Spring 2021 – Epigenetics and Systems Biology
Discussion Session (Epigenetics)
Michael K. Skinner – Biol 476/576
Week 5 (February 18)

Epigenetics (History / Molecular Processes/ Genomics)

Primary Papers

1. Singer J, et al. (1977) J Biol Chem. 10;252(15):5509-13. (Classic) (PMID: 195953)
2. Holliday and Pugh (1975) Science 24;187(4173):226-32. (PMID: 1111098)
3. Wen, et al. (2012) BMC Genomics 13:566. (PMID: 23102236)
4. Kempfer and Pombo (2020) Nat Rev Genet. Apr;21(4):207-226. (PMID: 31848476)

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 are the different methods for chromatin structure mapping?
- What are the functions 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.

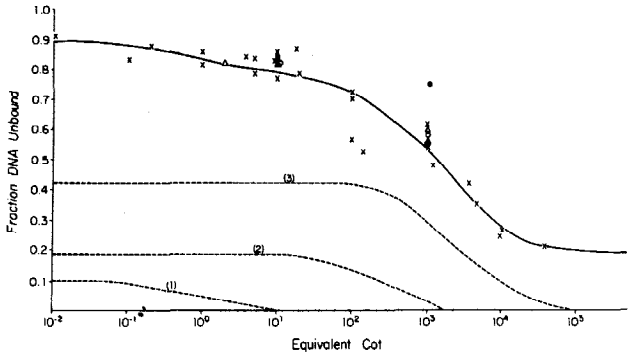


FIG. 1. C_0t curve of rat hepatoma DNA. DNA was purified, sheared to 400 nucleotides, incubated to the C_0t values indicated, then fractionated on hydroxylapatite columns as described under "Experimental Procedures." The experimental points were fit by computer to a theoretical curve with three components, 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.

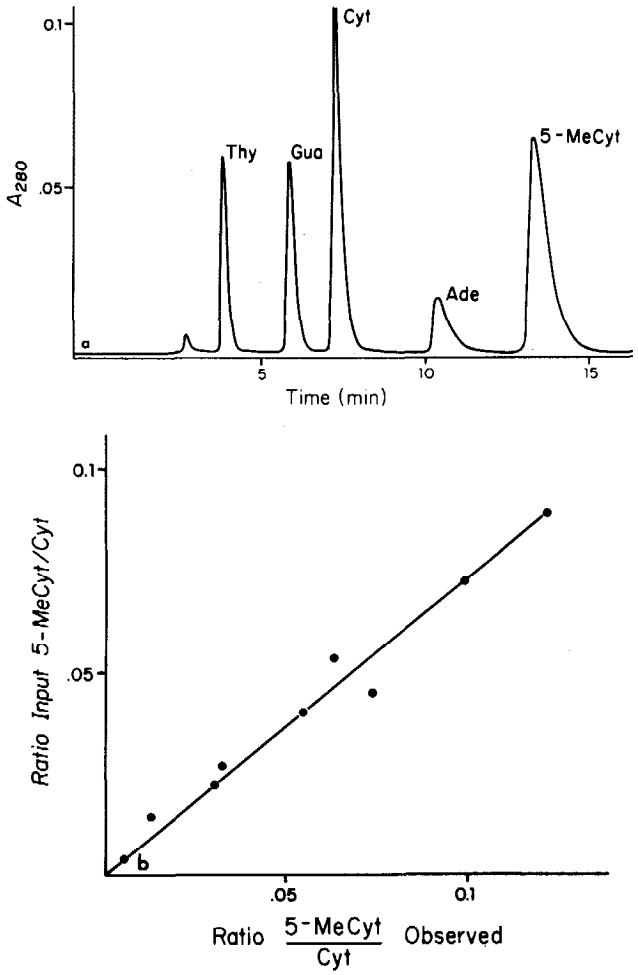


FIG. 2. *a*, elution profile of a base mixture. Ten microliters of a solution containing 1 μg each of thymine, guanine, cytosine, 5-methylcytosine, and adenine were applied to a Partisil SCX column (0.21 × 30 cm) and eluted with 0.045 M ammonium phosphate buffer, pH 2.3, as described under "Experimental 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.

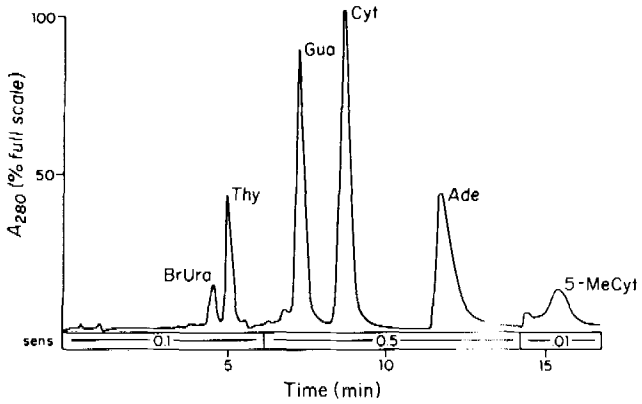


FIG. 3. Elution profile of DNA. Repetitive Component 2 DNA from Sample C (20% substituted in newly made strands with BrdUrd, Table III) was hydrolyzed, applied to a Partisil SCX column, and eluted as outlined under "Experimental Procedures." Changes in the sensitivity of the ultraviolet monitor are indicated.

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

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-

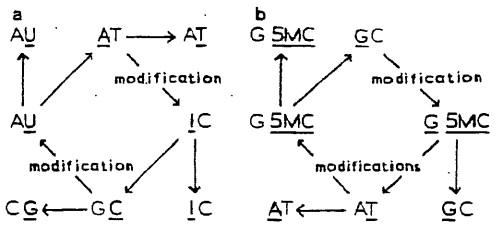


Fig. 1. (a) The transition from an A · T to a G · C base pair and the reverse transition, based on the scheme of Hawthorne (*5*). The modification of adenine occurs by the removal of an amino group at the 6-position to give inosine (I), which pairs with cytosine (C). The modification of cytosine occurs 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.

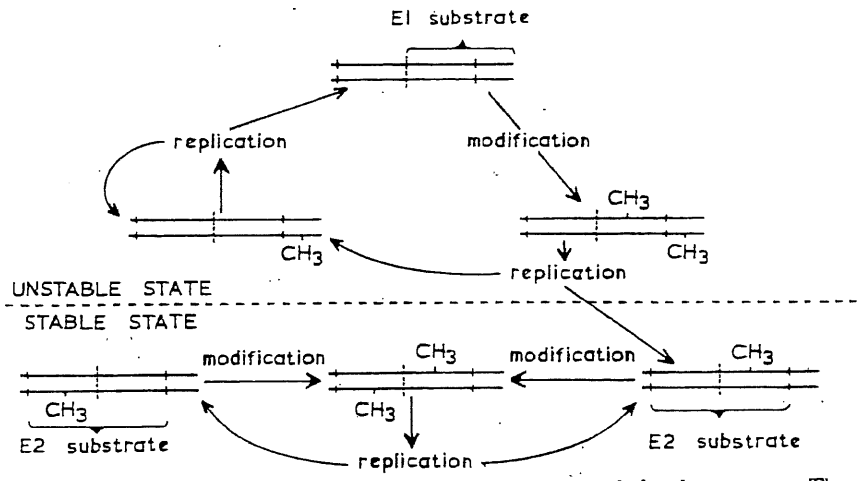


Fig. 2. The segregation of methylated DNA from an unmethylated precursor. The first modification enzyme, E1, methylates one-half of a palindromic sequence and an adjacent sequence in the complementary strand. Replication provides one substrate for the second enzyme, E2, which methylates the other half 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.

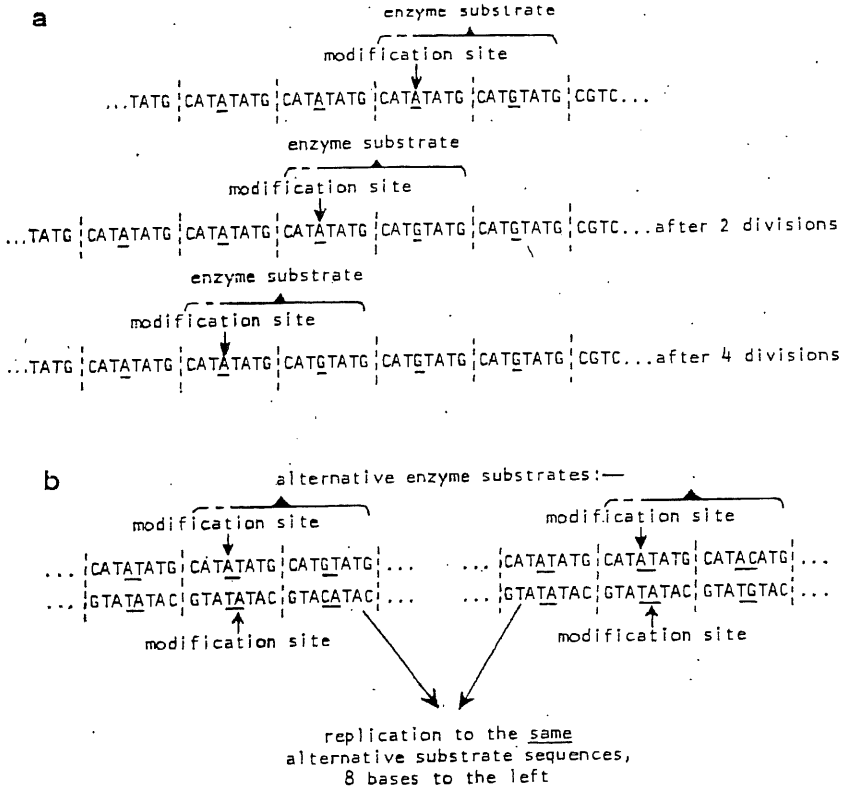


Fig. 3. (a) A mechanism for counting cell divisions based on the A · T → G · C transition in Fig. 1a. The modification enzyme recognizes the first sequence of eight bases, because it contains G at the 5-position, together with the whole or some part of the sequence to its left. The A at the 5-position 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

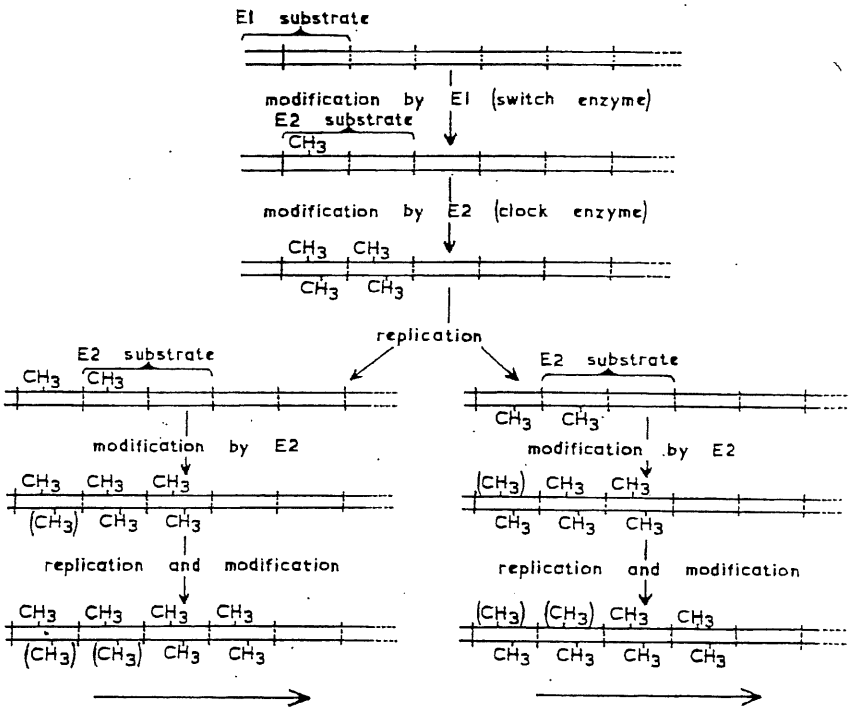


Fig. 4. A mechanism for counting cell divisions based on the methylation of palindromic controlling sequences. The first enzyme, E1, switches on the clock by recognizing a starter sequence, at the extreme left, which is adjacent to the first of the repeated sequences of the clock. One strand of this 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.)

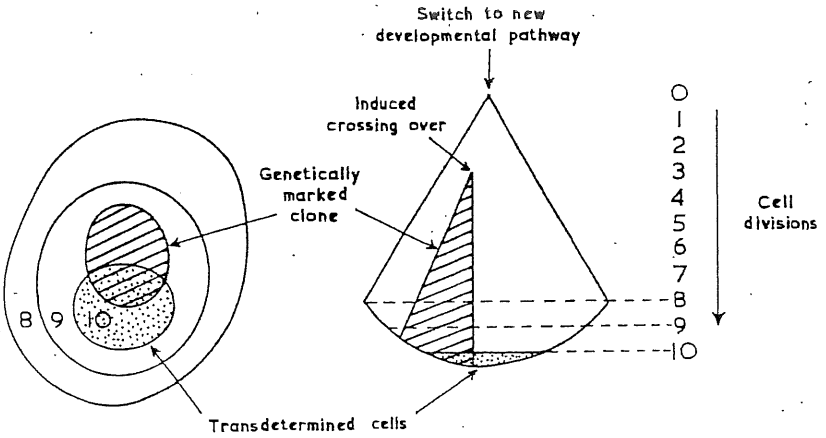


Fig. 5. An explanation for transdetermination based on an event in a single cell. The diagram on the right represents a clone derived from a cell in which a developmental clock has been triggered. It takes ten divisions for the clock to induce transdetermination in a group of cells. The induction of 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

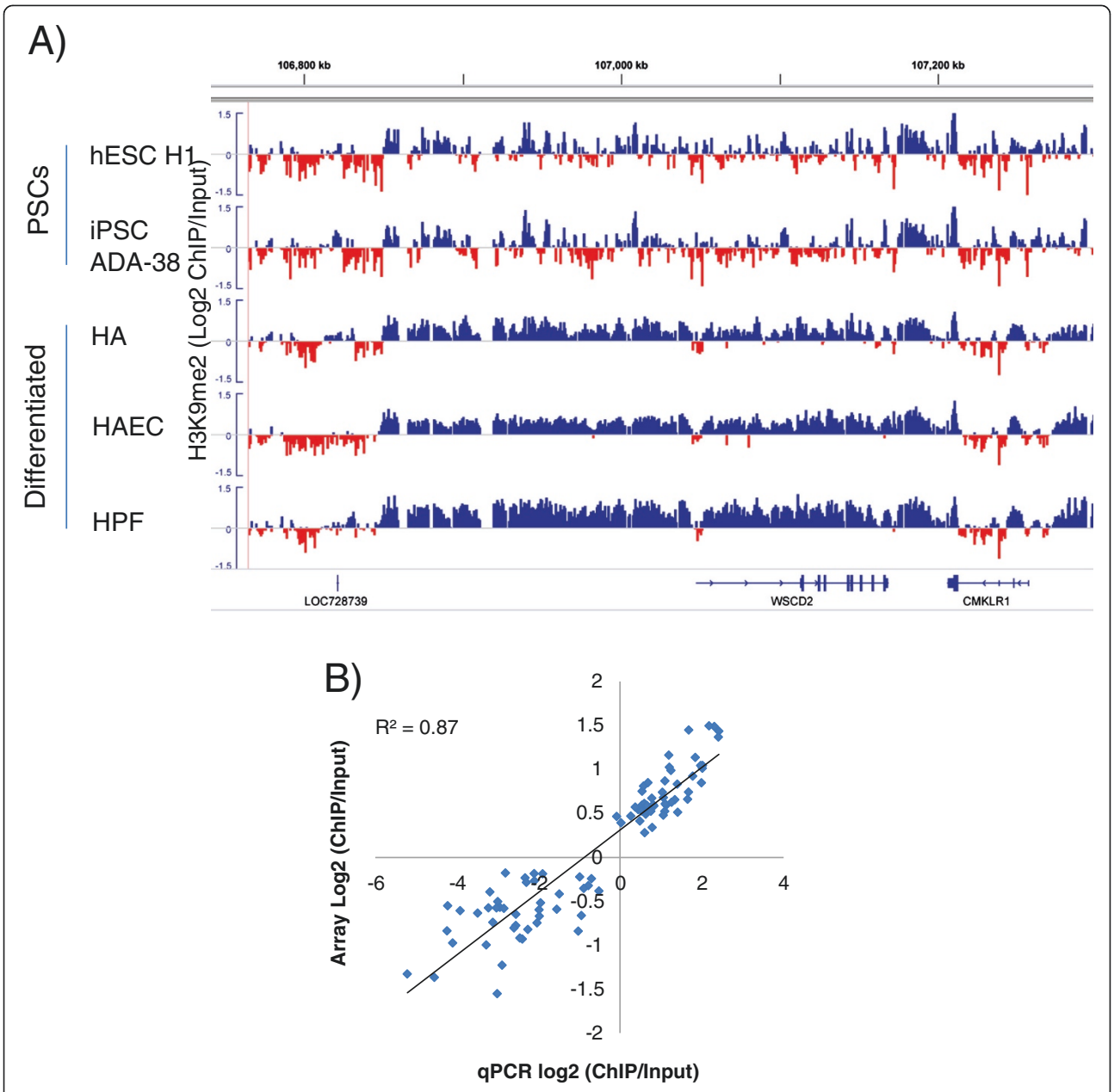


Figure 1 Global pattern of H3K9me2 in human pluripotent and differentiated cells. A) A representative region that shows LOCKs in differentiated cells (HA, HAEC and HPF) but not PSCs (H1 and ADA-38). Shown are H3K9me2 signals of ChIP-chip experiments in a ~500 kb long region containing *WSCD2* gene. Positive 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

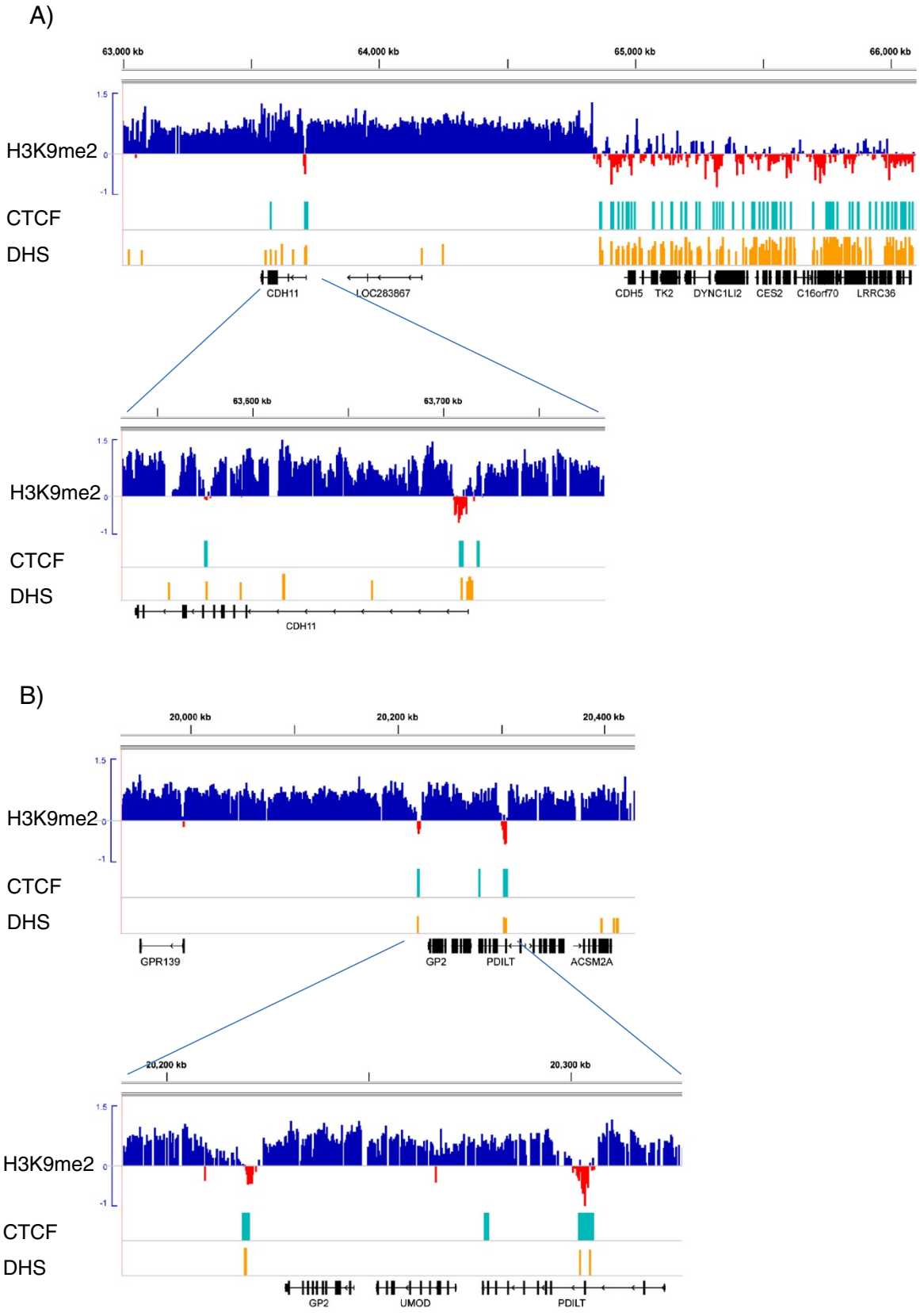


Figure 2 (See legend on next page.)

(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.

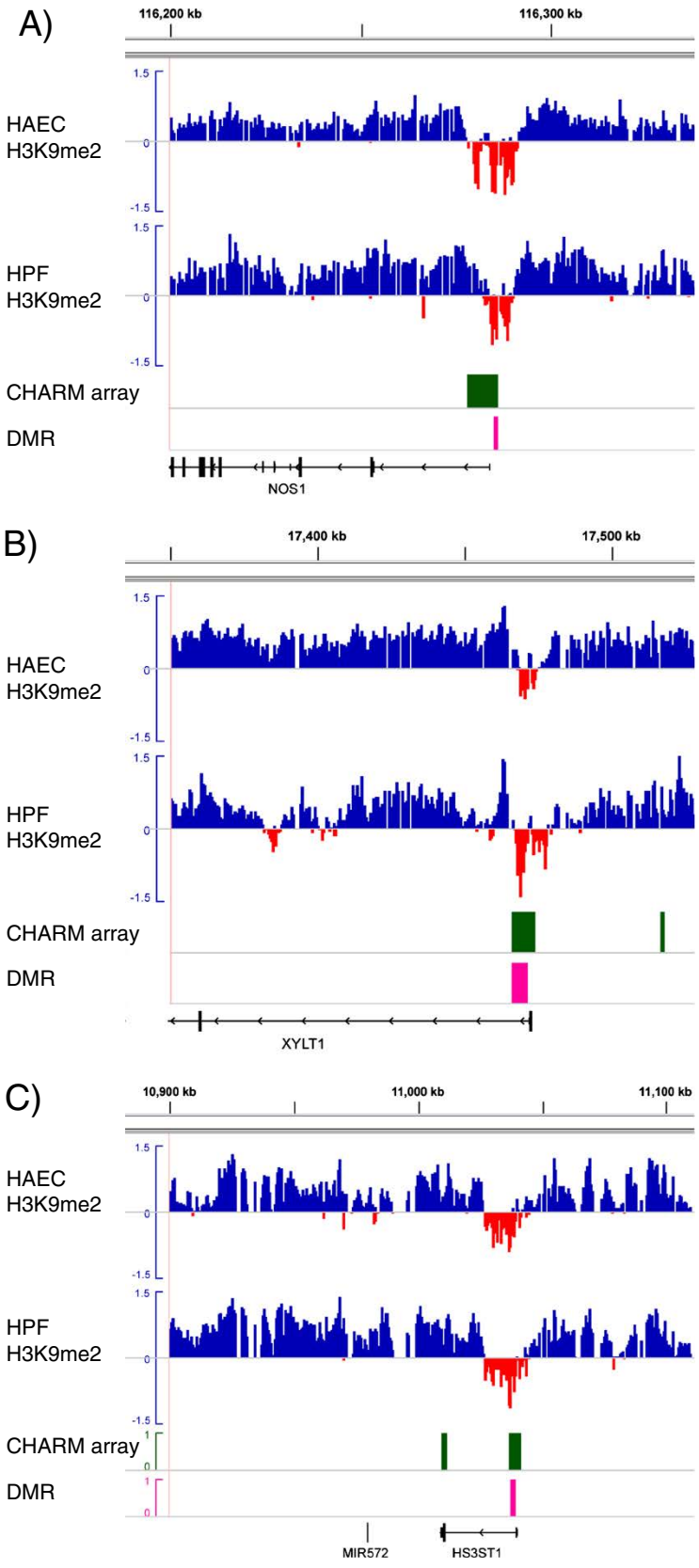


Figure 3 (See legend on next page.)

(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.

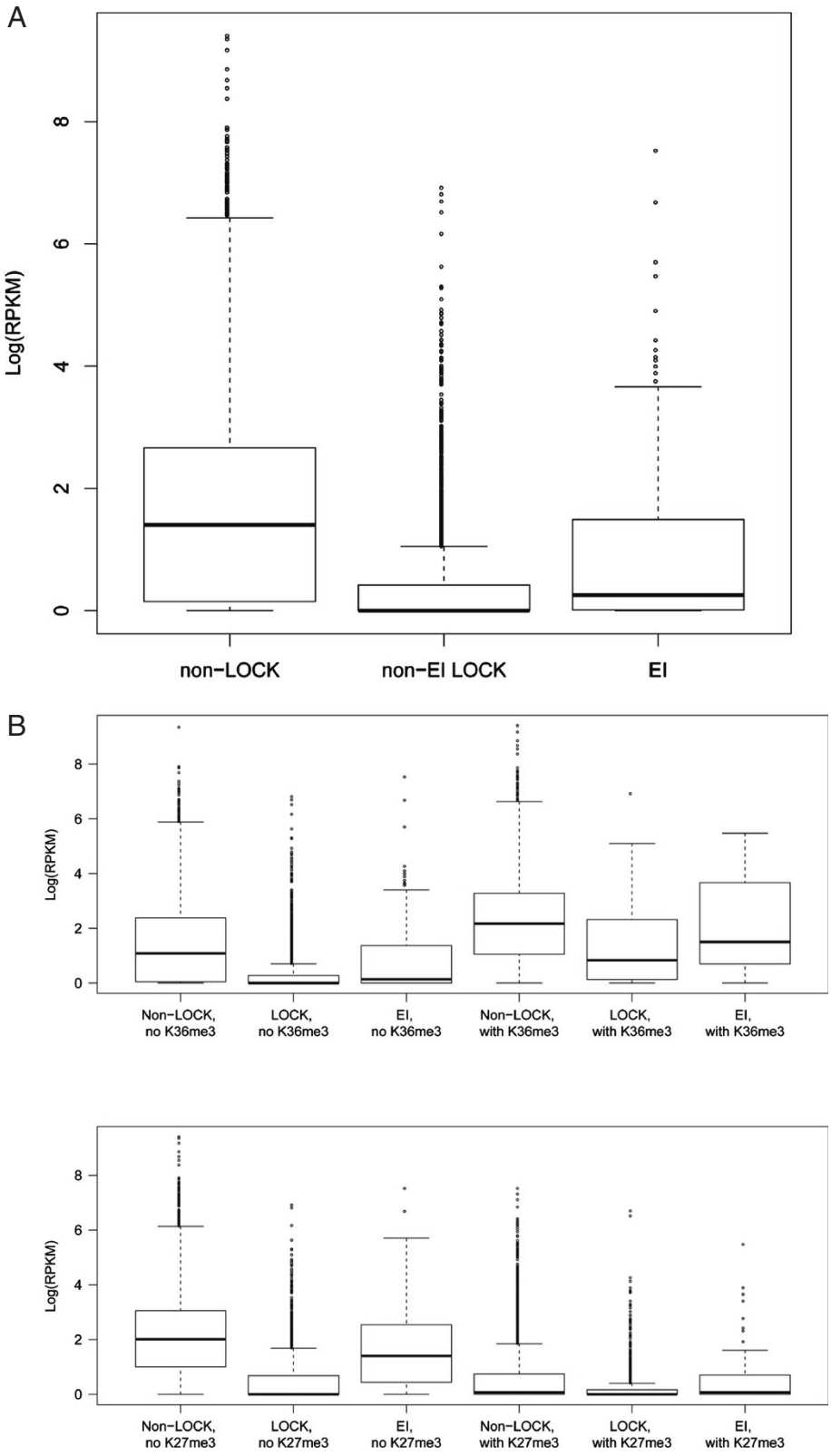


Figure 4 Expression of genes associated with EIs. We compared expression levels for genes with TSS at different regions. Expression values are RPKM (read per kb per million reads) for lung fibroblast IMR90 [7]. **A**) boxplot of expression level of genes with TSS 1) overlapping EIs; 2) overlapping LOCKs 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

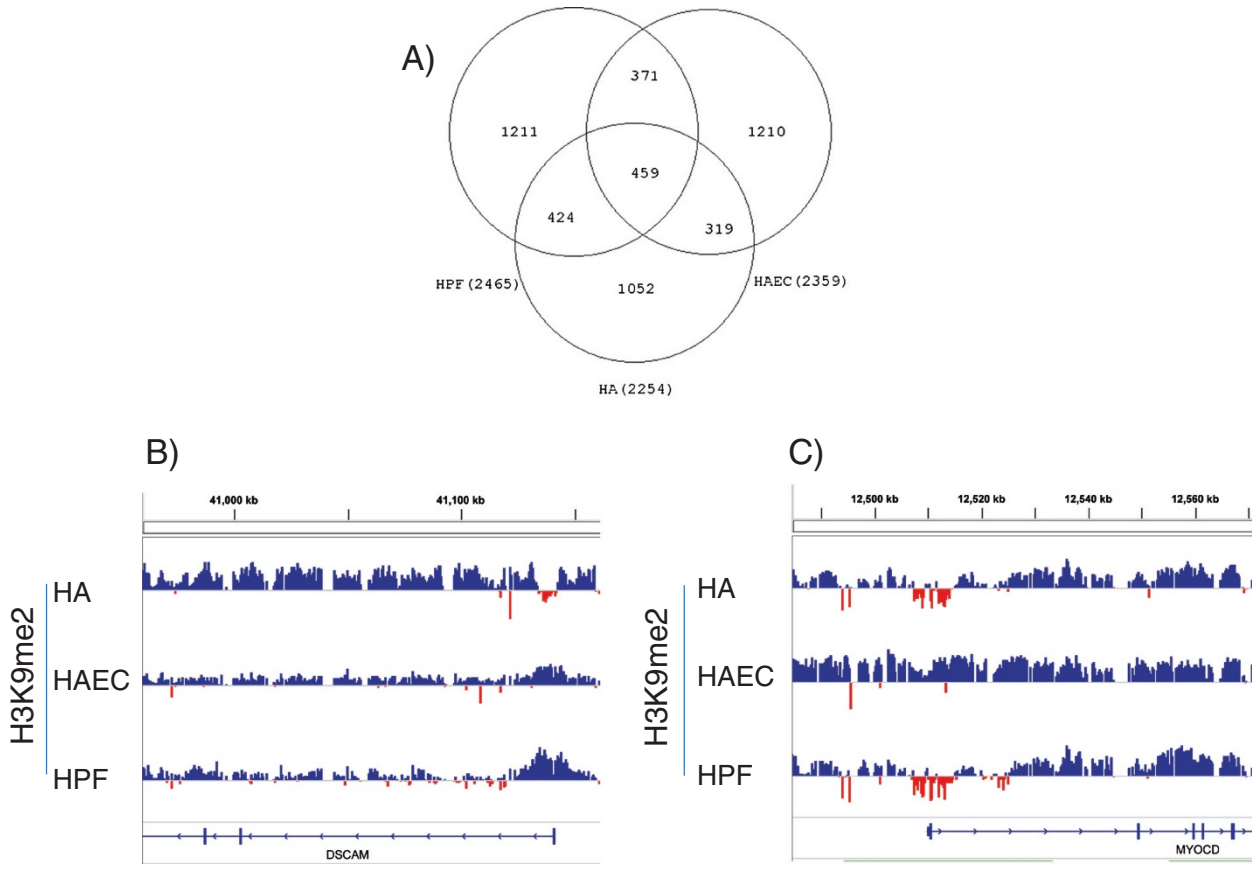


Figure 5 Cell type specific EIs. A) The Venn diagram shows overlaps of EIs among three differentiated cell lines (HA, HAEC and HPF). **B**-**C**) Shown are examples of EIs which differ among cell types. The annotation of tracks is the same as in Figures 2 and 3. An EI is found near the 5′ end of the 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.

Methods for mapping 3D chromosome architecture

Rieke Kempfer[1,2]* *and Ana Pombo*[1,2]*

Abstract | Determining how chromosomes are positioned and folded within the nucleus is critical to understanding the role of chromatin topology in gene regulation. Several methods are available for studying chromosome architecture, each with different strengths and limitations. Established imaging approaches and proximity ligation-based chromosome conformation capture (3C) techniques (such as DNA-FISH and Hi-C, respectively) have revealed the existence of chromosome territories, functional nuclear landmarks (such as splicing speckles and the nuclear lamina) and topologically associating domains. Improvements to these methods and the recent development of ligation-free approaches, including GAM, SPRITE and ChIA-Drop, are now helping to uncover new aspects of 3D genome topology that confirm the nucleus to be a complex, highly organized organelle.

Chromosome territories
The nuclear volumes occupied by each specific chromosome. Chromosomes tend to interact predominantly within themselves and occupy distinct regions within the interphase nucleus.

Chromosomal compartments
Chromosomes fold into distinct subcompartments, which correlate with transcriptional activity (A compartment) or repression (B compartment). The A and B compartments are defined by Hi-C contact frequencies.

[1]*Epigenetic Regulation and Chromatin Architecture Group, Berlin Institute for Medical Systems Biology, Max-Delbrück Centre for Molecular Medicine, Berlin, Germany.*

[2]*Institute for Biology, Humboldt University of Berlin, Berlin, Germany.*

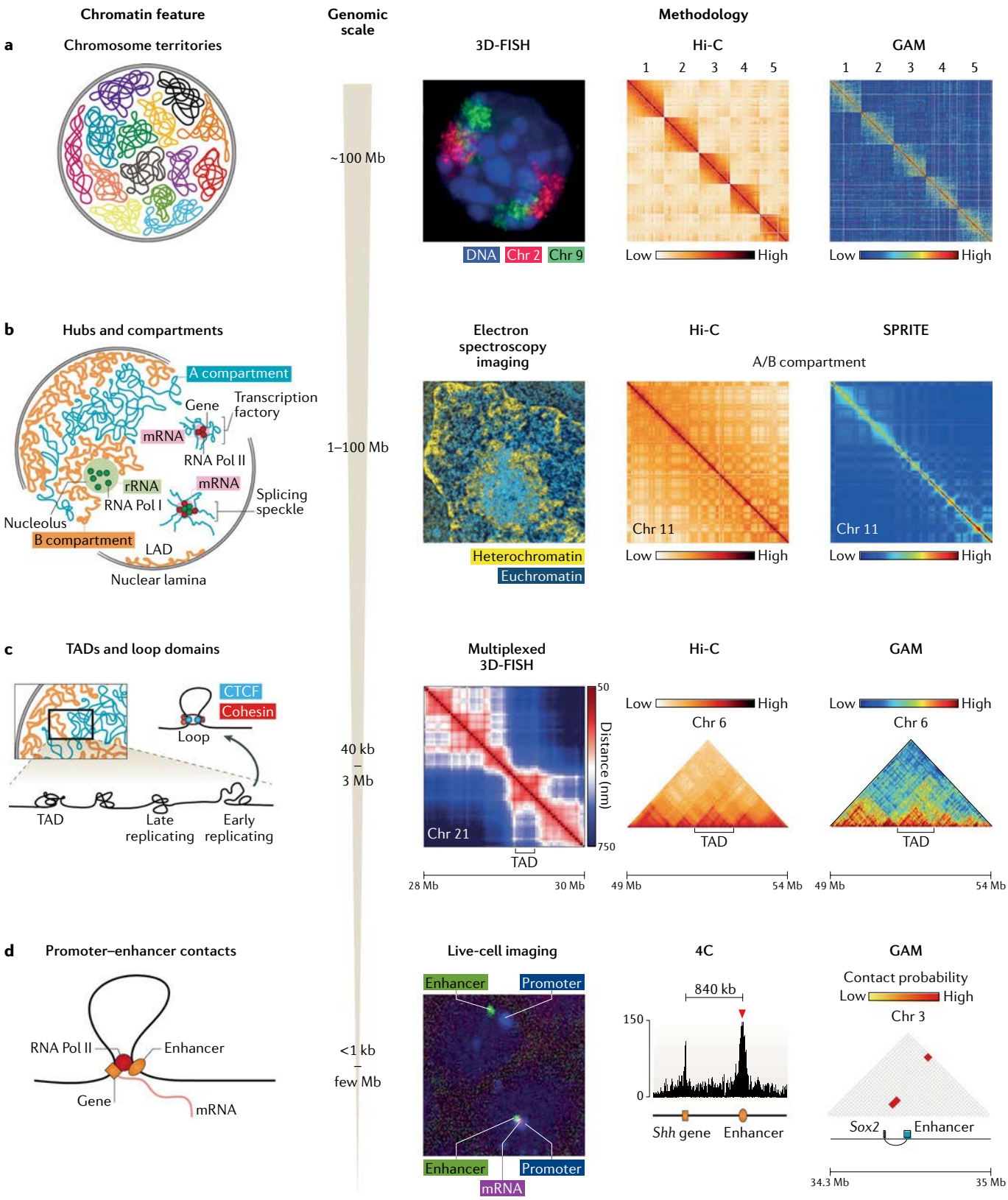
**e-mail: rieke.kempfer@mdc-berlin.de; ana.pombo@mdc-berlin.de*

https://doi.org/10.1038/s41576-019-0195-2

The nucleus of human cells harbours 46 densely packed chromosomes. Chromosomes are folded into hierarchical domains at different genomic scales, which likely enable efficient packaging and organize the genome into functional compartments. Chromosomes occupy distinct positions within the nucleus, called chromosome territories, which are partitioned into chromosomal compartments, and further partitioned into topologically associating domains (TADs) and chromatin loops mediated by CCCTC-binding factor (CTCF) or enhancer–promoter contacts (FIG. 1). Chromatin folding is a major feature of gene regulation and dynamically changes in development and disease[1–4]. Transcriptional control is mediated through physical contacts between enhancers and target genes, which occur via loop formation between the respective DNA elements. Functional loops between regulatory regions and genes are thought to occur predominantly within TADs. The expression of genes can also be influenced by their positioning relative to spatial landmarks inside the nucleus that are enriched for specific biochemical activities, such as the nuclear lamina. The disruption of enhancer–gene contacts and alteration of nuclear subcompartments play important roles in disease, including congenital disorders and cancer. Importantly, many disease-associated mutations of the linear genomic sequence can only be understood by considering their 3D conformation in nuclear space.

Advances in our understanding of chromosome folding have been restricted by a lack of approaches that can map chromatin contacts genome-wide while simultaneously retrieving spatial information, such as molecular distances between different genomic regions or between genomic regions and distinct nuclear compartments. Until recently, studies of 3D genome folding were limited to two main technologies: imaging, particularly fluorescence in situ hybridization of DNA (DNA-FISH); and approaches based on chromosome conformation capture (3C), namely Hi-C (high-throughput chromosome conformation capture). DNA-FISH was a revolutionary approach, which allowed visualization of the spatial organization of chromosomes and genes in the nucleus[5,6]. The approach provides single-cell information, but typically has a limited throughput that allows only a small number of genomic loci to be analysed at a time. 3C-based approaches, which depend on proximity ligation of DNA ends involved in a chromatin contact, have helped identify enhancer–promoter contacts. High-throughput derivatives, such as Hi-C, map chromatin contacts genome-wide at a length scale of hundreds of kilobases to a few megabases.

More recently, improvements in imaging techniques have increased the number of loci that can be analysed in parallel[7] and have extended the approach to live cells[8,9]. Orthogonal ligation-free approaches have also emerged, namely genome architecture mapping (GAM)[10], split-pool recognition of interactions by tag extension (SPRITE)[11] and chromatin-interaction analysis via droplet-based and barcode-linked sequencing (ChIA-Drop)[12], which have started to reveal novel aspects of chromatin organization. GAM, SPRITE and ChIA-Drop map chromatin contacts genome-wide and identify topological domains but also robustly detect a previously unappreciated level of high-complexity chromatin contacts that involve three or more DNA fragments and uncover specific contacts that span tens of megabases.



Here, we review the main approaches currently used in 3D genome research, highlighting their major advantages and caveats. To recognize the strengths of each technique, it is important to understand the principles and experimental details underlying each method, their intrinsic biases and their power to capture specific aspects of 3D genome architecture (TABLE 1). We discuss major features of 3D genome organization that have emerged, at the kilobase scale and above, through the application of these different

◀ Fig. 1 | **Methods for studying the major features of 3D chromatin folding across different genomic scales. a** | Chromosomes occupy discrete territories in the nucleus, which were first detected using imaging techniques. The 3D-fluorescence in situ hybridization (3D-FISH) image shows the positions of the chromosome territories of chromosome 2 (red) and chromosome 9 (green) within DAPI-stained nuclei (blue) from mouse embryonic stem cells (ESCs). Chromosome territories are also detected as regions of high-frequency intrachromosomal interactions on contact maps generated by chromosome conformation capture (3C)-based methods (such as Hi-C (high-throughput chromosome conformation capture)) and ligation-free approaches (such as genome architecture mapping (GAM)). **b** | DNA inside the nucleus separates into hubs of active (A compartment) and inactive (B compartment) chromatin, clustering around the nucleolus, splicing speckles, transcription factories and other nuclear bodies not represented here. Electron spectroscopy imaging of the mouse epiblast shows the distribution of heterochromatin (yellow) around the nucleolus (light blue) and at the nuclear periphery. Decondensed euchromatin (dark blue) is positioned more centrally in the nucleus. Nucleic acid-based structures are stained yellow, protein-based structures blue. Hi-C and split-pool recognition of interactions by tag extension (SPRITE) contact maps of mouse chromosome 11 show the separation of chromatin into discrete contact hubs (A and B compartments), which are visible as checkerboard-like contact patterns. **c** | At shorter genomic length scales, chromatin folds into topologically associating domains (TADs), which overlap with domains of early and late replication, and DNA loops, that arise from cohesin-mediated interactions between paired CCCTC-binding factor (CTCF) proteins. Multiplexed FISH of consecutive DNA segments in a 2-Mb region in the human genome shows the emergence of TADs in the population-average distance map. In Hi-C and GAM contact maps, TADs are represented by regions of high internal interaction frequencies and demarcated by a drop in local interactions at their boundaries. **d** | Contacts between a gene and its *cis*-regulatory elements occur via loop formation between the enhancer bound by RNA polymerase II (Pol II) and the gene promoter. These contacts can be detected by live-cell imaging; shown are contacts between the enhancer (green) and promoter (blue) of the *eve* gene in a *Drosophila melanogaster* embryo, with simultaneous imaging of *eve* mRNA expression (red). The circular chromosome conformation capture (4C)-sequencing track shows the interactions between the *Shh* gene promoter and the ZRS (a limb-specific enhancer of the *Shh* gene) in the anterior forelimb in mice. GAM data can be processed using the mathematic model statistical inference of co-segregation (SLICE) to extract the most significant enhancer–promoter contacts from the data set, resulting in a contact matrix with only the high-probability interactions[10]. The most significant interaction at the *Sox2* locus can be found between the *Sox2* gene and one of its well-studied enhancers[189]. For parts **a**, **b** and **c**: HiGlass[190] was used to generate contact maps for previously published Hi-C data from mouse ESCs[191]; heat maps for GAM and SPRITE were generated from normalized published matrix files from previously published mouse ESC data (for GAM[10], for SPRITE[11]). LAD, lamina-associated domain; rRNA, ribosomal RNA. 3D-FISH image reprinted from REF.[187], CC BY 2.0 (https://creativecommons.org/licenses/by/2.0/). Electron spectroscopy image reprinted from REF.[188], CC BY 3.0 (https://creativecommons.org/licenses/by/3.0/). Part **c** reprinted with permission from REF.[32], Science. Part **d** adapted from REF.[85], Springer Nature Limited, and from REF.[48], CC BY 3.0 (https://creativecommons.org/licenses/by/3.0/).

Topologically associating domains
(TADs). Chromosomal regions that fold into self-associating domains, with high internal interaction frequencies, demarcated by a clear drop of local interactions with neighbouring regions at their boundaries.

Chromatin loops
Local regions of high interaction frequency between two genomic loci indicate that these regions form the basis of a DNA loop. Loops often form between regions with divergent CCCTC-binding factor (CTCF) sites, or between enhancers and their target promoters.

technologies, and highlight discrepancies between approaches. We will not cover chromatin folding at the level of nucleosomes, which has been reviewed previously[13].

Imaging-based detection of contacts

The visualization of nuclear structures and specific genomic sequences is key to understanding how chromatin is organized in the nucleus. Various light microscopy and electron microscopy techniques can be used to identify nuclear compartments or image the physical positions of specific genomic loci in the nucleus of fixed or live cells. The most commonly used imaging technique for detecting chromatin contacts in fixed cells is DNA-FISH. Contacts can be visualized in live cells using insertions of DNA binding site arrays (such as the Lac operator-repressor[14,15], Tet operator-repressor[16] and ANCHOR[17] systems) or, more recently, using CRISPR-based approaches (FIG. 2).

Measuring contacts with DNA-FISH. FISH uses fluorescently tagged DNA sequences (such as oligonucleotides) as probes to hybridize to complementary target regions of interest in the genome (FIG. 2). For hybridization to occur, a single-stranded probe needs to be able to enter the nucleus, which is usually achieved by permeabilizing the cell with a detergent or organic solvent, such as methanol. To ensure the probe can bind to its target, the DNA is most often denatured by heat and formamide treatment. The genomic regions highlighted by the hybridized fluorescent probes are then visualized by microscopy.

DNA-FISH is typically used to measure the physical distances between two or a few differentially labelled genomic regions of interest. A chromatin contact is often defined by a distance threshold, which is usually set arbitrarily according to the scale of genomic distances between the regions of interest and the resolution of the microscope. Thus, chromatin contacts have been inferred when fluorescent signals co-localize within a spatial distance of 50 nm to 1 μm (REFS[18–21]), although it is not clear whether distances at the top end of this range (which are close to the diameter of a whole chromosome) represent true interactions or indirect non-random positioning. DNA-FISH can also be used to visualize chromatin compaction[22] or positioning of genomic regions with respect to nuclear structures, such as the nuclear lamina[23]. The overall distributions of spatial distances between loci or relative to the nuclear periphery found across the cell population are usually summarized by the frequency of co-localization (that is, the frequency with which chromatin contacts are detected in the cell population), but other metrics, such as mean or median distances, are also used. The data are compared with the physical distances between other (control) genomic regions (which are often separated by similar genomic distances to the experimental loci) or, in some cases, with the nuclear diameter or volume. These metrics can help distinguish specific chromosomal conformations but can also be ambiguous, depending on the choice of control probes or if allelic differences or other forms of heterogeneity are present within the cell population.

The accuracy and power to detect different nuclear structures or contacts also depend on how well the organization of the target DNA and nuclear compartments is preserved during the FISH procedure, on the resolution of the microscope and on the size of the target genomic sequence. FISH experiments use probes made of a collection of small DNA fragments that are either synthesized (oligos) or produced by nick-translation from larger DNA molecules (plasmids, fosmids, bacterial artificial chromosomes or whole mammalian chromosomes), resulting in overlapping fragments of 100–500 bp. The probes often cover genomic sequences ranging in length from 30 kb up to entire chromosomes. The signal-to-noise ratio for locus detection increases with the target length due to increased local fluorescence and higher target specificity. Thus, with standard 3D

Table 1 | **Comparison of methods used to detect chromatin contacts**

Assay	Description	Number of contacts per experiment	Multiplicity of contacts	Single-cell information	Number of cells	Detectable contacts	Protocol
3C-based methods							
3C	Proximity ligation and selection of target regions with primers, detection by quantitative PCR	One versus one	Pairwise	No	100 million[192]	Protein-mediated	192
4C	Proximity ligation and enrichment for contacts with one bait region by inverse PCR, detection by sequencing	One versus all	Pairwise	No	Robust: 10 million[193], low input: 340,000 (REF.[194])	Protein-mediated	193
5C	Proximity ligation and enrichment for larger target region with primers, detection by sequencing	Many versus many	Pairwise	No	Robust: 50–70 million[195], low input: 2 million[196]	Protein-mediated	195,196
Hi-C	Proximity ligation and enrichment for all ligated contact pairs, detection by sequencing	All versus all	Pairwise	No	Robust: 2–5 million[64], low input: 100,000–500,000 (REFS[70,197])	Protein-mediated	64,197
TCC	Tethered proximity ligation and enrichment for all ligated contact pairs, detection by sequencing	All versus all	Pairwise	No	25 million[57]	Protein-mediated	57
PLAC-seq, ChIA-PET	Proximity ligation and pull-down of specific protein-mediated contacts, detection by sequencing	Many versus many	Pairwise	No	Robust: 100 million[198], low input: 500,000 (REF.[81])	Protein-mediated (specific)	81,198
Capture-C, C-HiC	Proximity ligation and target enrichment using probes for genomic regions of interest, detection by sequencing	Many versus all	Pairwise	No	Robust: 100,000 (REF.[199]), low input: 10,000–20,000 (REF.[97])	Protein-mediated	199
Single-cell Hi-C	Proximity ligation and enrichment for all ligated contact pairs, detection by sequencing	All versus all	Pairwise	Yes	Hundreds	Protein-mediated	71
Imaging							
2D-FISH	Fixation to flatten cells, hybridization of fluorescent probes to target regions, measurement of 2D spatial distances	Between 2 and 52 regions[a]	Pairwise or more	Yes	Hundreds	All in spatial proximity	200
3D-FISH	Fixation of cells, hybridization of fluorescent probes for target regions, measurement of 3D spatial distances	Between 2 and 52 regions[a]	Pairwise or more	Yes	Hundreds	All in spatial proximity	201
Cryo-FISH	Fixation of cells, cryosectioning, hybridization of fluorescent probes for target regions, measurement of 2D spatial distances	Between 2 and 52 regions[a]	Pairwise or more	Yes	Hundreds	All in spatial proximity	101
Live-cell imaging	Fluorescent labelling of genomic loci in living cells, measurement of spatial distances over time	Between 2 and 12 regions	Pairwise or more	Yes	Hundreds	All in spatial proximity	9,39,202,203
Ligation-free methods							
GAM	Cryosectioning of fixed cells, DNA extraction from nuclear sections and sequencing, inferring spatial distances from co-segregation of genomic regions in nuclear sections	All versus all	Pairwise or more	Yes	Hundreds[10]	All in spatial proximity	10
SPRITE	Fixation of cells, identification of crosslinked chromatin fragments by split-pool barcoding and sequencing	All versus all	Many	No	10 million[11]	Protein-mediated	11
ChIA-Drop	Fixation of cells, identification of crosslinked chromatin fragments by droplet-based and barcode-linked sequencing	All versus all	Many	No	10 million[12]	Protein-mediated	12

3C, chromosome conformation capture; 4C, circular chromosome conformation capture; 5C, chromosome conformation capture carbon copy; ChIA-Drop, chromatin-interaction analysis via droplet-based and barcode-linked sequencing; ChIA-PET, chromatin interaction analysis by paired-end tag sequencing; C-HiC, capture HiC; FISH, fluorescence in situ hybridization; GAM, genome architecture mapping; Hi-C, high-throughput chromosome conformation capture; PLAC-seq, proximity ligation-assisted chromatin immunoprecipitation sequencing; SPRITE, split-pool recognition of interactions by tag extension; TCC, tethered chromosome capture. [a]Classical FISH experiments rarely distinguish between more than 2–5 differentially labelled regions simultaneously[204]. Cycles of probe hybridization can increase this number up to 52 (REF.[205]).

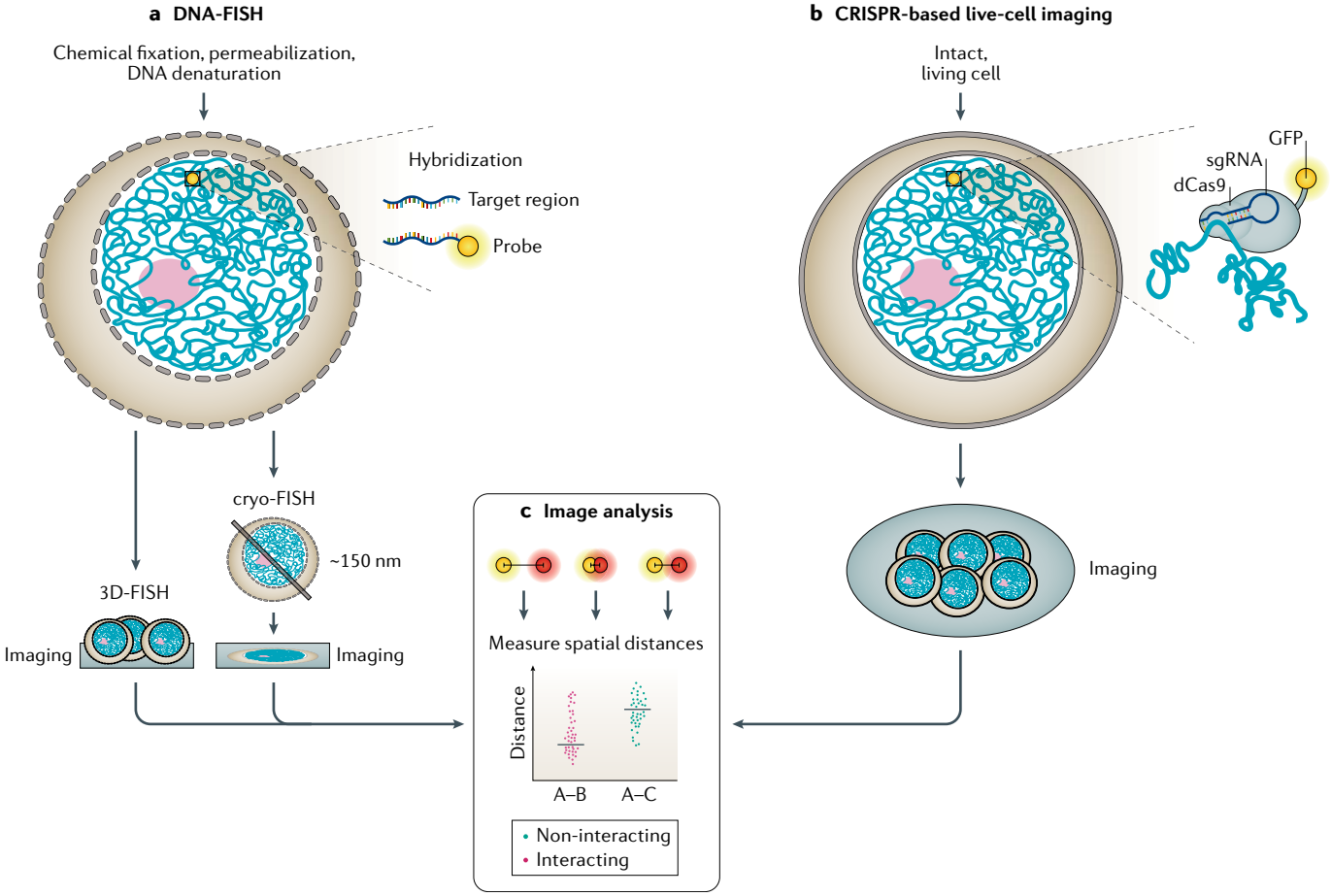


Fig. 2 | **Imaging-based approaches to visualize chromatin contacts.** **a** | Fluorescence in situ hybridization of DNA (DNA-FISH) uses fluorescently labelled probes that hybridize to specific genomic loci in the nucleus. Typically, cells are fixed and permeabilized and, upon denaturing of the DNA, FISH probes hybridize to their complementary target region (yellow circle). The FISH procedure can be performed in whole cells, embryos, thick tissue slices (3D-FISH) or thin cryosections of cells (cryo-FISH). The nucleolus is represented in pink. **b** | CRISPR-based live-cell imaging can be performed in the intact, living cell. Typically, dead Cas9 (dCas9) is fused to green fluorescent protein (GFP) and the fusion protein is recruited to the target region by small guide RNAs (sgRNAs), which are complementary to the region of interest. **c** | For all of the techniques, chromatin contacts are assessed by the spatial distances between the fluorophores targeting the regions of interest (yellow and red circles). To determine the specificity of a contact, spatial distances between interacting loci should be compared with distances between non-interacting loci in numerous cells. The distribution of distances, the mean distance and the median distance can all inform about the quality and abundance of the contact in a cell population.

CCCTC-binding factor
(CTCF). A transcription factor with 11 conserved zinc-finger (ZF) domains. This nuclear protein is able to use different combinations of the ZF domains to bind different DNA target sequences and proteins. CTCF is enriched at topologically associating domain (TAD) borders, where its binding can be important to specify TAD border definition.

Chromatin
The combination of DNA, RNA and protein that constitutes the chromosomes in eukaryotic cells. Broadly, heterochromatin is associated with transcriptional repression and euchromatin is associated with transcriptional activity.

FISH long-range contacts within large genomic regions, such as between TADs[10,24] or in whole chromosomes[25], can be accurately detected. However, short-range interactions between chromosomal regions that are less than 100 kb apart are difficult to detect, making it harder to quantify fine-scale chromatin folding below the TAD level, such as enhancer–promoter interactions.

High-resolution imaging of chromatin contacts can be achieved using cryo-FISH, in which standard FISH probes are hybridized to thin (~100–200 nm) cryosections from cells fixed using conditions optimized to preserve the nuclear ultrastructure; the signal is then visualized using fluorescence or electron microscopy[10,19,25–27]. More recently, the short length and high specificity of fluorophore-tagged oligonucleotides known as Oligopaints[28] have made it possible to target 15-kb loci using conventional microscopy[29] or 5-kb regions using super-resolution microscopy (when combined with a second labelling step to enhance the fluorescence signal)[30]. Oligopaints are not derived from cloned genomic regions but are instead generated from synthetic libraries of short (~60–100 bp) oligonucleotides, which are produced by massively parallel synthesis[31]. Once generated, the library pool can be amplified in a flexible manner, using different primer pairs to give rise to different sets of FISH probes. The ease of design of Oligopaints has opened new possibilities for the study of chromatin folding, such as being able to visualize chromatin in different epigenetic states at a resolution of tens of nanometres[22]. Oligopaint-based FISH has also been used in combination with high-throughput imaging to generate low-resolution contact maps (for example, at the TAD level) of whole chromosomes[7] and high-resolution (30-kb) contact maps for stretches of DNA 1.2–2.5 Mb in length[32]. In addition, molecular beacon FISH probes have emerged as a way to target genomic regions as short as 2.5 kb (REF.[33]). In an unbound state, these probes form a hairpin loop that minimizes the off-target fluorescent signal by bringing together the fluorescent label and a quencher.

By reducing the background signal from the unbound probe, the technique improves the visualization of small genomic regions.

Live-cell imaging of nuclear structures. Chromosome folding is a highly dynamic process that varies greatly throughout the cell cycle[34,35]. Our ability to study these chromatin dynamics has been revolutionized by technologies based on genome editing that allow specific genomic loci to be targeted in live cells. Early iterations of this approach were rather laborious; cell lines needed to be created in which the target locus was tagged with DNA binding site arrays that recruit a fluorescently tagged cognate DNA binding protein (such as the Lac operator-repressor[32,33], Tet operator-repressor[34] and ANCHOR[35] systems). Now, loci can be targeted in live cells with a version of the CRISPR system that uses an endonuclease-deficient form of Cas9 (dead-Cas9 (dCas9)) fused with a fluorescent protein[36]. The tagged dCas9 is recruited to the genomic locus of interest via its interactions with sequence-specific small guide RNAs (FIG. 2). For simultaneous labelling of two genomic regions, small guide RNAs can be differentially modified to act as scaffolds that bring fluorescent proteins to the target loci. For example, fusion proteins that comprise a fluorescent protein and either tandem dimer MS2 coat-binding protein (tdMCP) or tandem dimer PP7 coat-binding protein (tdPCP) can be directed to target loci by guide RNAs containing MS2 or PP7 aptamers, respectively. As both proteins have a comparably high exchange rate, which compensates for photobleaching, this approach is also well suited to long-term live-cell imaging[37–39]. However, most CRISPR-based methods are currently limited to the detection of repetitive sequences because they rely on a single species of guide RNA, which hybridizes to identical genomics sequences, to direct simultaneous binding of dozens of copies of the fluorescent protein to achieve a strong fluorescent signal. A notable exception is the chimeric array of gRNA oligonucleotides (CARGO); by delivering 12 different guide RNAs into a single cell, this technique was able to efficiently label a non-repetitive 2-kb genomic region[40].

Ligation-based detection of contacts

3C-based methods extract chromatin interaction frequencies between genomic loci via chromatin crosslinking and proximity ligation (FIG. 3). Following formaldehyde fixation to capture protein-mediated and RNA-mediated contacts, chromatin is fragmented using a restriction enzyme, and the crosslinked restriction fragments are ligated[41]. The purified ligation fragments are called a 3C library. The ligation frequency between two loci of interest can be quantified by PCR using appropriate primer pairs. Thus, 3C focuses on interactions between two loci ('one versus one') and requires prior knowledge of the targets of interest. However, the 3C library contains all ligation products for the genome investigated and the 3C workflow can therefore be adapted to enable genome-wide analysis of chromatin contacts. Chromosome conformation capture-on-chip[27] or circular chromosome conformation capture[42], both called 4C, enrich for interactions of one region with the remaining genome ('one versus all'). Chromosome conformation capture carbon copy (5C)[43] captures contacts of a larger genomic stretch at high resolution ('many versus many'). Finally, Hi-C[44] captures all ligation events across the entire genome ('all versus all'). Workflows and differences between these techniques have been described elsewhere in great detail[45]. Here, we focus on the most commonly used versions (FIG. 3; TABLE 1).

Mapping all contacts at a single locus with 4C. A straightforward and cost-effective method to obtain additional information from a 3C library is 4C. Here, primers for a region of interest (such as a promoter) are used to amplify all ligation partners of the locus under investigation (called the 'viewpoint') (FIG. 3). The amplified ligation products are sequenced (to a depth of 1–5 million reads per library[46]) and used to analyse genome-wide interaction partners of the region of interest at a resolution of a few kilobases. 4C has been widely used to investigate *cis*-regulatory landscapes of genes, especially in development and disease[47]. It is well suited for detecting short-range regulatory interactions[48], but has also been applied to detect contacts spanning long genomic distances, including whole chromosomes[27,49].

Mapping all contacts occurring within a large genomic region with 5C. In 5C, large genomic regions spanning up to several megabases are amplified from the 3C library using an elegant, yet complex, mix of forward and reverse primers. For example, 5C analysis of a 4.5-Mb chromosomal region around the *Xist* gene revealed the presence of TADs[24]. 5C has the advantage of producing high-resolution data at an affordable sequencing depth (~60 million reads per library to obtain resolution of 15–20 kb for a 1-Mb region)[50]. However, the resolution of 5C is dependent on the ability to design forward and reverse primers for all possible restriction fragments across a given locus; in the absence of appropriate primers, some mappable fragments will be excluded from the contact map.

Mapping all contacts at one or more loci with capture-based methods. A 3C library can be enriched for one or more genomic targets of interest using capture-based methods, such as Capture-C[51], Capture Hi-C[52] and CAPTURE[53]. In these approaches, biotinylated oligonucleotides complementary to a genomic region of interest are used to pull-down specific ligation products from the library, which are then amplified and sequenced. These approaches can be used to detect interactions of one viewpoint but also of entire genomic regions[47] or groups of targets[54,55].

Mapping all genome-wide contacts with Hi-C and its derivatives. Hi-C is the most commonly used genome-wide approach to map chromatin contacts from a 3C chromatin preparation[44]. In this approach, the ends of crosslinked DNA restriction fragments are labelled with biotin and then ligated. After ligation, the exonuclease activity of T4 DNA polymerase is used to remove the biotin label from the ends of unligated fragments. Ligated

Nuclear lamina
A protein mesh, consisting of lamins and other membrane-associated proteins, at the inner nuclear membrane that contributes to nuclear structure and function. Chromatin in the proximity of the lamina tends to be heterochromatic and transcriptionally repressed.

Fluorescence in situ hybridization
A technique that can be used to visualize the location of nucleic acid sequences within the nucleus using sequence-specific fluorescent probes that hybridize to the regions of interest, combined with microscopy.

Chromosome conformation capture
(3C). A technique used to detect the frequency of interactions between any specified two loci in the genome. Interactions between loci are captured by formaldehyde fixation, followed by restriction enzyme digestion and ligation. The frequencies of interactions between loci are determined by quantitative real-time PCR.

Hi-C
(High-throughput chromosome conformation capture). A genome-wide version of chromosome conformation capture that allows all chromatin interactions in the genome to be mapped simultaneously. The frequencies of interactions between loci are determined by paired end sequencing.

Proximity ligation
Fixation of cells, followed by fragmentation of chromatin and ligation of nearby, crosslinked DNA fragments.

Genome architecture mapping
(GAM). A genome-wide approach to detect chromatin contacts based on their physical distances within the nucleus. DNA loci are detected in thin nuclear slices by DNA extraction and sequencing. Chromatin contacts are inferred from co-segregation frequencies of pairs of DNA loci across a large (400–1,000) collection of nuclear slices.

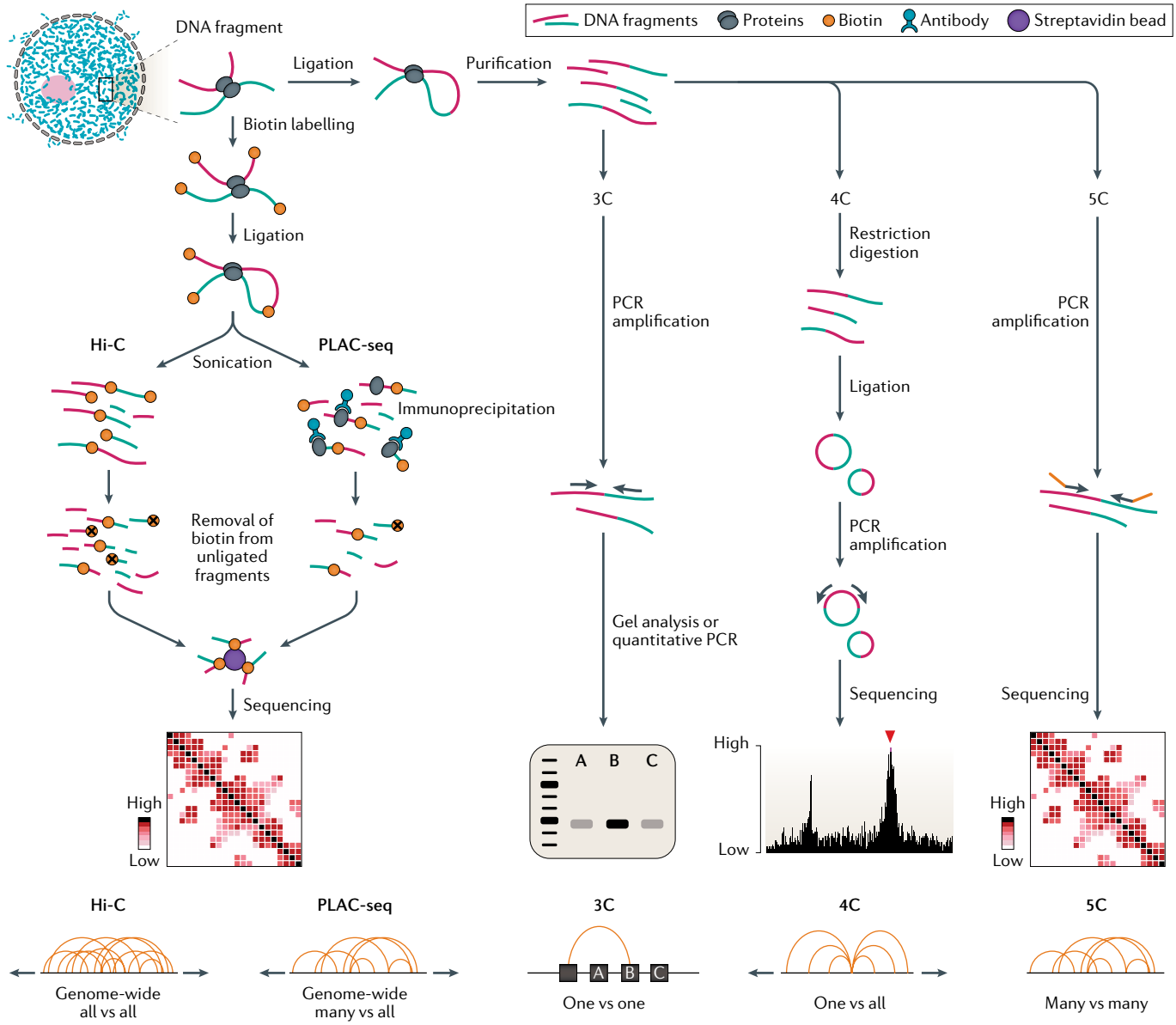


Fig. 3 | **3C and its derivatives.** Chromosome conformation capture (3C)-based assays measure contact frequencies of pairs of DNA loci by proximity ligation of crosslinked and fragmented chromatin. All 3C-based assays involve fixation of the chromatin, isolation of nuclei and DNA fragmentation (for example, with a restriction enzyme). The obtained crosslinked chromatin fragments are then processed for 3C, circular chromosome conformation capture (4C) or chromosome conformation capture carbon copy (5C), which map chromatin contacts for preselected regions, or for genome-wide assays, such as high-throughput chromosome conformation capture (Hi-C) and proximity ligation-assisted chromatin immunoprecipitation sequencing (PLAC-seq). In 3C, 4C and 5C, the crosslinked chromatin fragments are ligated and the DNA is purified. In 3C, the interactions between two chosen genomic regions are detected by PCR amplification with primers specific to the two regions of interest. PCR products are analysed semi-quantitatively on an agarose gel or by real-time quantitative PCR. Interactions are defined by higher ligation frequencies compared with control regions of similar genomic distance. In 4C, interactions of one viewpoint with the whole genome are measured. The ligated and purified DNA is fractionated with a secondary restriction digest, and the digested, smaller DNA fragments are circularized and amplified with primers facing outwards from the viewpoint. The PCR products are sequenced by paired end sequencing, providing the sequence information and frequency of every chromatin contact of the viewpoint. In 5C, the ligated and purified DNA is directly amplified using primers for all restriction fragments within a consecutive genomic region, usually hundreds of kilobases up to several megabases. The PCR products are sequenced and provide information about the ligation frequencies of all fragments within the region of interest. In Hi-C and PLAC-seq, digested DNA fragments are labelled with biotin, ligated and then fragmented further by sonication. In PLAC-seq, DNA fragments bound to a protein of interest are pulled-down by immunoprecipitation. Then, in PLAC-seq and Hi-C, the DNA is purified, biotinylated nucleotides are removed from unligated fragment ends and all ligated DNA fragments are pulled-down with streptavidin beads. After pull-down, DNA fragments are sequenced and provide information about the interaction frequencies of all pairs of loci in the genome (Hi-C) or the interactions mediated by a protein of interest (PLAC-seq).

Split-pool recognition of interactions by tag extension
(SPRITE). A ligation-free approach to detect chromatin interactions by tagging crosslinked chromatin complexes. The DNA (and RNA) molecules within an individual chromatin complex are identified after sequencing by their unique combination of barcodes that have been sequentially added using a split-pool strategy.

Chromatin immunoprecipitation
(ChIP). A method used to determine whether a given protein binds to, or is localized to, specific chromatin loci in vivo, detected after (native or crosslinked) chromatin purification and immunoprecipitation, followed by DNA detection by PCR, microarray hybridization or sequencing.

fragments, which retain the biotin label, are enriched using streptavidin beads to minimize the number of unligated DNA molecules in the sequencing library. Depending on the enrichment efficiency, about 50–70% of sequencing reads map to pairs of ligated restriction fragments in Hi-C libraries[56]. In tethered chromosome capture (TCC)[57], an early modification of Hi-C, the detection of unspecific ligation events between non-crosslinked material is minimized by tethering the crosslinked and biotinylated chromatin to streptavidin beads before ligation. This approach detects more long-range intrachromosomal contacts and contacts between chromosomes than standard 3C-technologies[57]. By contrast, genome conformation capture (GCC)[58], an approach developed at the same time as Hi-C, sequences all DNA present in the 3C library, without preselection of ligated fragments. Although currently much more expensive, especially for large genomes, GCC has the advantage of allowing direct normalization of DNA abundance, thereby controlling for biases in sequencing and for the presence of genomic alterations, such as copy number variations. Methods for detection and normalization of copy number variations have also recently been developed for Hi-C[59–61].

Many other variants of genome-wide 3C-methods have been reported, ranging from technical optimizations of the original Hi-C protocol (such as DNase Hi-C[62,63] and in situ Hi-C[64]) and advances to improve resolution (such as Micro-C)[65–67], to protocols based on the enrichment of contacts mediated by specific proteins or open chromatin regions (open chromatin enrichment and network Hi-C (OCEAN-C)[68]). Currently, the most commonly used version is in situ Hi-C. In the original Hi-C protocol, sodium dodecyl sulfate (SDS) is used to disrupt the nuclear membrane and ligation of crosslinked DNA therefore occurs partially in solution. In situ Hi-C omits this SDS step, allowing ligation of chromatin fragments within the presumably more native environment of the intact nucleus. As a result, the number of random ligation events is reduced and signal-to-noise ratios are improved, thereby reducing the sequencing depth and enabling higher-resolution contact maps. However, detailed analyses of the nuclear fragments that contribute to contacts in the original version of Hi-C showed that large portions of the chromatin were thought to remain inside the partially digested nucleus during ligation[69]. Nonetheless, the in situ Hi-C protocol is faster and easier than the original version[64], mainly because it does not require extensive dilution of the crosslinked chromatin prior to DNA ligation. Consequently, all subsequent steps can be conducted in smaller volumes, allowing more efficient ligation and DNA extraction. Easy Hi-C is another recent approach to simplify Hi-C[70]. It avoids biotin enrichment and can be used with lower cell numbers than standard Hi-C (TABLE 1).

Mapping genome-wide contacts in single cells with single-cell Hi-C. Standard Hi-C generates average contact maps from millions of cells, without any possibility to understand heterogeneity of the cell population. Single-cell Hi-C overcomes this limitation by allowing Hi-C contact maps to be produced from individual cells isolated during the process of generating Hi-C libraries[71,72]. This approach allows rare cell types to be studied[73] and helps chromosome structures to be determined at specific stages of the cell cycle[74]. The single-cell Hi-C protocol involves in situ proximity ligation of crosslinked and digested chromatin, followed by isolation of single nuclei from the cell suspension and generation of sequencing libraries from each nucleus[71,74]. Single-cell combinatorial indexed Hi-C (sciHi-C) adopts a different approach; instead of isolating single cells, DNA within each nucleus is tagged with a unique combination of barcodes[75]. First, cells are fixed, lysed and digested with a restriction enzyme. Then, the cell suspension of digested, but intact, nuclei is split into 96-well plates, indexed with individual barcodes, pooled and split again. After several rounds of indexing, in situ proximity ligation and library preparation are performed on pooled nuclei, allowing high-throughput generation of single-cell Hi-C libraries.

One of the major challenges in single-cell Hi-C is the efficient recovery of contacts: inefficient digestion and ligation and incomplete recovery of input material result in contact maps that represent only a proportion of the contacts that may exist in a single cell. Modifications of the original protocol increased the average number of contacts detected in one cell from ten thousands up to hundreds of thousands[34,74], but this remained a fraction (2–5%) of the possible contacts in the genome. Recently, the development of Dip-C (diploid chromatin conformation capture) has increased the number of detectable contacts to an average of 1 million per cell by omitting biotin incorporation and including a whole-genome amplification step in the protocol[76].

Combining 3C-based approaches with chromatin immunoprecipitation. 3C-based methods can be used to study chromatin contacts mediated by specific proteins, such as chromatin modifiers, architectural proteins, members of the transcription machinery or cell type-specific transcription factors. To explore contacts that coincide with chromatin occupancy of specific proteins, Hi-C libraries can be enriched by chromatin immunoprecipitation (ChIP) before ligation. Early methods, such as ChIP-loop[77] and enhanced 4C-ChIP (e4C)[78], required that chromatin be solubilized to enable specific immunoprecipitation before ligation. However, standard 3C conditions often do not fully solubilize chromatin, as nuclei stay mostly intact after SDS treatment[69], resulting in low signal-to-noise ratios. Other approaches, such as chromatin interaction analysis by paired-end tag sequencing (ChIA-PET), included sonication of the nuclei, as is more typically used for ChIP[79]. Although sonication allows efficient precipitation of chromatin, its influence on the outcome of the subsequent proximity ligation remains unclear. Challenges in implementing ChIA-PET have led to other strategies for combining ChIP with Hi-C, namely Hi-ChIP[80] and proximity ligation-assisted chromatin immunoprecipitation sequencing (PLAC-seq)[81]. Instead of performing protein pull-down followed by ligation of DNA fragments, Hi-ChIP and PLAC-seq perform in situ Hi-C and proximity ligation before sonication and immunoprecipitation. In this order, the ligation occurs in intact nuclei under optimal conditions, before chromatin contacts specific to the protein of interest are

enriched. Regardless of these increased efficiencies, the results from immunoprecipitated 3C-libraries should be interpreted carefully because of the bias introduced by enriching for genomic regions that are bound by the protein of interest[82].

Genomic resolution of genome-wide 3C-methods. A major consideration for any genome-wide technique is genomic resolution. Hi-C data represent interaction frequencies between genomic regions in a contact matrix, consisting of equally sized genomic bins. The bin size (resolution) depends almost entirely on the sequencing depth. Resolutions of 30 kb or lower are often preferred to study the chromatin domain and compartments but also long-range contacts between large genomic regions (such as TADs); using standard Hi-C, this requires sequencing depths of approximately 200–400 million reads in mammalian genomes. However, billions of reads become necessary for high-resolution (1-kb) data sets of the human genome that can provide detailed insights into 3D genome topology[64]. Recently, a computational approach, HiCPlus, applied deep learning to infer high-resolution contact matrices from low-resolution Hi-C data, which reduced the sequencing depth required to obtain a given resolution by a factor of 16 (REF.[83]).

Ligation-free detection of contacts

The reliance of 3C-based approaches on the ligation of the ends of DNA fragments found in a cluster of contacts favours the detection of 'simple' chromatin contacts which involve two or a few genomic regions. This bias occurs because each DNA fragment can ligate with only one or two other fragments, so not all instances of every interaction in a complex cluster are detected[84]. Thus, the full interactome of each DNA fragment is diluted by the choice of only one or two other fragments during ligation. Recently, three ligation-free approaches have been developed for genome-wide mapping of chromatin contacts: GAM[10], SPRITE[11] and ChIA-Drop[12]. These methods are orthogonal to ligation-based approaches and are starting to provide new insights into 3D genome topology. Other ligation-free approaches — tyramide signal amplification (TSA-seq)[85] and DNA adenine methyltransferase identification (DamID)[86–88] — map chromatin with respect to nuclear landmarks (such as the nuclear lamina or various nuclear bodies), thereby helping to define chromatin positions in 3D space.

Mapping contacts with nuclear structures with DamID and TSA-seq. DamID is an in vivo genome-wide method for detecting interaction sites between a protein of interest and DNA. The DNA binding domain of the protein of interest (for example, RNA polymerase II (Pol II)) is fused to the DNA adenine methyltransferase (Dam) protein from *Escherichia coli*[86–88], which specifically methylates adenines in the sequence GATC. When the fusion protein is expressed at low levels in cells, GATC sequences within or close to DNA binding sites of the protein of interest are marked by methylation. After DNA extraction, the methylated GATC sites are cut with a methylation-sensitive restriction enzyme and adapters are added to the restriction fragments to ensure only methylated binding sites are amplified and sequenced. In an interesting adaptation called targeted DamID (TaDa)[88], expression of the Dam fusion protein is restricted to a specific cell type of interest, using targeted expression systems (such as the Gal4–UAS system), which allows detection of DNA–protein interactions, in a cell type-specific manner without prior isolation or sorting of cells. DamID has been successfully used to study DNA interactions with proteins such as Lamin B1, which resulted in the genome-wide mapping of lamina-associated domains and provided spatial information about chromatin with regard to the nuclear periphery[89,90]. However, interactions between chromatin and other nuclear compartments, such as splicing speckles, are not readily detected with DamID because most of the DNA surrounding these compartments does not directly bind to the tagged proteins[91].

TSA-seq addresses this problem using tyramide signal amplification to measure the distances between chromatin and nuclear compartments[92]. In this approach, horseradish peroxidase (HRP) is conjugated to an antibody that binds to a protein specific to the nuclear compartment of interest, where it catalyses the production of biotin-conjugated tyramide free radicals, which diffuse and bind to nearby macromolecules — including DNA. Biotin-labelled DNA can be subsequently selected by biotin pull-down and sequenced to identify all genomic regions that were close enough to the protein of interest to be labelled. TSA-seq has been used to map genome-wide the distances between all genes and their nearest splicing speckle[92].

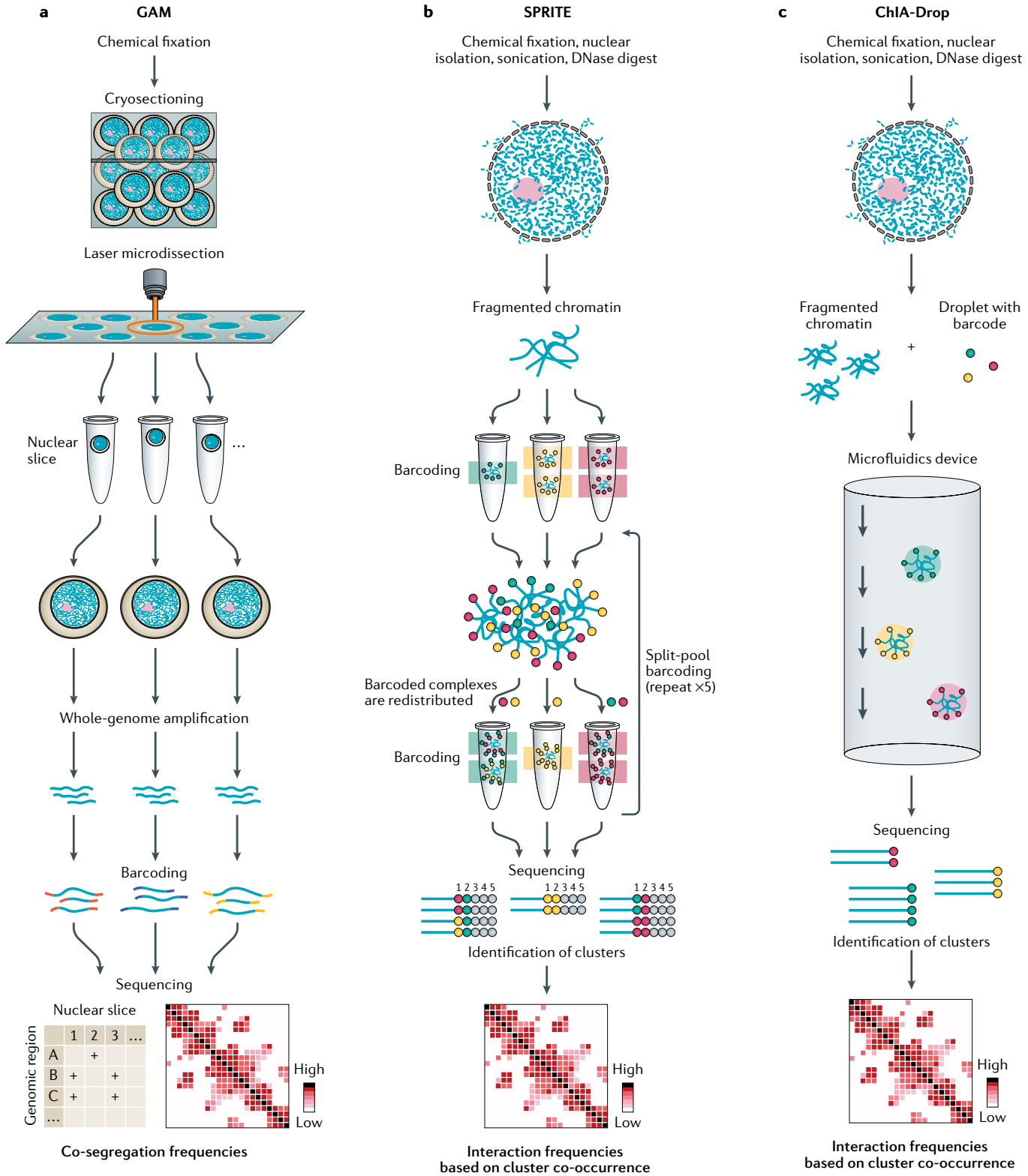
Another recent adaptation of DamID, called DamC, detects 4C-like contacts between a target region and the surrounding DNA regions, up to distances of a few hundred kilobases[93]. In DamC, Dam is fused with the reverse tetracycline receptor (rTetR), which binds to Tet operator sites inserted at the genomic region of interest. The Dam fusion protein methylates the target and its interaction partners in vivo. When combined with high-throughput sequencing, DamC reveals chromatin contacts independently of crosslinking or ligation, but unlike the other 3C-methods and ligation-free approaches it requires engineering of the cells of interest. Comparison of DamC data with 4C and Hi-C data showed high similarities at the level of TADs and CTCF loops at many genomic sites; however, some differences at loops and sub-TAD structures could also be observed[93].

Mapping all genome-wide contacts with GAM. In GAM, nuclei are sectioned in random orientations from a population of fixed and sucrose-embedded cells using ultra-thin cryosectioning (220 nm thickness). Single nuclear slices are then isolated directly from the cryosection by laser microdissection. GAM thus avoids cell extraction or sorting, both of which can disrupt cellular and nuclear structures, which can be especially important when analysing complex tissues. The DNA from every slice is extracted, whole-genome amplification is performed and indexed sequencing adapters are added before the DNA from all slices is pooled for sequencing (FIG. 4). From the sequencing data for several hundred nuclear sections, each from a single cell,

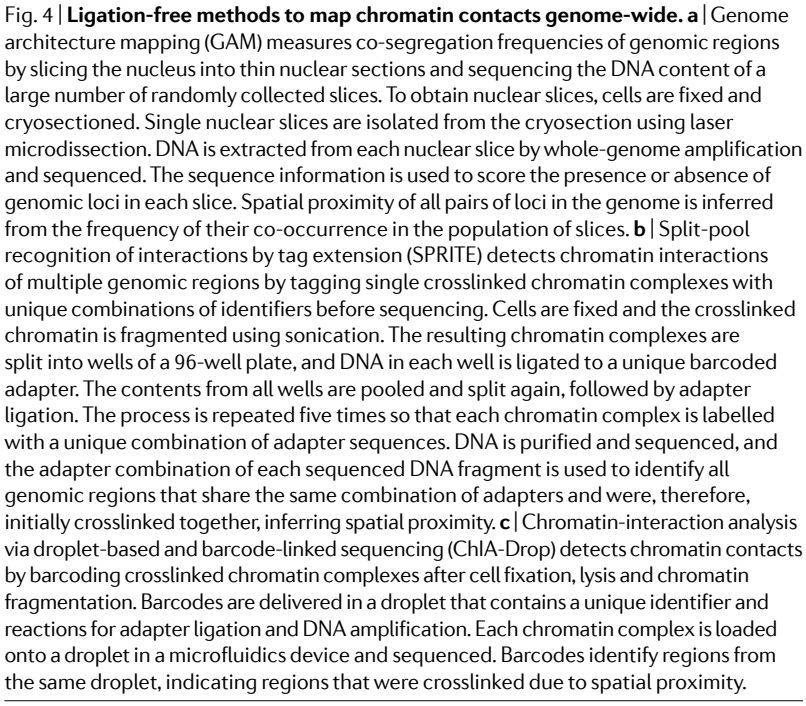
Genomic resolution
The size of the window (often in the range of kilobases) when, for most assays, reads after sequencing are mapped to the genome and then binned into equally sized genomic windows (bins).

Sequencing depth
The average number of reads representing a given nucleotide in the reconstructed sequence. A 10× sequence depth means that each nucleotide of the transcript was sequenced, on average, 10 times.

Nuclear bodies
Membrane-less compartments in the nucleus with high concentrations of DNA binding proteins, chromatin modifiers or RNAs that can be involved in shaping chromatin structure and modulating gene regulation. Nuclear bodies include the nucleolus, splicing speckles and Polycomb bodies.



chromatin contacts between pairs of DNA loci can be inferred by counting their co-segregation frequency (that is, how often the two loci are contained in the same nuclear sections). Genomic regions that are closer in 3D space are more frequently found in the same nuclear slice. To detect statistically significant interactions, GAM was combined with a mathematical model, statistical inference of co-segregation (SLICE)[10]. The most specific chromatin contacts detected with SLICE were found to contain active genomic regions, such as active enhancers and actively transcribed genes, with these contacts extending over megabases up to entire chromosomes[10]. SLICE separately models the random interactions that depend on genomic distance and



◀ Fig. 4 | **Ligation-free methods to map chromatin contacts genome-wide. a** | Genome architecture mapping (GAM) measures co-segregation frequencies of genomic regions by slicing the nucleus into thin nuclear sections and sequencing the DNA content of a large number of randomly collected slices. To obtain nuclear slices, cells are fixed and cryosectioned. Single nuclear slices are isolated from the cryosection using laser microdissection. DNA is extracted from each nuclear slice by whole-genome amplification and sequenced. The sequence information is used to score the presence or absence of genomic loci in each slice. Spatial proximity of all pairs of loci in the genome is inferred from the frequency of their co-occurrence in the population of slices. **b** | Split-pool recognition of interactions by tag extension (SPRITE) detects chromatin interactions of multiple genomic regions by tagging single crosslinked chromatin complexes with unique combinations of identifiers before sequencing. Cells are fixed and the crosslinked chromatin is fragmented using sonication. The resulting chromatin complexes are split into wells of a 96-well plate, and DNA in each well is ligated to a unique barcoded adapter. The contents from all wells are pooled and split again, followed by adapter ligation. The process is repeated five times so that each chromatin complex is labelled with a unique combination of adapter sequences. DNA is purified and sequenced, and the adapter combination of each sequenced DNA fragment is used to identify all genomic regions that share the same combination of adapters and were, therefore, initially crosslinked together, inferring spatial proximity. **c** | Chromatin-interaction analysis via droplet-based and barcode-linked sequencing (ChIA-Drop) detects chromatin contacts by barcoding crosslinked chromatin complexes after cell fixation, lysis and chromatin fragmentation. Barcodes are delivered in a droplet that contains a unique identifier and reactions for adapter ligation and DNA amplification. Each chromatin complex is loaded onto a droplet in a microfluidics device and sequenced. Barcodes identify regions from the same droplet, indicating regions that were crosslinked due to spatial proximity.

the specific interactions that occur at a given physical distance (for example, below 100 nm)[10]; it interrogates which pairs of loci co-segregate more often in the collection of slices than expected from random contacts, and quantifies the frequency of specific interaction in the cell population. GAM also allows genome-wide interactions between three or more DNA loci to be detected simultaneously, and has detected long-range contacts between TADs containing super-enhancers and highly-transcribed TADs[10]. The resolution of GAM data sets depends on the number of nuclear slices collected. With 400 nuclear slices, sequenced with ~1 million reads per slice, it was possible to achieve a resolution of 30 kb for pairwise chromatin contacts[10], comparable with a Hi-C library with similar sequencing depth[94]. Larger GAM data sets comprising a few thousand nuclear slices will help define the maximal resolution that can be practically afforded by GAM.

Mapping all genome-wide contacts with SPRITE and ChIA-Drop. SPRITE[11] and ChIA-Drop[12] detect chromatin interactions by tagging crosslinked chromatin complexes. Similar to 3C-based approaches, these methods rely on mild fixation and fragmentation of chromatin inside the nucleus – but unlike 3C-based approaches, they do not use proximity ligation. Instead, in SPRITE, the crosslinked chromatin fragments are split across a 96-well plate, where each well contains a unique barcode (FIG. 4). The indexed chromatin complexes are re-pooled, followed by sequential rounds of splitting, barcoding and pooling. The DNA (and RNA) molecules within an individual chromatin complex are identified after sequencing by their unique combination of barcodes added using this split-pool strategy; only DNA fragments that were crosslinked with each other will display the same combinations of barcodes. In ChIA-Drop, crosslinked and fragmented chromatin is separated into single chromatin complexes by droplet formation using a microfluidics device. Each droplet contains reagents for barcoding and amplification, and barcoded complexes are pooled and sequenced, as in SPRITE. SPRITE detects TADs and loop domains, both of which are features of Hi-C contact maps. However, SPRITE also detects additional genome-wide features of nuclear architecture, such as the association of specific genomic regions with nucleoli and splicing speckles. The predominant chromatin hubs around these nuclear bodies contain genomic regions from different chromosomes, an observation that is in agreement with single-cell imaging[95] but that had not been made using 3C-based assays. SPRITE also detects long-range contacts between regions containing active genes and super-enhancer regions that were first recognized as being multiway-specific interactions in a study using GAM[10].

Comparing approaches

Fundamental differences exist between current approaches for mapping 3D genome folding, including how the chromatin is fixed and prepared, their power to detect multiple chromatin contacts or contacts with different spatial distances and protein occupancy, and their ability to detect long-range contacts within the same (FIG. 5) or different chromosomes. These differences have sometimes led to observations that can be difficult to reconcile between the different approaches.

Fixation and chromatin preparation. With the exception of live-cell methods (such as DAM-based and CRISPR-based approaches), all chromatin folding techniques start by crosslinking DNA–protein complexes to stabilize nuclear structures (TABLE 2). Chemical fixation using formaldehyde is the most common approach for crosslinking, but concentrations, buffers and fixation times vary widely; for example, 1% formaldehyde is typically used for 3C-based methods, 4% for most DNA-FISH experiments in whole cells and 8% for GAM or cryo-FISH in nuclear slices. Other fixatives include solvent-based precipitation using ethanol, methanol or acetone. A recent imaging study compared the effects of formaldehyde fixation and cryofixation on nuclear structure using partial wave spectroscopy[96]. It revealed that weaker fixatives (such as 4% formaldehyde in PBS) introduce larger structural distortions than stronger fixatives (such as glutaraldehyde, often used for electron microscopy). However, the distinction between condensed and decondensed chromatin remains detectable at the population level[96], which is consistent with the ability of all current chromatin folding methods to successfully map euchromatin and heterochromatin. The effect of varying crosslinking conditions (from no fixation to 5% formaldehyde fixation) has been examined in Capture-C experiments; similar short-range interactions were detected under all conditions, but formaldehyde concentrations below 2% improved the efficiency of detection[97]. Our own previous work showed that the organization of the active form of Pol II, which marks transcription sites, can be highly disrupted with weaker fixatives, but not with the fixation regimen used for GAM or cryo-FISH[98]. In FISH, denaturation of the DNA

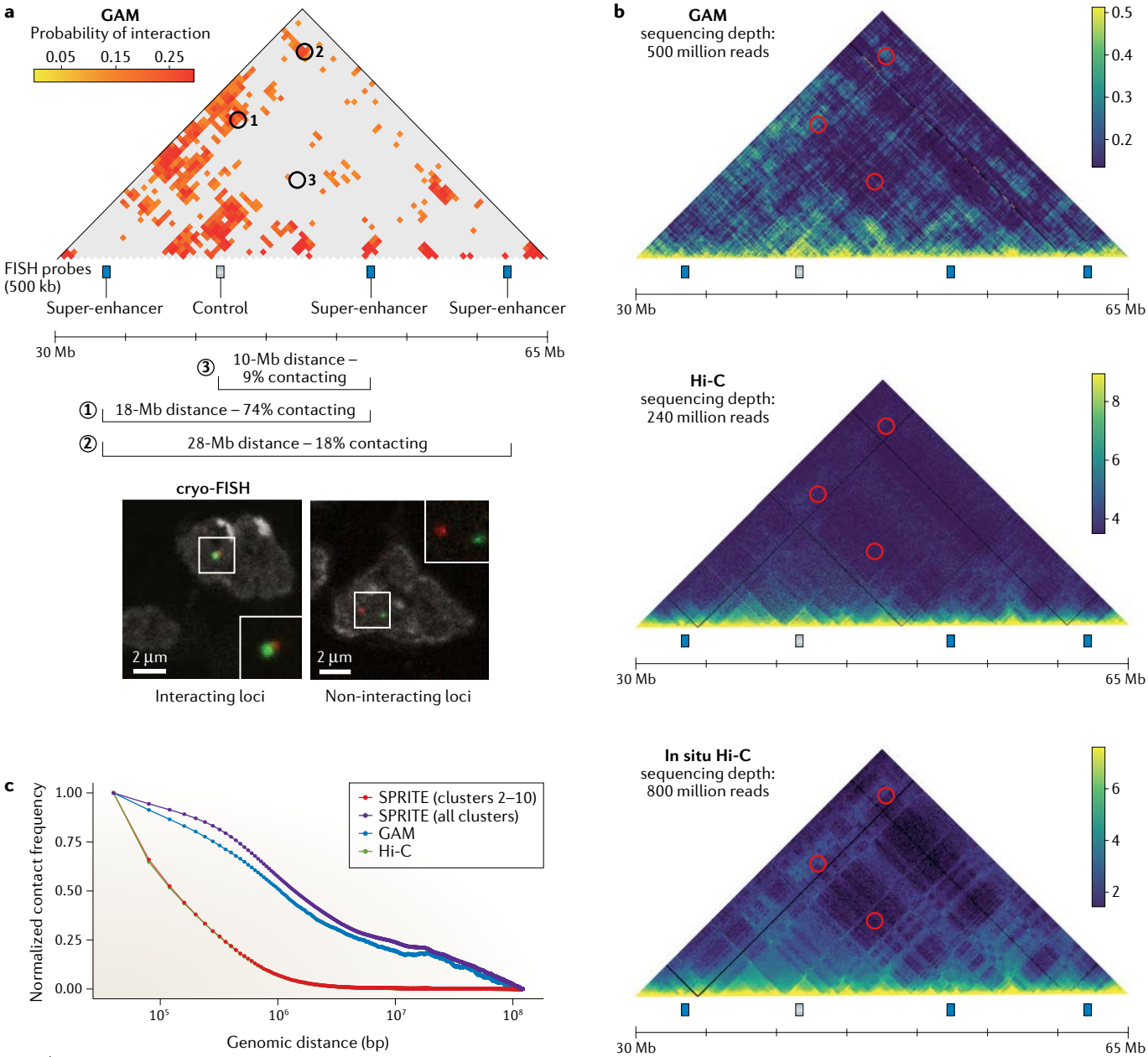


Fig. 5 | **Comparison of long-range chromatin contacts across methods.** **a** | Genome architecture mapping (GAM) detects significant interactions between super-enhancers (circles 1 and 2) that span large genomic distances (18 Mb and 28 Mb). The heat map shows GAM interaction probabilities for chromosome 11 (region 30–65 Mb) in mouse embryonic stem cells (ESCs) at 500-kb resolution[10]. Contacts between the super-enhancer regions can also be detected using cryo-fluorescence in situ hybridization (cryo-FISH), which confirmed their interactions in a high percentage of cells in the population: the 18-Mb distant super-enhancer contact and the 28-Mb distant contact were detected in 74% and 18% of cells, respectively. By comparison, contacts between a super-enhancer and a 10-Mb distant control region (circle 3) not detected by GAM were detected in only 9% of cells. The images show DAPI-stained cryosections with interacting and non-interacting 500-kb FISH probes. **b** | Long-range super-enhancer contacts can also be found when looking at GAM contacts without filtering for the most significant interactions (~500 million reads, 40-kb resolution, mouse ESCs, data from REF.[10]), and although they are not readily detected in Hi-C data with an average sequencing depth (~240 million reads, 40-kb resolution, mouse ESCs, data from REF.[94]), they start to emerge in deep-sequenced in situ Hi-C data (~800 million reads, 50-kb resolution, mouse ESCs, data from REF.[191]). Heat maps were generated by Christophe Thieme from the published, normalized matrix files. Scores are colour-coded, where the colour-code range (maximum and minimum cutoffs) is determined by the mean value of the bin distances 1–20 and −50 to −30 from the diagonal, respectively. **c** | The plot shows the distribution of contact frequencies detected by high-throughput chromosome conformation capture (Hi-C), GAM and split-pool recognition of interactions by tag extension (SPRITE) (all clusters or clusters with 2–10 reads) along the linear genomic distance of chromosome 11 in mouse ESCs, scaled to the maximum observed value in each data set. The ligation-free methods GAM and SPRITE detect similar ranges of chromatin contacts, which can extend over large genomic distances. By contrast, Hi-C contacts typically extend over shorter genomic distances. However, SPRITE data can be sorted based on the number of interactions within one chromatin complex. When considering only small SPRITE clusters with fewer than 10 genomic regions in the same chromatin cluster, the range of detection between Hi-C and SPRITE is comparable, indicating that Hi-C favours less complex short-range contacts over long-range interactions involved in chromatin hubs with many interaction partners. The plot was generated using the same data for GAM (REF.[10]) and Hi-C (REF.[94]) as used in part **b**, and data provided by Sofia Quinodoz for normalized SPRITE clusters for chromosome 11, according to figure 3B of REF.[11]. Part **a** adapted from REF.[10], Springer Nature Limited. Heat maps in part **b** and plot in part **c** courtesy of Christoph Thieme, Max Delbrueck Center, Germany.

Table 2 | **Experimental differences between chromatin contact assays and their effects on nuclear structures**

Assay	Fixation	Chromatin preparation	Effects on chromatin
2D-FISH	Hypotonic treatment, followed by methanol–acetic acid for 10 min	Permeabilization, HCl treatment, heat and formamide denaturation, physical flattening of cells through gravity and drying	The fixation process flattens cells, alters the shape of the nucleus and the structure of chromocentres, and changes the association of some loci with their chromosome territories[206–209]; denaturation can lead to loss of DNA[210]
3D-FISH	2–4% depolymerized PFA for 10 min (fixatives are buffered with PBS)	Permeabilization, HCl treatment, heat and formamide denaturation	Fixation can cause nuclei to increase in size, accompanied by a change in the positions of distant chromatin domains; overall, nuclear structures and the distribution of chromatin in the nucleus remain intact; denaturation disrupts the nuclear membrane, accompanied by loss of some heterochromatic regions and redistribution of histone H2B[89]
Cryo-FISH and GAM	4% depolymerized PFA for 10 min, followed by 8% PFA for 2 h (fixatives are buffered with 0.25 M HEPES)	Sucrose-embedding for cryoprotection, cryosectioning. For cryo-FISH, slices are permeabilized, treated with HCl, heat and formamide denaturation	Stringent, electron microscopy-grade fixation preserves nuclear and cytoplasmic ultrastructures and is indistinguishable from glutaraldehyde fixation[98]; robust fixation renders nuclear structures more resilient to the negative effects of heat denaturation for FISH, with good preservation of nuclear compartments before and after FISH[25,101]
3C-based methods, SPRITE and ChIA-Drop	1% formaldehyde for 10 min (fixatives are buffered with cell culture medium or PBS)	Cell lysis, SDS treatment, fragmentation of the genome (for example, restriction digest, DNase treatment, sonication)	Mild fixation is required to allow subsequent restriction digest, but can result in redistribution of nuclear proteins (similar fixation studied in REF.[98]); after chromatin preparation, restriction enzymes disrupt nuclear structures, nuclei swell and chromatin is distributed more uniformly[69]; DNA-FISH on digested nuclei shows that chromosome territories seem larger and less condensed compared with fixed cells before digestion, but DNA is maintained in territories[69]

3C, chromosome conformation capture; ChIA-Drop, chromatin-interaction analysis via droplet-based and barcode-linked sequencing; DNA-FISH, fluorescence in situ hybridization of DNA; FISH, fluorescence in situ hybridization; GAM, genome architecture mapping; HCl, hydrogen chloride; PFA, paraformaldehyde; SDS, sodium dodecyl sulfate; SPRITE, split-pool recognition of interactions by tag extension.

by heat and formamide induces fine structural changes in chromatin folding, such as slight distortions of the interchromatin space[99,100]. However, 3D-FISH preserves the organization of centromeres seen by imaging the same cells before and after hybridization[100], and cryo-FISH retains the organization of active Pol II sites[101]. In another method, resolution after single-strand exonuclease resection (RASER)-FISH, heat denaturation of the DNA is avoided and DNA accessibility is achieved by exonuclease digestion, thereby reducing the effects of DNA denaturation[102].

Multiplicity of chromatin contacts. The dependency of 3C-based methods on DNA end ligation results in preferential detection of low-multiplicity contacts that involve only a few genomic regions[84]. However, every 3C library also includes interaction events that occur at complex clusters involving more than two DNA fragments, albeit at a low representation. Current methods to capture these higher-complexity ligation events include multi-contact 4C (MC-4C)[103], which uses long-read sequencing (such as nanopore sequencing) of 4C libraries to capture three-way contacts of a region of interest, and chromosomal walks (C-walks)[104], which implement multiple ligation steps followed by dilution and barcoding of the isolated ligation products. Alternatively, methods such as the concatemer ligation assay (COLA)[105] and Tri-C[106] generate 3C libraries with a restriction enzyme that cuts small DNA fragments, which increases the frequency of detecting multiple ligation events in one sequencing read. Estimates based on direct comparison of pairwise and multiway ligation events indicate that only 17% of chromatin contacts in mouse embryonic stem cells (ESCs) are pairwise contacts and, therefore, the majority of the genome is involved in higher-order contacts between more than two genomic loci[104]. These observations are supported by a recent study using SPRITE, which showed that classical ligation-dependent methods under-represent higher complexity contacts[11] (FIG. 5c). Assays that do not depend on ligation detect DNA fragments that are in spatial proximity regardless of the number of interacting genomic loci. For example, long-range multiway contacts between genomic regions harbouring super-enhancers were readily found in GAM[10], FISH[10] and SPRITE[11] data, but had not previously been detected with Hi-C. Furthermore, analyses of triplet interactions between TADs in GAM showed that multiple interactions between super-enhancer regions and active genes are a common feature of genome conformation in mouse ESCs[10].

Spatial distance between contacting genomic regions. The spatial distance between genomic loci is thought to influence the probability of ligation irrespective of the frequency of contacts. Whereas cryo-FISH and SPRITE have readily detected abundant interchromosomal

contacts in human, mouse and Drosophila cells[25,76,107,108], 3C-based methodologies are more often used to explore specific contacts within chromosomes, with some exceptions[27,76,109–111]. Recent CRISPR–Cas9 live-cell imaging of a small number of chromatin contacts within and between chromosomes showed that interchromosomal contacts display spatial distances in the range of ~280 nm, in contrast to distances of ~190 nm for intrachromosomal interactions[20]. Interestingly, only the intrachromosomal contacts could be observed in matching Hi-C data, indicating a dependency of close spatial distances for successful proximity ligation.

Protein-mediated interactions versus bystander contacts. GAM and all imaging-based techniques collect all possible spatial relationships between genomic regions, regardless of their involvement in a protein-mediated interaction, and allow sampling of the whole range of spatial distances within the interphase nucleus. Thus, these methods also detect bystander contacts. However, it is possible to identify the most specific contacts through effective sampling to take into account all behaviours of all genomic regions at all linear distances across the cell population. In this regard, GAM currently has more statistical power than FISH as it samples all possible combinations, whereas FISH remains limited to the analyses of a subset of regions or chromosomes.

Levels of concordance between different methods. The validation of results obtained by 3C-based methods often entails the use of DNA-FISH on a few selected loci. Many examples show agreement between 3C interaction frequencies and spatial distances measured by FISH, especially at large genomic distances[44,64,112–115]. Loci in the same TAD are often closer in nuclear distance than loci in different TADs[24,94], and interaction frequencies obtained from Hi-C correlate with spatial distances at and above the TAD level[115]. A linear relationship between Hi-C contacts and FISH distances was found by investigating the physical distances between all TADs along a chromosome[115]. An overall correlation between Hi-C interactions and the median spatial distance measured by high-throughput FISH have recently been shown for 90 pairs of loci. However, the range of physical distances between genomic regions containing Hi-C interactors (with high ligation frequency) and non-interactors (with low ligation frequency) overlap extensively, with about 20% of distances being closer between two non-interactors than two interactors[21]. Thus, Hi-C captures spatial proximity but Hi-C interactions are not easily translated into physical distances. Other comparisons between FISH and 3C-based methods have also found non-trivial relationships between physical distance distributions and population-average interaction frequencies[113] and show that contact frequency is distinct from average spatial distance, both in polymer simulations and in experimental data[116].

The use of FISH to validate Hi-C results has helped investigate false positives in Hi-C data, assuming FISH is correct, but is not a valid strategy for an unbiased search for contacts that are missed by Hi-C (that is, false negatives). Thus, any under-represented contacts in Hi-C data have so far not been systematically studied. The development of orthogonal genome-wide ligation-free approaches, such as GAM and SPRITE, have been able to identify new aspects of 3D genome folding that had not been detected by Hi-C but which are fully validated by FISH[10,11]. The first and relatively small GAM data set, combined with the mathematical model SLICE, identified specific long-range contacts across genomic distances that span tens of megabases, which involve active and enhancer-rich genomic regions (FIG. 5a,b). One promising outcome of the emergence of these orthogonal approaches is the development of analysis tools that use the information they generate about such long-range contacts to discover the same contacts in Hi-C data. In this regard, it is interesting to note that CTCF depletion in human cells results in the detection by Hi-C of long-range contacts between super-enhancers[117], which raises the possibility that CTCF-mediated contacts may be preferentially detected by Hi-C in normal conditions, but once CTCF-dependent interactions are lost, other underlying folding patterns, including long-range contacts, become easier to detect.

The first SPRITE data set has also highlighted novel aspects of 3D folding that are not readily captured by Hi-C[11]. By discriminating contacts according to their multiplicity, SPRITE shows a contact decay with genomic distance that is Chromosome territories and interchromosomal very similar to Hi-C when considering only low-complexity SPRITE clusters (2–10 genomic regions per contact hub; FIG. 5c). By contrast, SPRITE shows a striking abundance of long-range contacts when considering also higher-order contacts, which confirms early theoretical predictions that ligation-based approaches are biased to the detection of more simple 3D chromatin contacts[84]. Although GAM and SPRITE are orthogonal methodologies, their frequency of contacts relative to genomic distance are remarkably concordant[10,11] (FIG. 5c).

Limitations and applications of different methodologies. Methods that use proximity ligation are limited by the low efficiency of ligation, and are also potentially affected by the local distance between, or the topology of, the two DNA ends within the cluster of contacting DNA fragments. SPRITE also depends on ligation of a small oligo to each DNA end in a contact cluster; however, it is no longer dependent on the physical distance between two DNA fragments in the cluster, which allows mapping of all contacts within one chromatin complex. In 3C-based methods and SPRITE, detection of contacts depends on the efficiency of the fragmentation step to expose the DNA end. In GAM, there is no DNA restriction digest or ligation, and the detection of DNA depends on its extractability and sequencing depth. 3C-based methods, GAM, SPRITE and FISH can be applied directly to cells, tissues or organisms, whereas insertions of DNA binding site arrays (such as the Lac operator system), CRISPR-based imaging and Dam-related methods require genetic engineering of cell lines or whole organisms, and will not be suitable for the analyses of most human biopsies.

Each of the assays discussed here has different limitations and applications, and thus contributes to our

current understanding of 3D genome folding in different ways. 3C-based techniques have the advantage of providing enormous amounts of chromatin contact information in one comparably simple biochemical experiment, although they may require high-depth sequencing when aiming for high resolution. 3C-based methods, and in particular proximity ligation itself, also have important limitations that favour the detection of more simple contacts over higher-order chromatin contacts, which can lead to misunderstanding the importance and abundance of certain interactions. However, 3C-based techniques are well suited for studying local chromatin folding within the range of kilobases up to a few megabases.

Imaging and ligation-free methods have the ability to detect chromatin contacts at all scales of chromosome folding, including contacts between chromosomes. GAM and SPRITE can be readily used for sequence-unbiased genome-wide explorations, whereas detection of contacts with DNA-FISH remains limited to preselected loci and is most often used to validate findings from genome-wide techniques. Imaging fluorescently labelled chromatin loci in live cells with CRISPR-based techniques will improve our understanding of possible artefacts resulting from chromatin preparation or fixation. Other developments based on cryo-focused ion beam (cryo-FIB) milling of intact, frozen cells[118] or cryolysis[119] also hold the potential of devising fixation-free versions of GAM and SPRITE that sample fractionated frozen nuclei.

Insights into chromatin organization

Each of the techniques available for studying 3D chromatin folding has provided important structural and functional insight into the different hierarchical levels of chromatin organization.

Chromosome territories and interchromosomal contacts. FISH imaging shows that specific chromosomes occupy discrete non-random nuclear spaces during interphase, termed chromosome territories[120] (FIG. 1). Chromosome territories show cell type-dependent preferences in terms of both their radial position within the nucleus and their position relative to other chromosomes[25,107,121]. Specific contacts can be detected at the interface between chromosome territories[20,122,123]; overall, an estimated 20% of the volume of chromosome territories intermingles with other chromosome territories, often at their peripheries, both in human primary lymphocytes[24] and in *Drosophila melanogaster* cells[25,108]. The extent of intermingling between chromosome territories directly correlates with translocation probabilities upon ionizing radiation damage, highlighting that the physical proximity between chromosomes affects their stability in response to DNA damage[25,124,125]. The organization of chromosomes into discrete territories is also inferred from 3C-based and ligation-free approaches, as higher interaction frequencies are detected within chromosomes than between them (FIG. 1). 3C-based technologies have also detected contacts between chromosomes, and these have been successfully validated by imaging[52,74,78,110,122,126–128].

Chromatin hubs and compartments. The organization of chromosomes into subchromosomal domains has been extensively studied. For example, in mammalian cells, chromatin domains were observed in relation to replication origins, which contain many replicons and maintain their domain co-association across subsequent cell cycles[129]. The compartmentalization of chromosomes into early and late replicating domains[130–132] was also shown to be linked to transcriptional activity, with sites of active transcription occurring predominantly in early replicating domains[133]. More recently, these observations have been largely confirmed by genome-wide assays to map replication and transcription, in which transcriptionally active and early replicating chromatin domains organize into separate subcompartments, distinct from late replicating domains[134–136]. Early analyses of nuclear organization by electron and confocal microscopy had shown that chromatin occurs in highly condensed (heterochromatic) and less condensed (euchromatic) states[137], and revealed that transcription occurs in euchromatic areas of the nucleus[138]. With the emergence of whole-genome 3C-based methodologies, such as Hi-C, the mapping of active and repressed chromatin states has become possible at the genome-wide scale, providing powerful insights into how gene expression relates to chromatin compaction. Application of principal component analysis to Hi-C data revealed a strong segregation of ligation events into two distinct compartments (A and B compartments) according to the activity state of the genomic regions[44]. These compartments can also be seen in contact maps generated by ligation-free approaches[10,11] (FIG. 1). Comparisons with linear maps of protein occupancy on chromatin helped reveal a strong relationship between the A compartment and transcriptionally active, open chromatin, as defined by DNase hypersensitivity, and the B compartment with closed chromatin, defined by repressive epigenetic marks of heterochromatin[44]. Increased depth of Hi-C data sets has since allowed smaller subcompartments to be detected, which capture fine differences in replication timing as well as preferred associations with the nucleolus or the nuclear lamina[64].

Nuclear compartments or domains. Nuclear compartments are membrane-free organelles enriched for specific nuclear proteins and RNAs, which often have preferred associations with specific genomic regions and thereby influence the large-scale organization of chromosomes during interphase. They include the nucleolus, nuclear lamina, splicing speckles, paraspeckles, Cajal bodies, promyelocytic leukaemia bodies, Polycomb bodies, replication factories and transcription factories, which have all been described initially using microscopy[139,140] (FIG. 1). For example, active ribosomal gene clusters are localized in the nucleolus, where the large ribosomal RNAs are transcribed, processed and assembled into pre-ribosomes[141]. Splicing speckles occupy internal nuclear positions, separate from the nuclear lamina and nucleoli, and bring together gene-dense regions[91,142,143]. Association between specific genes at splicing speckles has been shown using imaging techniques, and has been confirmed at the genome-wide

level with SPRITE, which revealed that regions from different chromosomes come together at the same speckles[11]. Genome-wide mapping of gene association with speckles has also recently been achieved by TSA-seq[92]. Fluorescence microscopy and electron microscopy showed that transcription itself occurs at discrete sites in the nucleus, termed transcription factories, which may organize active transcription units[144–146], with only a small proportion of transcriptional activity (~5–10%) being found immediately adjacent to the most prominent splicing speckles[146]. Interestingly, the fraction of the genome that associates closely with slicing speckles has been shown by TSA-seq to contain highly transcribed genes and super-enhancers[92], in keeping with previous imaging data[142]. Co-expressed genes can share the same transcription factory, which may be compatible with mechanisms of coordinated gene regulation via chromatin folding[26,78,147,148], but it remains unclear whether transcription factories are strictly specialized. Recent findings show that several factors involved in the transcription process, such as Pol II[149] or transcriptional co-activators BRD4 and MED1 (REF.[150]), can form condensates by liquid–liquid phase separation, a process that may concentrate transcription factors and generate transcription factories. Moreover, the formation of nuclear condensates has been suggested as a general principle of nuclear body formation[151]. Clustering of distant genomic regions is not only mediated by transcription, but also occurs in the context of gene repression. Chromatin contacts at Polycomb bodies, which are repressive nuclear compartments, are a prominent example of gene clustering. In *D. melanogaster*, Polycomb-repressed Hox genes come together over a genomic distance of 10 Mb when they interact with a Polycomb body[152]. Other studies have reported long-range intrachromosomal and interchromosomal contacts between Polycomb-bound genes in human teratocarcinoma cells[153] and in mouse ESCs[52].

Understanding how the preferential associations of genomic regions with specific nuclear domains relate to 3C-derived chromatin contacts remains a major challenge. Comparisons of genome-wide maps of lamina-associated domains[90] and Hi-C contacts show a strong coincidence between the transcriptionally inactive B compartment and the nuclear lamina[64,94,154] or late replication domains[136]. Repressive histone marks that define the heterochromatic B compartment are also strongly enriched at genomic regions that associate with the nucleolus[155], suggesting that the compacted B compartment is both situated at the nuclear periphery and clustered around the more central nucleoli, separated by the active, open A compartment. However, the bimodal separation of the A and B compartments derived from 3C-technologies should not be naively inferred as strictly active or silent chromatin. Genes can be activated in all areas of the nucleus, including at the nuclear lamina[156] or at centromeric regions[157], and gene positioning at the periphery does not always lead to gene inactivation[158,159]. Heterochromatin domains also contain active sites of transcription[160]. Consequently, a strict separation of the A and B compartments, as defined by 3C-approaches, seems unlikely, especially considering that contacts between compartments can be found in Hi-C maps[94] and that they are found even more robustly using orthogonal methods, such as GAM, that do not rely on weak fixations[10]. These observations suggest that long-range gene-regulation mechanisms are complex, and not only depend on pairwise contacts between genomic regions but may also be influenced by the local nuclear environment where each region is located[95]. The ongoing challenge of disentangling the direct functional relationships between the positions of genomic regions in the nucleus, their local and long-range contacts, and the state of gene activity is being addressed by analysing chromatin contacts at the single-cell level and with allele specificity, for example, using DNA-FISH[21] or single-cell Hi-C[73,74].

TADs and loop domains. At smaller scales, chromosomes fold into self-associating chromatin domains, termed TADs[24,94,161] (FIG. 1). Chromatin domains had been previously identified by microscopy but their detailed genomic composition was unclear. Since the discovery of TADs, the segmentation of the genome into megabase-sized domains has been extensively studied in several organisms and with different methodologies, leading to major breakthroughs in the discovery of mechanisms of disease caused by congenital genomic rearrangements[3,47,162,163]. TADs often enclose clusters of co-regulated enhancers and promoters[164,165]. Their size has been re-examined with the increasing resolution afforded by improved 3C-based assays, and found to vary from 40 kb to 3 Mb in the human genome[64], leading to the proposal of smaller loop domains as a substructure of TADs. Loop domains had been detected by microscopy before the emergence of 3C-technologies as DNA loops between transcriptionally active regions[166]. Loop domains derived from 3C-based technologies often coincide with pairs of convergent CTCF binding sites, indicating that CTCF binding can contribute to the partition of specific regions of the genome into self-associating domains[64,167–169]. Higher-order contacts between TADs have also been investigated, leading to the identification of metaTADs, which bring together distant TADs in cell type-specific patterns that relate to gene activity[154,170].

It has been debated whether TADs represent domains that exist predominantly across the cell population or represent an average of individual preferred contacts. Although interactions observed in single cells by single-cell Hi-C and imaging do not often identify whole TADs, the contacts detected frequently occur with the TAD coordinates defined by population Hi-C[32,34,72]. However, this preference might not be as strong as anticipated. Imaging of chromatin contacts in mouse ESCs and oocytes showed that in 40% of cases 3D physical distances between regions that flank TAD borders are shorter than distances between regions within TADs[73], leading to highly variable contact clusters in individual cells that do not coincide with the positions of TADs in the cell population. This observation agrees with the detection of chromatin contacts between regions separated by TAD borders in single cells, often found at similar frequencies to regions within TADs[21]. However, it is particularly noteworthy that combining the single-cell Hi-C data results in the same TAD coordinates

observed in bulk population Hi-C, which supports the idea that TADs represent contact preferences of a cell population, rather than compact domains of chromatin in single cells[73,171].

Chromatin contacts between cis-regulatory elements. Physical contacts between enhancers and promoters are essential for the transcription of genes[85] and can occur over distances ranging from less than 1 kb up to several megabases[172–176] (FIG. 1). Genome-wide maps of candidate promoter–enhancer contacts can be created using high-resolution 3C-based methodologies that enrich for contacts mediated by Pol II or promoter histone marks, or that preferentially capture promoter-based contacts[52,80,81]. Direct pairwise contacts between gene promoters and enhancers have become the most prominent concept of enhancer function, possibly as a result of the increased power of 3C-based technologies to detect local pairwise contacts rather than higher-order conformations. However, other mechanisms for regulating enhancer function are also emerging, which can involve formation of chromatin hubs, tethering of genes to active chromatin or nuclear environments[156,158,177,178] and phase separation[179,180]. An interesting study in budding yeast suggests homologue pairing as a mechanism for gene activation[181]. In the diploid yeast genome, upon glucose deprivation of the cell, both copies of the genomic locus containing the gene *TDA1* are relocalized to the nuclear periphery, where the homologues associate with each other and *TDA1* expression is activated. A more classical concept of gene regulation can be observed at developmental loci, where *cis*-regulatory contacts between enhancers and promoters are thought to occur most commonly within TADs[48,162,182]. Although regulatory landscapes within TADs seem to be a common mechanism, genes themselves also contact each other across TAD boundaries over large genomic distances[10,152,153,183]. Ligation-free methods, such as FISH, GAM and SPRITE, all detect long-range contacts across TAD borders[10,11,154], and detailed analyses of Hi-C ligation frequencies also identify ligation events across TADs, over tens of megabases, that are statistically different from random contacts[154]. The functional relevance of these contacts is a compelling question that is beginning to be addressed by developments that allow ectopic chromatin contacts to be engineered in the cell[184,185]. The spatial and functional relationship between gene promoters that contact each other also remains poorly understood. Deletions of several gene promoters in the mouse ESC genome altered the expression of nearby genes[186]. This observation suggests that genes themselves may act as enhancers for other genes, possibly by recruiting *cis*-regulatory signals, and supports the concept that clustering of genes in transcription factories has regulatory functions.

Conclusions

The development of genome-wide approaches for studying 3D genome folding have revolutionized our ability to understand the regulatory content of the linear genomic sequence. Alongside 3C-based methods, the recent development of orthogonal technologies to map chromatin contacts brings us closer to uncovering 3D genome folding architectures with unprecedented detail, at all genomic scales and with single-cell resolution. The ongoing revolution in live-cell imaging, including improvements to the number of genomic loci that can be tagged simultaneously, will provide a deeper mechanistic understanding of how 3D folding structures are formed and disassembled, and how they contribute to genome stability gene expression, and of homeostatic changes in cell states in response to stimuli. Ultimately, the ability to detect changes in chromosome topology will open new avenues for disease diagnostics, disease target discovery and many other applications.

Published online 17 December 2019

1. Pombo, A. & Dillon, N. Three-dimensional genome architecture: players and mechanisms. *Nat. Rev. Mol. Cell Biol.* **16**, 245–257 (2015).
2. Dekker, J. et al. The 4D nucleome project. *Nature* **549**, 219–226 (2017).
3. Spielmann, M., Lupiáñez, D. G. & Mundlos, S. Structural variation in the 3D genome. *Nat. Rev. Genet.* **19**, 453–467 (2018).
This review explains the influence of disrupted chromatin folding on gene regulation in disease with examples from developmental disorders.
4. Krijger, P. H. & de Laat, W. Regulation of disease-associated gene expression in the 3D genome. *Nat. Rev. Mol. Cell Biol.* **17**, 771–782 (2016).
5. Gall, J. G. & Pardue, M. L. Formation and detection of RNA–DNA hybrid molecules in cytological preparations. *Proc. Natl Acad. Sci. USA* **63**, 378–383 (1969).
6. Speicher, M. R., Gwyn Ballard, S. & Ward, D. C. Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nat. Genet.* **12**, 368–375 (1996).
7. Wang, S. et al. Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **353**, 598–602 (2016).
8. Ma, H., Reyes-Gutierrez, P. & Pederson, T. Visualization of repetitive DNA sequences in human chromosomes with transcription activator-like effectors. *Proc. Natl Acad. Sci. USA* **110**, 21048–21053 (2013).
9. Ma, H. et al. Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nat. Biotechnol.* **34**, 528–530 (2016).
10. Beagrie, R. A. et al. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519–524 (2017).
This study introduces GAM, a ligation-free technique to map chromatin contacts genome-wide. GAM confirms the presence of TADs and reveals multiway contacts between super-enhancers spanning tens of megabases.
11. Quinodoz, S. A. et al. Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell* **174**, 744–757.e24 (2018).
This paper describes SPRITE, a ligation-free approach to map chromatin interactions and DNA–RNA contacts across the entire genome. SPRITE detects chromosomal hubs at nuclear bodies that bring together regions from different chromosomes.
12. Zheng, M. et al. Multiplex chromatin interactions with single-molecule precision. *Nature* **566**, 558–562 (2019).
13. Cutter, A. R. & Hayes, J. J. A brief review of nucleosome structure. *FEBS Lett.* **589**, 2914–2922 (2015).
14. Robinett, C. C. et al. In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *J. Cell Biol.* **135**, 1685–1700 (1996).
15. Belmont, A. S. & Straight, A. F. In vivo visualization of chromosomes using lac operator-repressor binding. *Trends Cell Biol.* **8**, 121–124 (1998).
16. Lucas, J. S., Zhang, Y., Dudko, O. K. & Murre, C. 3D trajectories adopted by coding and regulatory DNA elements: first-passage times for genomic interactions. *Cell* **158**, 339–352 (2014).
17. Germier, T., Sylvain, A., Silvia, K., David, L. & Kerstin, B. Real-time imaging of specific genomic loci in eukaryotic cells using the ANCHOR DNA labelling system. *Methods* **142**, 16–23 (2018).
18. Barutcu, A. R., Maass, P. G., Lewandowski, J. P., Weiner, C. L. & Rinn, J. L. A TAD boundary is preserved upon deletion of the CTCF-rich Firre locus. *Nat. Commun.* **9**, 1444–1455 (2018).
19. Barbieri, M. et al. Active and poised promoter states drive folding of the extended HoxB locus in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* **24**, 515–524 (2017).
20. Maass, P. G., Barutcu, A. R., Weiner, C. L. & Rinn, J. L. Inter-chromosomal contact properties in live-cell imaging and in Hi-C. *Mol. Cell* **69**, 1039–1045.e3 (2018).
21. Finn, E. H. et al. Extensive heterogeneity and intrinsic variation in spatial genome organization. *Cell* **176**, 1502–1515.e10 (2019).
22. Boettiger, A. N. et al. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529**, 418–422 (2016).
23. Luperchio, T. R. et al. Chromosome conformation paints reveal the role of lamina association in genome organization and regulation. Preprint at *bioRxiv* https://doi.org/10.1101/122226 (2017).
24. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
This study uses 5C to identify the compartmentalization of chromosomes into TADs.
25. Branco, M. R. & Pombo, A. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent

associations. *PLOS Biol.* **4**, e1380780–e1380788 (2006).
This microscopy study of contacts between chromosomes uses cryo-FISH to show the previously underestimated extent of chromosome intermingling between chromosome territories.
26. Ferrai, C. et al. Poised transcription factories prime silent uPA gene prior to activation. *PLOS Biol.* **8**, e1000270 (2010).
27. Simonis, M. et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C). *Nat. Genet.* **38**, 1348–1354 (2006).
28. Beliveau, B. J. et al. Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl Acad. Sci. USA* **109**, 21301–21306 (2012).
29. Boyle, S., Rodesch, M. J., Halvensleben, H. A., Jeddeloh, J. A. & Bickmore, W. A. Fluorescence in situ hybridization with high-complexity repeat-free oligonucleotide probes generated by massively parallel synthesis. *Chromosome Res.* **19**, 901–909 (2011).
30. Beliveau, B. J. et al. Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nat. Commun.* **6**, 7147–7160 (2015).
31. Gnirke, A. et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* **27**, 182–189 (2009).
32. Bintu, B. et al. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, eaau1783 (2018).
33. Ni, Y. et al. Super-resolution imaging of a 2.5 kb non-repetitive DNA in situ in the nuclear genome using molecular beacon probes. *eLife* **6**, 21660 (2017).
This elegant study uses super-resolution microscopy and DNA-FISH to fine-map enhancer–promoter contacts.
34. Stevens, T. J. et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59–64 (2017).
35. Gibcus, J. H. et al. A pathway for mitotic chromosome formation. *Science* **359**, eaao6135 (2018).
36. Chen, B. et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–1491 (2013).
37. Shao, S. et al. Long-term dual-color tracking of genomic loci by modified sgRNAs of the CRISPR/Cas9 system. *Nucleic Acids Res.* **44**, e86–e98 (2016).
38. Fu, Y. et al. CRISPR–dCas9 and sgRNA scaffolds enable dual-colour live imaging of satellite sequences and repeat-enriched individual loci. *Nat. Commun.* **7**, 11707–11714 (2016).
39. Wang, S., Su, J. H., Zhang, F. & Zhuang, X. An RNA-aptamer-based two-color CRISPR labeling system. *Sci. Rep.* **6**, 26857–26863 (2016).
40. Gu, B. et al. Transcription-coupled changes in nuclear mobility of mammalian *cis*-regulatory elements. *Science* **359**, 1050–1055 (2018).
41. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
42. Zhao, Z. et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* **38**, 1341–1347 (2006).
43. Dostie, J. et al. Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* **16**, 1299–1309 (2006).
44. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
This article describes Hi-C, a 3C-based approach to map chromatin contacts genome-wide using selection of ligated DNA fragments.
45. Denker, A. & de Laat, W. The second decade of 3C technologies: detailed insights into nuclear organization. *Genes Dev.* **30**, 1357–1382 (2016).
46. van de Werken, H. J. et al. Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat. Methods* **9**, 969–972 (2012).
47. Franke, M. et al. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265–269 (2016).
48. Symmons, O. et al. The Shh topological domain facilitates the action of remote enhancers by reducing the effects of genomic distances. *Dev. Cell* **39**, 529–543 (2016).
49. Loviglio, M. N. et al. Chromosomal contacts connect loci associated with autism, BMI and head circumference phenotypes. *Mol. Psychiatry* **22**, 836–849 (2017).
50. Kundu, S. et al. Polycomb repressive complex 1 generates discrete compacted domains that change during differentiation. *Mol. Cell* **65**, 432–446.e5 (2017).
51. Hughes, J. R. et al. Analysis of hundreds of *cis*-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat. Genet.* **46**, 205–212 (2014).
52. Mifsud, B. et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **47**, 598–606 (2015).
53. Liu, X. et al. In situ capture of chromatin interactions by biotinylated dCas9. *Cell* **170**, 1028–1043.e19 (2017).
54. Andrey, G. et al. Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding. *Genome Res.* **27**, 223–233 (2017).
55. Schoenfelder, S., Javierre, B. M., Furlan-Magaril, M., Wingett, S. W. & Fraser, P. Promoter capture Hi-C: high-resolution, genome-wide profiling of promoter interactions. *J. Vis. Exp.* **136**, e57320 (2018).
56. Belton, J. M. et al. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276 (2012).
57. Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F. & Chen, L. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat. Biotechnol.* **30**, 90–98 (2011).
58. Rodley, C. D. M., Bertels, F., Jones, B. & O'Sullivan, J. M. Global identification of yeast chromosome interactions using genome conformation capture. *Fungal Genet. Biol.* **46**, 879–886 (2009).
This paper describes GCC, a 3C-based approach to map chromatin contacts genome-wide without selection of ligated DNA fragments.
59. Servant, N., Varoquaux, N., Heard, E., Barillot, E. & Vert, J.-P. Effective normalization for copy number variation in Hi-C data. *BMC Bioinformatics* **19**, 313–313 (2018).
60. Dixon, J. R. et al. Integrative detection and analysis of structural variation in cancer genomes. *Nat. Genet.* **50**, 1388–1398 (2018).
61. Vidal, E. et al. OneD: increasing reproducibility of Hi-C samples with abnormal karyotypes. *Nucleic Acids Res.* **46**, e49–e58 (2018).
62. Ma, W. et al. Fine-scale chromatin interaction maps reveal the *cis*-regulatory landscape of human lincRNA genes. *Nat. Methods* **12**, 71–78 (2015).
63. Ma, W. et al. Using DNase Hi-C techniques to map global and local three-dimensional genome architecture at high resolution. *Methods* **142**, 59–73 (2018).
64. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
This article describes a high-resolution contact map of the human genome, which reveals the presence of loop domains and introduces the concept of loop formation between divergent CTCF sites.
65. Hsieh, T.-Han S. et al. Mapping nucleosome resolution chromosome folding in yeast by micro-C. *Cell* **162**, 108–119 (2015).
66. Hsieh, T. S., Fudenberg, G., Goloborodko, A. & Rando, O. J. Micro-C XL: assaying chromosome conformation from the nucleosome to the entire genome. *Nat. Methods* **13**, 1009–1011 (2016).
67. Hsieh, T.-H. S. et al. Resolving the 3D landscape of transcription-linked mammalian chromatin folding. Preprint at *bioRxiv* https://doi.org/10.1101/638775 (2019).
68. Li, T., Jia, L., Cao, Y., Chen, Q. & Li, C. OCEAN-C: mapping hubs of open chromatin interactions across the genome reveals gene regulatory networks. *Genome Biol.* **19**, 54–68 (2018).
69. Gavrilov, A. A. et al. Disclosure of a structural milieu for the proximity ligation reveals the elusive nature of an active chromatin hub. *Nucleic Acids Res.* **41**, 3563–3575 (2013).
This paper presents electron microscopy and confocal microscopy images showing the changes that occur in chromatin ultrastructure during a proximity-ligation assay.
70. Lu, L., Liu, X., Peng, J., Li, Y. & Jin, F. Easy Hi-C: a simple efficient protocol for 3D genome mapping in small cell populations. Preprint at *bioRxiv* https://doi.org/10.1101/245688 (2018).
71. Nagano, T. et al. Single-cell Hi-C for genome-wide detection of chromatin interactions that occur simultaneously in a single cell. *Nat. Protoc.* **10**, 1986–2003 (2015).
72. Nagano, T. et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).
73. Flyamer, I. M. et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* **544**, 110–114 (2017).
This study uses single-cell Hi-C and DNA-FISH to study differences in 3D genome folding between oocytes and zygotes, and reveals the absence of chromatin compartments in the oocyte, as well as single-cell heterogeneity in TAD organization.
74. Nagano, T. et al. Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61–67 (2017).
This paper describes the use of single-cell Hi-C to map chromatin contacts throughout the cell cycle in thousands of individual cells, which shows the emergence of chromatin structures after mitosis and the temporal dynamics of different chromatin topologies.
75. Ramani, V. et al. Massively multiplex single-cell Hi-C. *Nat. Methods* **14**, 263–266 (2017).
76. Tan, L., Xing, D., Chang, C.-H., Li, H. & Xie, X. S. Three-dimensional genome structures of single diploid human cells. *Science* **361**, 924–928 (2018).
77. Horike, S., Cai, S., Miyano, M., Cheng, J. F. & Kohwi-Shigematsu, T. Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat. Genet.* **37**, 31–40 (2005).
78. Schoenfelder, S. et al. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat. Genet.* **42**, 53–61 (2010).
79. Fullwood, M. J. et al. An oestrogen-receptor-α-bound human chromatin interactome. *Nature* **462**, 58–64 (2009).
80. Mumbach, M. R. et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* **13**, 919–922 (2016).
81. Fang, R. et al. Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Res.* **26**, 1345–1348 (2016).
82. Davies, J. O., Oudelaar, A. M., Higgs, D. R. & Hughes, J. R. How best to identify chromosomal interactions: a comparison of approaches. *Nat. Methods* **14**, 125–134 (2017).
83. Zhang, Y. et al. Enhancing Hi-C data resolution with deep convolutional neural network HiCPlus. *Nat. Commun.* **9**, 750–758 (2018).
84. O'Sullivan, J. M., Hendy, M. D., Pichugina, T., Wake, G. C. & Langowski, J. The statistical-mechanics of chromosome conformation capture. *Nucleus* **4**, 390–398 (2013).
85. Chen, H. et al. Dynamic interplay between enhancer–promoter topology and gene activity. *Nat. Genet.* **50**, 1296–1303 (2018).
In this study, live-cell imaging shows that transcription directly depends on contact between enhancers and promoters.
86. van Steensel, B. & Henikoff, S. Identification of in vivo DNA targets of chromatin proteins using tethered Dam methyltransferase. *Nat. Biotechnol.* **18**, 424–428 (2000).
87. Vogel, M. J., Peric-Hupkes, D. & van Steensel, B. Detection of in vivo protein–DNA interactions using DamID in mammalian cells. *Nat. Protoc.* **2**, 1467–1478 (2007).
88. Marshall, O. J., Southall, T. D., Cheetham, S. W. & Brand, A. H. Cell-type-specific profiling of protein–DNA interactions without cell isolation using targeted DamID with next-generation sequencing. *Nat. Protoc.* **11**, 1586–1598 (2016).
89. Peric-Hupkes, D. et al. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell* **38**, 603–613 (2010).
90. Guelen, L. et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).
This article describes DamID, an elegant approach to map chromatin contacts at the nuclear lamina.
91. Spector, D. L. & Lamond, A. I. Nuclear speckles. *Cold Spring Harb. Perspect. Biol.* **3**, a000646 (2011).
92. Chen, Y. et al. Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *J. Cell Biol.* **270**, 4025–4048 (2018).
This article introduces TSA-seq, a genome-wide sequencing technique, to map spatial distances between DNA and nuclear bodies, such as splicing speckles.

93. Redolfi, J. et al. DamC reveals principles of chromatin folding in vivo without crosslinking and ligation. *Nat. Struct. Mol. Biol.* **26**, 471–480 (2019). **This article introduces DamC, an approach that allows mapping of chromatin contacts in vivo and without crosslinking and ligation.**
94. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012). **In this article, Hi-C experiments reveal the compartmentalization of chromosomes into TADs.**
95. Pombo, A. & Branco, M. R. Functional organisation of the genome during interphase. *Curr. Opin. Genet. Dev.* **17**, 451–455 (2007).
96. Li, Y. et al. The effects of chemical fixation on the cellular nanostructure. *Exp. Cell Res.* **358**, 253–259 (2017).
97. Oudelaar, A. M., Davies, J. O. J., Downes, D. J., Higgs, D. R. & Hughes, J. R. Robust detection of chromosomal interactions from small numbers of cells using low-input Capture-C. *Nucleic Acids Res.* **45**, e184–e192 (2017).
98. Guillot, P. V., Xie, S. Q., Hollinshead, M. & Pombo, A. Fixation-induced redistribution of hyperphosphorylated RNA polymerase II in the nucleus of human cells. *Exp. Cell Res.* **295**, 460–468 (2004).
99. Solovei, I. et al. Spatial preservation of nuclear chromatin architecture during three-dimensional fluorescence in situ hybridization (3D-FISH). *Exp. Cell Res.* **276**, 10–23 (2002).
100. Markaki, Y. et al. The potential of 3D-FISH and super-resolution structured illumination microscopy for studies of 3D nuclear architecture: 3D structured illumination microscopy of defined chromosomal structures visualized by 3D (immuno)-FISH opens new perspectives for studies of nuclear architecture. *Bioessays* **34**, 412–426 (2012).
101. Xie, S. Q., Lavitas, L. M. & Pombo, A. CryoFISH: fluorescence in situ hybridization on ultrathin cryosections. *Methods Mol. Biol.* **659**, 219–230 (2010).
102. Brown, J. M. et al. A tissue-specific self-interacting chromatin domain forms independently of enhancer–promoter interactions. *Nat. Commun.* **9**, 3849–3863 (2018).
103. Allahyar, A. et al. Enhancer hubs and loop collisions identified from single-allele topologies. *Nat. Genet.* **50**, 1151–1160 (2018).
104. Olivares-Chauvet, P. et al. Capturing pairwise and multi-way chromosomal conformations using chromosomal walks. *Nature* **540**, 296–300 (2016).
105. Darrow, E. M. et al. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proc. Natl Acad. Sci. USA* **113**, E4504–E4512 (2016).
106. Oudelaar, A. M. et al. Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains. *Nat. Genet.* **50**, 1744–1751 (2018).
107. Branco, M. R., Branco, T., Ramirez, F. & Pombo, A. Changes in chromosome organization during PHA-activation of resting human lymphocytes measured by cryo-FISH. *Chromosome Res.* **16**, 413–426 (2008).
108. Rosin, L. F., Nguyen, S. C. & Joyce, E. F. Condensin II drives large-scale folding and spatial partitioning of interphase chromosomes in Drosophila nuclei. *PLOS Genet.* **14**, e1007393 (2018).
109. Loviglio, M. N. et al. Chromosomal contacts connect loci associated with autism, BMI and head circumference phenotypes. *Mol. Psychiatry* **22**, 836–849 (2016).
110. Spilianakis, C. G., Lalioti, M. D., Town, T., Lee, G. R. & Flavell, R. A. Interchromosomal associations between alternatively expressed loci. *Nature* **435**, 637–645 (2005).
111. Monahan, K., Horta, A. & Lomvardas, S. LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. *Nature* **565**, 448–453 (2019).
112. Hakim, O. et al. Diverse gene reprogramming events occur in the same spatial clusters of distal regulatory elements. *Genome Res.* **21**, 697–706 (2011).
113. Giorgetti, L. & Heard, E. Closing the loop: 3C versus DNA FISH. *Genome Biol.* **17**, 215–223 (2016).
114. Tang, Z. et al. CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell* **163**, 1611–1627 (2015).
115. Wang, X. T., Dong, P. F., Zhang, H. Y. & Peng, C. Structural heterogeneity and functional diversity of topologically associating domains in mammalian genomes. *Nucleic Acids Res.* **43**, 7237–7246 (2015). **This paper describes a chromosome-wide effort to map TADs using DNA-FISH. In addition to revealing the single-cell behaviour of TADs, the paper shows that the imaging data have a high correlation with Hi-C data.**
116. Fudenberg, G. & Imakaev, M. FISH-ing for captured contacts: towards reconciling FISH and 3C. *Nat. Methods* **14**, 673–678 (2017).
117. Rao, S. S. P. et al. Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320.e24 (2017).
118. Mahamid, J. et al. Visualizing the molecular sociology at the HeLa cell nuclear periphery. *Science* **351**, 969–972 (2016).
119. Aitchison, J. D. & Rout, M. P. The interactome challenge. *J. Cell Biol.* **211**, 729 (2015).
120. Cremer, T. & Cremer, M. Chromosome territories. *Cold Spring Harb. Perspect. Biol.* **2**, a003889 (2010).
121. Parada, L. & Misteli, T. Chromosome positioning in the interphase nucleus. *Trends Cell Biol.* **12**, 425–432 (2002).
122. Hacisuleyman, E. et al. Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat. Struct. Mol. Biol.* **21**, 198–206 (2014).
123. Maass, P. G. et al. Reorganization of inter-chromosomal interactions in the 2q37-deletion syndrome. *EMBO J.* **37**, e96257 (2018).
124. Maharana, S. et al. Chromosome intermingling — the physical basis of chromosome organization in differentiated cells. *Nucleic Acids Res.* **44**, 5148–5160 (2016).
125. Zhang, Y. et al. Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* **148**, 908–921 (2012).
126. Lomvardas, S. et al. Interchromosomal interactions and olfactory receptor choice. *Cell* **126**, 403–413 (2006).
127. Cairns, J. et al. CHiCAGO: robust detection of DNA looping interactions in capture Hi-C data. *Genome Biol.* **17**, 127–143 (2016).
128. Fanucchi, S., Shibayama, Y., Burd, S., Weinberg, M. S. & Mhlanga, M. M. Chromosomal contact permits transcription between coregulated genes. *Cell* **155**, 606–620 (2013).
129. Jackson, D. A. & Pombo, A. Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. *J. Cell Biol.* **140**, 1285–1295 (1998).
130. Zink, D., Bornfleth, H., Visser, A., Cremer, C. & Cremer, T. Organization of early and late replicating DNA in human chromosome territories. *Exp. Cell Res.* **247**, 176–188 (1999).
131. Visser, A. E. et al. Spatial distributions of early and late replicating chromatin in interphase chromosome territories. *Exp. Cell Res.* **243**, 398–407 (1998).
132. Ferreira, J., Paolella, G., Ramos, C. & Lamond, A. I. Spatial organization of large-scale chromatin domains in the nucleus: a magnified view of single chromosome territories. *J. Cell Biol.* **139**, 1597–1610 (1997).
133. Sadoni, N. et al. Nuclear organization of mammalian genomes. *J. Cell Biol.* **146**, 1211–1226 (1999).
134. Hiratani, I. et al. Global reorganization of replication domains during embryonic stem cell differentiation. *PLOS Biol.* **6**, e245 (2008).
135. Schwaiger, M. et al. Chromatin state marks cell-type- and gender-specific replication of the Drosophila genome. *Genes. Dev.* **23**, 589–601 (2009).
136. Pope, B. D. et al. Topologically associating domains are stable units of replication-timing regulation. *Nature* **515**, 402–405 (2014).
137. Monneron, A. & Bernhard, W. Fine structural organization of the interphase nucleus in some mammalian cells. *J. Ultrastruct. Res.* **27**, 266–288 (1969).
138. Verschure, P. J., van der Kraan, I., Manders, E. M. M. & van Driel, R. Spatial relationship between transcription sites and chromosome territories. *J. Cell Biol.* **147**, 13–24 (1999).
139. Dundr, M. & Misteli, T. Biogenesis of nuclear bodies. *Cold Spring Harb. Perspect. Biol.* **2**, a000711 (2010).
140. Mao, Y. S., Zhang, B. & Spector, D. L. Biogenesis and function of nuclear bodies. *Trends Genet.* **27**, 295–306 (2011).
141. Pederson, T. The nucleolus. *Cold Spring Harb. Perspect. Biol.* **3**, a000638 (2011).
142. Shopland, L. S., Johnson, C. V., Byron, M., McNeil, J. & Lawrence, J. B. Clustering of multiple specific genes and gene-rich R-bands around SC-35 domains: evidence for local euchromatic neighborhoods. *J. Cell Biol.* **162**, 981–990 (2003).
143. Brown, J. M. et al. Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. *J. Cell Biol.* **182**, 1083–1097 (2008).
144. Iborra, F. J., Pombo, A., Jackson, D. A. & Cook, P. R. Active RNA polymerases are localized within discrete transcription "factories' in human nuclei. *J. Cell Sci.* **109**, 1427–1436 (1996).
145. Pombo, A. et al. Regional specialization in human nuclei: visualization of discrete sites of transcription by RNA polymerase III. *EMBO J.* **18**, 2241–2253 (1999).
146. Xie, S. Q., Martin, S., Guillot, P. V., Bentley, D. L. & Pombo, A. Splicing speckles are not reservoirs of RNA polymerase II, but contain an inactive form, phosphorylated on serine2 residues of the C-terminal domain. *Mol. Biol. Cell* **17**, 1723–1733 (2006).
147. Osborne, C. S. et al. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.* **36**, 1065–1071 (2004).
148. Osborne, C. S. & Eskiw, C. H. Where shall we meet? A role for genome organisation and nuclear sub-compartments in mediating interchromosomal interactions. *J. Cell Biochem.* **104**, 1553–1561 (2008).
149. Boehning, M. et al. RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat. Struct. Mol. Biol.* **25**, 833–840 (2018).
150. Sabari, B. R. et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, aar3958 (2018).
151. Banani, S. F. et al. Compositional control of phase-separated cellular bodies. *Cell* **166**, 651–663 (2016).
152. Bantignies, F. et al. Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. *Cell* **144**, 214–226 (2011).
153. Tiwari, V. K., Cope, L., McGarvey, K. M., Ohm, J. E. & Baylin, S. B. A novel 6C assay uncovers Polycomb-mediated higher order chromatin conformations. *Genome Res.* **18**, 1171–1179 (2008).
154. Fraser, J. et al. Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Mol. Syst. Biol.* **11**, 852–865 (2015).
155. Nemeth, A. et al. Initial genomics of the human nucleolus. *PLOS Genet.* **6**, e1000889 (2010).
156. Kumaran, R. I. & Spector, D. L. A genetic locus targeted to the nuclear periphery in living cells maintains its transcriptional competence. *J. Cell Biol.* **180**, 51–65 (2008).
157. Sabbattini, P., Georgiou, A., Sinclair, C. & Dillon, N. Analysis of mice with single and multiple copies of transgenes reveals a novel arrangement for the λ5–VpreB1 locus control region. *Mol. Cell Biol.* **19**, 671–679 (1999).
158. Finlan, L. E. et al. Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLOS Genet.* **4**, e1000039 (2008). **This article demonstrates that the nuclear environment influences gene expression; bringing genes into the repressive context of the nuclear lamina can lead to downregulation of their expression.**
159. Wang, H. et al. CRISPR-mediated programmable 3D genome positioning and nuclear organization. *Cell* **175**, 1405–1417.e14 (2018).
160. Grewal, S. I. & Elgin, S. C. Transcription and RNA interference in the formation of heterochromatin. *Nature* **447**, 399–406 (2007).
161. Sexton, T. et al. Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458–472 (2012).
162. Lupianez, D. G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene–enhancer interactions. *Cell* **161**, 1012–1025 (2015).
163. Hnisz, D. et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454–1458 (2016).
164. Shen, Y. et al. A map of the *cis*-regulatory sequences in the mouse genome. *Nature* **488**, 116–120 (2012).
165. Symmons, O. et al. Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* **24**, 390–400 (2014).
166. Jackson, D. A., Bartlett, J. & Cook, P. R. Sequences attaching loops of nuclear and mitochondrial DNA to underlying structures in human cells: the role of transcription units. *Nucleic Acids Res.* **24**, 1212–1219 (1996).
167. de Wit, E. et al. CTCF binding polarity determines chromatin looping. *Mol. Cell* **60**, 676–684 (2015).
168. Gomez-Marin, C. et al. Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. *Proc. Natl Acad. Sci. USA* **112**, 7542–7547 (2015).
169. Vietri Rudan, M. et al. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.* **10**, 1297–1309 (2015).

170. Weinreb, C. & Raphael, B. J. Identification of hierarchical chromatin domains. *Bioinformatics* **32**, 1601–1609 (2016).
171. Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
172. Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. & de Laat, W. Looping and interaction between hypersensitive sites in the active β-globin locus. *Mol. Cell* **10**, 1453–1465 (2002).
173. Lettice, L. A. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12**, 1725–1735 (2003).
174. Nobrega, M. A., Ovcharenko, I., Afzal, V. & Rubin, E. M. Scanning human gene deserts for long-range enhancers. *Science* **302**, 413–413 (2003).
175. Qin, Y. et al. Long-range activation of Sox9 in Odd Sex (Ods) mice. *Hum. Mol. Genet.* **13**, 1213–1218 (2004).
176. Javierre, B. M. et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* **167**, 1369–1384.e19 (2016).
177. Zullo, J. M. et al. DNA sequence-dependent compartmentalization and silencing of chromatin at the nuclear lamina. *Cell* **149**, 1474–1487 (2012).
178. Reddy, K. L., Zullo, J. M., Bertolino, E. & Singh, H. Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* **452**, 243–247 (2008).
179. Nott, T. J. et al. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol. Cell* **57**, 936–947 (2015).
180. Strom, A. R. et al. Phase separation drives heterochromatin domain formation. *Nature* **547**, 241–245 (2017).
This article introduces phase separation as a model for heterochromatin formation in the nucleus by showing liquid–liquid phase separation of the heterochromatic protein HPa1 in vitro.
181. Kim, S. et al. The dynamic three-dimensional organization of the diploid yeast genome. *eLife* **6**, 23623–23645 (2017).
182. Chetverina, D., Aoki, T., Erokhin, M., Georgiev, P. & Schedl, P. Making connections: insulators organize eukaryotic chromosomes into independent *cis*-regulatory networks. *Bioessays* **36**, 163–172 (2014).
183. Schoenfelder, S. et al. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res.* **25**, 582–597 (2015).
184. Deng, W. et al. Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell* **158**, 849–860 (2014).
185. Kim, J. H. et al. LADL: light-activated dynamic looping for endogenous gene expression control. *Nat. Methods* **16**, 633–639 (2019).
186. Engreitz, J. M. et al. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**, 452–455 (2016).
This study shows that regulation of genes can occur though the promoters of nearby genes, highlighting the functional importance of transcription factories.
187. Mayer, R. et al. Common themes and cell type specific variations of higher order chromatin arrangements in the mouse. *BMC Cell Biol.* **6**, 44–65 (2005).
188. Ahmed, K. et al. Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. *PLOS ONE* **5**, e10531 (2010).
189. Li, Y. et al. CRISPR reveals a distal super-enhancer required for Sox2 expression in mouse embryonic stem cells. *PLOS ONE* **9**, e114485 (2014).
190. Kerpedjiev, P. et al. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol.* **19**, 125–136 (2018).
191. Bonev, B. et al. Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572.e24 (2017).
192. Naumova, N., Smith, E. M., Zhan, Y. & Dekker, J. Analysis of long-range chromatin interactions using chromosome conformation capture. *Methods* **58**, 192–203 (2012).
193. van de Werken, H. J. et al. 4C technology: protocols and data analysis. *Methods Enzymol.* **513**, 89–112 (2012).
194. Schwartzman, O. et al. UMI-4C for quantitative and targeted chromosomal contact profiling. *Nat. Methods* **13**, 685–691 (2016).
195. Dostie, J. & Dekker, J. Mapping networks of physical interactions between genomic elements using 5C technology. *Nat. Protoc.* **2**, 988–1002 (2007).
196. Kim, J. H. et al. 5C-ID: increased resolution chromosome-conformation-capture-carbon-copy with in situ 3C and double alternating primer design. *Methods* **142**, 39–46 (2018).
197. Belaghzal, H., Dekker, J. & Gibcus, J. H. Hi-C 2.0: an optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. *Methods* **123**, 56–65 (2017).
198. Li, X. et al. Long-read ChIA-PET for base-pair-resolution mapping of haplotype-specific chromatin interactions. *Nat. Protoc.* **12**, 899–915 (2017).
199. Davies, J. O. et al. Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat. Methods* **13**, 74–80 (2016).
200. Croft, J. A. et al. Differences in the localization and morphology of chromosomes in the human nucleus. *J. Cell Biol.* **145**, 1119–1131 (1999).
201. Cremer, M. et al. Multicolor 3D fluorescence in situ hybridization for imaging interphase chromosomes. *Methods Mol. Biol.* **463**, 205–239 (2008).
202. Chen, B. et al. Expanding the CRISPR imaging toolset with *Staphylococcus aureus* Cas9 for simultaneous imaging of multiple genomic loci. *Nucleic Acids Res.* **44**, e75–e87 (2016).
203. Takei, Y., Shah, S., Harvey, S., Qi, L. S. & Cai, L. Multiplexed dynamic imaging of genomic loci by combined CRISPR imaging and DNA sequential FISH. *Biophys. J.* **112**, 1773–1776 (2017).
204. Shimizu, N., Maekawa, M., Asai, S. & Shimizu, Y. Multicolor FISHs for simultaneous detection of genes and DNA segments on human chromosomes. *Chromosome Res.* **23**, 649–662 (2015).
205. Müller, S., Neusser, M. & Wienberg, J. Towards unlimited colors for fluorescence in-situhybridization (FISH). *Chromosome Res.* **10**, 223–232 (2002).
206. Hepperger, C., Otten, S., von Hase, J. & Dietzel, S. Preservation of large-scale chromatin structure in FISH experiments. *Chromosoma* **116**, 117–133 (2007).
207. Volpi, E. V. et al. Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *J. Cell Sci.* **113**, 1565–1576 (2000).
208. Williams, R. R., Broad, S., Sheer, D. & Ragoussis, J. Subchromosomal positioning of the epidermal differentiation complex (EDC) in keratinocyte and lymphoblast interphase nuclei. *Exp. Cell Res.* **272**, 163–175 (2002).
209. Chambeyron, S. & Bickmore, W. A. Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev.* **18**, 1119–1130 (2004).
210. Raap, A. K., Marijnen, J. G., Vrolijk, J. & van der Ploeg, M. Denaturation, renaturation, and loss of DNA during in situ hybridization procedures. *Cytometry* **7**, 235–242 (1986).

Acknowledgements
The authors thank the Helmholtz Association (Germany) for support, C. Thieme (our laboratory) for help plotting GAM and SPRITE contact maps in Figs 1 and 5a,b, and S. Quinodoz (M. Guttman laboratory) for sharing SPRITE contact cluster data (Figs 1 and 5b). A.P. acknowledges support from the National Institutes of Health Common Fund 4D Nucleome Program grant U54DK107977. The authors apologize to the many scientists whose studies were not discussed in our review due to length constraints.

Author contributions
All authors contributed to all aspects of the article.

Competing interests
A.P. has filed a patent application on GAM: Pombo, A., Edwards, P. A. W., Nicodemi, M., Beagrie, R. A. & Scialdone, A. Patent application on 'Genome Architecture Mapping'. Patent PCT/EP2015/079413 (2015).

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.