Dissecting chromatin - mediated gene regulation and epigenetic memory through mathematical modelling

Leonie Ringrose[1] and Martin Howard[2]

1) Corresponding author.

Humboldt-Universität zu Berlin, Philippstrasse 13, Haus 22, 10115 Berlin, Germany

Email: leonie.ringrose@hu-berlin.de

Tel: +49 30 2093 49772

2) Computational and Systems Biology, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK

Email: martin.howard@jic.ac.uk

Tel: +44 1603 450892

Abstract

The application of mathematical modelling to chromatin - mediated gene regulation is gaining momentum, but is still surprisingly rare. Here we review examples in which the combination of quantitative experimentation and mathematical modelling has given mechanistic insights into the processes involved. Examples include recruitment of epigenetic regulators, the establishment and maintenance of epigenetic memory, the dynamic cell cycle - dependent changes in chromatin binding of epigenetic regulators, and the contribution of 3D genome architecture to cell identity. The successful combination of theory and experiment requires tractable experimental systems in which quantitative measurements and precise perturbations are possible. The advent of single cell technologies and genome editing presents an unprecedented opportunity for combining quantitative experiments, precise perturbation and modelling, that in future will enable new epigenetic data to be embedded in a coherent theoretical framework.

Introduction

If publication rate is an indicator of scientific activity, then the field of epigenetics has seen a dramatic increase in activity over the last 20 years (Figure 1). Epigenetic processes have been implicated in many human diseases [1-5] and “epigenetic drugs” are entering the clinics www.insightpharmareports.com/. The field is moving rapidly, but does this mean we are progressing towards understanding? Surprisingly, despite the vast amounts of epigenetic data that have been generated, we are still far from a quantitative mechanistic understanding of many epigenetic phenomena.

Although the term “epigenetics” is itself still the subject of healthy debate [6], there are two definitions that are relevant for the purposes of this review. The definition proposed by Ptashne and Gann [7] of epigenetic regulation as “a change in the state of expression of a gene that does not involve a mutation, but that is nevertheless inherited in the absence of the signal (or event) that caused that change“ has informed a large body of work on modelling epigenetic memory and switching. A broader definition proposed by Bird [8] of epigenetics as “the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states”, also provides a useful framework within which to consider models for targeting epigenetic regulators, and large scale chromosome architecture. Within both of these definitions, epigenetic regulatory systems share several key features: they are **complex,** comprising multiple molecular components that regulate many genomic targets; they are **dynamic,** allowing flexibility in reaction to environmental, developmental or disease signals; and they involve **stochastic processes**, such that the output of a given epigenetic regulatory event can vary from cell to cell, over time, and from individual to individual.

Given these properties, the application of mathematical modelling to epigenetics clearly has immense potential. Indeed, we have reached a stage at which it is very difficult to make sense of the biology of epigenetics without a coherent theoretical framework. The crisp formalism of mathematics imposes a requirement for clear conceptualisation, which is greatly needed in the field of epigenetics. However, the number of papers in which modelling is applied to epigenetic questions, although growing steadily, has not increased at the same explosive rate as those in the field of epigenetics in general (Figure 1). Interestingly, the idea of modelling epigenetic phenomena is almost as old as the concept of epigenetics itself [9, 10]. In this review we cover some of the key recent advances in the application of mechanistic mathematical modelling to quantitative experimental data in epigenetics. We highlight examples in which the insights gained would not

have been possible without modelling, and identify future directions in which the combination of theory and experiment will greatly accelerate our progress in understanding epigenetics. The following sections give examples of different types of modelling that have given insights into various epigenetic questions, summarised in Figure 2.

Genome- wide distribution: from correlation to mechanism

The advent of “omics” technologies has provided a vast amount of data enabling correlations to be drawn between DNA sequence, chromatin immunoprecipitation (ChIP) - based profiling of specific factors and histone modifications, and transcription status, to address the questions of what is where in the genome. Machine learning approaches have been extensively used to segment chromatin states beyond the original microscopy based classification into eu - and heterochromatin [11-14]. Further developments have seen data mining and modelling combined to generate predictive testable hypotheses, giving mechanistic insight into why chromatin binding proteins or chromatin modifications are where they are, and what are the consequences of their being there [15,16].

For example, [15] distinguishes direct from indirect interactions in large ChIP- seq data sets. Genome wide profiles for related chromatin binding proteins are usually highly correlated. For example for a complex consisting of three members (A, B, and C) that binds a histone modification (D), the four ChIP profiles (A-D) will be highly correlated. However if (A) binds directly to (D) then the correlation between the profiles of (A) and (D) will be higher than for any other pair. Complete and partial correlations were used to classify interactions as likely to be direct or indirect for a large set of chromatin binding proteins and modifiers. Several of the interactions thus identified were confirmed by previous studies, and several newly predicted interactions were confirmed experimentally, for example a new link between H4K20me1 and specific Polycomb group proteins. Thus by elegant analysis of large data sets, mechanistic properties beyond correlation can be discovered.

Several studies have used machine learning to extract DNA sequence principles that are predictive for binding of epigenetic regulators [17-21], reviewed in [22]. A very valuable aspect of DNA-based predictions, in which particular motifs contribute to a high predictive score, is that they are amenable to precise experimental tests. For example, a motif (GTGT) predicted computationally to be important for the function of *Drosophila* Polycomb Response elements (PREs) [17], was shown in transgenic assays to be essential for PRE function [23]. Later genetic and biochemical assays in *Drosophila* identified the protein (Combgap) that binds directly to this motif and is indeed essential for PRE regulation [24]. This example, involving several labs over a period of more than 10 years, illustrates the full circle of computational prediction, perturbation and experimental validation, leading to mechanistic insights.

The GTGT motif may also provide a bridge to the elusive question of sequence principles of mammalian PREs. The long-held belief that vertebrate Polycomb group proteins are recruited to CpG islands via the GC rich sequences they contain (reviewed in [22]) was challenged recently by a combined computational and experimental analysis of DNA sequences at sites that recruit PRC2 (Polycomb Repressive Complex 2) in *Xenopus* embryos [20]. These sites were found to be GC poor, and to contain many other motifs, including GTGT repeats. The important feature in common between the frog and mammalian sites is that they are unmethylated. Thus vertebrate PcG proteins need an unmethylated island to bind, and the question of which DNA features specifically recruit them is still open (reviewed in [22]). However, although sites of potential epigenetic

regulation are defined directly or indirectly by DNA sequence [22, 25, 26], whether a given regulator does or does not bind at a given site depends on the developmental and transcriptional changes that make one cell different from another. Thus, models that integrate information over several stages of differentiation may give a more complete picture. For example in [21], DNA sequences and transcription factors that target mouse PcG proteins were identified by their dynamic recruitment of PcG proteins during differentiation of ESCs into neural lineages.

The above studies [17, 20, 21] have in common that they started with a computational prediction based on DNA sequence, and validated predictions in functional reporter assays. Several independent studies have recently questioned the reliability of data generated by ChIP-seq, showing that active promoters can give misleadingly high signals and urging caution in interpretation [27-30]. This underlines the importance of independent functional assays for any given sequence of interest, and reminds us that the only unambiguous genome-wide data set is DNA sequence itself.

Locus-specific regulation: stochastic models for memory and switching

The mechanistic nature of epigenetic memory storage at specific loci is a central issue in the field but one which has proved difficult to elucidate. In part this has been due to the prevalence of whole genome analyses. Such approaches typically generate correlative information, often based on datasets which include both memory-storing and non-memory storing loci, making firm conclusions about memory elusive. In particular, it has been difficult to assess which local factors could be causative memory elements, as opposed to being downstream consequences of particular transcriptional states [31, 32]. Recent experiments on the Polycomb Repressive Complex 2 (PRC2) silencing system at a whole chromosome level [33], and using dual fluorescent imaging at a single locus [34], have shown that epigenetic memory can be stored locally in the chromatin in cis. However, exactly which features are strictly required to generate stable cis-based memory has remained unclear.

In this context, mathematical modelling has been of vital importance in probing how stable memory can be created. Although epigenetic dynamics can be dauntingly complex, modelling excels in reducing such dynamics to their core ingredients. Such appropriately simplified models of epigenetic dynamics mediated by histone modifications were first introduced by [35]. Epigenetic dynamics at a single locus were abstracted down to the stochastic dynamics of activating (A) or silencing marks (M), together with an unmarked intermediate U. As originally applied to the yeast mating type locus, M marks represented the histone modification H3K9me. Feedback from the marks were then introduced such that M marks tended to add more M marks nearby and also removed A marks. The A marks were assumed to possess similar self-reinforcing dynamics. With the added ingredient of long-range interactions between marks, such dynamics were shown to generate bistable (ON/OFF) states, either predominantly M covered or A covered. Importantly such states were highly resistant to fluctuations, particularly at DNA replication where the coverage of marks is (on average) halved representing random partitioning of the marks onto the two daughter DNA strands. Because of the inherent feedbacks in the system, newly inserted unmarked U intermediates are rapidly converted into either M or A, whichever previously predominated, thereby stably re-establishing the memory state. Variations of the above model have been extensively studied theoretically [36, 37].

Conceptually similar models have also been applied to the establishment and inheritance of DNA methylation patterns [38-40]. These studies show that by adapting an existing model to a new biological question, novel insights can arise. DNA methylation at individual CpGs was thought to be independent of other CpGs. However, by comparing simulations to data on the genome wide distribution of CpG methylation, the authors could explain the observed patterns only by introducing cooperativity and dynamic feedback between methylated (or unmethylated) CpGs [38, 39]. Furthermore the observed relationships between CpG cluster size and methylation status could be accounted for by including nucleosome occupancy in the models [40]. These conceptually simple models thus make experimentally testable mechanistic predictions in a field that has traditionally built hypotheses in the absence of theoretical analysis.

Locus-specific models have also been applied more directly to experiments [41-43]. However, perhaps the best established application of modelling, and the first to try to fuse modelling directly with experiments, was in the PRC2-based plant vernalization system [44]. Exposure to winter cold had long been known to generate a quantitative, epigenetically-stable downregulation of the floral repressor gene *FLC*. Importantly, a combined experimental-modelling approach, showed that this quantitative downregulation was actually a population level effect, with individual loci adopting a digital ON/OFF status with the fraction of PRC2-silenced OFF loci increasing with cold duration. Modelling of this system, where the M mark now represented the H3K27me3 silencing modification, also allowed various aspects of the switching process to be elucidated, with switching from one epigenetic state to another mediated through a nucleation and spreading mechanism [44, 45]. The *FLC* system is continuing to generate powerful insights, both experimentally, for example through the dual labelling approach mentioned above [34] and also by inspiring new models. In the latter category, it has recently been proposed that transcription itself may act to antagonise the M marks of the silenced state without the need for activating A marks [46]. Such models also allow long-ranged interactions to be dispensed with, potentially important in preventing ‘runaway’ spreading of marks across whole genomes [47]. Integrating transcription and cis feedbacks in a single framework also permits the simultaneous study of cis-based bistable dynamics together with more traditional, continuously-varying transcription factor based regulation.

The *FLC* system is an excellent model for epigenetic memory for many reasons, one of which is that the stimulus (cold) and the switch in epigenetic status at a population level occur over long time periods (weeks to months) and so can be captured experimentally with time course ChIP experiments. An alternative system to study epigenetic switching over faster time scales has recently been described [43]. The authors used a reporter gene expressing a fluorescently tagged protein in combination with induced recruitment of different epigenetic regulators, to image epigenetic memory and switching in real time in individual mammalian cells. A phenomenological 3-state model based on “active” (A), “reversible silent” (R), and “irreversible silent” (I) states was implemented. This model is conceptually different from the three state models of [35, 44, 45], where the A, U and M marks represent histone modifications, fundamental components of the mechanism of activation or silencing. In contrast, the three states in the model of [43] represent the active or silent states themselves, without explicit reference to the activating or silencing mechanism. This model is nevertheless sufficient to recapitulate the data and enabled extraction of the conversion rates between these states for the different regulators, with timescales ranging from hours to several days. This synthetic system offers an elegant and tractable framework within which to perform controlled experiments and quantitative modelling.

Both this study [43] and the *FLC* studies described above [34, 44] show experimentally that epigenetically regulated reporter genes have the property of an “all or none” response. Quantitative

responses to stimuli (in the case of *FLC*, cold exposure, and in the case of the mammalian study, experimentally manipulated levels of chromatin modifying enzyme recruitment), are manifested as the proportion of cells in which the gene of interest is on or off. This raises the question of how other genes that are subject to epigenetic memory (for example, the *Hox* genes in *Drosophila*) achieve exquisite developmental precision both in terms of their expression level and their spatial patterning [48]. Future models taking spatial patterning into account may help to address these issues.

Global dynamics during the cell cycle

The above studies give insights into locus- specific effects of epigenetic regulation. At the other end of the scale are models that address the global behaviour of epigenetic regulators and how their interaction with chromatin changes during the cell cycle. Replication and mitosis represent the biggest molecular obstacles to epigenetic memory [48]. The events of mitosis are particularly tractable to live imaging studies and have been addressed by modelling in combination with quantitative kinetic analysis [49, 50]. These models lack information on specific loci but have the advantage that large-scale changes in kinetic properties can be addressed in living cells, and in living animals, giving mechanistic insights and generating testable hypotheses. For example, in [49, 50], absolute quantification was combined with kinetic analysis and ODE (ordinary differential equation) modelling for members of the Polycomb (PcG) and Trithorax (TrxG) groups of proteins in living *Drosophila.* This analysis revealed cell-type specific regulated mitotic dissociation, and predicted the existence of active and regulated mechanisms to displace and retain PcG and TrxG proteins on mitotic chromosomes. Perturbation experiments by tissue specific RNAi [49, 50], and mutational analysis of the PcG and TrxG proteins themselves, uncovered molecular mechanisms [49, 50], and showed that mitotic attachment of the TrxG protein ASH1 is required for cell identity and viability (Steffen et al., unpublished). These studies reveal that a simple model containing only three components (a protein, its target site, and the complex between them) can nevertheless reveal unexpected insights when combined with absolute quantification.

The third and fourth dimensions: 3 dimensional genome architecture

Epigenetic regulation occurs within a highly folded genome, in which distant loci may be brought close together in 3-dimensional space, and chromosomal segments are constantly moving [51, 52]. Interestingly, despite the topological and structural disruptions of replication and mitosis, many of these long -range contacts are highly conserved between cell types and species, whereas others are highly cell-type specific [53, 54]. Nuclear localization can have profound effects on gene expression [55] and disruption of essential contacts can lead to disease [56, 57].

Many properties of DNA and chromosomes can be understood by modelling the DNA or chromatin fibre as a flexible polymer [58]. Although this is by no means a new concept [59], the last decade or so has seen a rapid expansion of techniques to measure chromosome conformation in whole genomes at ever increasing resolution [60], and an accompanying increase in the application of biophysical polymer models to the problem of deconvolving these data sets [61]. In the case of “Hi-C” techniques, static data are generated from a large population of cells, based on crosslinking and quantifying DNA sequences that are physically close to each other. Stochastic polymer based models have then been developed that generate an ensemble of polymer conformations able to reproduce the observed “Hi-C” contact frequency maps. These models aim to discern which local features of the polymer are required to explain the data [53, 62-64]. The stochastic nature of these

models is again crucial: for example, predictions, and experiments using DNA FISH to allow single cell resolution, have demonstrated that long-range contacts occur in the context of fluctuating structures rather than stable loops. Indeed, several groups have developed imaging based techniques to track long- range interactions of specific tagged loci [65-67] *(*reviewed in [68-70]) or labeled histones [52, 71, 72]*,* either in real time or in fixed single cells at high resolution [73]. Polymer based simulations have then been used to capture the dynamic nature of the processes involved, and to infer properties of the chromatin fiber that change during gene activation [64], replication [74], localization in the nucleus or within a topological domain [71], or a change in epigenetic status [73].

In summary, the application of modelling to 3 dimensional chromatin configurations is currently one of the most active areas of synergy between mechanistic modelling and experiments. This expansion has been to a large extent driven by the availability of new types of data to which the long established field of polymer physics can be productively applied.

Outlook

What does the future hold? We hope that there will soon be an acceleration of activity in the application of mechanistic modelling to a wide range of epigenetic questions. There has never been a better time. New technologies enabling single cell and single molecule analysis are now delivering quantitative data in real time at unprecedented resolution [75-80]. This will enable the field to move away from population averages and to measure single events in living, developing organisms. Furthermore, the advent of technologies for precisely editing the genome and the epigenome should now allow precisely designed perturbations of experimental systems [81]. The potential for modelling combined with quantitative experiments and precise perturbation is at an unprecedented level. However, without a coherent theoretical framework, we risk entering an era of single cell omics and editomics without a compass.

Figure legends.

Figure 1. Number of publications per year (including reviews) retrieved with the Pubmed search term “epigenetic’ (light grey) or “epigenetic AND mathematical model” (dark grey).

Figure 2. Four aspects of epigenetic and chromatin mediated regulation and the associated questions are covered in this review. For each we give examples of the successful combination of modelling and quantitative experiments.

References

1. Laugesen, A. and K. Helin, *Chromatin repressive complexes in stem cells, development, and cancer.* Cell Stem Cell, 2014. **14**(6): p. 735-51.
2. Zeller, P., J. Padeken, R. van Schendel, V. Kalck, M. Tijsterman, and S.M. Gasser, *Histone H3K9 methylation is dispensable for Caenorhabditis elegans development but suppresses RNA:DNA hybrid-associated repeat instability.* Nat Genet, 2016. **48**(11): p. 1385-1395.
3. van der Knaap, J.A. and C.P. Verrijzer, *Undercover: gene control by metabolites and metabolic enzymes.* Genes Dev, 2016. **30**(21): p. 2345-2369.

4. Nestler, E.J., C.J. Pena, M. Kundakovic, A. Mitchell, and S. Akbarian, *Epigenetic Basis of Mental Illness.* Neuroscientist, 2016. **22**(5): p. 447-63.
5. Morandini, A.C., C.F. Santos, and O. Yilmaz, *Role of epigenetics in modulation of immune response at the junction of host-pathogen interaction and danger molecule signaling.* Pathog Dis, 2016. **74**(7).
6. Ptashne, M., *Epigenetics: core misconcept.* Proc Natl Acad Sci U S A, 2013. **110**(18): p. 7101-3.
7. Ptashne, M.a.G., A., *Genes and Signals*. 2002: CHSL Press
8. Bird, A., *Perceptions of epigenetics.* Nature, 2007. **447**(7143): p. 396-8.
9. Sauvan, J., *[Applications of a cybernetic model to physiology; metastable system with multiple stationary states: epigenetic hypothesis of various biological behaviors].* Presse Med, 1959. **67**(25): p. 1023-5.
10. Kauffman, S., *Control circuits for determination and transdetermination: interpreting positional information in a binary epigenetic code.* Ciba Found Symp, 1975. **0**(29): p. 201-21.
11. Filion, G.J., J.G. van Bemmel, U. Braunschweig, W. Talhout, J. Kind, L.D. Ward, W. Brugman, I.J. de Castro, R.M. Kerkhoven, H.J. Bussemaker, and B. van Steensel, *Systematic protein location mapping reveals five principal chromatin types in Drosophila cells.* Cell, 2010. **143**(2): p. 212-24.
12. Ernst, J. and M. Kellis, *Discovery and characterization of chromatin states for systematic annotation of the human genome.* Nat Biotechnol, 2010. **28**(8): p. 817-25.
13. Hoffman, M.M., O.J. Buske, J. Wang, Z. Weng, J.A. Bilmes, and W.S. Noble, *Unsupervised pattern discovery in human chromatin structure through genomic segmentation.* Nat Methods, 2012. **9**(5): p. 473-6.
14. van Bemmel, J.G., G.J. Filion, A. Rosado, W. Talhout, M. de Haas, T. van Welsem, F. van Leeuwen, and B. van Steensel, *A network model of the molecular organization of chromatin in Drosophila.* Mol Cell, 2013. **49**(4): p. 759-71.
15. Perner, J., J. Lasserre, S. Kinkley, M. Vingron, and H.R. Chung, *Inference of interactions between chromatin modifiers and histone modifications: from ChIP-Seq data to chromatin-signaling.* Nucleic Acids Res, 2014. **42**(22): p. 13689-95.
16. Moore, B.L., S. Aitken, and C.A. Semple, *Integrative modeling reveals the principles of multi-scale chromatin boundary formation in human nuclear organization.* Genome Biol, 2015. **16**: p. 110.
17. Ringrose, L., M. Rehmsmeier, J.M. Dura, and R. Paro, *Genome-wide prediction of Polycomb/Trithorax response elements in Drosophila melanogaster.* Dev Cell, 2003. **5**(5): p. 759-71.
18. Hauenschild, A., L. Ringrose, C. Altmutter, R. Paro, and M. Rehmsmeier, *Evolutionary plasticity of polycomb/trithorax response elements in Drosophila species.* PLoS Biol, 2008. **6**(10): p. e261.
19. Zeng, J., B.D. Kirk, Y. Gou, Q. Wang, and J. Ma, *Genome-wide polycomb target gene prediction in Drosophila melanogaster.* Nucleic Acids Res, 2012. **40**(13): p. 5848-63.
20. van Heeringen, S.J., R.C. Akkers, I. van Kruijsbergen, M.A. Arif, L.L. Hanssen, N. Sharifi, and G.J. Veenstra, *Principles of nucleation of H3K27 methylation during embryonic development.* Genome Res, 2014. **24**(3): p. 401-10.

*** By showing that methylation status and not GC content is important for PRC2 recruitment this paper challenges current models in which CpG islands are thought to recruit vertebrate Polycomb group proteins.*

21. Arnold, P., A. Scholer, M. Pachkov, P.J. Balwierz, H. Jorgensen, M.B. Stadler, E. van Nimwegen, and D. Schubeler, *Modeling of epigenome dynamics identifies transcription factors that mediate Polycomb targeting.* Genome Research, 2013. **23**(1): p. 60--73.
22. Bauer, M., J. Trupke, and L. Ringrose, *The quest for mammalian Polycomb response elements: are we there yet?* Chromosoma, 2016. **125**(3): p. 471-96.
23. Okulski, H., B. Druck, S. Bhalerao, and L. Ringrose, *Quantitative analysis of polycomb response elements (PREs) at identical genomic locations distinguishes contributions of PRE sequence and genomic environment.* Epigenetics Chromatin, 2011. **4**: p. 4.
24. Ray, P., S. De, A. Mitra, K. Bezstarosti, J.A. Demmers, K. Pfeifer, and J.A. Kassis, *Combgap contributes to recruitment of Polycomb group proteins in Drosophila.* Proc Natl Acad Sci U S A, 2016. **113**(14): p. 3826-31.
25. Kassis, J.A. and J.L. Brown, *Polycomb group response elements in Drosophila and vertebrates.* Adv Genet, 2013. **81**: p. 83-118.
26. Quante, T. and A. Bird, *Do short, frequent DNA sequence motifs mould the epigenome?* Nat Rev Mol Cell Biol, 2016. **17**(4): p. 257-62.
27. Park, D., Y. Lee, G. Bhupindersingh, and V.R. Iyer, *Widespread misinterpretable ChIP-seq bias in yeast.* PLoS One, 2013. **8**(12): p. e83506.
28. Teytelman, L., D.M. Thurtle, J. Rine, and A. van Oudenaarden, *Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins.* Proc Natl Acad Sci U S A, 2013. **110**(46): p. 18602-7.
29. Ramachandran, P., G.A. Palidwor, and T.J. Perkins, *BIDCHIPS: bias decomposition and removal from ChIP-seq data clarifies true binding signal and its functional correlates.* Epigenetics Chromatin, 2015. **8**: p. 33.
30. Jain, D., S. Baldi, A. Zabel, T. Straub, and P.B. Becker, *Active promoters give rise to false positive 'Phantom Peaks' in ChIP-seq experiments.* Nucleic Acids Res, 2015. **43**(14): p. 6959-68.

*** (27-30). Several studies show that ChIP-seq data are prone to systematic false positive artifacts, and recommend means of identifying true signals.*

31. Henikoff, S. and A. Shilatifard, *Histone modification: cause or cog?* Trends Genet, 2011. **27**(10): p. 389-96.
32. Steffen, P.A., J.P. Fonseca, and L. Ringrose, *Epigenetics meets mathematics: towards a quantitative understanding of chromatin biology.* Bioessays, 2012. **34**(10): p. 901-13.
33. Gaydos, L.J., W. Wang, and S. Strome, *Gene repression. H3K27me and PRC2 transmit a memory of repression across generations and during development.* Science, 2014. **345**(6203): p. 1515-8.
34. Berry, S., M. Hartley, T.S. Olsson, C. Dean, and M. Howard, *Local chromatin environment of a Polycomb target gene instructs its own epigenetic inheritance.* Elife, 2015. **4**.

*** A dual fluorescent labelling approach was taken to demonstrate that epigenetic memory at a single PRC2-target locus is stored locally in the chromatin in cis.*

35. Dodd, I.B., M.A. Micheelsen, K. Sneppen, and G. Thon, *Theoretical analysis of epigenetic cell memory by nucleosome modification.* Cell, 2007. **129**(4): p. 813-22.
36. David-Rus, D., S. Mukhopadhyay, J.L. Lebowitz, and A.M. Sengupta, *Inheritance of epigenetic chromatin silencing.* J Theor Biol, 2009. **258**(1): p. 112-20.
37. Sneppen, K. and I.B. Dodd, *A simple histone code opens many paths to epigenetics.* PLoS Comput Biol, 2012. **8**(8): p. e1002643.

38. Haerter, J.O., C. Lovkvist, I.B. Dodd, and K. Sneppen, *Collaboration between CpG sites is needed for stable somatic inheritance of DNA methylation states.* Nucleic Acids Res, 2014. **42**(4): p. 2235-44.
39. Lovkvist, C., I.B. Dodd, K. Sneppen, and J.O. Haerter, *DNA methylation in human epigenomes depends on local topology of CpG sites.* Nucleic Acids Res, 2016. **44**(11): p. 5123-32.
40. Sneppen, K. and I.B. Dodd, *Nucleosome dynamics and maintenance of epigenetic states of CpG islands.* Phys Rev E, 2016. **93**(6): p. 062417.
41. Muller-Ott, K., F. Erdel, A. Matveeva, J.P. Mallm, A. Rademacher, M. Hahn, C. Bauer, Q. Zhang, S. Kaltofen, G. Schotta, T. Hofer, and K. Rippe, *Specificity, propagation, and memory of pericentric heterochromatin.* Mol Syst Biol, 2014. **10**: p. 746.
42. Obersriebnig, M.J., E.M. Pallesen, K. Sneppen, A. Trusina, and G. Thon, *Nucleation and spreading of a heterochromatic domain in fission yeast.* Nat Commun, 2016. **7**: p. 11518.

** 41, 42. Two recent combined experimental/modelling studies focusing on heterochromatin dynamics in mammalian and yeast systems.*

43. Bintu, L., J. Yong, Y.E. Antebi, K. McCue, Y. Kazuki, N. Uno, M. Oshimura, and M.B. Elowitz, *Dynamics of epigenetic regulation at the single-cell level.* Science, 2016. **351**(6274): p. 720-4.

*** A manipulable synthetic reporter system combined with modelling reveals inherently different characteristics of specific epigenetic regulators in real time.*

44. Angel, A., J. Song, C. Dean, and M. Howard, *A Polycomb-based switch underlying quantitative epigenetic memory.* Nature, 2011. **476**(7358): p. 105-8.
45. Angel, A., J. Song, H. Yang, J.I. Questa, C. Dean, and M. Howard, *Vernalizing cold is registered digitally at FLC.* Proc Natl Acad Sci U S A, 2015. **112**(13): p. 4146-51.
46. Berry, S., Dean, C., and Howard, M. , *Noise filtering by Polycomb target genes requires slow chromatin dynamics.* Cell Systems 2017 **in press**

*** A new theoretical model for epigenetic memory where transcription now directly antagonizes Polycomb silencing. The model allows integration of trans-regulatory signals with bistable chromatin states.*

47. Dodd, I.B. and K. Sneppen, *Barriers and silencers: a theoretical toolkit for control and containment of nucleosome-based epigenetic states.* J Mol Biol, 2011. **414**(4): p. 624-37.
48. Steffen, P.A. and L. Ringrose, *What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory.* Nat Rev Mol Cell Biol, 2014. **15**(5): p. 340-56.
49. Fonseca, J.P., P.A. Steffen, S. Müller, J. Lu, A. Sawicka, C. Seiser, and L. Ringrose, *In vivo Polycomb kinetics and mitotic chromatin binding distinguish stem cells from differentiated cells.* Genes Dev, 2012. **26**(8): p. 857-871.
50. Steffen, P.A., J.P. Fonseca, C. Gänger, E. Dworschak, T. Kockmann, C. Beisel, and L. Ringrose, *Quantitative in vivo analysis of chromatin binding of Polycomb and Trithorax group proteins reveals retention of ASH1 on mitotic chromatin.* Nucleic Acids Res, 2013. **41**(10): p. 5235-5250.
51. Liang, Z., D. Zickler, M. Prentiss, F.S. Chang, G. Witz, K. Maeshima, and N. Kleckner, *Chromosomes Progress to Metaphase in Multiple Discrete Steps via Global Compaction/Expansion Cycles.* Cell, 2015. **161**(5): p. 1124-37.

52. Maeshima, K., S. Ide, K. Hibino, and M. Sasai, *Liquid-like behavior of chromatin.* Curr Opin Genet Dev, 2016. **37**: p. 36-45.
53. Fraser, J., C. Ferrai, A.M. Chiariello, M. Schueler, T. Rito, G. Laudanno, M. Barbieri, B.L. Moore, D.C. Kraemer, S. Aitken, S.Q. Xie, K.J. Morris, M. Itoh, H. Kawaji, I. Jaeger, Y. Hayashizaki, P. Carninci, A.R. Forrest, F. Consortium, C.A. Semple, J. Dostie, A. Pombo, and M. Nicodemi, *Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation.* Mol Syst Biol, 2015. **11**(12): p. 852.
54. Ghavi-Helm, Y., F.A. Klein, T. Pakozdi, L. Ciglar, D. Noordermeer, W. Huber, and E.E. Furlong, *Enhancer loops appear stable during development and are associated with paused polymerase.* Nature, 2014. **512**(7512): p. 96-100.
55. Gonzalez-Sandoval, A. and S.M. Gasser, *On TADs and LADs: Spatial Control Over Gene Expression.* Trends Genet, 2016. **32**(8): p. 485-95.
56. Lupianez, D.G., K. Kraft, V. Heinrich, P. Krawitz, F. Brancati, E. Klopocki, D. Horn, H. Kayserili, J.M. Opitz, R. Laxova, F. Santos-Simarro, B. Gilbert-Dussardier, L. Wittler, M. Borschiwer, S.A. Haas, M. Osterwalder, M. Franke, B. Timmermann, J. Hecht, M. Spielmann, A. Visel, and S. Mundlos, *Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions.* Cell, 2015. **161**(5): p. 1012-25.
57. Spielmann, M. and S. Mundlos, *Looking beyond the genes: the role of non-coding variants in human disease.* Hum Mol Genet, 2016. **25**(R2): p. R157-R165.
58. Langowski, J., *Polymer chain models of DNA and chromatin.* Eur Phys J E Soft Matter, 2006. **19**(3): p. 241-9.
59. Hagerman, P.J., *Flexibility of DNA.* Annu Rev Biophys Biophys Chem, 1988. **17**: p. 265-86.
60. Fraser, J., I. Williamson, W.A. Bickmore, and J. Dostie, *An Overview of Genome Organization and How We Got There: from FISH to Hi-C.* Microbiol Mol Biol Rev, 2015. **79**(3): p. 347-72.
61. Nicodemi, M. and A. Pombo, *Models of chromosome structure.* Curr Opin Cell Biol, 2014. **28**: p. 90-5.
62. Barbieri, M., M. Chotalia, J. Fraser, L.M. Lavitas, J. Dostie, A. Pombo, and M. Nicodemi, *Complexity of chromatin folding is captured by the strings and binders switch model.* Proc Natl Acad Sci U S A, 2012. **109**(40): p. 16173-8.
63. Sewitz, S. and K. Lipkow, *Systems Biology Approaches for Understanding Genome Architecture.* Methods Mol Biol, 2016. **1431**: p. 109-26.
64. Giorgetti, L., R. Galupa, E.P. Nora, T. Piolot, F. Lam, J. Dekker, G. Tiana, and E. Heard, *Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription.* Cell, 2014. **157**(4): p. 950-63.
65. Hajjoul, H., S. Kocanova, I. Lassadi, K. Bystricky, and A. Bancaud, *Lab-on-Chip for fast 3D particle tracking in living cells.* Lab Chip, 2009. **9**(21): p. 3054-8.
66. Hajjoul, H., J. Mathon, H. Ranchon, I. Goiffon, J. Mozziconacci, B. Albert, P. Carrivain, J.M. Victor, O. Gadal, K. Bystricky, and A. Bancaud, *High-throughput chromatin motion tracking in living yeast reveals the flexibility of the fiber throughout the genome.* Genome Res, 2013. **23**(11): p. 1829-38.
67. Belton, J.M., B.R. Lajoie, S. Audibert, S. Cantaloube, I. Lassadi, I. Goiffon, D. Bau, M.A. Marti-Renom, K. Bystricky, and J. Dekker, *The Conformation of Yeast Chromosome III Is Mating Type Dependent and Controlled by the Recombination Enhancer.* Cell Rep, 2015. **13**(9): p. 1855-67.
68. Bystricky, K., *Chromosome dynamics and folding in eukaryotes: Insights from live cell microscopy.* FEBS Lett, 2015. **589**(20 Pt A): p. 3014-22.

69. Lassadi, I., A. Kamgoue, I. Goiffon, N. Tanguy-le-Gac, and K. Bystricky, *Differential chromosome conformations as hallmarks of cellular identity revealed by mathematical polymer modeling.* PLoS Comput Biol, 2015. **11**(6): p. e1004306.
70. Huet, S., C. Lavelle, H. Ranchon, P. Carrivain, J.M. Victor, and A. Bancaud, *Relevance and limitations of crowding, fractal, and polymer models to describe nuclear architecture.* Int Rev Cell Mol Biol, 2014. **307**: p. 443-79.
71. Shinkai, S., T. Nozaki, K. Maeshima, and Y. Togashi, *Dynamic Nucleosome Movement Provides Structural Information of Topological Chromatin Domains in Living Human Cells.* PLoS Comput Biol, 2016. **12**(10): p. e1005136.

    ** Sparse fluorescent labelling of nucleosomes combined with superresolution imaging enables visualisation of single nucleosomes in live cells. Interpretation of the data using polymer modelling provides insights into the dynamic behaviour of topological domains.*

72. Maeshima, K., K. Kaizu, S. Tamura, T. Nozaki, T. Kokubo, and K. Takahashi, *The physical size of transcription factors is key to transcriptional regulation in chromatin domains.* J Phys Condens Matter, 2015. **27**(6): p. 064116.
73. Boettiger, A.N., B. Bintu, J.R. Moffitt, S. Wang, B.J. Beliveau, G. Fudenberg, M. Imakaev, L.A. Mirny, C.T. Wu, and X. Zhuang, *Super-resolution imaging reveals distinct chromatin folding for different epigenetic states.* Nature, 2016. **529**(7586): p. 418-22.

    *** Uses superresolution imaging to study specific genomic sites with known epigenetic status, revealing for example, that packing density increases with domain length in Polycomb-repressed domains.*

74. Pichugina, T., T. Sugawara, A. Kaykov, W. Schierding, K. Masuda, J. Uewaki, R.S. Grand, J.R. Allison, R.A. Martienssen, P. Nurse, M. Ueno, and J.M. O'Sullivan, *A diffusion model for the coordination of DNA replication in Schizosaccharomyces pombe.* Sci Rep, 2016. **6**: p. 18757.
75. Lambert, T.J. and J.C. Waters, *Navigating challenges in the application of superresolution microscopy.* J Cell Biol, 2016.
76. Schwartzman, O. and A. Tanay, *Single-cell epigenomics: techniques and emerging applications.* Nat Rev Genet, 2015. **16**(12): p. 716-26.
77. Clark, S.J., H.J. Lee, S.A. Smallwood, G. Kelsey, and W. Reik, *Single-cell epigenomics: powerful new methods for understanding gene regulation and cell identity.* Genome Biol, 2016. **17**: p. 72.
78. Levi, V. and E. Gratton, *Exploring dynamics in living cells by tracking single particles.* Cell Biochem Biophys, 2007. **48**(1): p. 1-15.
79. Morisaki, T., W.G. Muller, N. Golob, D. Mazza, and J.G. McNally, *Single-molecule analysis of transcription factor binding at transcription sites in live cells.* Nat Commun, 2014. **5**: p. 4456.
80. Zhan, H., R. Stanciauskas, C. Stigloher, K.K. Dizon, M. Jospin, J.L. Bessereau, and F. Pinaud, *In vivo single-molecule imaging identifies altered dynamics of calcium channels in dystrophin-mutant C. elegans.* Nat Commun, 2014. **5**: p. 4974.
81. Thakore, P.I., J.B. Black, I.B. Hilton, and C.A. Gersbach, *Editing the epigenome: technologies for programmable transcription and epigenetic modulation.* Nat Methods, 2016. **13**(2): p. 127-37.

Figure 1: Epigenetics and mathematical modelling

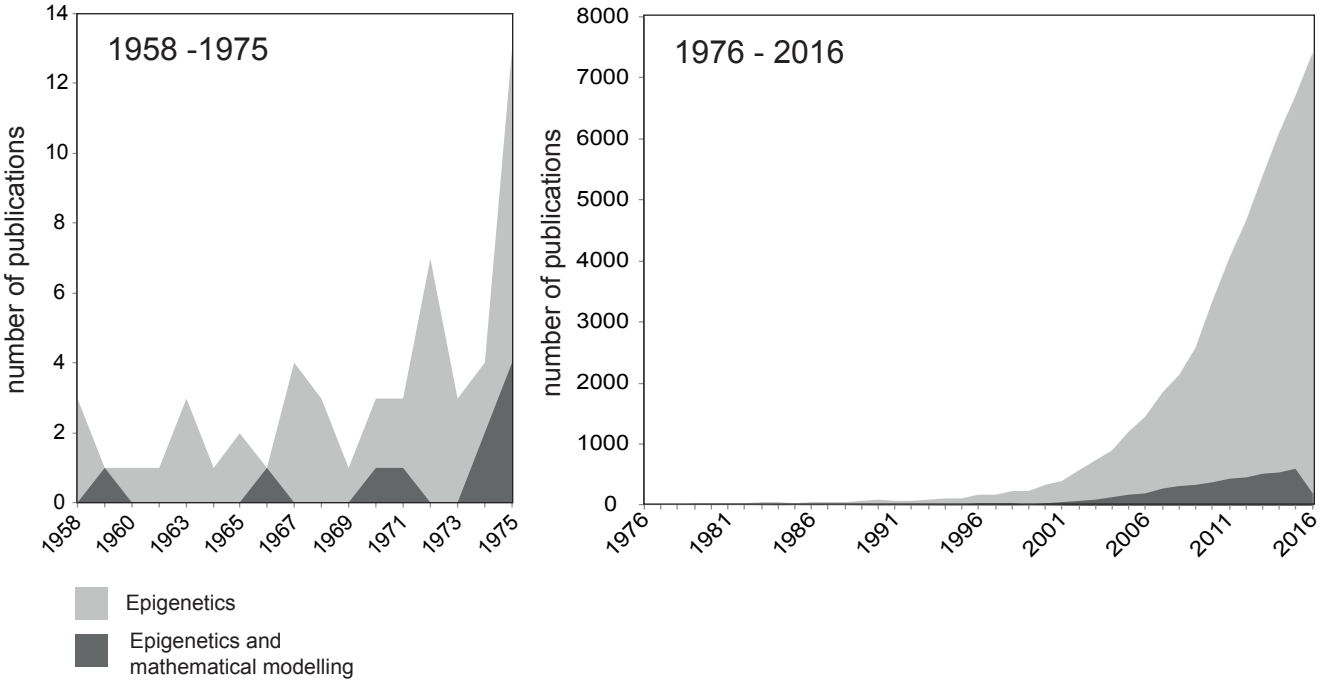


Figure 2.
Epigenetic questions that have been addressed by modelling

Genome - wide distribution

How are proteins and histone modifications distributed in the genome?

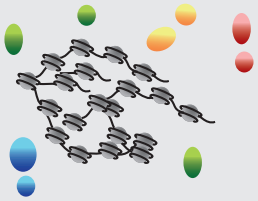
How did they get there?

Locus- specfic regulation

How is epigenetic memory stored?

How does a locus switch memory states?

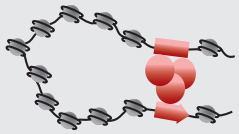
Global dynamics



How do epigenetic regulators interact with chromatin?

How and why does this change during the cell cycle ?

Genome architecture



How is the genome folded in 3 dimensions?

How is this linked to transcription?

Does 3D structure have memory?