cnrs.fr (B.S.), luciano.dicroce@crg.eu (L.D.C.), giacomo.cavalli@igh.cnrs.fr (G.C.)
http://dx.doi.org/10.1016/j.cell.2017.08.002

Polycomb (PcG) and Trithorax (TrxG) group proteins are evolutionarily conserved chromatin-modifying factors originally identified as part of an epigenetic cellular memory system that maintains repressed or active gene expression states. Recently, they have been shown to globally control a plethora of cellular processes. This functional diversity is achieved by their ability to regulate chromatin at multiple levels, ranging from modifying local chromatin structure to orchestrating the three-dimensional organization of the genome. Understanding this system is a fascinating challenge of critical relevance for biology and medicine, since misexpression or mutation of multiple PcG components, as well as of TrxG members of the COMPASS family and of the SWI/SNF complex, is implicated in cancer and other diseases.

Epigenetic regulation of gene expression is one of the key mechanisms regulating cell-fate choices and cell identity during development. One of the most prominent and enigmatic epigenetic regulatory systems involves the evolutionarily conserved Polycomb group (PcG) and Trithorax group (TrxG) components, acting antagonistically to orchestrate the expression of key genes in cell differentiation and developmental processes. Seventy years ago, the first PcG gene, *Polycomb (PC)*, was discovered in *Drosophila melanogaster* by Pamela Lewis (Lewis, 1947). Later, Ed Lewis determined that *Polycomb* mutations transform anterior embryonic segments into more posterior ones due to the ectopic expression of *Homeotic* (*Hox*) genes (Lewis, 1978). Subsequent genetic screens identified other genes whose mutations resulted in phenotypes similar to the loss of function of *Pc* mutations, leading to the definition of the Polycomb group (PcG) proteins. A few years after the discovery of PcG, the first TrxG gene, *Trithorax,* was isolated as a regulator of *Hox* gene expression. Its mutation causes embryonic segments to be transformed into more anterior ones by antagonizing PcG proteins (Ingham, 1983; Ingham, 1985b; Struhl and Akam, 1985). After additional genes were discovered that had mutant phenotypes counteracting PcG mutations and consistent with the loss of *Hox* function, the overall group was defined as the Trithorax group (TrxG) of proteins (Kennison and Tamkun, 1988). The observation that PcG and TrxG factors are required to maintain *Hox* gene expression after their initial transcriptional regulators disappear from the embryo gave rise to the hypothesis that PcG and TrxG proteins act as a cellular memory system (Ingham, 1985a). Soon after these early discoveries, some PcG and TrxG members were linked to proliferation, senescence (Jacobs et al., 1999a), and cancer (Djabali et al., 1992; van Lohuizen et al., 1991). Extensive research over the past decade has revealed that PcG and TrxG proteins are much more than epigenetic gatekeepers of *Hox* gene expression. They regulate a plethora of cellular processes, including X chromosome inactivation, genomic imprinting, cell cycle control, stem cell biology, and cancer. This functional diversity is achieved by a variety of PcG and TrxG complexes that are assembled in a developmental-stage- and cell-specific manner to modify chromatin at their target genes via histone-modifying or chromatin-remodeling activities. Although PcG and TrxG target genes are conserved during evolution (Schuettengruber et al., 2007), the way these complexes are targeted to chromatin can significantly differ between species. In *Drosophila*, Polycomb response elements (PREs) and Trithorax response elements (TREs) target PcG and TrxG complexes to chromatin, thus driving the epigenetic inheritance of silent or active chromatin states throughout development. In mammals, hypomethylated CpG islands (CGIs) represent PRE-like sequences that can recruit PcG and TrxG complexes. Research in the last decade has shown that PcG and TrxG complexes can regulate their target genes at multiple levels, from modifying local chromatin structure to regulating higher-order chromatin organization and global genome architecture. PcG and TrxG proteins are currently under intense study and, during the last decade, major progress was made toward understanding (1) their diversity and biochemical interactions, (2) their recruitment to chromatin, (3) the mechanisms by which they regulate genome function, (4) their role and mechanisms in conveying inheritance of chromatin states, and (5) their implications in normal and pathogenic cellular processes, particularly embryonic stem cell (ESC) biology and cancer (Figure S1).

PcG Complexes

Historically, based on biochemical purification experiments from *Drosophila melanogaster*, the PcG machinery has been

Table 1. Overview of PcG and TrxG Complexes and Their Core Subunits, Enzymatic Functions, and Functional/Structural Domains

PcG Complex Components		Characteristic Domain	(Epigenetic) Function
Mammals	Flies		
Core PRC1 Complex (common to all PRC1 Complexes)			
RING1A/B	dRing/Sce	RING finger	H2AK119 ubiquitylation
PCGF1–PCGF6	Psc/Suz(2)	RING finger, UBL (RAWUL)	H2AK119 ubiquitylation, oligomerization
Canonical PRC1			
CBX2, CBX4, CBX6–CBX8	Pc	chromodomain	H3K27me3 binding
PHC1–PHC3	Ph	sterile alpha motif (SAM)	oligomerization/protein-protein interaction
SCMH1/2	Scm	SAM	oligomerization/protein-protein interaction
Non-canonical PRC1			
RYBP/YAF2	dRybp	zinc finger	DNA binding
KDM2B	dKdm2	JmjC, CxxC	H3K36 demethylase, DNA binding
DCAF7	Wap/n.i.	WD40	scaffold factors
WDR5	Wds/n.i.	WD40	scaffold factors
Core PRC2 Complex			
EZH1/2	E(z)	SET, SANT	H3K27 methyltransferase, histone binding
SUZ12	Suz(12)	zinc finger	RNA/DNA binding
EED164	Esc/Escl	WD40	H3K27me binding
RBBP4/7	Nurf55/Caf1	WD40	H3K36me3 binding
PRC2 Accessory Proteins			
PCL1–PCL3	Pcl	Tudor, PHD finger	H3K36me3 binding
JARID2	Jarid2	zinc finger, ARID	H2Aub binding, RNA binding
AEBP2	Jing	zinc finger	DNA binding, H2Aub binding
EPOP/C17orf96	n.i.	N/A	modulating enzymatic activity
LCOR/C10orf12	n.i.	N/A	N/A
Core PR-DUB			
BAP1	Calypso	ubiquitin C-terminal hydrolase (UCH) N-terminal catalytic	ubiquitin C-terminal hydrolase
ASXL1/2	Asx	N/A	chromatin binding
PR-DUB Accessory Proteins			
FOXK1/2	FoxK/n.i.	Forkhead box	DNA binding
OGT1	Sxc/n.i.	N/A	O-GlcNAcylation
KDM1B	dLsd1/n.i.	amine oxidase	histone demethylation
MBD5/6	Sba/n.i.	methyl binding	DNA binding

TrxG Complex Components		Protein Domain	(Epigenetic) Function
Mammals	Flies		
Core COMPASS Components common to all COMPASS (-like) Complexes			
WDR5	Wds	WD40	histone binding
ASH2L	Ash2	zinc finger	DNA binding
RBBP5	Rbbp5	WD40	histone binding
DPY30	Dpy30	N/A	N/A
SET1/COMPASS			
SET1A/B	dSet1	SET	H3K4 methyltransferase
HCF1	Hcf1	Kelch	N/A

(*Continued on next page*)

Table 1. *Continued*

TrxG Complex Components		Protein Domain	(Epigenetic) Function
WDR82	Wdr82	WD40	histone binding
CFP1	Cfp1	CxxC	DNA binding
MLL1/2 COMPASS-like			
MLL1/2	Trx	SET d	H3K4 methyltransferase
HCF1	Hcf1	Kelch	N/A
MENIN	Menin	N/A	N/A
MLL3/4 COMPASS-like			
MLL3/4	Trr	SET	H3K4 methyltransferase
NCOA6	Ncoa6	N/A	N/A
PA1	Pa1	N/A	N/A
UTX	Utx	JmjC	H3K27 demethylase
PTIP	Ptip	BRCT	N/A
ASH1			
ASH1L	Ash1	SET, bromodomain	H3K36 methyltransferase
CBP	dCbp	HAT, bromodomain	H3K27 acetyltransferase
SWI/SNF (BAF and PBAF) Complex			
BRM/BRG1	Brm	helicase, bromodomain	ATPase-dependent chromatin remodeling
BAF250A/B	Osa	ARID	possible DNA binding
BAF155/170	Mor	SWIRM, SANT, chromodomain	possible DNA and histone binding
BAF47	Snr1	winged helix	possible DNA binding
BAF45A-D	Sayp	PHD-finger	possible DNA binding
BAF53A/B	Bap55	actin-like	–
BAF180/BAF200	Bap180	polybromodomain	histone binding
BAF60A-C	Bap60	Swi-B	–
BAF57	Bap111	HMG	possible DNA binding
beta-ACTIN	Actin5C	N/A	N/A
BCL7A-C	Bcl7-like/n.i.	N/A	N/A
BRD7/9	CG7154/n.i.	N/A	N/A

Note that only TrxG complexes with a major function in counteracting PcG function are shown. n.i. indicates that homologous proteins exist, but they have not been purified as part of PcG/TrxG complexes.

subdivided into two main complexes: Polycomb repressive complex 1 (PRC1) and PRC2 (Table 1). This view has been greatly clarified and expanded in the last decade revealing that the PcG system is much more diverse in mammals than in *Drosophila*, especially for PRC1, which is subdivided into canonical (cPRC1) and non-canonical complexes (ncPRC1) (for a review, see Blackledge et al., 2015). All PRC1 complexes share a protein core that is conserved in the five major animal lineages and in plants, but not in fungi (Figure 1A). In mammals, this core is composed of RING1 proteins (RING1A or RING1B), which have E3 ubiquitin ligase activity mediating ubiquitylation of histone H2A on lysine 119 (H2AK119ub), and one of the six Polycomb group ring-finger domain proteins (PCGF1–PCGF6). cPRC1 complexes are assembled around PCGF2/4 and are specified by the presence of one chromobox protein (CBX2, CBX4, and CBX6–CBX8) that binds H3K27me3 and a Polyhomeotic (Ph) homologous protein (PHC1–PHC3), which contains a sterile alpha motif (SAM) domain essential for PcG-mediated repression (Robinson et al., 2012). In contrast, ncPRC1 possesses the zinc-finger domain and YY1-binding protein (RYBP) or its paralog, YAF2 (Wilkinson et al., 2010), which associates with PCGF1, PCGF3/5, or PCGF6 to form ncPRC1.1, ncPRC1.3/PRC1.5, or ncPRC1.6, respectively (Figure 1B). ncPRC1.1, which contains the H3K36-specific histone demethylase KDM2B, is the only non-canonical complex also purified in *Drosophila* (dRAF) (Lagarou et al., 2008). The enzymatic activity of each complex is defined by the choice of PCGF protein, which is stimulated by the presence of RYBP (Rose et al., 2016). The WD40 domain proteins DCAF7 and WDR5 act as central scaffold proteins of the ncPRC1.3/1.5/1.6 complexes (Hauri et al., 2016), whereas an additional set of subunits specific to each complex can modulate their DNA-binding affinities or regulatory functions. Interestingly, RYBP/YAF2 and KDM2B, but not the CBX, PHC, or SCMH proteins, are present in choanoflagellates, the closest unicellular relatives of animals, suggesting that ncPRC1 appeared earlier during evolution than cPRC1 (Figure 1A).

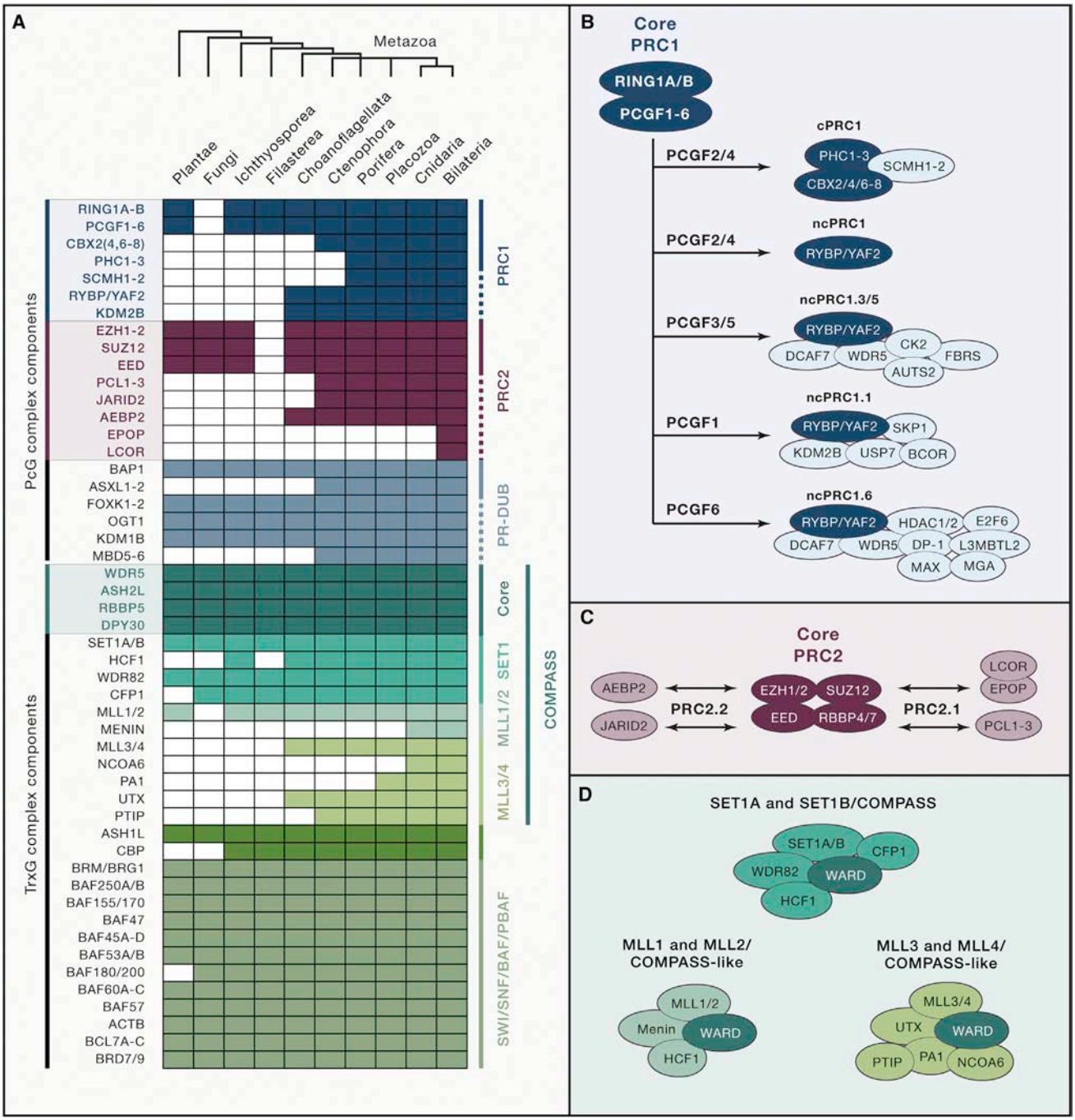


Figure 1. Composition and Evolution of PcG and TrxG Complexes

(A) Phylogenetic distribution of PcG and TrxG complexes. A broad spectrum of eukaryotes has been investigated, with an emphasis on holozoans, among which are the five major metazoans lineages (see Supplemental Information for experimental details on homolog searches). The presence of predicted proteins orthologous to bilaterian PcG/TrxG complex subunits is indicated by filled boxes. Accessory proteins (or non-canonical PRC1 components) are represented by lateral dashed lines. Note that the broadly conserved DCAF7 (ncPRC1), RBBP4/7 (core PRC2), and RBBP5 (COMPASS) WD proteins are not included in this analysis.

(B) Composition of PRC1 complexes. The core of PRC1 can associate with different PCGF proteins (PCGF1–PCGF6) defining canonical PRC1 complexes (cPRC1) and non-canonical (ncPRC1) PRC1 complexes. cPRC1 contains PHC and CBX proteins and the more loosely associated SCMH proteins. ncPRC1 complexes are defined by the presence of RYBP/YAF2 that associate with additional proteins.

(C) Composition of PRC2 complexes. The core of the PRC2 complex associates with different accessory proteins to define the PRC2.1 and PRC2.2 complexes, respectively.

(D) Composition of COMPASS(-like) complexes. A conserved core (WARD) composed of the proteins WDR5, ASH2, RBBP5, and DPY30 associates with the H3K4 methyltransferases SET1A/B, MLL1/2, and MLL3/4 as well as with additional regulatory proteins to produce COMPASS(-like) complexes.

The functional core of the mammalian PRC2 essential to mediate histone methyltransferase activity (for a review, see Blackledge et al., 2015) is composed of one of the SET-domain-containing histone methyltransferases enhancer of zeste (EZH2 or EZH1), embryonic ectoderm development (EED), suppressor of zeste (SUZ12), and the CAF1 histone-binding proteins RBBP4 and RBBP7 (Table 1). This PRC2 core is broadly conserved in the five major animal lineages as well as in plants and some fungi, including *N. crassa*, and the yeast *Cryptococcus neoformans* (Dumesic et al., 2015; Jamieson et al., 2013) (Figure 1A). PRC2 compositional diversity is achieved through association with non-stoichiometric subunits (hereafter called accessory proteins) that can modulate its enzymatic activities or chromatin target sites (Ciferri et al., 2012; Li et al., 2010; Pasini et al., 2010; Shen et al., 2009). Importantly, the emergence of accessory subunits of PRC2 follows the increasing cell and tissue complexity of metazoans (Figure 1A), suggesting that the regulation of PRC2 activity by these different regulatory subunits might be key for specifying cell identity during development and differentiation. A systematic proteomics approach in human cells revealed two fundamental alternative assemblies linked to the PRC2 core complex (Hauri et al., 2016) (Figure 1C). PRC2.1 is defined by its mutually exclusive binding of one of the three Polycomb-like homologs (PCLs) PHF1, PHF19, or MTF2. PHD finger protein 1 (PHF1) and its *Drosophila* homolog, Pcl, stimulate efficient trimethylation activity of EZH2 toward the H3K27me2 substrate (Nekrasov et al., 2007; Sarma et al., 2008). PRC2.2 is defined by the presence of the zinc-finger proteins AEBP2 and JARID2, which enhance enzymatic activity and regulate chromatin binding of the PRC2 complex. Finally, the two mammalian-specific proteins, C10ORF12/LCOR and C17ORF96/EPOP, also co-purify with PRC2 complexes (Beringer et al., 2016; Kloet et al., 2016).

Genetic screens in *Drosophila* further identified PcG proteins that are not components of PRC1/2 but constitute additional PcG complexes (Table 1). Pho-repressive complex (PhoRC) is composed of the zinc-finger protein Pleiohomeotic (Pho) and dSfmbt (Scm-related gene containing four MBT domains), which can bind to H3K9me1 and H3K20me2 (Klymenko et al., 2006). No enzymatic activity is associated with PhoRC. The Polycomb repressive deubiquitinase complex (PR-DUB) was initially identified in *Drosophila* and contains the ubiquitin C-terminal hydrolase Calypso (Bap1) and additional sex combs (Asx) (Scheuermann et al., 2010). It possesses H2A-specific deubiquitinase activity that is paradoxically required for PcG-mediated repression, suggesting that a tight balance between H2A ubiquitination and deubiquitination is essential for PcG-dependent silencing. Mammalian homologs of PR-DUB can be divided into PR-DUB1 and PR-DUB2, depending on whether the BAP1 partner is ASXL1 or ASXL2, respectively (Hauri et al., 2016). Both complexes share a similar set of accessory proteins, including FOXK1/2, the histone demethylase KDM1B, methyl-binding domain proteins MBD5/6, and the O-GlcNAc transferase OGT1. Notably, the *Drosophila* homolog Ogt/Sxc was previously identified as a PcG protein (Gambetta et al., 2009) but had been not co-purified with *Drosophila* PR-DUB. In summary, considerable progress has been made in determining the composition of PcG complexes in various species. However, much remains to be done in this field. In particular, the functional importance of many of the identified subunits—purified from diverse and sometimes highly aneuploid cell lines—in a developing organism has yet to be determined.

TrxG (COMPASS, SWI/SNF) Complexes

The complexity of TrxG complexes matches that of their PcG counterparts. TrxG is a heterogeneous group that plays a widespread role in transcriptional activation. Here, we will focus on TrxG complexes (the SWI/SNF complex and COMPASS family) that were shown to be key players in specifically counteracting PcG-mediated repression. Genetic studies in *Drosophila* first identified antagonistic links between PcG genes and the SWI/SNF (switch/sucrose non-fermentable) complexes (Kennison and Tamkun, 1988), originally purified and characterized in yeast. Subsequently, homologous complexes having ATP-dependent chromatin-remodeling activities were identified in flies and mammals (Dingwall et al., 1995; Khavari et al., 1993). Elegant biochemical, genetic, and molecular studies led to the discovery that mammalian homologs of SWI/SNF, BRM, and BRG1 form two distinct complexes, BAF and PBAF, each of which contains up to 15 additional subunits that are well conserved during evolution (Table 1; Figure 1A) and which regulate chromatin structure of a large number of genes implicated in the cell cycle, signaling, and proliferation and are often dysregulated in cancer (reviewed in Hodges et al., 2016; Kadoch and Crabtree, 2015).

A second group of TrxG complexes is associated with histone-modifying activities, including histone acetylation and methylation, thereby counteracting the action of the repressive marks deposited by PcG complexes. The histone methyltransferase complex SET1/COMPASS (Piunti and Shilatifard, 2016) was first isolated in yeast and is well conserved across all pre-metazoan and major animal lineages (Figure 1A). However, during evolution, this complex diverged to give rise to COMPASS-like complexes containing different SET-domain methyltransferases with unique functionalities (reviewed in Piunti and Shilatifard, 2016). The well-conserved protein core of all COMPASS complexes is required for their function and includes the proteins WDR5, ASH2, RBBP5, and DPY30 (sometimes abbreviated as WARD) (Figure 1D). SET1/COMPASS additionally contains HCF1, WDR82, and the DNA-binding protein CXXC1 (CFP1) and mediates bulk trimethylation of K4 on histone H3 (H3K4me3) (Figure 2B). MLL1/MLL2 COMPASS-like additionally contains the proteins MENIN, HCF1, and either MLL1 or MLL2 in a mutually exclusive manner (Hu et al., 2013). MLL2 is responsible for H3K4 trimethylation at bivalent promoters in ESCs (Figure 2B) (Denissov et al., 2014), whereas MLL1 is required for H3K4 trimethylation at only a small subset of genes, including *HOX* genes, emphasizing the non-redundant function of these two histone methyltransferases. The *Drosophila* homolog of MLL1, Trx, is dispensable for genome-wide H3K4me3 levels but has recently been found to dimethylate H3K4 at PREs to maintain the developmental expression pattern of its target genes (Rickels et al., 2016). In addition to the WARD subunits, MLL3/4 COMPASS-like includes the proteins NCOA6 and PA1 and the histone demethylase KDM6/UTX, which can remove the PcG-mediated repressive H3K27me3 mark (Figure 2A). Like the *Drosophila* homolog Trr, MLL3/4 are the major methyltransferases mediating monomethylation of H3K4 at enhancers (Figure 2B) (reviewed in Piunti

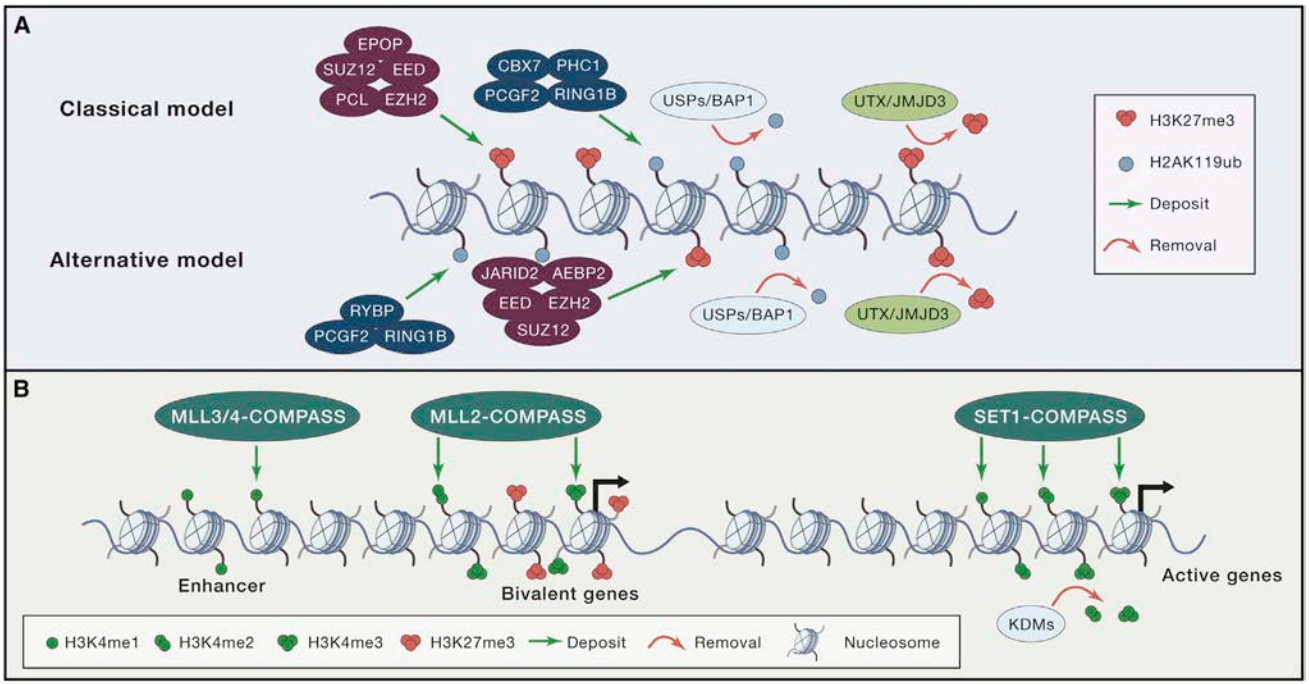


Figure 2. Deposition and Removal of PcG and TrxG-Mediated Histone Marks Contributes to Their Balanced Action on Chromatin

(A) According to the "classical model," targeting of the PRC2 complex leads to the methylation of H3K27. This modification is recognized by the Cbx subunit of the canonical PRC1 complex, which in turn catalyzes monoubiquitination of histone H2A at lysine 119 (H2AK119ub). An alternative recruitment mechanism posits that ncPRC1 is targeted to unmethylated CpG islands through the FBXL10/KDM2B subunit. ncPRC1 seems to be responsible for the majority of the H2AK119ub modifications present at PcG-target sites, which might facilitate the recruitment of PRC2. H3K27 methyl marks can be erased by histone demethylases, such as UTX/KDM6A and JMJD3/KDM6B, while several deuquibitinases (including BAP1, USP16 and USP21) can remove the monoubiquitin moiety from histone H2A. (B) Set1A/B COMPASS complexes catalyze mono-, di-, and trimethylation on H3K4 at active promoters. The activity of the partially redundant COMPASS complexes containing MLL3/KMT2C and MLL4/KMT2D leads to the deposition of H3K4me1 at enhancers, facilitating the recruitment of other activators such as CBP/p300. The deposition of methyl marks on H3K4 at bivalent regions is performed by MLL2/COMPASS. Multiple histone demethylases, including members of the KDM1/LSD, KDM2A/FBXL11, and KDM5/JARID families, are implicated in the removal of methyl groups on H3K4.

and Shilatifard, 2016). Intriguingly, the choanoflagellate *S. rosetta* (but not *M. brevicolis*) possesses MLL3/4 and UTX proteins (Figure 1A). This is of special interest, as *S. rosetta* can exist either as single cells or as multicellular rosette-shaped colonies, and it possesses homologs of several genes required for animal cell signaling and adhesion (Dayel et al., 2011). Further, distal enhancer elements are only present in metazoans that possess MLL3/4; in the filasterean *Capsaspora*, in which MLL3/4 are absent, regulatory regions are close to transcription start sites (Sebé-Pedrós et al., 2016). The correlation of the presence of MLL3/4 in metazoans with a corresponding increase in *cis*-regulatory genome complexity suggests that the emergence of enhancer-specific MLL3/4 activity was a key event in cell differentiation processes during pre-metazoan evolution. An additional histone demethylase (KDM) specific for H3K4me3 (little imaginal discs [Lid]) has been classified as a TrxG protein in *Drosophila* (Eissenberg et al., 2007; Lee et al., 2007b; Secombe et al., 2007), and its mammalian homolog, JARID1d, associates with PCL proteins regulating transcriptional initiation through H3K4 demethylation (Figure 2B) (Lee et al., 2007a).

Recruitment of PcG Components

Although PcG proteins and their target genes are highly conserved during evolution (Schuettengruber et al., 2007), the sequence requirements underlying PcG recruitment can diverge significantly. Early work in *Drosophila* identified PREs as the DNA regulatory elements that recruit PcG factors to chromatin and mediate epigenetic inheritance of silent chromatin states throughout development (Fauvarque and Dura, 1993; Müller and Bienz, 1991; Simon et al., 1993). *Drosophila* PREs frequently contain DNA motifs for sequence-specific DNA-binding proteins like Pho, Phol, Trl/GAF, Spps, Dsp1, and Cg. While these transcription factors (TFs) play important roles in PcG recruitment in *Drosophila*, none of them are sufficient to recruit PcG complexes on their own (reviewed in Entrevan et al., 2016). Initially, a hierarchical recruitment model was proposed in which TFs recruit PRC2, which subsequently recruits PRC1 via the interaction of Polycomb (Pc/CBX) with the PRC2-deposited H3K27me3 mark (Figure 2A) (Wang et al., 2004). However, genome-wide identification of PREs (Schuettengruber et al., 2009; Schwartz et al., 2006; Tolhuis et al., 2006) revealed that they are usually nucleosome-depleted regions and that PRC1-bound regions devoid of H3K27me3 exist (Loubière et al., 2016; Schaaf et al., 2013; Schwartz et al., 2006). This context-dependent relationship among TFs, PRC1, and PRC2 is cooperative rather than hierarchical (Figure 3A). In particular, PRC1 can stabilize binding of PhoRC by a positive feedback loop (Kahn et al., 2014; Schuettengruber et al., 2014) and supports PRC2 binding at a large

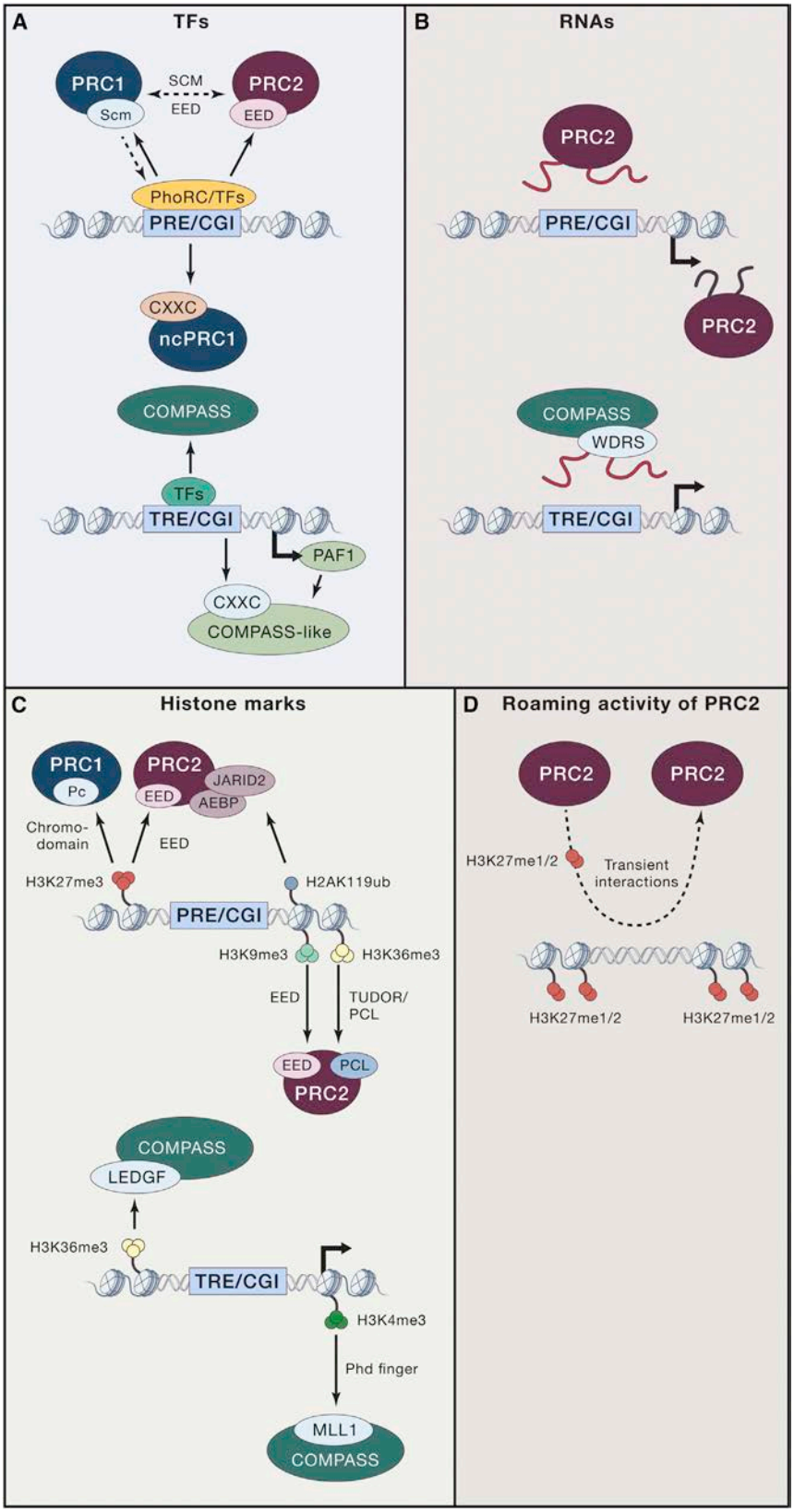


Figure 3. Analogous Factors Regulate the Recruitment and Stable Binding of PcG and TrxG Complexes

(A) Transcription factor (TF)-mediated recruitment. Cooperative interactions (indicated by the dashed arrows) between TFs and PRC1 and PRC2 complexes contribute to stable PcG binding. The PRC1 accessory protein Scm is a key protein in mediating PRC1 interaction and PhoRC. Similarly, SCMH1/2 and EED might mediate interactions between PRC1 and PRC2 complexes, thereby facilitating stable binding to chromatin. KDM2B, which is part of ncPRC1.1, binds to CGIs through its zinc-finger CxxC domain. Similarly, COMPASS(-like) complexes can bind sequence specifically to their target sites via the CxxC domain of MLL1/2 or can be recruited by TFs such as CFP1 or FOXA1 or via their interaction with PAF1.

(B) RNA-mediated recruitment. ncRNAs (like XIST, HOTAIR, or KCNQ1, shown in red), or short RNAs transcribed from repressed genes that form stem-loop structures (in black), can induce recruitment of PcG complexes. In a similar manner, HOTTIP ncRNA interacts with the WDR5 subunit of COMPASS to recruit it to the *HOXA* locus.

(C) Chromatin-mediated recruitment. H3K27me3 can increase the affinity of PRC1 to chromatin via interaction with the chromodomain of Pc/CBX. The EED subunit of PRC2 can bind to methylated H3K9 and H3K27, thereby stabilizing its binding to chromatin. H2AK119ub can induce PRC2 recruitment mediated by AEBP2 and JARID2. H3K36me3 can stabilize binding of PRC2 via the TUDOR domain of PCL proteins. TrxG complexes containing MLL1 can be recruited to their target sites via their interaction of the PHD finger with H3K4me3. H3K36me2, deposited by ASH1L, can promote binding of COMPASS complexes via the epigenetic reader protein LEDGF.

(D) In the absence of PRE-like sequences and PcG-recruiting components, PRC2 associates only transiently to mediate mono- and dimethylation of H3K27 (H3K27me1/me2).

subset of PREs. Moreover, both complexes can be recruited independently to chromatin in certain genome contexts (Kahn et al., 2016). PRC2-independent recruitment of PRC1 might be mediated by the PRC1 accessory component Scm, which acts as a molecular bridge connecting PhoRC and PRC1 via the oligomerization ability of the SAM domain (Frey et al., 2016).

In mammals, genome-wide mapping studies of PcG proteins have revealed a strong correlation with hypomethylated CGIs (Boyer et al., 2006; Tanay et al., 2007). This, together with the observation that an artificial GC-rich element devoid of activating TF-binding sites can ectopically recruit PcG proteins, gave rise to the idea that CGIs represent mammalian PREs (Farcas et al., 2012; Mendenhall et al., 2010; Riising et al., 2014). KDM2B binds to CGIs through its zinc-finger CxxC domain, contributing to PRC1.1 recruitment. However, most PcG-binding sites are not affected by loss of KDM2B, suggesting that PcG complexes can bind to CGIs in alternative ways. In addition, although KDM2B is bound to virtually all CGIs, it induces recruitment of PRCs at only ~30% of mammalian CGIs in ESCs (Bernstein et al., 2006a; Ku et al., 2008). These PcG-bound CGIs correspond to repressed promoters, and inhibition of transcription induces recruitment of PcG proteins to newly silenced CGIs (Riising et al., 2014). This supports a "chromatin sampling" model (Klose et al., 2013), which proposes that PcG proteins weakly interact with all potential binding sites but that their stable binding is blocked by active transcription or the presence of activating TFs. However, even after inhibition of transcription, a significant number of CGIs remains unbound by PcG proteins, and PcG proteins have been shown to be targeted to sites of active transcription, arguing for an alternative mechanism that complements chromatin sampling by PcG proteins. This might entail interaction of PcG proteins with TFs, such as E2F6, MGA/MAX, REST, SNAIL, RUNX1, JARID2, AEBP2, and YY1. Individually, these TFs induce recruitment only in specific circumstances and to a subset of their target sites. It is therefore possible that the PcG machinery is recruited to target genes via multiple interactions with diverse TFs, similar to the transcription machinery that can induce transcription via interactions with multiple enhancer-bound TFs.

Long non-coding RNAs (lncRNAs) also regulate recruitment of PcG complexes (Figure 3B). After the initial discovery that XIST plays a role in targeting PcG to the inactive X chromosome (Plath et al., 2003), additional work implicated other lncRNAs, like HOTAIR and KCNQ1, in targeting PcG proteins to *HOX* or imprinted gene loci (reviewed in Davidovich and Cech, 2015), although the direct role of lncRNAs in PcG recruitment remains disputed (Cerase et al., 2014; Portoso et al., 2017). In addition, short RNAs transcribed from repressed genes that form stem-loop structures resembling PRC2-binding sites in XIST can interact with PRC2 in *cis*, suggesting a general role for RNAs in PRC2 targeting (Kanhere et al., 2010). The complex relationship between RNAs and PRC2 is underscored by the observations that PRC2-RNA interactions can also prevent PRC2 recruitment to chromatin or inhibit its HMT activity (Beltran et al., 2016; Cifuentes-Rojas et al., 2014; Herzog et al., 2014). In summary, the current data suggest that a general RNA-PRC2 interaction counteracts silencing, whereas selective RNAs might favor PRC2-mediated silencing in the presence of specific cofactors.

A third factor that regulates PcG recruitment is chromatin marks deposited by other histone-modifying complexes or by PcG proteins themselves (Figure 3C). Whereas the PRC2-dependent mark H3K27me3 increases the affinity for Pc/CBX-containing PRC1 complexes to chromatin (Fischle et al., 2003; Min et al., 2003), recent evidence in mammals suggests the existence of a recruiting step upstream of this canonical hierarchy (Figure 2A). Tethering KDM2B induces the recruitment of ncPRC1 complexes in a PRC2-independent manner, which results in the deposition of H2AK119ub and promotes binding of PRC2 (Blackledge et al., 2014; Cooper et al., 2014). Moreover, a recent study showed that the ncPRC1.3/5 complex initiates recruitment of PRC1 and PRC2 to the inactive X chromosome via H2A ubiquitination, demonstrating a function of ncPRC1 and H2A ubiquitination for the initiation of Polycomb domains in a physiological context (Almeida et al., 2017). H2AK119ub-mediated recruitment of PRC2 might depend on AEBP2 and JARID2, which bind to H2AK119ub in vitro (Cooper et al., 2016; Kalb et al., 2014). Intriguingly, however, the E3 ligase activity of RING1B is not essential for PcG function in mouse (Illingworth et al., 2015), and H2A ubiquitination in flies seems to be largely dispensable for PRC2 recruitment (Kahn et al., 2016; Pengelly et al., 2015). Direct interactions between PRC1 and PRC2 complexes mediated by the proteins SCM or EED might support their co-binding to target genes, providing a potential explanation for why histone modifications are dispensable for PcG recruitment in many cases (Cao et al., 2014; Frey et al., 2016; Kang et al., 2015). The H3K36me3 modification, usually associated with active genes or DNA damage sites (Musselman et al., 2013), can also contribute to PcG targeting. The PRC2-associated factor Polycomb-like binds to H3K36me3 via its Tudor domain (Ballaré et al., 2012), facilitating the intrusion of PRC2 into active chromatin regions to promote silencing. Finally, multiple links have been established between H3K9 methylation and Polycomb. EED can bind H3K9me3 (Margueron et al., 2009), and PRC2 physically interacts with the H3K9 methyltransferases GLP/EHMT1 and G9A, whose loss results in reduced PRC2 binding to chromatin (Mozzetta et al., 2014).

In addition to H3K27me3, PRC2 catalyzes H3K27me1 and H3K27me2, which cover the large majority of the euchromatic genome (Ferrari et al., 2014; Lee et al., 2015). PRC2 is not detected in regions marked by H3K27me2, suggesting that its binding to these regions is transient. As PRC2 prefers lower methylated states (i.e., H3K27me0 or H3K27me1) over H3K27me2 for its methylation reaction (McCabe et al., 2012), this suggests a model in which PRC2 associates with most of the genomic chromatin through weak and transient interactions, which lead to mono- and dimethylation of H3K27. In contrast, additional components, including TFs, non-coding RNAs (ncRNAs), and specific chromatin-protein interactions, stabilize the chromatin targeting of PRC2 to a subset of regions where it can deposit H3K27me3 and facilitate cPRC1 targeting. Likewise, multiple targeting components might induce PRC1-specific chromatin binding in the absence of H3K27me3. The overarching picture is thus that PcG complexes have relatively poor DNA sequence specificity, but they can be specifically recruited to or evicted from selected regions by specific TFs, nascent RNAs, ncRNAs, and chromatin modifications.

Recruitment of TrxG Components

Compared to the wealth of data concerning PcG complexes, much less is known about the recruitment of TrxG complexes, although this is an equally important subject. In flies, TrxG complexes co-occupy a fraction of PREs with PcG complexes and are potentially recruited by a similar set of TFs implicated in PcG complex recruitment (Beisel et al., 2007). Whereas Trx is recruited independently of the activation status of its target gene, Ash1 and Brm are targeted to chromatin in an activation-dependent manner (Dejardin and Cavalli, 2004; Papp and Müller, 2006). In analogy to KDM2B-mediated recruitment of PRC1 to CGIs, sequence-specific DNA binding of COMPASS-like complexes to CGIs can be mediated by the CxxC domain of MLL1/2 (Ayton et al., 2004; Hu et al., 2017) (Figure 3A). COMPASS-like complexes can also be recruited to CGIs via their interaction with PAF1, which directly interacts with motifs flanking the CxxC domain of MLL1/2 (Muntean et al., 2010). Moreover, SET1/COMPASS complexes can be recruited to CGIs via their interaction with another CxxC-domain-containing protein, CFP1 (also known as CXXC1), that specifically associates with non-methylated CGIs (Thomson et al., 2010) (Figure 3A). Recently, the pioneering factor FOXA1 was shown to recruit MLL3 to enhancers, thereby mediating H3K4me1/2 (Jozwik et al., 2016). lncRNAs can also contribute to the recruitment of COMPASS to chromatin. *HOXA* transcript at the distal tip (HOTTIP) interacts with the COMPASS subunit WDR5 (Wang et al., 2011; Yang et al., 2014), resulting in COMPASS recruitment to the *HOXA* locus (Figure 3B). However, similar to the role of lncRNAs in PcG recruitment, the question of a more general role of lncRNAs in TrxG complex targeting to chromatin remains elusive. Finally, COMPASS can also be targeted to chromatin by preexisting histone marks (Figure 3C). MLL1 binds H3K4me3 via its PHD finger, thereby stabilizing its association with the *HOXA9* target gene locus (Milne et al., 2010). The COMPASS subunit WDR5 can contribute to chromatin binding by binding to H3K4 tails, irrespective of their methylation status. Another histone mark that contributes to targeting COMPASS to chromatin is the H3K36me2 mark deposited by ASH1L, which promotes binding of the epigenetic reader lens epithelium-derived growth factor (LEDGF) and MLL at key leukemia target genes. Conversely, overexpression of the demethylase KDM2A reduces H3K36me2, leading to the replacement of LEDGF/MLL and reduced expression of MLL target genes (Zhu et al., 2016). These data suggest that multifactorial contributions might stabilize the TrxG/COMPASS complex targeting to a subset of chromosomal regions, similar to the case of PcG complexes.

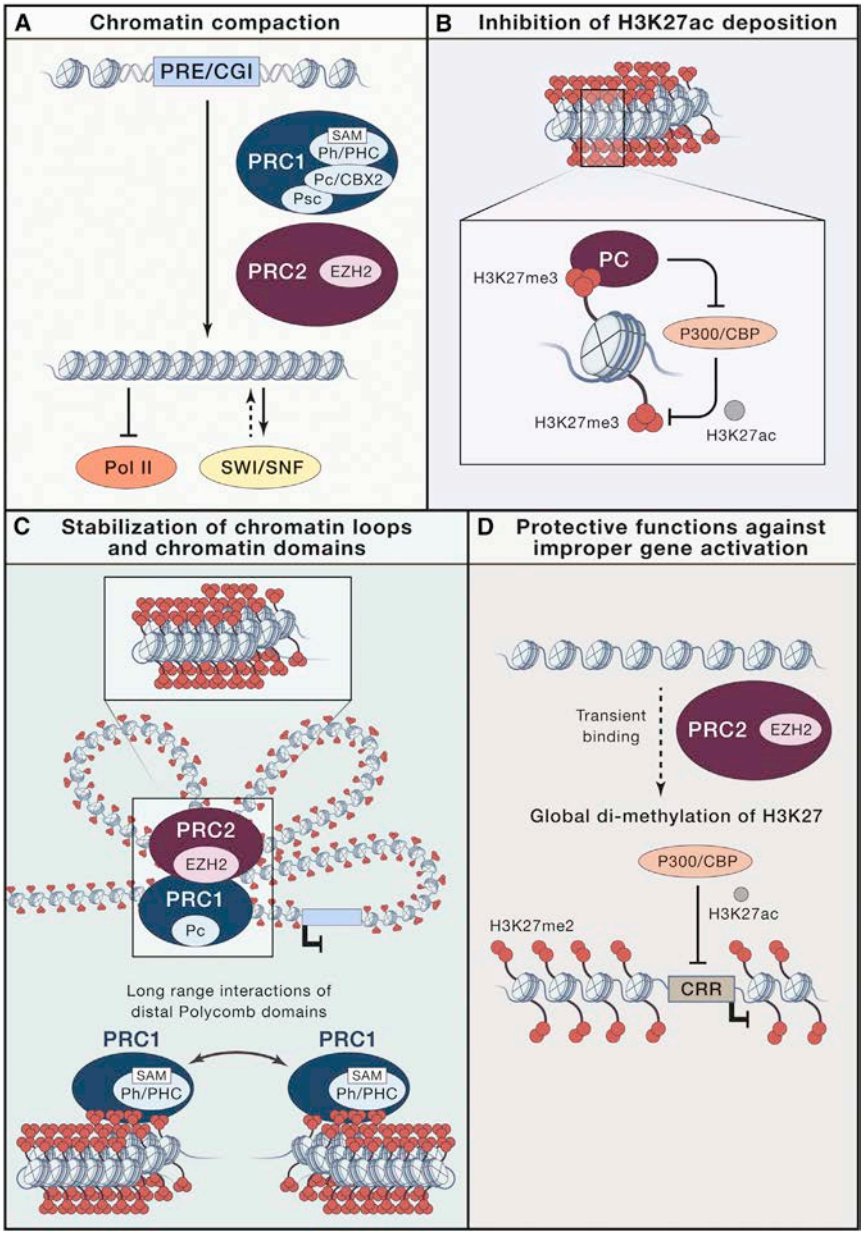
Mechanisms of PcG- and TrxG-Mediated Chromatin Regulation

PcG-mediated repression is considered to be the default state, with gene activation mediated by the counteracting activity of TrxG proteins (Klymenko and Müller, 2004). Intriguingly, both in flies and mammals, PcG and TrxG complexes extensively colocalize at chromatin, regardless of the activity state of the target gene (Beisel et al., 2007; Enderle et al., 2011; Papp and Müller, 2006). One key question is therefore how PcG and TrxG complexes exert their repressive or activating function and what tips the balance between PcG-mediated repression and TrxG-dependent gene activation.

One of the principal mechanisms of PcG-mediated repression is chromatin compaction, rendering chromatin inaccessible to the transcription machinery and inhibiting SWI/SNF-mediated chromatin remodeling (Shao et al., 1999) (Figure 4A). Importantly, different PRC1 variants might exert different levels of chromatin compaction that is mediated via a highly positively charged "compaction region" present in fly Psc (Grau et al., 2011) and vertebrate CBX2 bridging adjacent nucleosomes to compact chromatin (Lau et al., 2017). Moreover, PRC2 itself can lead to chromatin compaction (Margueron et al., 2008; Terranova et al., 2008). Together with the ability of PRC2 to bind H3K27 trimethylated nucleosomes (Margueron et al., 2009) and with the allosteric stimulation of catalysis by dense chromatin templates (Jiao and Liu, 2015; Yuan et al., 2012), this activity might lead to cooperation of PRC2 and PRC1 toward chromatin condensation and inhibition of transcription. The SWI/SNF complex, which was recently shown to be constantly involved in a dynamic competition with PRC1 thanks to a SMARCA4/BRG1-dependent PRC1 eviction activity (Kadoch et al., 2017; Stanton et al., 2017), might be key to switch this repressive chromatin toward activation. The outcome of this competition at individual loci might depend on the amount of PRC1 being recruited, which may lead to different net levels of PcG complexes on specific target genes and drive each locus into an open or closed chromatin state. This corroborates classic genetic studies showing that the SWI/SNF subunit Brm suppresses Polycomb mutations (Tamkun et al., 1992). Interestingly, beyond regulating Hox genes, the mammalian SWI/SNF complex is critically important in cell-fate decisions, such as neuronal differentiation and mitotic exit (Victor et al., 2014; Yoo et al., 2009, 2011). It will be interesting to see whether this role involves the activation of Polycomb target genes.

Polycomb can also act as a repressive agent by inhibiting the acetyltransferase activity of dCBP associated with regulatory regions to block H3K27 acetylation, thereby favoring H3K27 methylation (Tie et al., 2016) (Figure 4B). On the activating side, Trx associates with dCBP and is required for H3K27 acetylation to block PRC2-mediated methylation (Tie et al., 2009). Similarly, Trx-catalyzed H3K4me3 and Ash1-mediated methylation of H3K36 can inhibit PRC2 activity (Schmitges et al., 2011; Yuan et al., 2011) and Trx-dependent dimethylation on histone H3K4 marks PREs and contributes to the maintenance of activation (Rickels et al., 2016). The molecular trigger to switch PREs between silencing and activation might be mediated by non-coding transcription through PREs, affecting PRC2 activity (Herzog et al., 2014) rather than displacement of PcG proteins (Erokhin et al., 2015).

PRC1-mediated H2A ubiquitination was initially suggested to play an important role in PcG-mediated repression by interfering with multiple steps of the transcription process (Nakagawa et al., 2008; Stock et al., 2007; Zhou et al., 2008). However, more recent data have demonstrated that H2A ubiquitination is not required for PcG repression in flies and mammals (Illingworth et al., 2015; Pengelly et al., 2015), suggesting that PRC2 might be a key effector of PcG-mediated silencing. Indeed, flies carrying a point mutation in lysine 27 of histone H3 fail to repress PRC2-target genes, mimicking PRC2 mutant phenotypes and indicating that H3K27 is the PRC2 substrate relevant for PcG



Figures 4. PcG Proteins Use Multiple, Cooperative Mechanisms to Silence Their Target Genes

(A) PRC1 can induce chromatin compaction via a positively charged compacting region present in mammalian CBX2 or fly Psc. In addition, Pc/CBX (via H3K27me3-binding through its chromodomain) and Ph/PHC (via polymerization through its SAM domain) can contribute to chromatin compaction, which can interfere with SWI/SNF-mediated chromatin remodeling or PolII recruitment. PRC2 can also contribute to the compaction of nucleosomal arrays.

(B) Trimethylation of H3K27 (H3K27me3) can directly block acetylation of H3K27 (H3K27ac), which is involved in gene activation. Further, Pc/CBX can inhibit the acetyltransferase activity of CBP, therefore favoring methylation of H3K27.

(C) Distribution of the H3K27me3 mark over large genomic regions might stabilize chromatin-looping interactions between PcG-binding sites and gene regulatory regions, thereby contributing to stably locking genes in a repressed state. Oligomerization of the SAM domain of Ph/PHC is essential for PcG-mediated repression and can mediate long-range interactions between distal Polycomb domains.

(D) Dimethylation of H3K27 (H3K27me2) exerts protective functions by preventing acetylation of *cis* regulatory regions (CRRs), such as enhancers or promoters, thereby inhibiting their inappropriate activation.

repression (Pengelly et al., 2013). It seems that H3K27me3 stabilizes PcG-mediated repression but is not the initial cue (Kahn et al., 2016). This mark might also stabilize chromatin looping interactions of PRE-bound PcG complexes with neighboring gene regulatory elements, thereby contributing to locking target genes in a repressed state and promoting the formation of large repressive genomic domains (Figure 4C). Furthermore, the PRC1 subunit Ph cooperates with H3K27me3 to lock genes in a repressed state by mediating long-range interactions between distal PcG domains via oligomerization of its SAM domain, thereby establishing nuclear subcompartments enriched in PcG proteins (Wani et al., 2016). While PRC1 and PRC2 can and do cooperate at multiple target genes, the long-standing dogma that PcG repression requires both complexes has been challenged by evidence for an uncoupling of these complexes in gene regulation for at least a subset of PcG target genes that depend on PRC1, but not PRC2, for their silencing (Loubière et al., 2016; Schaaf et al., 2013). However, PRC1-dependent transcriptional repression is weaker at these genes than at canonical genes, suggesting that the cooperation between the PRC1 and PRC2 classes of complexes ensures more robust gene silencing.

Finally, the repressive function of PcG complexes might be mediated by the widespread H3K27 dimethylation, which can suppress pervasive chromatin opening and transcriptional activities by preventing H3K27 acetylation (Ferrari et al., 2014; Lee et al., 2015) (Figure 4D).

Higher-Order Genome Regulation

In metazoans, chromatin is organized in a hierarchical series of 3D architectures. At the lower level, nucleosome fibers fold into chromatin loops, which build topologically associating domains (TADs). TADs interact to form active and inactive chromosome compartments, which organize into chromosome territories (for reviews, see Bonev and Cavalli, 2016; Entrevan et al., 2016; Schwartz and Cavalli, 2017). PcG proteins display a punctate distribution in the cell nucleus, both in flies and in mammals (Buchenau et al., 1998; Saurin et al., 1998). These

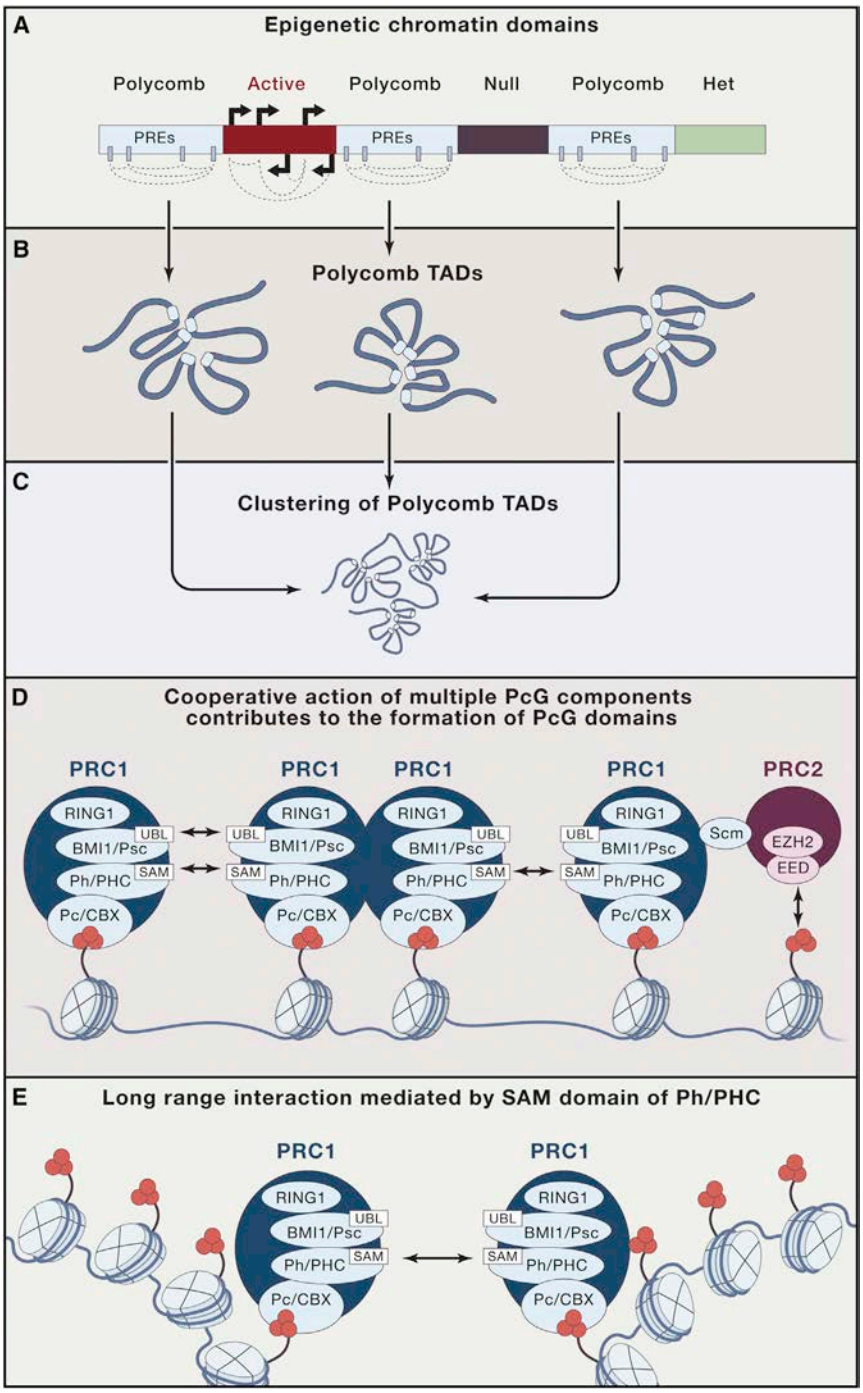


Figure 5. Roles of PcG Proteins in Regulating Higher-Order Chromosome Organization

(A) Chromatin domains are defined by a specific combination of chromatin marks. Intradomain chromatin contacts are much more frequent than contacts with regions outside domains.

(B) PcG components contribute to multi-looped higher-order structures that lead to the formation of repressive chromosomal domains or Polycomb TADs.

(C) Multiple Polycomb TADs can cluster, forming nuclear repressive compartments that stabilize silencing.

(D) EED and Pc/CBX can bind to H3K27me3, which might contribute to the propagation of PcG complexes in *cis* along the chromatin fiber. The PRC1 accessory protein Scm/SCM can contribute to the spreading of PcG complexes by acting as a molecular bridge connecting PRC1 and PRC2 via its SAM domain. In addition, the SAM domain of Ph/PHC and the UBL domain of PCGF4/BMI1 can contribute to PRC1 oligomerization, which might stabilize PcG complexes bound to their target sites. Similarly, *Drosophila* Psc has been show to bind to itself.

(E) Ph/PHC oligomerization mediated by the SAM domain can also facilitate binding of PcG complexes to extended chromatin regions by mediating long-range interactions between distal PRC1 binding sites.

PcG foci were shown to be the physical manifestation of PcG-mediated silencing (Grimaud et al., 2006a; Terranova et al., 2008), and recent developments have begun to unravel the molecular basis of PcG-mediated 3D genome regulation (Figure 5). At the level of oligonucleosomes and chromatin fibers, new technology has begun to pinpoint differences between Polycomb and other types of chromatin (Risca et al., 2017). In mammals, flies, and plants, several PcG components have been shown to be involved in regulatory chromatin loops (for a review, see Entrevan et al., 2016) as well as in the regulation of higher-order chromosome structure and function. Genome-wide mapping of a large set of DNA-binding proteins and histone marks revealed that the genome can be partitioned into epigenetic domains characterized by a specific combination of active or repressive chromatin marks (Filion et al., 2010; Kharchenko et al., 2011; Thurman et al., 2012) (Figure 5A). These states are distributed in chromosome domains, and the analysis of their 3D organization by Hi-C has shown that they form TADs, in which intradomain chromatin contacts have a much higher frequency than contacts with regions outside the domain (Dixon et al., 2012; Nora et al., 2012; Rao et al., 2014; Sexton et al., 2012). PcG components organize the 3D architecture of Polycomb-containing TADs (Figure 5B). Many studies focused on *HOX* gene clusters, which are regulated by PcG, TrxG, and insulator proteins in insects as well as in mammals. In flies, PcG- and TrxG-binding sites juxtapose insulator elements, and this proximity is important for *Hox* regulation (Schwartz and Pirrotta, 2007; Singh and Mishra, 2015). Within these domains, PcG binding sites establish preferential contacts, suggesting that silencing might involve the formation of a multi-looped structure (Lanzuolo et al., 2007). In mammals, repressed *HOX* clusters also

form a repressive domain spatially segregated from adjacent chromosome regions (Noordermeer et al., 2011). 3D contacts between independent PcG binding sites may support spreading of PcG complexes in *cis*, leading to the establishment of the highly condensed chromatin architecture described in these species. Local looping interactions and spreading in *cis* is likely mediated by the cooperative action of PRC2 (Eed) and PRC1 (Ph/PHC, Psc/BMI1, and Scm) subunits (Figure 5D) (Kundu et al., 2017; Wani et al., 2016). Mammalian *HOX* domains are progressively opened during activation, and 4C (circular chromosome conformation capture) studies showed that the active part of the cluster segregates away from the Polycomb-repressed domain (Noordermeer et al., 2011). The CTCF insulator protein regulates this process and is required for appropriate regulation of *HOX* genes (Narendra et al., 2015, 2016). These data thus suggest that regulation of 3D architecture of chromosome domains plays a critical role in specification of the body plan.

At a higher level of folding, Polycomb domains cluster to contact each other in the cell nucleus (Figure 5C). Again, this phenomenon is conserved in flies, plants, and mammalian systems (Bantignies et al., 2011; Rosa et al., 2013; Schoenfelder et al., 2015; Vieux-Rochas et al., 2015). Furthermore, Polycomb-mediated long-range interactions are dynamic. In flies, activation of a subset of the *Hox* genes prevents their interaction with their Polycomb-silenced counterparts (Bantignies et al., 2011), and in mouse ESCs, PRC2-dependent interactions are established during the transition from the naive to the primed state (Joshi et al., 2015). Both PRC2 and PRC1 have been shown to be required for long-range interactions of PcG-regulated regions in ESCs (Boettiger et al., 2016; Denholtz et al., 2013; Schoenfelder et al., 2015; Wani et al., 2016). However, PRC2 does not seem to be critical for establishing TADs (Nora et al., 2012), and its depletion has only limited effects on long-range interactions of Polycomb domains (Denholtz et al., 2013; Schoenfelder et al., 2015). In contrast, PRC1 plays a critical role, both in local chromatin condensation of Polycomb domains and in the establishment of their long-range interactions (Schoenfelder et al., 2015; Wani et al., 2016). In particular, oligomerization mediated by the SAM domain of the Ph/PHC1 subunit of cPRC1 (Figures 5D and 5E) is crucial for the condensation of individual Polycomb domains (Kundu et al., 2017) as well as for mediating long-range Polycomb domain interactions (Isono et al., 2013; Wani et al., 2016). Of note, not only Polycomb-silenced genes but also many active gene promoters interact in space with their enhancers and other promoters (Javierre et al., 2016; Schoenfelder et al., 2015). Whether COMPASS complexes are causally linked to 3D interactions, and whether this architecture affects function, are important questions for future research.

Equally important is to understand whether the 3D organization of PcG proteins plays a regulatory role or whether it is a consequence of the silencing process. In addition to a reciprocal stabilization of PRC1 and PRC2 complexes, PRC1 stabilizes the binding of the key recruiter protein Pho (Kahn et al., 2014; Schuettengruber et al., 2014), suggesting that, after initial PcG recruitment by DNA-binding factors, PcG complexes may form nuclear subcompartments that stabilize their binding as well as that of their own recruiters. This scenario suggests that 3D organization contributes to Polycomb-mediated silencing (Schuettengruber et al., 2014), yet how can one test this hypothesis? Mutations of PRC2 components such as EED, or of PRC1 components such as RING1 and Ph/PHC1, induce changes in gene expression and in 3D chromosome architecture, such that effects due to the loss of local PcG repressive components cannot be easily disentangled from those due to higher-order chromatin organization (Bantignies et al., 2011; Denholtz et al., 2013; Joshi et al., 2015; Schoenfelder et al., 2015; Wani et al., 2016). To tackle this, Wijchers et al. (2016) ectopically targeted EZH2 to naive loci containing lacO-binding sites, which led to local deposition of H3K27me3 and the establishment of new long-range contacts with other Polycomb domains, yet these changes had little impact on gene expression. In contrast, evidence for a direct role of 3D organization was obtained in *Drosophila* by deleting a DNA element in one of the *Hox* clusters (the BITHORAX complex). This mutation induced loss of contacts with the other *Hox* cluster (the ANTENNAPEDIA complex), which is located 10 Mb away along the same chromosome arm. In this configuration, genes in the ANTENNAPEDIA complex were upregulated, suggesting that 3D contacts contribute to the stability of Polycomb-dependent gene silencing (Bantignies et al., 2011). Similar genome engineering studies (i.e., manipulating the 3D neighborhood of the tested loci while preserving their *cis*-regulatory sequences) will be required to understand the relationship between nuclear organization and gene expression changes on a genome-wide scale.

The Role of PcG and TrxG Proteins in Somatic and Transgenerational Epigenetic Inheritance

PcG and TrxG proteins were originally studied for their ability to maintain the memory of *HOX* gene expression patterns throughout development (Cavalli and Paro, 1998; Grimaud et al., 2006b; Margueron and Reinberg, 2011; Poux et al., 2002; Schwartz and Pirrotta, 2007), but how powerful is this memory function, and what are its mechanisms? The two most challenging processes for chromatin inheritance during cell proliferation are DNA replication and mitosis/meiosis. PcG components were shown to remain bound to chromatin templates during replication, both in vitro (Francis et al., 2009; Lengsfeld et al., 2012; Lo et al., 2012) and in vivo (Hansen et al., 2008; Petruk et al., 2012). Although H3K27me3 is diluted during DNA replication, it can mediate short-term memory of repressed chromatin states (Coleman and Struhl, 2017), and post-replicative PRC2 activity is induced by binding to H3K27me3 on parental nucleosomes to restore original H3K27me3 levels (Alabert et al., 2015). However, sequence-specific targeting of PRC2 to PREs is required for efficient propagation of H3K27me3 levels over multiple cell divisions, indicating that the whole PcG machinery (including PRE sequences, chromatin-modifying activities, and their associated histone marks) acts as a epigenetic memory system that is needed for long-term memory of repressive chromatin states (Coleman and Struhl, 2017; Laprell et al., 2017). Likewise, Trx remains associated during DNA replication, and H3K4me3 is rapidly reconstituted after DNA replication (Alabert et al., 2015; Petruk et al., 2012), suggesting that Trx/COMPASS proteins use similar mechanisms to propagate active chromatin states.

Much less is known about the maintenance of PcG and TrxG components during mitosis. Quantitative in vivo analysis and genome-wide mapping studies in mitotic cells revealed the presence of Psc and Pc as well as the TrxG member Ash1 on mitotic chromatin (Follmer et al., 2012; Fonseca et al., 2012; Steffen et al., 2013). Likewise, both MLL and PcG components were detected on mammalian mitotic chromosomes, suggesting that at least a part of the PcG and TrxG machinery may bookmark mitotic chromosomes to inherit chromatin states. However, the mitotic chromosome distribution of both mammalian MLL components and fly Psc differs at least partly from that in interphase cells, and the mechanism of this targeting dynamic is still unknown (Blobel et al., 2009; Follmer et al., 2012).

Polycomb-mediated memory of chromatin states can last for long periods of time, as illustrated by the phenomenon of vernalization, in which exposure to prolonged cold accelerates flowering after a return to milder temperatures. In *Arabidopsis*, this phenomenon depends on PRC2-mediated silencing of the *FLC* locus, which progressively and quantitatively increases with longer periods of cold exposure (Angel et al., 2011; Bastow et al., 2004). In every generation, the chromatin state of the *FLC* locus is reset by the H3K27-specific demethylase ELF6, which prevents transgenerational inheritance of previous exposures to cold (Crevillén et al., 2014). Other observations suggest the possibility that inter- and transgenerational inheritance of chromatin states may occur under certain circumstances. Recent data suggest that intergenerational inheritance of maternally provided nucleosomes carrying H3K27me3 are propagated in early embryos to prevent inappropriate activation of lineage-specific enhancers during development (Zenk et al., 2017). PcG and TrxG proteins were also shown to induce transgenerational inheritance of alternative states of transgene expression in flies (Bantignies et al., 2003; Cavalli and Paro, 1998). In *C. elegans*, PRC2 proteins had been suggested to transmit inter- and transgenerational silencing of the X chromosome via both sperm and oocytes (Gaydos et al., 2014). In vertebrates, the mechanisms underlying transgenerational epigenetic inheritance are hotly debated (Blake and Watson, 2016; Heard and Martienssen, 2014; Hollick, 2017). This is mostly because specific pathways are responsible for resetting the state of the epigenome during germline development. DNA methylation is vastly reprogrammed, and most methylated DNA sites do not survive in the next generation (Smallwood and Kelsey, 2012). However, a small subset of loci can preserve their DNA methylation status (Surani, 2015). Improvement of chromatin-immunoprecipitation-sequencing (ChIP-seq) technologies allowed histone marks to be mapped in gametes and the early stages of mouse development. The data showed that both sperm and oocytes bear regions endowed with H3K4me3 and H3K27me3, which are extensively reprogrammed upon fertilization and during the early cell division cycles of the embryo (Dahl et al., 2016; Liu et al., 2016; Zhang et al., 2016; Zheng et al., 2016). However, a significant number of these regions are maintained during early embryogenesis (Zheng et al., 2016), and both PRC1 components and MLL2 are required in the oocyte for the correct specification of gene expression patterns in embryos of the subsequent generation (Andreu-Vieyra et al., 2010; Posfai et al., 2012). Likewise, the alteration of H3K4me3 levels by overexpression or inhibition of the H3K4 demethylase KDM5B induces severe early embryonic defects (Dahl et al., 2016; Liu et al., 2016; Zhang et al., 2016). Mouse and frog studies directly suggested the possibility of a role of PcG and TrxG proteins in transgenerational inheritance (Siklenka et al., 2015; Teperek et al., 2016). Finally, the analysis of a system in which the nuclear organization of a transgene was transiently perturbed demonstrated the PRC2-dependent transgenerational inheritance of different levels of H3K27me3 in *Drosophila.* Remarkably, inheritance could be transmitted through both the female and the male germline, suggesting that it is robust even to the partial replacement of histones by protamines during spermatogenesis (Ciabrelli et al., 2017). It will be important to define the extent to which this phenomenon can regulate endogenous genome function and its underlying molecular mechanisms, including the interplay with other components, such as ncRNAs, that have been linked to transgenerational inheritance (Sharma, 2017; Sharma et al., 2016).

Role in Stem Cell Renewal and Differentiation

While mutations of PcG proteins lead to homeotic transformations and other developmental defects in *Drosophila*, knockout of most PRC2 or PRC1 subunits in mice causes embryonic lethality at the gastrulation stage (Aloia et al., 2013). These genetic studies highlight a key role of PcG proteins in the specification and maintenance of cell fate rather than defects in the viability of early pluripotent cells (Montgomery et al., 2005). Indeed, genome-wide mapping of PRC1 and PRC2 subunits in pluripotent stem cells (Morey et al., 2012), more differentiated progenitors, and terminally differentiated cells (Morey et al., 2015) supports the key role of PcG proteins in cell identity and the maintenance of a proper differentiation program during development. In both embryonic and adult stem cells, a large number of PcG-bound regions are decorated by both the H3K27me3 repressive mark and the H3K4me3 activation mark (Shema et al., 2016). Those modifications, which are catalyzed by PRC2 and the MLL2 complex (Piunti and Shilatifard, 2016), respectively, contribute to set the promoter of genes implicated in cell-fate determination and development into a poised bivalent state. Such ''bivalent promoters'' are transcribed at very low levels and can be either activated or repressed, depending on developmental signals. How the correct balance between PcG and MLL2 occupancy is regulated at bivalent regions is still unclear, and the contributions of histone post-translational modifications (PTMs) in modulating promoter activity and chromatin organization of bivalent regions remain to be determined.

Bivalent regions and, more generally, the set of PcG-repressed genes differ depending on the state of the cells. In pluripotent ESCs, PcG complexes repress all lineage-specific genes, while genes necessary to maintain the proliferative and undifferentiated state of ESCs are continuously transcribed. Repression of specific lineage genes is relieved when ESCs are induced to differentiate. This occurs concomitantly with the silencing of pluripotency genes. Under this condition, PcG proteins still maintain the repression of alternative cell-fate genes, thus allowing cells to proceed to one specific cell type and preventing cells from de-differentiation or trans-differentiation. This exquisite regulation and the genome-wide redistribution of both PRC1 and PRC2 relies on the assembly and disassembly of PcG

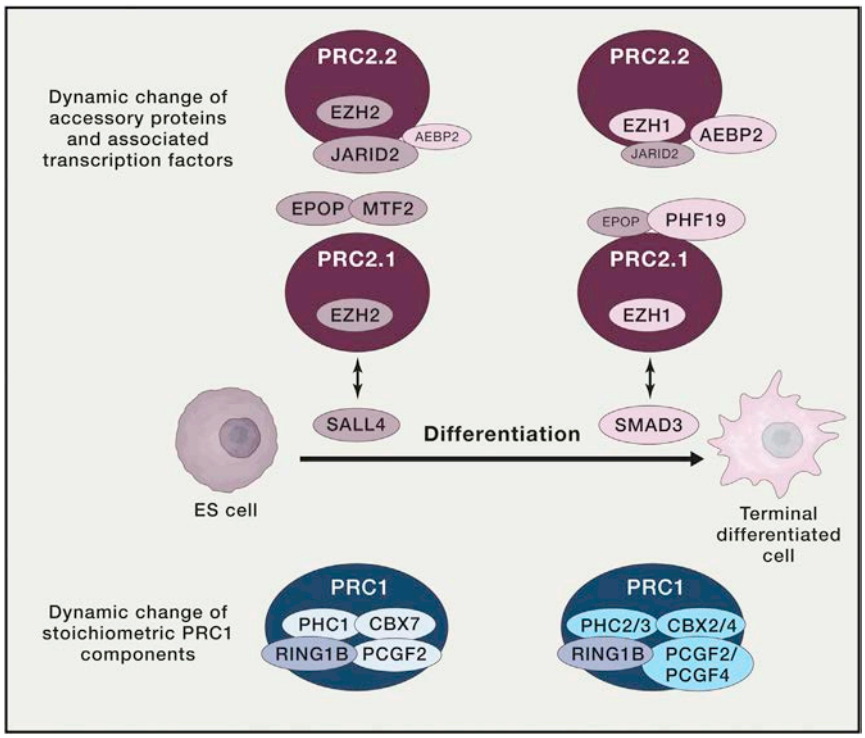


Figure 6. Dynamic Assembly of PRC1 and PRC2 during ES Cell Differentiation

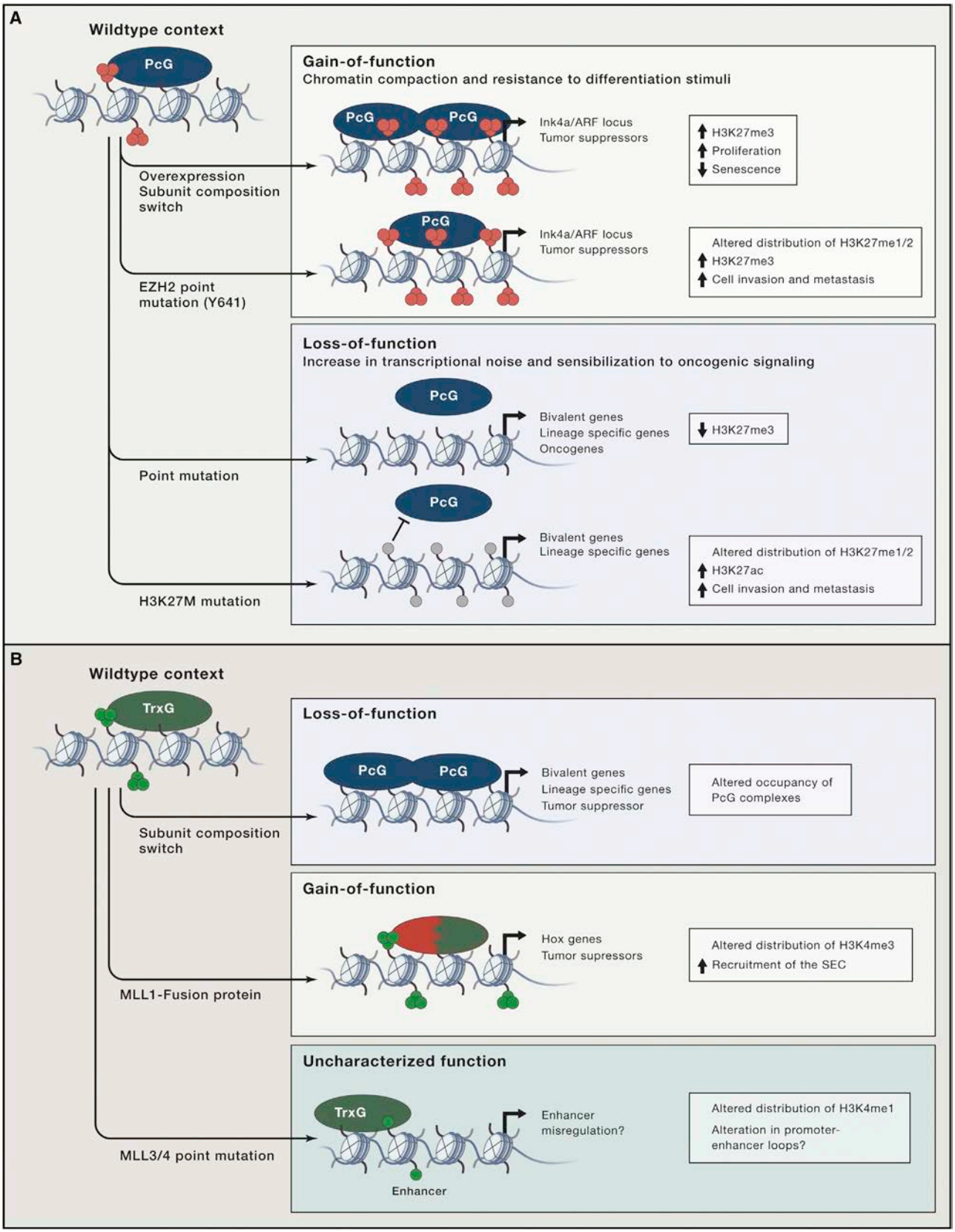
Different paralogs of PRC1 core components, such as PCGF or CBX proteins, are incorporated into PRC1 in ESCs or differentiated cells to potentially regulate PRC1 function. Similarly, different ancillary proteins assemble with PRC2 in ESCs or differentiated cells to modulate its function and localization during differentiation.

complexes for the proper balance between pluripotency and lineage commitment (Figure 6). In pluripotent ESCs, CBX7 is the most abundant CBX family member. To allow differentiation, CBX7 has to be replaced by other paralogous family members, such as CBX2, CBX4, and CBX8. Failure to replace CBX7 with other CBX family members causes hyper-proliferation of hematopoietic stem cells and thus leukemia (Klauke et al., 2013), while the function of CBX6 in gene regulation and in ESC differentiation is still enigmatic. Recently, several reports unveiled the specific role of PCGF family members in ESC maintenance (Endoh et al., 2017), mesodermal differentiation (Morey et al., 2015), and reprogramming (Zdzieblo et al., 2014). Thus, variation in assembly contributes to the dynamic targeting of PRC1 to different promoters (via CBX proteins) and provides specific biochemical functions that can regulate cell identity and cell-fate choices (via PCGF proteins). We know much less about how this occurs in adult stem cells or in differentiated cell types. In the past few years, several studies demonstrated an essential role for PRC1 and PRC2 in the patterning and homeostasis of the adult mouse intestine, skin, craniofacial structures, and hematopoietic systems (Chiacchiera et al., 2016; Ezhkova et al., 2009; Hidalgo et al., 2012; Minoux et al., 2017; Xie et al., 2014). Indeed, deletion of PcG proteins compromises the self-renewal capacity of adult stem cells, often triggering senescence and unscheduled activation of the differentiation program.

The situation is a bit simpler for the PRC2 core components, with an exchange of EZH2 with EZH1 during the switch between proliferation and differentiation. Although EZH1 can partially complement EZH2 functions in ESCs (Shen et al., 2008), biochemical data suggest that EZH2-PRC2 is enzymatically more active, while EZH1-PRC2 is more efficient in chromatin compaction assays (Margueron et al., 2008). However, most of the variation of the PRC2 complex relies on its dynamic interactions with accessory proteins (Figure 6); whereas EPOP, MTF2, PHF19, and JARID2 are more strongly enriched in undifferentiated cells, PHF1 and AEBP2 are preferentially associated with PRC2 core components in differentiated cells (Kloet et al., 2016; Oliviero et al., 2016). Clearly, these proteins can contribute to PRC2 targeting during cell-fate transitions via direct recognition of histone modifications (Ballaré et al., 2012) or by providing affinity for DNA sequences (Pasini et al., 2010). Indeed, depletion of PCLs or of JARID2 compromises the association of PRC2 with specific target genes as well as ESC differentiation. In contrast, EPOP, which is only present in the PCL-PRC2 complex, does not affect PRC2 occupancy (Beringer et al., 2016; Liefke et al., 2016). Together, these data suggest a scenario in which EPOP/PCLs and JARID2 associate with PRC2 in a mutually exclusive fashion (Grijzenhout et al., 2016; Hauri et al., 2016). Interestingly, EPOP also brings Elongins B/C into the PCL-PRC2 complex, opening the possibility of direct communication between PRC2 and the Pol II machinery. Since the expression pattern of PRC2-associated factors varies dramatically among different cell types, this adds a further layer of complexity to the assembly and function of the PRC2 complexes. To date, it is unclear how the expression of PRC1 and PRC2 core components and accessory proteins is regulated in different cell types or how the assembly of the different PcG variations is regulated. It is likely that post-translational modifications play an important role.

PcG and TrxG Proteins in Cancer

The role of PcG proteins in both cell identity and development is mainly achieved by the transcription control of pluripotency genes and linage-specific genes. Several years ago, van Lohuizen and coworkers reported that PcG can cooperate with c-MYC to generate mouse lymphomas via direct silencing of the *CDKN2a* locus (Jacobs et al., 1999a), thus suggesting that PcG proteins can play an oncogenic role by modulating cell proliferation and senescence. Later, the Chinnaiyan lab showed that elevated levels of EZH2 and the H3K27me3 mark in prostate tumors often correlates with a poor prognosis (Varambally et al., 2002). Along the same line, several subunits of PRC1 and PRC2 have been found to be overexpressed in both solid



(legend on next page)

cancers and leukemias and to be required for proliferation of cancer cells. This was then followed up by the identification of mutations in the demethylase UTX (KDM6A) in myeloma and renal cell carcinoma. Finally, a gain-of-function mono-allelic mutation in the catalytic pocket of EZH2 (Y641) has also been identified in lymphoma (Morin et al., 2010). The missense mutation renders the EZH2 Y641 incapable of either mono- or dimethylating H3K27, but it enhances the subsequent catalytic step (i.e., trimethylation to H3K27me3). Mechanistically, EZH2 expressed from the wild-type allele provides the enzymatic activity for the initial conversion of unmethylated H3K27 into H3K27me1/2, while mutated EZH2 activity leads to aberrantly high global levels of H3K27me3 in tumor samples. All these data suggest that hypermethylation at H3K27 acts as a driver in several human cancers (reviewed in Comet et al., 2016) and that it can initiate cell invasion and metastasis (Figure 7A). Based on these observations, several small-molecule inhibitors of EZH2 have been developed and are either being tested in clinical trials or already in use in clinical practice (Helin and Dhanak, 2013).

Unexpectedly, inactivating mutations in EZH2, EED, and SUZ12 have also been reported in myelodysplastic syndrome and malignant peripheral nerve sheath tumors (MPNST) (Lee et al., 2014), thus challenging the concept that PRC2 acts as an oncogene. Indeed, these loss-of-function mutations lead to reduced levels of H3K27me2/3 (Figure 7A). Likewise, a lysine-to-methionine mutation at position 27 of histone H3 (H3K27M) has been identified in diffuse intrinsic pontine glioma (DIPG). This mutation also results in reduced levels of H3K27me2/3 (Lewis et al., 2013). Although this was initially correlated to the capacity of H3K27me2/3 to strongly bind and sequester the PRC2 complex, recent data seem to support a model in which H3K27M excludes PRC2 binding, facilitating an aberrant accumulation of acetylated H3K27 (Figure 7A) on the other wild-type copies of histone H3 in heterotypic nucleosomes (Piunti et al., 2017). This has important clinical implications, since inhibition of proteins that bind to acetylated lysines, such as members of the bromodomain (BRD) family of proteins, offer new therapeutic avenues for DIPG. Moreover, recent crystal structural analyses of PcG complexes have provided important information about the mechanisms of histone modification catalysis and specificity. These studies propose a mechanism for the oncogenic effects of H3K27M, which explains how H3K27me3 binding to the PRC2 subunit EED enhances PRC2 activity and suggests possibilities for designing drugs that could interfere with PRC2 function (Brooun et al., 2016; Jiao and Liu, 2015; Justin et al., 2016; McGinty et al., 2014).

Should EZH2 inhibitors thus be considered for cancer treatments? In other words, do PcGs function as oncogenes, tumor suppressors, or both, depending on the cellular context? It is likely that they can be both, leaving us with a highly complex puzzle of understanding the PcG function in each cellular content. What is clear is that an unbalance in cellular levels of H3K27me3 can have devastating effects, compromising both cellular identity and homeostasis. PcG loss of function leads not only to the reactivation of PcG-silenced genes but also to increased transcriptional noise (Kar et al., 2017), which might sensitize the cell to oncogenic signaling. In contrast, PcG gain of function causes diffuse genome compaction, which is likely to be unresponsive to differentiation stimuli, thereby promoting uncontrolled growth.

A further interesting aspect to consider is that TrxG proteins are also often mutated in cancer (Figure 7B). Several SWI/SNF subunits, including SNF5/SMARCB1 and ARID1, are highly mutated in ovarian clear cell carcinomas and malignant rhabdoid tumors (Zinzalla, 2016). Mutations in the ATPase domain of SMARCA4 (also known as BRG1) are frequent in renal, ovarian, medulloblastoma, rhabdoid, colorectal, and lung tumors and lead to increased PcG occupancy on unmethylated CGI promoters, indicating a role for the SWI/SNF complex in displacing PcG complexes from chromatin (reviewed in Kadoch and Crabtree, 2015). COMPASS members are also heavily mutated in cancer. *MLL* genes were originally identified in a region that has frequently translocations in leukemia, and ELL elongation factor was the first reported fusion partner of MLL1 (Ziemin-van der Poel et al., 1991). Surprisingly, in the chimeric MLL-ELL protein, the catalytic SET domain of MLL is not retained, suggesting that altered levels of H3K4 methylation are not implicated in leukemogenesis. These initial observations were corroborated by the characterization of several other MLL fusion partners also implicated in transcription elongation, including AFF1, AF9, ELL, and ENL. Biochemical studies from the Shilatifard laboratory demonstrated that aberrant activation via the recruitment of the super elongation complex (SEC) is a common theme for all MLL translocations (Smith et al., 2011), thus exposing a potential Achilles' heel for treatment. In contrast, the other family members are often mutated rather than translocated in human tumors. *MLL4* missense mutations have been found in lymphomas and medulloblastomas, while *MLL3* mutations have been reported in bladder and kidney neoplasia (Piunti and Shilatifard, 2016). Given the dedicated role of MLL3 and MLL4 in methylating H3K4 at enhancers, it is intriguing to speculate that enhancer malfunction is the direct cause of these pathologies (Figure 7B). However, the underlying mechanisms are still unclear. Alterations in the chromatin loops between those enhancers affected by *MLL3/4* mutations with gene promoters are a likely possibility. Identifying direct

Figure 7. The Function of PcG and TrxG Proteins in Cancer

A summary of common mutations in PcG and TrxG subunits that affect gene expression and cellular functions.

(A) Gain-of-function mutations in PcG due to overexpression and changes in PcG complex composition or due to dominant-negative point mutation in the catalytic subunit (top) often leads to aberrant deposition of H3K27me3 at tumor suppressor loci. Loss-of-function mutations, such as missense mutations (bottom) or mutation in lysine 27 of histone H3, cause reactivation of PcG-repressed genes through depletion of H3K27me3 and accumulation of H3K27ac at lineage-specific genes and oncogenes.

(B) Similarly, deletion, mutation, and truncation of TrxG subunits results in reduced stability and/or activity of TrxG complexes, which in turn affects PcG genome-wide distribution (top). Gain of function, such as chromosome translocations, also are implicated in carcinogenesis processes (middle) through aberrant recruitment of the super elongation complex (SEC). Recently identified mutations in MLL3/4 (bottom panel) could possibly results in changes in enhancers' histone modifications and thus affect interactions with promoters.

target genes using Hi-C technology might shed some light on this.

The roles of PRC1 components in cancer are likely to reflect their functions in stem cells. CBX7 overexpression has been reported in human leukemias, prostate cancer, and ovarian clear cell carcinoma. Results obtained using animal models indicate that it acts as an oncogene by silencing the *INK4A/ARF* locus and preventing senescence and/or apoptosis. Conversely, no clear functional data have been reported so far for an oncogenic role of the other CBX paralogs, although correlative studies have shown their misexpression in several cancers. It is important to note that CBX family members have different affinity for H3K27me3, with CBX7 having the highest binding capability (Bernstein et al., 2006b). Other CBX proteins have been shown to recognize additional histone modifications as well as ncRNA. Understanding the exact mechanism of CBX chromatin recognition is a key challenge to properly understand the roles of this protein family in human cancer. Finally, specific inhibitors for each member could provide important tools to block cancer progression.

Similarly, BMI1 is highly expressed in hematopoietic stem cells but is replaced within the PRC1 complex by MEL18 during differentiation (Morey et al., 2015). Mechanistic insight provided by many studies suggests that BMI1 promotes cell proliferation by orchestrating the repression of *CDKN2*a and *PTEN* and by modulating the AKT pathway (Jacobs et al., 1999b), and its deletion in animal models confirms a key role in the initial steps of the carcinogenesis process. In contrast, MEL18 seems to act as a tumor suppressor by modulating the expression of several oncogenes, including *c-MYC* (Guo et al., 2007). Moreover, in several human tumors, *MEL18* is often downregulated, and its depletion in mouse models does not cause increased tumor formation. Since BMI1 and MEL18 have opposite effects on PRC1 activity (Cao et al., 2005), it is tempting to speculate that the switch of subunits might affect the deposition of the ubiquitin group and therefore gene expression programs. Whether this is a key aspect behind the oncogenic differences between BMI1 and MEL18 remains to be demonstrated, especially since the impact of H2Aub on gene regulation is still unclear.

Conclusions and Future Perspectives

Despite the boom in the PcG and TrxG research field, many open questions remain. Among the future developments, we suggest six major areas to watch. (1) PcG and TrxG members have now gained a reputation as important players in development, physiology, and cancer etiology. We also now have a reasonably good inventory of their target genes in a subset of cell types. However, we do not know how the PcG and TrxG complexes are targeted at specific subsets of genes in each cell type or which targets are critical for the regulation of specific functions in each cell type. While classical research showed, for instance, that *Hox* genes are crucial for regulating the anteroposterior body plan in flies, much more work will have to be devoted to understand critical PcG- and TrxG-regulated processes in the other cases. (2) PcG and TrxG proteins have an increasing evolutionary complexity with the gradual acquisition of more paralogs in more complex species, but what are the functions of individual paralogs? Changing the balance between paralogs in different protein complexes could restrict, expand, or even change the repertoire of target genes, and hence the regulatory outputs, in different cell types or in response to various cues. (3) Over the last decade, a clear link between PcG and TrxG proteins and 3D genome regulation has been established, and 3D contact networks have been shown to be established between active or PcG-silenced genes. However, these findings almost always described correlations. Are the architectural functions required, or are they a consequence of gene regulatory processes? A combination of analytical studies using microscopy and 3C-type technologies, as well as dedicated genome engineering studies, will be required to disentangle causes and consequences. (4) Recent studies have begun to decipher the molecular mechanisms that allow repressive or active chromatin states to be propagated through cell division. Furthermore, it is becoming clear that some PcG and TrxG components can convey transmission of inheritance not only throughout mitosis but also across multiple generations of an organism. On the other hand, additional work suggests that these components not only are able to carry out maintenance functions but also could be used dynamically to change gene regulatory states. It will be important to understand how these dynamic functions differ from memory functions and how transitions between labile, reprogrammable states and more stable, transmissible states might be regulated. Furthermore, in a specific context, these proteins can switch their function. For instance, PcG proteins may be involved in gene activation. They can associate with active genes (Brookes et al., 2012; Frangini et al., 2013; Kaneko et al., 2013; Mousavi et al., 2012; van den Boom et al., 2016) and can activate gene expression via PcG protein association with specific cofactors, positive regulation of PolII activity, recruitment of co-activators, and modulation of chromosome architecture (Arora et al., 2015; Frangini et al., 2013; Gao et al., 2014; Hauri et al., 2016; Lv et al., 2016; Morey et al., 2015). Moreover, a recent report demonstrated that a subset of *Drosophila* PREs can function as developmental enhancers in vivo, activating transcription in specific spatial domains (Erceg et al., 2017). Whether this dual functionality of PREs is conserved in mammals is unclear, and much remains to be done in this field. (5) The points above require a full, in-depth understanding of how PcG and TrxG functions are regulated, such as by post-translational modifications of individual components or by regulation of their gene expression. While multiple post-translational modifications of these components have been described (Gambetta and Müller, 2014; Gonzalez et al., 2014; Niessen et al., 2009) and subunit switching has also been described in various systems and cell types, the mechanistic regulation of these processes has not been investigated in detail. (6) Finally, new and unexpected discoveries will surely be made and will have to be taken into consideration, such as the role of Polycomb components in mitochondrial function (Liu et al., 2009) and the cytoplasm (Bodega et al., 2017; Chen et al., 2017). Future research is thus likely to deepen and expand the understanding the biological functions of epigenetic components in normal physiology and disease to provide the community with new prognostic and diagnostic tools and, finally, allow for the development of new therapeutic treatments for many cancers and other diseases.

SUPPLEMENTAL INFORMATION

Supplemental Information includes one figure and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2017.08.002.

ACKNOWLEDGMENTS

We thank all members of the Cavalli and Di Croce laboratories for helpful discussions. We thank Veronica A. Raker for editing the manuscript. Work in the Di Croce laboratory is supported by grants from the Spanish "Ministerio de Educación y Ciencia" (BFU2016-75008-P), AGAUR, and La Marato TV3. Work in the Cavalli laboratory is supported by grants from the CNRS, the FP7 European Network of Excellence EpiGeneSys, the European Union's Horizon 2020 research and innovation programme (grant 676556 MuG), the Agence Nationale de la Recherche (ANR-15-CE12-0006-01), the Fondation pour la Recherche Médicale (N. 144186), INSERM and ITMO Cancer (MM&TT), and the Laboratory of Excellence EpiGenMed.

REFERENCES

Alabert, C., Barth, T.K., Reverón-Gómez, N., Sidoli, S., Schmidt, A., Jensen, O.N., Imhof, A., and Groth, A. (2015). Two distinct modes for propagation of histone PTMs across the cell cycle. Genes Dev. *29*, 585–590.

Almeida, M., Pintacuda, G., Masui, O., Koseki, Y., Gdula, M., Cerase, A., Brown, D., Mould, A., Innocent, C., Nakayama, M., et al. (2017). PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation. Science *356*, 1081–1084.

Aloia, L., Di Stefano, B., and Di Croce, L. (2013). Polycomb complexes in stem cells and embryonic development. Development *140*, 2525–2534.

Andreu-Vieyra, C.V., Chen, R., Agno, J.E., Glaser, S., Anastassiadis, K., Stewart, A.F., and Matzuk, M.M. (2010). MLL2 is required in oocytes for bulk histone 3 lysine 4 trimethylation and transcriptional silencing. PLoS Biol. *8*, e1000453.

Angel, A., Song, J., Dean, C., and Howard, M. (2011). A Polycomb-based switch underlying quantitative epigenetic memory. Nature *476*, 105–108.

Arora, M., Packard, C.Z., Banerjee, T., and Parvin, J.D. (2015). RING1A and BMI1 bookmark active genes via ubiquitination of chromatin-associated proteins. Nucleic Acids Res. *44*, 2136–2144.

Ayton, P.M., Chen, E.H., and Cleary, M.L. (2004). Binding to nonmethylated CpG DNA is essential for target recognition, transactivation, and myeloid transformation by an MLL oncoprotein. Mol. Cell. Biol. *24*, 10470–10478.

Ballaré, C., Lange, M., Lapinaite, A., Martin, G.M., Morey, L., Pascual, G., Liefke, R., Simon, B., Shi, Y., Gozani, O., et al. (2012). Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. Nat. Struct. Mol. Biol. *19*, 1257–1265.

Bantignies, F., Grimaud, C., Lavrov, S., Gabut, M., and Cavalli, G. (2003). Inheritance of Polycomb-dependent chromosomal interactions in Drosophila. Genes Dev. *17*, 2406–2420.

Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J., Tixier, V., Mas, A., and Cavalli, G. (2011). Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. Cell *144*, 214–226.

Bastow, R., Mylne, J.S., Lister, C., Lippman, Z., Martienssen, R.A., and Dean, C. (2004). Vernalization requires epigenetic silencing of FLC by histone methylation. Nature *427*, 164–167.

Beisel, C., Buness, A., Roustan-Espinosa, I.M., Koch, B., Schmitt, S., Haas, S.A., Hild, M., Katsuyama, T., and Paro, R. (2007). Comparing active and repressed expression states of genes controlled by the Polycomb/Trithorax group proteins. Proc. Natl. Acad. Sci. USA *104*, 16615–16620.

Beltran, M., Yates, C.M., Skalska, L., Dawson, M., Reis, F.P., Viiri, K., Fisher, C.L., Sibley, C.R., Foster, B.M., Bartke, T., et al. (2016). The interaction of PRC2 with RNA or chromatin is mutually antagonistic. Genome Res. *26*, 896–907.

Beringer, M., Pisano, P., Di Carlo, V., Blanco, E., Chammas, P., Vizán, P., Gutiérrez, A., Aranda, S., Payer, B., Wierer, M., and Di Croce, L. (2016). EPOP functionally links Elongin and Polycomb in pluripotent stem cells. Mol. Cell *64*, 645–658.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006a). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315–326.

Bernstein, E., Duncan, E.M., Masui, O., Gil, J., Heard, E., and Allis, C.D. (2006b). Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. Mol. Cell. Biol. *26*, 2560–2569.

Blackledge, N.P., Farcas, A.M., Kondo, T., King, H.W., McGouran, J.F., Hanssen, L.L., Ito, S., Cooper, S., Kondo, K., Koseki, Y., et al. (2014). Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. Cell *157*, 1445–1459.

Blackledge, N.P., Rose, N.R., and Klose, R.J. (2015). Targeting Polycomb systems to regulate gene expression: modifications to a complex story. Nat. Rev. Mol. Cell Biol. *16*, 643–649.

Blake, G.E., and Watson, E.D. (2016). Unravelling the complex mechanisms of transgenerational epigenetic inheritance. Curr. Opin. Chem. Biol. *33*, 101–107.

Blobel, G.A., Kadauke, S., Wang, E., Lau, A.W., Zuber, J., Chou, M.M., and Vakoc, C.R. (2009). A reconfigured pattern of MLL occupancy within mitotic chromatin promotes rapid transcriptional reactivation following mitotic exit. Mol. Cell *36*, 970–983.

Bodega, B., Marasca, F., Ranzani, V., Cherubini, A., Della Valle, F., Neguembor, M.V., Wassef, M., Zippo, A., Lanzuolo, C., Pagani, M., and Orlando, V. (2017). A cytosolic Ezh1 isoform modulates a PRC2-Ezh1 epigenetic adaptive response in postmitotic cells. Nat. Struct. Mol. Biol. *24*, 444–452.

Boettiger, A.N., Bintu, B., Moffitt, J.R., Wang, S., Beliveau, B.J., Fudenberg, G., Imakaev, M., Mirny, L.A., Wu, C.T., and Zhuang, X. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. Nature *529*, 418–422.

Bonev, B., and Cavalli, G. (2016). Organization and function of the 3D genome. Nat. Rev. Genet. *17*, 661–678.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349–353.

Brookes, E., de Santiago, I., Hebenstreit, D., Morris, K.J., Carroll, T., Xie, S.Q., Stock, J.K., Heidemann, M., Eick, D., Nozaki, N., et al. (2012). Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. Cell Stem Cell *10*, 157–170.

Brooun, A., Gajiwala, K.S., Deng, Y.L., Liu, W., Bolaños, B., Bingham, P., He, Y.A., Diehl, W., Grable, N., Kung, P.P., et al. (2016). Polycomb repressive complex 2 structure with inhibitor reveals a mechanism of activation and drug resistance. Nat. Commun. *7*, 11384.

Buchenau, P., Hodgson, J., Strutt, H., and Arndt-Jovin, D.J. (1998). The distribution of polycomb-group proteins during cell division and development in Drosophila embryos: impact on models for silencing. J. Cell Biol. *141*, 469–481.

Cao, R., Tsukada, Y., and Zhang, Y. (2005). Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. Mol. Cell *20*, 845–854.

Cao, Q., Wang, X., Zhao, M., Yang, R., Malik, R., Qiao, Y., Poliakov, A., Yocum, A.K., Li, Y., Chen, W., et al. (2014). The central role of EED in the orchestration of polycomb group complexes. Nat. Commun. *5*, 3127.

Cavalli, G., and Paro, R. (1998). The Drosophila Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. Cell *93*, 505–518.

Cerase, A., Smeets, D., Tang, Y.A., Gdula, M., Kraus, F., Spivakov, M., Moindrot, B., Leleu, M., Tattermusch, A., Demmerle, J., et al. (2014). Spatial separation of Xist RNA and polycomb proteins revealed by superresolution microscopy. Proc. Natl. Acad. Sci. USA *111*, 2235–2240.

Chen, R., Kong, P., Zhang, F., Shu, Y.N., Nie, X., Dong, L.H., Lin, Y.L., Xie, X.L., Zhao, L.L., Zhang, X.J., and Han, M. (2017). EZH2-mediated α-actin

methylation needs lncRNA TUG1, and promotes the cortex cytoskeleton formation in VSMCs. Gene *616*, 52–57.

Chiacchiera, F., Rossi, A., Jammula, S., Piunti, A., Scelfo, A., Ordóñez-Morán, P., Huelsken, J., Koseki, H., and Pasini, D. (2016). Polycomb complex PRC1 preserves intestinal stem cell identity by sustaining Wnt/β-catenin transcriptional activity. Cell Stem Cell *18*, 91–103.

Ciabrelli, F., Comoglio, F., Fellous, S., Bonev, B., Ninova, M., Szabo, Q., Xuéreb, A., Klopp, C., Aravin, A., Paro, R., et al. (2017). Stable Polycomb-dependent transgenerational inheritance of chromatin states in Drosophila. Nat. Genet. *49*, 876–886.

Ciferri, C., Lander, G.C., Maiolica, A., Herzog, F., Aebersold, R., and Nogales, E. (2012). Molecular architecture of human polycomb repressive complex 2. eLife *1*, e00005.

Cifuentes-Rojas, C., Hernandez, A.J., Sarma, K., and Lee, J.T. (2014). Regulatory interactions between RNA and polycomb repressive complex 2. Mol. Cell *55*, 171–185.

Coleman, R.T., and Struhl, G. (2017). Causal role for inheritance of H3K27me3 in maintaining the OFF state of a Drosophila HOX gene. Science *356*, eaai8236.

Comet, I., Riising, E.M., Leblanc, B., and Helin, K. (2016). Maintaining cell identity: PRC2-mediated regulation of transcription and cancer. Nat. Rev. Cancer *16*, 803–810.

Cooper, S., Dienstbier, M., Hassan, R., Schermelleh, L., Sharif, J., Blackledge, N.P., De Marco, V., Elderkin, S., Koseki, H., Klose, R., et al. (2014). Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. Cell Rep. *7*, 1456–1470.

Cooper, S., Grijzenhout, A., Underwood, E., Ancelin, K., Zhang, T., Nesterova, T.B., Anil-Kirmizitas, B., Bassett, A., Kooistra, S.M., Agger, K., et al. (2016). Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. Nat. Commun. *7*, 13661.

Crevillén, P., Yang, H., Cui, X., Greeff, C., Trick, M., Qiu, Q., Cao, X., and Dean, C. (2014). Epigenetic reprogramming that prevents transgenerational inheritance of the vernalized state. Nature *515*, 587–590.

Dahl, J.A., Jung, I., Aanes, H., Greggains, G.D., Manaf, A., Lerdrup, M., Li, G., Kuan, S., Li, B., Lee, A.Y., et al. (2016). Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. Nature *537*, 548–552.

Davidovich, C., and Cech, T.R. (2015). The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2. RNA *21*, 2007–2022.

Dayel, M.J., Alegado, R.A., Fairclough, S.R., Levin, T.C., Nichols, S.A., McDonald, K., and King, N. (2011). Cell differentiation and morphogenesis in the colony-forming choanoflagellate Salpingoeca rosetta. Dev. Biol. *357*, 73–82.

Dejardin, J., and Cavalli, G. (2004). Chromatin inheritance upon Zeste-mediated Brahma recruitment at a minimal cellular memory module. EMBO J. *23*, 857–868.

Denholtz, M., Bonora, G., Chronis, C., Splinter, E., de Laat, W., Ernst, J., Pellegrini, M., and Plath, K. (2013). Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. Cell Stem Cell *13*, 602–616.

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. Development *141*, 526–537.

Dingwall, A.K., Beek, S.J., McCallum, C.M., Tamkun, J.W., Kalpana, G.V., Goff, S.P., and Scott, M.P. (1995). The Drosophila snr1 and brm proteins are related to yeast SWI/SNF proteins and are components of a large protein complex. Mol. Biol. Cell *6*, 777–791.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Djabali, M., Selleri, L., Parry, P., Bower, M., Young, B.D., and Evans, G.A. (1992). A trithorax-like gene is interrupted by chromosome 11q23 translocations in acute leukaemias. Nat. Genet. *2*, 113–118.

Dumesic, P.A., Homer, C.M., Moresco, J.J., Pack, L.R., Shanle, E.K., Coyle, S.M., Strahl, B.D., Fujimori, D.G., Yates, J.R., 3rd, and Madhani, H.D. (2015). Product binding enforces the genomic specificity of a yeast polycomb repressive complex. Cell *160*, 204–218.

Eissenberg, J.C., Lee, M.G., Schneider, J., Ilvarsonn, A., Shiekhattar, R., and Shilatifard, A. (2007). The trithorax-group gene in Drosophila little imaginal discs encodes a trimethylated histone H3 Lys4 demethylase. Nat. Struct. Mol. Biol. *14*, 344–346.

Enderle, D., Beisel, C., Stadler, M.B., Gerstung, M., Athri, P., and Paro, R. (2011). Polycomb preferentially targets stalled promoters of coding and noncoding transcripts. Genome Res. *21*, 216–226.

Endoh, M., Endo, T.A., Shinga, J., Hayashi, K., Farcas, A., Ma, K.W., Ito, S., Sharif, J., Endoh, T., Onaga, N., et al. (2017). PCGF6-PRC1 suppresses premature differentiation of mouse embryonic stem cells by regulating germ cell-related genes. eLife *6*, e27970.

Entrevan, M., Schuettengruber, B., and Cavalli, G. (2016). Regulation of genome architecture and function by Polycomb proteins. Trends Cell Biol. *26*, 511–525.

Erceg, J., Pakozdi, T., Marco-Ferreres, R., Ghavi-Helm, Y., Girardot, C., Bracken, A.P., and Furlong, E.E. (2017). Dual functionality of cis-regulatory elements as developmental enhancers and Polycomb response elements. Genes Dev. *31*, 590–602.

Erokhin, M., Elizar'ev, P., Parshikov, A., Schedl, P., Georgiev, P., and Chetverina, D. (2015). Transcriptional read-through is not sufficient to induce an epigenetic switch in the silencing activity of Polycomb response elements. Proc. Natl. Acad. Sci. USA *112*, 14930–14935.

Ezhkova, E., Pasolli, H.A., Parker, J.S., Stokes, N., Su, I.H., Hannon, G., Tarakhovsky, A., and Fuchs, E. (2009). Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells. Cell *136*, 1122–1135.

Farcas, A.M., Blackledge, N.P., Sudbery, I., Long, H.K., McGouran, J.F., Rose, N.R., Lee, S., Sims, D., Cerase, A., Sheahan, T.W., et al. (2012). KDM2B links the Polycomb repressive complex 1 (PRC1) to recognition of CpG islands. eLife *1*, e00205.

Fauvarque, M.O., and Dura, J.M. (1993). Polyhomeotic regulatory sequences induce developmental regulator-dependent variegation and targeted P-element insertions in Drosophila. Genes Dev. *7*, 1508–1520.

Ferrari, K.J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stützer, A., Fischle, W., Bonaldi, T., and Pasini, D. (2014). Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. Mol. Cell *53*, 49–62.

Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J., and van Steensel, B. (2010). Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell *143*, 212–224.

Fischle, W., Wang, Y., Jacobs, S.A., Kim, Y., Allis, C.D., and Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes Dev. *17*, 1870–1881.

Follmer, N.E., Wani, A.H., and Francis, N.J. (2012). A polycomb group protein is retained at specific sites on chromatin in mitosis. PLoS Genet. *8*, e1003135.

Fonseca, J.P., Steffen, P.A., Müller, S., Lu, J., Sawicka, A., Seiser, C., and Ringrose, L. (2012). In vivo Polycomb kinetics and mitotic chromatin binding distinguish stem cells from differentiated cells. Genes Dev. *26*, 857–871.

Francis, N.J., Follmer, N.E., Simon, M.D., Aghia, G., and Butler, J.D. (2009). Polycomb proteins remain bound to chromatin and DNA during DNA replication in vitro. Cell *137*, 110–122.

Frangini, A., Sjöberg, M., Roman-Trufero, M., Dharmalingam, G., Haberle, V., Bartke, T., Lenhard, B., Malumbres, M., Vidal, M., and Dillon, N. (2013). The aurora B kinase and the polycomb protein ring1B combine to regulate active promoters in quiescent lymphocytes. Mol. Cell *51*, 647–661.

Frey, F., Sheahan, T., Finkl, K., Stoehr, G., Mann, M., Benda, C., and Müller, J. (2016). Molecular basis of PRC1 targeting to Polycomb response elements by PhoRC. Genes Dev. *30*, 1116–1127.

Gambetta, M.C., and Müller, J. (2014). O-GlcNAcylation prevents aggregation of the Polycomb group repressor polyhomeotic. Dev. Cell *31*, 629–639.

Gambetta, M.C., Oktaba, K., and Müller, J. (2009). Essential role of the glycosyltransferase sxc/Ogt in polycomb repression. Science *325*, 93–96.

Gao, Z., Lee, P., Stafford, J.M., von Schimmelmann, M., Schaefer, A., and Reinberg, D. (2014). An AUTS2-Polycomb complex activates gene expression in the CNS. Nature *516*, 349–354.

Gaydos, L.J., Wang, W., and Strome, S. (2014). Gene repression. H3K27me and PRC2 transmit a memory of repression across generations and during development. Science *345*, 1515–1518.

Gonzalez, I., Mateos-Langerak, J., Thomas, A., Cheutin, T., and Cavalli, G. (2014). Identification of regulators of the three-dimensional polycomb organization by a microscopy-based genome-wide RNAi screen. Mol. Cell *54*, 485–499.

Grau, D.J., Chapman, B.A., Garlick, J.D., Borowsky, M., Francis, N.J., and Kingston, R.E. (2011). Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. Genes Dev. *25*, 2210–2221.

Grijzenhout, A., Godwin, J., Koseki, H., Gdula, M.R., Szumska, D., McGouran, J.F., Bhattacharya, S., Kessler, B.M., Brockdorff, N., and Cooper, S. (2016). Functional analysis of AEBP2, a PRC2 Polycomb protein, reveals a Trithorax phenotype in embryonic development and in ESCs. Development *143*, 2716–2723.

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

Grimaud, C., Nègre, N., and Cavalli, G. (2006b). From genetics to epigenetics: the tale of Polycomb group and trithorax group genes. Chromosome Res. *14*, 363–375.

Guo, W.J., Datta, S., Band, V., and Dimri, G.P. (2007). Mel-18, a polycomb group protein, regulates cell proliferation and senescence via transcriptional repression of Bmi-1 and c-Myc oncoproteins. Mol. Biol. Cell *18*, 536–546.

Hansen, K.H., Bracken, A.P., Pasini, D., Dietrich, N., Gehani, S.S., Monrad, A., Rappsilber, J., Lerdrup, M., and Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. Nat. Cell Biol. *10*, 1291–1300.

Hauri, S., Comoglio, F., Seimiya, M., Gerstung, M., Glatter, T., Hansen, K., Aebersold, R., Paro, R., Gstaiger, M., and Beisel, C. (2016). A high-density map for navigating the human Polycomb complexome. Cell Rep. *17*, 583–595.

Heard, E., and Martienssen, R.A. (2014). Transgenerational epigenetic inheritance: myths and mechanisms. Cell *157*, 95–109.

Helin, K., and Dhanak, D. (2013). Chromatin proteins and modifications as drug targets. Nature *502*, 480–488.

Herzog, V.A., Lempradl, A., Trupke, J., Okulski, H., Altmutter, C., Ruge, F., Boidol, B., Kubicek, S., Schmauss, G., Aumayr, K., et al. (2014). A strand-specific switch in noncoding transcription switches the function of a Polycomb/Trithorax response element. Nat. Genet. *46*, 973–981.

Hidalgo, I., Herrera-Merchan, A., Ligos, J.M., Carramolino, L., Nuñez, J., Martinez, F., Dominguez, O., Torres, M., and Gonzalez, S. (2012). Ezh1 is required for hematopoietic stem cell maintenance and prevents senescence-like cell cycle arrest. Cell Stem Cell *11*, 649–662.

Hodges, C., Kirkland, J.G., and Crabtree, G.R. (2016). The many roles of BAF (mSWI/SNF) and PBAF complexes in cancer. Cold Spring Harb. Perspect. Med. *6*, a026930.

Hollick, J.B. (2017). Paramutation and related phenomena in diverse species. Nat. Rev. Genet. *18*, 5–23.

Hu, D., Garruss, A.S., Gao, X., Morgan, M.A., Cook, M., Smith, E.R., and Shilatifard, A. (2013). The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1093–1097.

Hu, D., Gao, X., Cao, K., Morgan, M.A., Mas, G., Smith, E.R., Volk, A.G., Bartom, E.T., Crispino, J.D., Di Croce, L., et al. (2017). Not all H3K4 methylations are created equal: Mll2/COMPASS dependency in primordial germ cell specification. Mol. Cell *65*, 460–475.

Illingworth, R.S., Moffat, M., Mann, A.R., Read, D., Hunter, C.J., Pradeepa, M.M., Adams, I.R., and Bickmore, W.A. (2015). The E3 ubiquitin ligase activity of RING1B is not essential for early mouse development. Genes Dev. *29*, 1897–1902.

Ingham, P.W. (1983). Differential expression of bithorax complex genes in the absence of the extra sex combs and trithorax genes. Nature *306*, 591–593.

Ingham, P.W. (1985a). A clonal analysis of the requirement for the trithorax gene in the diversification of segments in Drosophila. J. Embryol. Exp. Morphol. *89*, 349–365.

Ingham, P.W. (1985b). Genetic control of the spatial pattern of selector gene expression in Drosophila. Cold Spring Harb. Symp. Quant. Biol. *50*, 201–208.

Isono, K., Endo, T.A., Ku, M., Yamada, D., Suzuki, R., Sharif, J., Ishikura, T., Toyoda, T., Bernstein, B.E., and Koseki, H. (2013). SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. Dev. Cell *26*, 565–577.

Jacobs, J.J., Kieboom, K., Marino, S., DePinho, R.A., and van Lohuizen, M. (1999a). The oncogene and Polycomb-group gene bmi-1 regulates cell proliferation and senescence through the ink4a locus. Nature *397*, 164–168.

Jacobs, J.J., Scheijen, B., Voncken, J.W., Kieboom, K., Berns, A., and van Lohuizen, M. (1999b). Bmi-1 collaborates with c-Myc in tumorigenesis by inhibiting c-Myc-induced apoptosis via INK4a/ARF. Genes Dev. *13*, 2678–2690.

Jamieson, K., Rountree, M.R., Lewis, Z.A., Stajich, J.E., and Selker, E.U. (2013). Regional control of histone H3 lysine 27 methylation in Neurospora. Proc. Natl. Acad. Sci. USA *110*, 6027–6032.

Javierre, B.M., Burren, O.S., Wilder, S.P., Kreuzhuber, R., Hill, S.M., Sewitz, S., Cairns, J., Wingett, S.W., Varnai, C., Thiecke, M.J., et al. (2016). Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. Cell *167*, 1369–1384.

Jiao, L., and Liu, X. (2015). Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. Science *350*, aac4383.

Joshi, O., Wang, S.Y., Kuznetsova, T., Atlasi, Y., Peng, T., Fabre, P.J., Habibi, E., Shaik, J., Saeed, S., Handoko, L., et al. (2015). Dynamic reorganization of extremely long-range promoter-promoter interactions between two states of pluripotency. Cell Stem Cell *17*, 748–757.

Jozwik, K.M., Chernukhin, I., Serandour, A.A., Nagarajan, S., and Carroll, J.S. (2016). FOXA1 directs H3K4 monomethylation at enhancers via recruitment of the methyltransferase MLL3. Cell Rep. *17*, 2715–2723.

Justin, N., Zhang, Y., Tarricone, C., Martin, S.R., Chen, S., Underwood, E., De Marco, V., Haire, L.F., Walker, P.A., Reinberg, D., et al. (2016). Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2. Nat. Commun. *7*, 11316.

Kadoch, C., and Crabtree, G.R. (2015). Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. Sci. Adv. *1*, e1500447.

Kadoch, C., Williams, R.T., Calarco, J.P., Miller, E.L., Weber, C.M., Braun, S.M., Pulice, J.L., Chory, E.J., and Crabtree, G.R. (2017). Dynamics of BAF-Polycomb complex opposition on heterochromatin in normal and oncogenic states. Nat. Genet. *49*, 213–222.

Kahn, T.G., Stenberg, P., Pirrotta, V., and Schwartz, Y.B. (2014). Combinatorial interactions are required for the efficient recruitment of pho repressive complex (PhoRC) to polycomb response elements. PLoS Genet. *10*, e1004495.

Kahn, T.G., Dorafshan, E., Schultheis, D., Zare, A., Stenberg, P., Reim, I., Pirrotta, V., and Schwartz, Y.B. (2016). Interdependence of PRC1 and PRC2 for recruitment to Polycomb response elements. Nucleic Acids Res. *44*, 10132–10149.

Kalb, R., Latwiel, S., Baymaz, H.I., Jansen, P.W., Müller, C.W., Vermeulen, M., and Müller, J. (2014). Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. Nat. Struct. Mol. Biol. *21*, 569–571.

Kaneko, S., Son, J., Shen, S.S., Reinberg, D., and Bonasio, R. (2013). PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1258–1264.

Kang, H., McElroy, K.A., Jung, Y.L., Alekseyenko, A.A., Zee, B.M., Park, P.J., and Kuroda, M.I. (2015). Sex comb on midleg (Scm) is a functional link between PcG-repressive complexes in Drosophila. Genes Dev. *29*, 1136–1150.

Kanhere, A., Viiri, K., Araújo, C.C., Rasaiyaah, J., Bouwman, R.D., Whyte, W.A., Pereira, C.F., Brookes, E., Walker, K., Bell, G.W., et al. (2010). Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. Mol. Cell *38*, 675–688.

Kar, G., Kim, J.K., Kolodziejczyk, A.A., Natarajan, K.N., Torlai Triglia, E., Mifsud, B., Elderkin, S., Marioni, J.C., Pombo, A., and Teichmann, S.A. (2017). Flipping between Polycomb repressed and active transcriptional states introduces noise in gene expression. Nat. Commun. *8*, 36.

Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia mutations in Drosophila. Proc. Natl. Acad. Sci. USA *85*, 8136–8140.

Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J., Larschan, E., Gorchakov, A.A., Gu, T., et al. (2011). Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature *471*, 480–485.

Khavari, P.A., Peterson, C.L., Tamkun, J.W., Mendel, D.B., and Crabtree, G.R. (1993). BRG1 contains a conserved domain of the SWI2/SNF2 family necessary for normal mitotic growth and transcription. Nature *366*, 170–174.

Klauke, K., Radulović, V., Broekhuis, M., Weersing, E., Zwart, E., Olthof, S., Ritsema, M., Bruggeman, S., Wu, X., Helin, K., et al. (2013). Polycomb Cbx family members mediate the balance between haematopoietic stem cell self-renewal and differentiation. Nat. Cell Biol. *15*, 353–362.

Kloet, S.L., Makowski, M.M., Baymaz, H.I., van Voorthuijsen, L., Karemaker, I.D., Santanach, A., Jansen, P.W.T.C., Di Croce, L., and Vermeulen, M. (2016). The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. Nat. Struct. Mol. Biol. *23*, 682–690.

Klose, R.J., Cooper, S., Farcas, A.M., Blackledge, N.P., and Brockdorff, N. (2013). Chromatin sampling–an emerging perspective on targeting polycomb repressor proteins. PLoS Genet. *9*, e1003717.

Klymenko, T., and Müller, J. (2004). The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. EMBO Rep. *5*, 373–377.

Klymenko, T., Papp, B., Fischle, W., Köcher, T., Schelder, M., Fritsch, C., Wild, B., Wilm, M., and Müller, J. (2006). A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. Genes Dev. *20*, 1110–1122.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet. *4*, e1000242.

Kundu, S., Ji, F., Sunwoo, H., Jain, G., Lee, J.T., Sadreyev, R.I., Dekker, J., and Kingston, R.E. (2017). Polycomb repressive complex 1 generates discrete compacted domains that change during differentiation. Mol. Cell *65*, 432–446.

Lagarou, A., Mohd-Sarip, A., Moshkin, Y.M., Chalkley, G.E., Bezstarosti, K., Demmers, J.A., and Verrijzer, C.P. (2008). dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. Genes Dev. *22*, 2799–2810.

Lanzuolo, C., Roure, V., Dekker, J., Bantignies, F., and Orlando, V. (2007). Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. Nat. Cell Biol. *9*, 1167–1174.

Laprell, F., Finkl, K., and Müller, J. (2017). Propagation of Polycomb-repressed chromatin requires sequence-specific recruitment to DNA. Science *356*, 85–88.

Lau, M.S., Schwartz, M.G., Kundu, S., Savol, A.J., Wang, P.I., Marr, S.K., Grau, D.J., Schorderet, P., Sadreyev, R.I., Tabin, C.J., and Kingston, R.E. (2017). Mutation of a nucleosome compaction region disrupts Polycomb-mediated axial patterning. Science *355*, 1081–1084.

Lee, M.G., Villa, R., Trojer, P., Norman, J., Yan, K.P., Reinberg, D., Di Croce, L., and Shiekhattar, R. (2007a). Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. Science *318*, 447–450.

Lee, N., Zhang, J., Klose, R.J., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2007b). The trithorax-group protein Lid is a histone H3 trimethyl-Lys4 demethylase. Nat. Struct. Mol. Biol. *14*, 341–343.

Lee, W., Teckie, S., Wiesner, T., Ran, L., Prieto Granada, C.N., Lin, M., Zhu, S., Cao, Z., Liang, Y., Sboner, A., et al. (2014). PRC2 is recurrently inactivated through EED or SUZ12 loss in malignant peripheral nerve sheath tumors. Nat. Genet. *46*, 1227–1232.

Lee, H.G., Kahn, T.G., Simcox, A., Schwartz, Y.B., and Pirrotta, V. (2015). Genome-wide activities of Polycomb complexes control pervasive transcription. Genome Res. *25*, 1170–1181.

Lengsfeld, B.M., Berry, K.N., Ghosh, S., Takahashi, M., and Francis, N.J. (2012). A Polycomb complex remains bound through DNA replication in the absence of other eukaryotic proteins. Sci. Rep. *2*, 661.

Lewis, P.H. (1947). New mutants report. Drosoph. Inf. Serv. *21*, 69.

Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. Nature *276*, 565–570.

Lewis, P.W., Müller, M.M., Koletsky, M.S., Cordero, F., Lin, S., Banaszynski, L.A., Garcia, B.A., Muir, T.W., Becher, O.J., and Allis, C.D. (2013). Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. Science *340*, 857–861.

Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B.E., and Reinberg, D. (2010). Jarid2 and PRC2, partners in regulating gene expression. Genes Dev. *24*, 368–380.

Liefke, R., Karwacki-Neisius, V., and Shi, Y. (2016). EPOP interacts with Elongin BC and USP7 to modulate the chromatin landscape. Mol. Cell *64*, 659–672.

Liu, J., Cao, L., Chen, J., Song, S., Lee, I.H., Quijano, C., Liu, H., Keyvanfar, K., Chen, H., Cao, L.Y., et al. (2009). Bmi1 regulates mitochondrial function and the DNA damage response pathway. Nature *459*, 387–392.

Liu, X., Wang, C., Liu, W., Li, J., Li, C., Kou, X., Chen, J., Zhao, Y., Gao, H., Wang, H., et al. (2016). Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. Nature *537*, 558–562.

Lo, S.M., Follmer, N.E., Lengsfeld, B.M., Madamba, E.V., Seong, S., Grau, D.J., and Francis, N.J. (2012). A bridging model for persistence of a polycomb group protein complex through DNA replication in vitro. Mol. Cell *46*, 784–796.

Loubière, V., Delest, A., Thomas, A., Bonev, B., Schuettengruber, B., Sati, S., Martinez, A.M., and Cavalli, G. (2016). Coordinate redeployment of PRC1 proteins suppresses tumor formation during Drosophila development. Nat. Genet. *48*, 1436–1442.

Lv, X., Han, Z., Chen, H., Yang, B., Yang, X., Xia, Y., Pan, C., Fu, L., Zhang, S., Han, H., et al. (2016). A positive role for polycomb in transcriptional regulation via H4K20me1. Cell Res. *26*, 529–542.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343–349.

Margueron, R., Li, G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C.L., Dynlacht, B.D., and Reinberg, D. (2008). Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. Mol. Cell *32*, 503–518.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., 3rd, Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V., et al. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762–767.

McCabe, M.T., Graves, A.P., Ganji, G., Diaz, E., Halsey, W.S., Jiang, Y., Smitheman, K.N., Ott, H.M., Pappalardi, M.B., Allen, K.E., et al. (2012). Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27). Proc. Natl. Acad. Sci. USA *109*, 2989–2994.

McGinty, R.K., Henrici, R.C., and Tan, S. (2014). Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. Nature *514*, 591–596.

Mendenhall, E.M., Koche, R.P., Truong, T., Zhou, V.W., Issac, B., Chi, A.S., Ku, M., and Bernstein, B.E. (2010). GC-rich sequence elements recruit PRC2 in mammalian ES cells. PLoS Genet. *6*, e1001244.

Milne, T.A., Kim, J., Wang, G.G., Stadler, S.C., Basrur, V., Whitcomb, S.J., Wang, Z., Ruthenburg, A.J., Elenitoba-Johnson, K.S., Roeder, R.G., and Allis,

C.D. (2010). Multiple interactions recruit MLL1 and MLL1 fusion proteins to the HOXA9 locus in leukemogenesis. Mol. Cell *38*, 853–863.

Min, J., Zhang, Y., and Xu, R.M. (2003). Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. Genes Dev. *17*, 1823–1828.

Minoux, M., Holwerda, S., Vitobello, A., Kitazawa, T., Kohler, H., Stadler, M.B., and Rijli, F.M. (2017). Gene bivalency at Polycomb domains regulates cranial neural crest positional identity. Science *355*, eaal2913.

Montgomery, N.D., Yee, D., Chen, A., Kalantry, S., Chamberlain, S.J., Otte, A.P., and Magnuson, T. (2005). The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation. Curr. Biol. *15*, 942–947.

Morey, L., Pascual, G., Cozzuto, L., Roma, G., Wutz, A., Benitah, S.A., and Di Croce, L. (2012). Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. Cell Stem Cell *10*, 47–62.

Morey, L., Santanach, A., Blanco, E., Aloia, L., Nora, E.P., Bruneau, B.G., and Di Croce, L. (2015). Polycomb regulates mesoderm cell fate-specification in embryonic stem cells through activation and repression mechanisms. Cell Stem Cell *17*, 300–315.

Morin, R.D., Johnson, N.A., Severson, T.M., Mungall, A.J., An, J., Goya, R., Paul, J.E., Boyle, M., Woolcock, B.W., Kuchenbauer, F., et al. (2010). Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. Nat. Genet. *42*, 181–185.

Mousavi, K., Zare, H., Wang, A.H., and Sartorelli, V. (2012). Polycomb protein Ezh1 promotes RNA polymerase II elongation. Mol. Cell *45*, 255–262.

Mozzetta, C., Pontis, J., Fritsch, L., Robin, P., Portoso, M., Proux, C., Margueron, R., and Ait-Si-Ali, S. (2014). The histone H3 lysine 9 methyltransferases G9a and GLP regulate polycomb repressive complex 2-mediated gene silencing. Mol. Cell *53*, 277–289.

Müller, J., and Bienz, M. (1991). Long range repression conferring boundaries of Ultrabithorax expression in the Drosophila embryo. EMBO J. *10*, 3147–3155.

Muntean, A.G., Tan, J., Sitwala, K., Huang, Y., Bronstein, J., Connelly, J.A., Basrur, V., Elenitoba-Johnson, K.S., and Hess, J.L. (2010). The PAF complex synergizes with MLL fusion proteins at HOX loci to promote leukemogenesis. Cancer Cell *17*, 609–621.

Musselman, C.A., Gibson, M.D., Hartwick, E.W., North, J.A., Gatchalian, J., Poirier, M.G., and Kutateladze, T.G. (2013). Binding of PHF1 Tudor to H3K36me3 enhances nucleosome accessibility. Nat. Commun. *4*, 2969.

Nakagawa, T., Kajitani, T., Togo, S., Masuko, N., Ohdan, H., Hishikawa, Y., Koji, T., Matsuyama, T., Ikura, T., Muramatsu, M., and Ito, T. (2008). Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation. Genes Dev. *22*, 37–49.

Narendra, V., Rocha, P.P., An, D., Raviram, R., Skok, J.A., Mazzoni, E.O., and Reinberg, D. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science *347*, 1017–1021.

Narendra, V., Bulajić, M., Dekker, J., Mazzoni, E.O., and Reinberg, D. (2016). CTCF-mediated topological boundaries during development foster appropriate gene regulation. Genes Dev. *30*, 2657–2662.

Nekrasov, M., Klymenko, T., Fraterman, S., Papp, B., Oktaba, K., Köcher, T., Cohen, A., Stunnenberg, H.G., Wilm, M., and Müller, J. (2007). Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. EMBO J. *26*, 4078–4088.

Niessen, H.E., Demmers, J.A., and Voncken, J.W. (2009). Talking to chromatin: post-translational modulation of polycomb group function. Epigenetics Chromatin *2*, 10.

Noordermeer, D., Leleu, M., Splinter, E., Rougemont, J., De Laat, W., and Duboule, D. (2011). The dynamic architecture of Hox gene clusters. Science *334*, 222–225.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Oliviero, G., Brien, G.L., Waston, A., Streubel, G., Jerman, E., Andrews, D., Doyle, B., Munawar, N., Wynne, K., Crean, J., et al. (2016). Dynamic protein interactions of the Polycomb repressive complex 2 during differentiation of pluripotent cells. Mol. Cell. Proteomics *15*, 3450–3460.

Papp, B., and Müller, J. (2006). Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. Genes Dev. *20*, 2041–2054.

Pasini, D., Cloos, P.A., Walfridsson, J., Olsson, L., Bukowski, J.P., Johansen, J.V., Bak, M., Tommerup, N., Rappsilber, J., and Helin, K. (2010). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature *464*, 306–310.

Pengelly, A.R., Copur, Ö., Jäckle, H., Herzig, A., and Müller, J. (2013). A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. Science *339*, 698–699.

Pengelly, A.R., Kalb, R., Finkl, K., and Müller, J. (2015). Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation. Genes Dev. *29*, 1487–1492.

Petruk, S., Sedkov, Y., Johnston, D.M., Hodgson, J.W., Black, K.L., Kovermann, S.K., Beck, S., Canaani, E., Brock, H.W., and Mazo, A. (2012). TrxG and PcG proteins but not methylated histones remain associated with DNA through replication. Cell *150*, 922–933.

Piunti, A., and Shilatifard, A. (2016). Epigenetic balance of gene expression by Polycomb and COMPASS families. Science *352*, aad9780.

Piunti, A., Hashizume, R., Morgan, M.A., Bartom, E.T., Horbinski, C.M., Marshall, S.A., Rendleman, E.J., Ma, Q., Takahashi, Y.H., Woodfin, A.R., et al. (2017). Therapeutic targeting of polycomb and BET bromodomain proteins in diffuse intrinsic pontine gliomas. Nat. Med. *23*, 493–500.

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. Science *300*, 131–135.

Portoso, M., Ragazzini, R., Brenčič, Ž., Moiani, A., Michaud, A., Vassilev, I., Wassef, M., Servant, N., Sargueil, B., and Margueron, R. (2017). PRC2 is dispensable for HOTAIR-mediated transcriptional repression. EMBO J. *36*, 981–994.

Posfai, E., Kunzmann, R., Brochard, V., Salvaing, J., Cabuy, E., Roloff, T.C., Liu, Z., Tardat, M., van Lohuizen, M., Vidal, M., et al. (2012). Polycomb function during oogenesis is required for mouse embryonic development. Genes Dev. *26*, 920–932.

Poux, S., Horard, B., Sigrist, C.J., and Pirrotta, V. (2002). The Drosophila trithorax protein is a coactivator required to prevent re-establishment of polycomb silencing. Development *129*, 2483–2493.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Rickels, R., Hu, D., Collings, C.K., Woodfin, A.R., Piunti, A., Mohan, M., Herz, H.M., Kvon, E., and Shilatifard, A. (2016). An evolutionary conserved epigenetic mark of Polycomb response elements implemented by Trx/MLL/COMPASS. Mol. Cell *63*, 318–328.

Riising, E.M., Comet, I., Leblanc, B., Wu, X., Johansen, J.V., and Helin, K. (2014). Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. Mol. Cell *55*, 347–360.

Risca, V.I., Denny, S.K., Straight, A.F., and Greenleaf, W.J. (2017). Variable chromatin structure revealed by in situ spatially correlated DNA cleavage mapping. Nature *541*, 237–241.

Robinson, A.K., Leal, B.Z., Chadwell, L.V., Wang, R., Ilangovan, U., Kaur, Y., Junco, S.E., Schirf, V., Osmulski, P.A., Gaczynska, M., et al. (2012). The growth-suppressive function of the polycomb group protein polyhomeotic is mediated by polymerization of its sterile alpha motif (SAM) domain. J. Biol. Chem. *287*, 8702–8713.

Rosa, S., De Lucia, F., Mylne, J.S., Zhu, D., Ohmido, N., Pendle, A., Kato, N., Shaw, P., and Dean, C. (2013). Physical clustering of FLC alleles during

Polycomb-mediated epigenetic silencing in vernalization. Genes Dev. *27*, 1845–1850.

Rose, N.R., King, H.W., Blackledge, N.P., Fursova, N.A., Ember, K.J., Fischer, R., Kessler, B.M., and Klose, R.J. (2016). RYBP stimulates PRC1 to shape chromatin-based communication between Polycomb repressive complexes. eLife *5*, e18591.

Sarma, K., Margueron, R., Ivanov, A., Pirrotta, V., and Reinberg, D. (2008). Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo. Mol. Cell. Biol. *28*, 2718–2731.

Saurin, A.J., Shiels, C., Williamson, J., Satijn, D.P., Otte, A.P., Sheer, D., and Freemont, P.S. (1998). The human polycomb group complex associates with pericentromeric heterochromatin to form a novel nuclear domain. J. Cell Biol. *142*, 887–898.

Schaaf, C.A., Misulovin, Z., Gause, M., Koenig, A., Gohara, D.W., Watson, A., and Dorsett, D. (2013). Cohesin and polycomb proteins functionally interact to control transcription at silenced and active genes. PLoS Genet. *9*, e1003560.

Scheuermann, J.C., de Ayala Alonso, A.G., Oktaba, K., Ly-Hartig, N., McGinty, R.K., Fraterman, S., Wilm, M., Muir, T.W., and Müller, J. (2010). Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. Nature *465*, 243–247.

Schmitges, F.W., Prusty, A.B., Faty, M., Stützer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., et al. (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. Mol. Cell *42*, 330–341.

Schoenfelder, S., Sugar, R., Dimond, A., Javierre, B.M., Armstrong, H., Mifsud, B., Dimitrova, E., Matheson, L., Tavares-Cadete, F., Furlan-Magaril, M., et al. (2015). Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. Nat. Genet. *47*, 1179–1186.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. Cell *128*, 735–745.

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.

Schuettengruber, B., Oded Elkayam, N., Sexton, T., Entrevan, M., Stern, S., Thomas, A., Yaffe, E., Parrinello, H., Tanay, A., and Cavalli, G. (2014). Cooperativity, specificity, and evolutionary stability of Polycomb targeting in Drosophila. Cell Rep. *9*, 219–233.

Schwartz, Y.B., and Cavalli, G. (2017). Three-dimensional genome organization and function in Drosophila. Genetics *205*, 5–24.

Schwartz, Y.B., and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. Nat. Rev. Genet. *8*, 9–22.

Schwartz, Y.B., Kahn, T.G., Nix, D.A., Li, X.Y., Bourgon, R., Biggin, M., and Pirrotta, V. (2006). Genome-wide analysis of Polycomb targets in Drosophila melanogaster. Nat. Genet. *38*, 700–705.

Sebé-Pedrós, A., Ballaré, C., Parra-Acero, H., Chiva, C., Tena, J.J., Sabidó, E., Gómez-Skarmeta, J.L., Di Croce, L., and Ruiz-Trillo, I. (2016). The dynamic regulatory genome of Capsaspora and the origin of animal multicellularity. Cell *165*, 1224–1237.

Secombe, J., Li, L., Carlos, L., and Eisenman, R.N. (2007). The Trithorax group protein Lid is a trimethyl histone H3K4 demethylase required for dMyc-induced cell growth. Genes Dev. *21*, 537–551.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458–472.

Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J.R., Wu, C.T., Bender, W., and Kingston, R.E. (1999). Stabilization of chromatin structure by PRC1, a Polycomb complex. Cell *98*, 37–46.

Sharma, A. (2017). Transgenerational epigenetics: Integrating soma to germline communication with gametic inheritance. Mech. Ageing Dev. *163*, 15–22.

Sharma, U., Conine, C.C., Shea, J.M., Boskovic, A., Derr, A.G., Bing, X.Y., Belleannee, C., Kucukural, A., Serra, R.W., Sun, F., et al. (2016). Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. Science *351*, 391–396.

Shema, E., Jones, D., Shoresh, N., Donohue, L., Ram, O., and Bernstein, B.E. (2016). Single-molecule decoding of combinatorially modified nucleosomes. Science *352*, 717–721.

Shen, X., Liu, Y., Hsu, Y.J., Fujiwara, Y., Kim, J., Mao, X., Yuan, G.C., and Orkin, S.H. (2008). EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. Mol. Cell *32*, 491–502.

Shen, X., Kim, W., Fujiwara, Y., Simon, M.D., Liu, Y., Mysliwiec, M.R., Yuan, G.C., Lee, Y., and Orkin, S.H. (2009). Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. Cell *139*, 1303–1314.

Siklenka, K., Erkek, S., Godmann, M., Lambrot, R., McGraw, S., Lafleur, C., Cohen, T., Xia, J., Suderman, M., Hallett, M., et al. (2015). Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. Science *350*, aab2006.

Simon, J., Chiang, A., Bender, W., Shimell, M.J., and O'Connor, M. (1993). Elements of the Drosophila bithorax complex that mediate repression by Polycomb group products. Dev. Biol. *158*, 131–144.

Singh, N.P., and Mishra, R.K. (2015). Specific combinations of boundary element and Polycomb response element are required for the regulation of the Hox genes in Drosophila melanogaster. Mech. Dev. *138*, 141–150.

Smallwood, S.A., and Kelsey, G. (2012). De novo DNA methylation: a germ cell perspective. Trends Genet. *28*, 33–42.

Smith, E., Lin, C., and Shilatifard, A. (2011). The super elongation complex (SEC) and MLL in development and disease. Genes Dev. *25*, 661–672.

Stanton, B.Z., Hodges, C., Calarco, J.P., Braun, S.M., Ku, W.L., Kadoch, C., Zhao, K., and Crabtree, G.R. (2017). Smarca4 ATPase mutations disrupt direct eviction of PRC1 from chromatin. Nat. Genet. *49*, 282–288.

Steffen, P.A., Fonseca, J.P., Gänger, C., Dworschak, E., Kockmann, T., Beisel, C., and Ringrose, L. (2013). Quantitative in vivo analysis of chromatin binding of Polycomb and Trithorax group proteins reveals retention of ASH1 on mitotic chromatin. Nucleic Acids Res. *41*, 5235–5250.

Stock, J.K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A.G., and Pombo, A. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nat. Cell Biol. *9*, 1428–1435.

Struhl, G., and Akam, M. (1985). Altered distributions of Ultrabithorax transcripts in extra sex combs mutant embryos of Drosophila. EMBO J. *4*, 3259–3264.

Surani, M.A. (2015). Human germline: a new research frontier. Stem Cell Reports *4*, 955–960.

Tamkun, J.W., Deuring, R., Scott, M.P., Kissinger, M., Pattatucci, A.M., Kaufman, T.C., and Kennison, J.A. (1992). brahma: a regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. Cell *68*, 561–572.

Tanay, A., O'Donnell, A.H., Damelin, M., and Bestor, T.H. (2007). Hyperconserved CpG domains underlie Polycomb-binding sites. Proc. Natl. Acad. Sci. USA *104*, 5521–5526.

Teperek, M., Simeone, A., Gaggioli, V., Miyamoto, K., Allen, G.E., Erkek, S., Kwon, T., Marcotte, E.M., Zegerman, P., Bradshaw, C.R., et al. (2016). Sperm is epigenetically programmed to regulate gene transcription in embryos. Genome Res. *26*, 1034–1046.

Terranova, R., Yokobayashi, S., Stadler, M.B., Otte, A.P., van Lohuizen, M., Orkin, S.H., and Peters, A.H. (2008). Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. Dev. Cell *15*, 668–679.

Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R., 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.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82.

Tie, F., Banerjee, R., Stratton, C.A., Prasad-Sinha, J., Stepanik, V., Zlobin, A., Diaz, M.O., Scacheri, P.C., and Harte, P.J. (2009). CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. Development *136*, 3131–3141.

Tie, F., Banerjee, R., Fu, C., Stratton, C.A., Fang, M., and Harte, P.J. (2016). Polycomb inhibits histone acetylation by CBP by binding directly to its catalytic domain. Proc. Natl. Acad. Sci. USA *113*, E744–E753.

Tolhuis, B., de Wit, E., Muijrers, I., Teunissen, H., Talhout, W., van Steensel, B., and van Lohuizen, M. (2006). Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in Drosophila melanogaster. Nat. Genet. *38*, 694–699.

van den Boom, V., Maat, H., Geugien, M., Rodríguez López, A., Sotoca, A.M., Jaques, J., Brouwers-Vos, A.Z., Fusetti, F., Groen, R.W., Yuan, H., et al. (2016). Non-canonical PRC1.1 targets active genes independent of H3K27me3 and is essential for leukemogenesis. Cell Rep. *14*, 332–346.

van Lohuizen, M., Frasch, M., Wientjens, E., and Berns, A. (1991). Sequence similarity between the mammalian bmi-1 proto-oncogene and the Drosophila regulatory genes Psc and Su(z)2. Nature *353*, 353–355.

Varambally, S., Dhanasekaran, S.M., Zhou, M., Barrette, T.R., Kumar-Sinha, C., Sanda, M.G., Ghosh, D., Pienta, K.J., Sewalt, R.G., Otte, A.P., et al. (2002). The polycomb group protein EZH2 is involved in progression of prostate cancer. Nature *419*, 624–629.

Victor, M.B., Richner, M., Hermanstyne, T.O., Ransdell, J.L., Sobieski, C., Deng, P.Y., Klyachko, V.A., Nerbonne, J.M., and Yoo, A.S. (2014). Generation of human striatal neurons by microRNA-dependent direct conversion of fibroblasts. Neuron *84*, 311–323.

Vieux-Rochas, M., Fabre, P.J., Leleu, M., Duboule, D., and Noordermeer, D. (2015). Clustering of mammalian Hox genes with other H3K27me3 targets within an active nuclear domain. Proc. Natl. Acad. Sci. USA *112*, 4672–4677.

Wang, L., Brown, J.L., Cao, R., Zhang, Y., Kassis, J.A., and Jones, R.S. (2004). Hierarchical recruitment of polycomb group silencing complexes. Mol. Cell *14*, 637–646.

Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A., et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature *472*, 120–124.

Wani, A.H., Boettiger, A.N., Schorderet, P., Ergun, A., Münger, C., Sadreyev, R.I., Zhuang, X., Kingston, R.E., and Francis, N.J. (2016). Chromatin topology is coupled to Polycomb group protein subnuclear organization. Nat. Commun. *7*, 10291.

Wijchers, P.J., Krijger, P.H., Geeven, G., Zhu, Y., Denker, A., Verstegen, M.J., Valdes-Quezada, C., Vermeulen, C., Janssen, M., Teunissen, H., et al. (2016). Cause and consequence of tethering a SubTAD to different nuclear compartments. Mol. Cell *61*, 461–473.

Wilkinson, F., Pratt, H., and Atchison, M.L. (2010). PcG recruitment by the YY1 REPO domain can be mediated by Yaf2. J. Cell. Biochem. *109*, 478–486.

Xie, H., Xu, J., Hsu, J.H., Nguyen, M., Fujiwara, Y., Peng, C., and Orkin, S.H. (2014). Polycomb repressive complex 2 regulates normal hematopoietic stem cell function in a developmental-stage-specific manner. Cell Stem Cell *14*, 68–80.

Yang, Y.W., Flynn, R.A., Chen, Y., Qu, K., Wan, B., Wang, K.C., Lei, M., and Chang, H.Y. (2014). Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency. eLife *3*, e02046.

Yoo, A.S., Staahl, B.T., Chen, L., and Crabtree, G.R. (2009). MicroRNA-mediated switching of chromatin-remodelling complexes in neural development. Nature *460*, 642–646.

Yoo, A.S., Sun, A.X., Li, L., Shcheglovitov, A., Portmann, T., Li, Y., Lee-Messer, C., Dolmetsch, R.E., Tsien, R.W., and Crabtree, G.R. (2011). MicroRNA-mediated conversion of human fibroblasts to neurons. Nature *476*, 228–231.

Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., and Zhu, B. (2011). H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. J. Biol. Chem. *286*, 7983–7989.

Yuan, W., Wu, T., Fu, H., Dai, C., Wu, H., Liu, N., Li, X., Xu, M., Zhang, Z., Niu, T., et al. (2012). Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation. Science *337*, 971–975.

Zdzieblo, D., Li, X., Lin, Q., Zenke, M., Illich, D.J., Becker, M., and Müller, A.M. (2014). Pcgf6, a polycomb group protein, regulates mesodermal lineage differentiation in murine ESCs and functions in iPS reprogramming. Stem Cells *32*, 3112–3125.

Zenk, F., Loeser, E., Schiavo, R., Kilpert, F., Bogdanović, O., and Iovino, N. (2017). Germ line-inherited H3K27me3 restricts enhancer function during maternal-to-zygotic transition. Science *357*, 212–216.

Zhang, B., Zheng, H., Huang, B., Li, W., Xiang, Y., Peng, X., Ming, J., Wu, X., Zhang, Y., Xu, Q., et al. (2016). Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. Nature *537*, 553–557.

Zheng, H., Huang, B., Zhang, B., Xiang, Y., Du, Z., Xu, Q., Li, Y., Wang, Q., Ma, J., Peng, X., et al. (2016). Resetting epigenetic memory by reprogramming of histone modifications in mammals. Mol. Cell *63*, 1066–1079.

Zhou, W., Zhu, P., Wang, J., Pascual, G., Ohgi, K.A., Lozach, J., Glass, C.K., and Rosenfeld, M.G. (2008). Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. Mol. Cell *29*, 69–80.

Zhu, L., Li, Q., Wong, S.H., Huang, M., Klein, B.J., Shen, J., Ikenouye, L., Onishi, M., Schneidawind, D., Buechele, C., et al. (2016). ASH1L links histone H3 lysine 36 dimethylation to MLL leukemia. Cancer Discov. *6*, 770–783.

Ziemin-van der Poel, S., McCabe, N.R., Gill, H.J., Espinosa, R., 3rd, Patel, Y., Harden, A., Rubinelli, P., Smith, S.D., LeBeau, M.M., Rowley, J.D., et al. (1991). Identification of a gene, MLL, that spans the breakpoint in 11q23 translocations associated with human leukemias. Proc. Natl. Acad. Sci. USA *88*, 10735–10739.

Zinzalla, G. (2016). A new way forward in cancer drug discovery: inhibiting the SWI/SNF chromatin remodelling complex. ChemBioChem *17*, 677–682.