CrossMark

Journal of Developmental Origins of Health and Disease (2018), **9**(3), 260–265.
© Cambridge University Press and the International Society for Developmental Origins of Health and Disease 2017
doi:10.1017/S2040174417000678
Themed Issue: NZ-Japan

REVIEW

Genome organization: connecting the developmental origins of disease and genetic variation

E. Jacobson, M. H. Vickers, J. K. Perry and J. M. O'Sullivan*

Liggins Institute, University of Auckland, Grafton, Auckland, New Zealand

An adverse early life environment can increase the risk of metabolic and other disorders later in life. Genetic variation can modify an individual's susceptibility to these environmental challenges. These gene by environment interactions are important, but difficult, to dissect. The nucleus is the primary organelle where environmental responses impact directly on the genetic variants within the genome, resulting in changes to the biology of the genome and ultimately the phenotype. Understanding genome biology requires the integration of the linear DNA sequence, epigenetic modifications and nuclear proteins that are present within the nucleus. The interactions between these layers of information may be captured in the emergent spatial genome organization. As such genome organization represents a key research area for decoding the role of genetic variation in the Developmental Origins of Health and Disease.

Received 20 February 2017; Revised 31 July 2017; Accepted 1 August 2017; First published online 29 August 2017

Key words: epigenetics, functional annotation, genome organization, GWAS, SNPs

Introduction

Early life adverse events can contribute to disease later in life, but not all individuals are affected to the same extent. These differences can be partially attributed to interactions between genetic variation and environmental risk factors such as maternal nutrition.[1–3] Investigating these gene by environment interactions can improve our understanding of non-communicable disease risk. This can be achieved by moving to a systems-wide view of the processes that are required to decode the information (e.g. genes) that is encoded within the linear sequence of the DNA. In effect, we must combine genomic and post-genomic approaches to interpret genome biology so that we can understand how developmental processes are affected by the combinatorial action of genetic variation and epigenetics. Here we will discuss recent attempts to link genetic risk factors to environmental responses and disease risk through the incorporation of the three-dimensional organization of the genome.

Genes are supervened on the genome organization

What is the nature of the information within the DNA sequence? Genes are an obvious candidate. Yet, the view that a gene is hard-coded in the DNA sequence[4–7] has a number of limitations. Notably, it is clear that genes are not fixed entities; rather they are supervened on the genome in a manner which is context dependent and programmable by the environment.[8] This is supported by observations that the functions of defined DNA sequences are context dependent.[9] For example, a promoter may become part of an intron resulting in production of a chimeric messenger RNA transcribed from groups of exons that were previously ascribed to different genes.[10] If one extends the definition of the gene to include the sequences that regulate transcription, then current evidence demonstrates that these elements are not fixed, nor necessarily in cis within the linear DNA sequence. Rather, the combinations are cell-type specific and this is reflected in the spatial organization of the DNA.[11–15]

Genome organization: a definition

When looking at a static microscopic image of a nucleus it is easy to forget that it is in a state of non-equilibrium, constantly exchanging its material constituents with the cytoplasm.[16] This non-equilibrium is most elegantly demonstrated by the formation of condensed chromosomes from interphase DNA as the cell enters metaphase of the cell cycle. Yet the DNA is spatially ordered within the nucleus throughout all phases of the cell cycle; chromosomes reside in regular domains within the nucleus known as chromosome territories. As such, the three-dimensional organization of a genome should be thought of as an emergent property of that particular genome in the context of the micro- (i.e. nuclear, intra-cellular) and macro-environments (inter- and extra-cellular) to which that genome is exposed. Notably, within a population absolute structure cannot be achieved, as there will always be a degree of stochasticity between the genome structure in identical cells exposed to identical conditions as a result of diffusion of molecules and random movement of loci (Brownian motion).[17] Nonetheless, if we capture the genome structure at any one moment in a particular cell, by definition it must have a single structure.

*Address for correspondence: J. M. O'Sullivan, Liggins Institute, University of Auckland, Auckland, 1142, New Zealand.
(Email justin.osullivan@auckland.ac.nz)

https://doi.org/10.1017/S2040174417000678 Published online by Cambridge University Press

Proximity ligation and modern microscopic approaches are capable of capturing genomes in the different spatial organizations that they assume. Despite the inherent limitations of these methods,[18] results from recent studies suggest that the genome and nucleus collectively forms a constrained system that is maintained on the boundary of order and chaos.[19] Within this constrained system, genomes are interleaved entities[20] that are spatially organized into hierarchically organized domains of different sizes (e.g. chromosome territories and topological associated domains).[14,21] The organization of these domains enables the rapid, simultaneous and appropriate accessing of hard-coded information within the DNA sequence as chromatin regions come in and out of contact.

Reproducible and directed changes to genome organization are observed throughout the cell cycle[18] and development.[12,15,22,23] For example, reprogramming of mouse pre-B cells, bone-marrow derived macrophages, neural stem cells and embryonic fibroblasts demonstrated that early passage induced pluripotent stem cells carry reproducibly acquired features of genome organization that are contingent on their cell of origin.[23] Assuming that genome organization emerges from the positioning of chromatin (Fig. 1), it is likely that metastable genome conformations are captured by the combined effects of environmentally signalled changes to the synthesis and degradation[31] of proteins and RNA[32] that occur during the reprogramming. These programmes of change are dependent upon the cell-of-origin composition of transcription factors, proteins and RNAs, and the environmental signals that the cell is exposed to. In such a scenario, genome organization is not deterministic. Rather, it captures the sum activity of the nuclear functions that are occurring at a moment in time, including patterns of gene regulation[33–37] and ultimately cell fate choices.[11,38] These choices often occur in early development, but can affect the activity of key metabolic organs for a lifetime.[39,40]

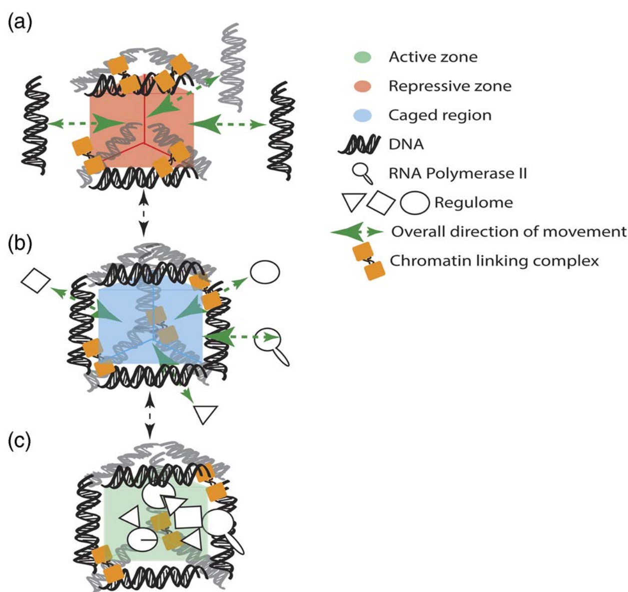

Fig. 1. Genomic structure emerges from the positioning of chromatin by either active or passive means to create phase separated subcompartments for stable gene regulation, repair and replication. (*a*) Chromatin is held in position by complexes (e.g. CTCF and cohesin[24–26]), which are continuously binding and releasing the DNA template. (*b*) The structured chromatin creates a region in which diffusible nuclear components become retarded (i.e. caged region). (*c*) Concentrations that effect phase transitions and promote nuclear functions are ultimately attained.[27] In this model, the retention within the caged region is promoted by high numbers of binding sites directly in the co-located chromatin loci or with other proteins bound to the chromatin.[28–30]

How does genome structure link to the developmental origins of disease?

Metabolic disorders such as obesity and diabetes are recognized as being highly heritable, but despite significant progress[41–48] their genetic basis has not been fully explained.[49,50] The majority of disease-associated single-nucleotide polymorphisms (SNPs) (daSNPs) are found in non-coding regions of the genome.[51] Traditionally, these intergenic or intronic daSNPs have been thought to act on the nearest gene, under the assumption that regulatory interactions involve cis acting sequences that are linked, or proximal, to the gene of interest.[52] Although this assumption is often correct, the three-dimensional nature of the genome allows regulatory sequences to interact with and modify the expression of distal genes; these may be many kilobases (kb) or megabases away on the same chromosome, or even on different chromosomes.[11,53,54]

Although the exons of a gene tend to occur in a linear order along the chromosome, the DNA elements that are necessary for the regulation of gene transcription can be located almost anywhere within the genome.[38,53] This includes distal intergenic regions[55,56] and the introns of other genes.[57,58] However, in order to contribute to the regulation of gene expression, at least a subset of these regulatory elements must physically associate with the target gene promoter. This is facilitated by the formation of DNA loops which allow the element to come into spatial proximity with the target gene.[59,60] A mutation in an enhancer element may disrupt this regulatory cluster, altering transcription of the target gene. Genetic variants that alter gene expression in this way are known as expression quantitative trait loci (eQTLs).[61]

eQTL analysis has proved valuable in assigning function to intergenic SNPs associated with disease in genome-wide association studies (GWAS).[61] Combining eQTL analyses with chromatin capture techniques [e.g. chromosome conformation capture,[62] circular chromosome conformation capture,[63] genome conformation capture,[64] high-throughput chromosome conformation capture (Hi-C)[15]], which detect spatial proximity of chromosomal loci, provides further evidence that an enhancer in which a SNP resides is spatially and functionally linked to the target gene.[54,56,57,65–67] Utilizing spatial proximity data to identify candidate regulatory targets increases the

https://doi.org/10.1017/S2040174417000678 Published online by Cambridge University Press

power of the study; fewer putative eQTLs are calculated and thus the statistical correction for multiple testing[68,69] is less severe.[56] For example, an obesity-associated locus on chromosome 16, identified from GWAS studies, was found to have no effect on transcript levels of the nearest gene (*FTO*).[57] Instead circular chromatin conformation capture followed by high-throughput sequencing (4C-seq) identified *IRX3*, a gene 300 kb away, as the target of the daSNPs.[57,58] These combined analyses help to interpret the effects of intergenic and non-coding SNPs by identifying the genes and genetic pathways that they affect. However, this approach relies upon the underlying assumption that intergenic and intronic daSNPs mark regulatory loci (e.g. enhancers, repressors, or modifiers of the aforementioned).

Intergenic SNPs are difficult to categorize, as they often fall outside conserved regions, non-coding RNAs, known enhancers, or distal regulatory elements. Chen and Tian[55] approached this issue by grouping all intergenic SNPs with their nearest regulatory element. They then predicted the target genes of each regulatory element using spatial proximity, epigenetic data and phylogenetic profiles.[55] This approach found that the predicted targets of the regulatory elements were often enriched for protein-coding genes associated with the investigated diseases. However, assigning SNPs to the closest regulatory element in cis, without evidence for a functional connection is a problematic assumption. In many respects this approach perpetuates our earlier practice of assigning SNPs to the closest protein-coding gene.

Combining information on the spatial organization and functional impact (e.g. eQTLs) of daSNPs to determine how they contribute to a phenotype is further complicated by the complexity of the regulatory circuits that exist within eukaryotic nuclei. For example, enhancers or repressors need not act individually. Rather, the elements are combinatorial and the tissue-specific manner in which they connect contributes to counteract stochastic variation in the regulation of the target gene. Consistent with this, Corradin *et al.*[70] found that within clusters of super-enhancers, isolated SNPs can have large effects on the disease risk in combination with known risk SNPs, even if one variant does not reach genome-wide significance or have a detectable spatial interaction with the target gene. Moreover, variants that alter epigenetic patterns can affect not just local gene regulation but large scale genome organization. For instance the CCCTC-binding factor (CTCF) is a key architectural protein,[71] holding together megabase scale regions of DNA.[72] These structures are known as topologically associated domains (TADs). It thought that TADs function to increase the incidence of contacts between loci within the TAD while simultaneously insulating genes in one TAD from the effects of enhancers in another.[72] CTCF binding varies greatly between cell types, and can be sensitive to DNA methylation.[73,74] Variants that affect methylation patterns (meQTLs)[75] could therefore cause widespread transcriptional changes by disrupting TAD boundaries.[76]

Future directions

Genome organization is a record of nuclear activity including gene regulation patterns.[22,54] These marks can be used to further our understanding of phenotypes. For example, genome organization informed-discovery of allele-specific enhancer, insulator or promoter activity using intergenic SNPs can be integrated into GWAS to help explain the environment-genotype component of missing human heritability.[52] However, accurate deconvolution of the nuclear activity requires accurate maps and contact-informed models of the genomic organization of different cell-types or tissues at different developmental or disease stages. The commonly used Hi-C technique requires hundreds of millions of reads in order to capture a representation of the interactions that are occurring in the genome.[15] However, due to the complexity of these libraries, specific interactions are rarely sequenced to a sufficient depth for interrogation.[37] Capture Hi-C is a method that enriches a Hi-C library for all interactions with, for example, gene promoters[37] or GWAS loci.[36,77] Use of this targeted approach enables the identification of all possible targets of non-coding risk loci identified by GWAS whilst overcoming limitations that are inherent to both microscopy and proximity ligation.[18,78,79]

A further limitation of both GWAS and Hi-C is that of resolution. GWAS can identify daSNPs, but they merely mark a locus that has potential regulatory effects associated with the phenotype of interest. The daSNP is typically in high linkage with one or more SNPs that are located within a linkage disequilibrium block. Similarly, Hi-C identifies an interacting region containing the tag SNP. However, linkage disequilibrium blocks can potentially cross several restriction fragments. Therefore, targeted methods such as Capture Hi-C must identify interactions that occur within the linkage disequilibrium block associated with the tag SNP – not simply the tag SNP itself.

It is currently not possible to bioinformatically determine the causal SNP within a region, but functional annotation can be used to prioritize SNPs for experimental follow-up.[80] The patterns of enhancers, methylation, histone modification, protein binding sequences and DNase hypersensitivity sites can all be used to predict plausible causal SNPs using large, publicly available datasets.[51,58,81,82] Information about the spatial organization of the genome can also contribute to this prediction, particularly if multiple restriction enzymes were used during proximity ligation, reducing the fragment size and identifying the interacting region with greater precision (Fig. 2). These predictions should then be tested using gene editing techniques, such as CRISPR/Cas9,[58] which enable the isolation of a specific SNP effect without losing the three-dimensional context of the interaction. Cell choice is essential in these types of study, due to the tissue specific nature of the genome organization.[11–15]

Furthermore, carefully designed studies are required to find variants that increase disease risk only under specific environmental conditions,[83] or variants that may contribute to a pathogenic environment such as hyperphagia.[84]

https://doi.org/10.1017/S2040174417000678 Published online by Cambridge University Press

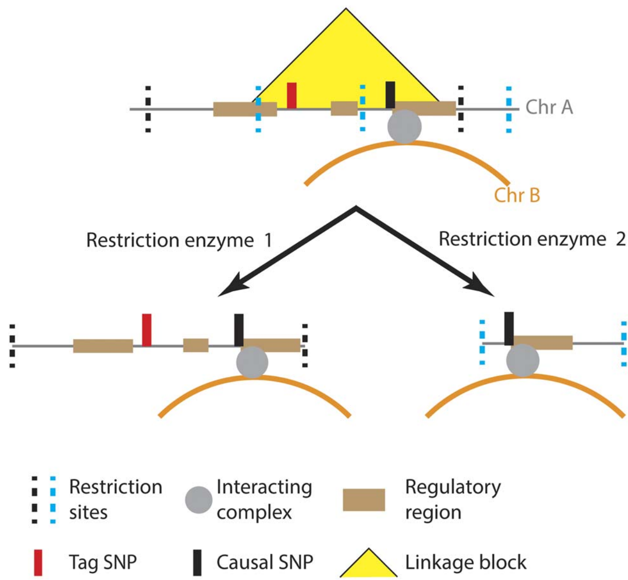

Fig. 2. Disease associated single-nucleotide polymorphisms (SNPs) identified by genome-wide association studies are often in linkage with one or more SNPs, any of which may be causal.[51] Functional information, such as epigenetic marks and experimentally validated enhancers, is used to identify regulatory regions which are more likely to contain causal SNPs.[58,81] Comparisons of genomic organizations captured by proximity ligation with different restriction enzymes can be used to refine the identification of the interacting regions and the causal SNP.

In multi-cellular organisms the nucleus is not a closed system and the genome is not a single entity. For example, interactions between the mitochondrial and nuclear genomes have been captured and linked to the control of gene expression, DNA repair and the cell cycle.[85,86] Therefore, inter-organelle DNA interactions likely form a highly specific component of intra-cellular communication. Future work should investigate the potential for inter-organelle DNA interactions to contribute directly to the regulatory mechanisms through which daSNPs located in the mitochondria, and other nucleated organelles, contribute to complex phenotypes.

Conclusion

Gene regulation and regulatory networks are a critical component of developmental processes and environmental responses. Genome structure acts in a read–write capacity capable of capturing the underlying action of the regulome or possibly even directly inducing changes under conditions of physical stress.[87] These interactions contribute to explaining how the various levels of nuclear control (structural, epigenetic and proteomic) come together to define genes and ensure cellular adaptation and selection through appropriate gene regulation, recombination and replication. Approaching the study of daSNPs from this viewpoint enables the interrogation of the genome as a complex organ[88] capable of permutations to define genes in response to environmental stimuli. Including information about the distribution and dynamic profiles of other epigenetic marks can further increase the power of these analyses by identifying the effects of gene by environment interactions on the epigenome. Further work to describe the interleaved genome promises to elucidate how epigenetics contributes to the control of developmental pathways.[89]

Acknowledgements

The authors would like to thank Phillip Smith, William Schierding and Tayaza Fadason for comments on this manuscript.

Financial Support

This work was supported by the Health Research Council New Zealand (grant number HRC 15/504 to JOS) and a University of Auckland Scholarship to E.J.

Conflicts of Interest

None.

References

1. Godfrey KM, Reynolds RM, Prescott SL, *et al.* Influence of maternal obesity on the long-term health of offspring. *Lancet Diabetes Endocrinol.* 2017; 5, 53–64.
2. O'Reilly JR, Reynolds RM. The risk of maternal obesity to the long-term health of the offspring. *Clin Endocrinol (Oxf)*. 2013; 78, 9–16.
3. Ong M-L, Lin X, Holbrook JD. Measuring epigenetics as the mediator of gene/environment interactions in DOHaD. *J Dev Orig Health Dis*. 2015; 6, 10–16.
4. Carlson EA. *The Gene; A Critical History*. 1966. Saunders: Philadelphia.
5. Everson T. *The Gene: A Historical Perspective*. 2007. Greenwood Press: Westport.
6. Fox Keller E. *The Century of the Gene*. 2000. Harvard University Press: Cambridge.
7. Gerstein MB, Bruce C, Rozowsky JS, *et al.* What is a gene, post-ENCODE? History and updated definition. *Genome Res*. 2007; 17, 669–681.
8. Lamm E. The metastable genome: a Lamarckian organ in a Darwinian world? In *Transformations of Lamarckism: From Subtle Fluids to Molecular Biology* (eds. Jablonka E, Gissis S), 2011; 480pp. MIT Press: Cambridge, Massachusetts.
9. Griffiths PE, Neumann-Held EM. The many faces of the gene. *Bioscience*. 1999; 49, 656–662.
10. Akiva P, Toporik A, Edelheit S, *et al.* Transcription-mediated gene fusion in the human genome. *Genome Res*. 2006; 16, 30–36.
11. Spilianakis CG, Lalioti MD, Town T, *et al.* Interchromosomal associations between alternatively expressed loci. *Nature*. 2005; 435, 637–645.
12. Dixon JR, Jung I, Selvaraj S, *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature*. 2015; 518, 331–336.

https://doi.org/10.1017/S2040174417000678 Published online by Cambridge University Press

13. Bouwman BAM, de Laat W. Getting the genome in shape: the formation of loops, domains and compartments. *Genome Biol*. 2015; 16, 154.
14. Fraser J, Ferrai C, Chiariello AM, *et al*. Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Mol Syst Biol*. 2015; 11, 852–852.
15. Rao SSP, Huntley MH, Durand NC, *et al*. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014; 159, 1665–1680.
16. Bischof M. Introduction to integrative biophysics. In *Integrative Biophysics* (eds. Popp F-A, Beloussov L), 2010; pp. 1–115. Springer-Science + Business Media: Dordrecht.
17. O'Sullivan J, Hendy M, Pichugina T, *et al*. The statistical-mechanics of chromosome conformation capture. *Nucleus*. 2013; 4, 1–9.
18. Grand RS, Gehlen LR, O'Sullivan JM. Methods for the investigation of chromosome organization. In *Advances in Genetics Research* (ed. Urbano KV), 2011; 5, 111–129. NOVA: Science publishers; ebook.
19. Kauffman SA. *The Origins of Order: Self Organization and Selection in Evolution*. 1993. Oxford University Press: New York.
20. Kapranov P, Willingham AT, Gingeras TR. Genome-wide transcription and the implications for genomic organization. *Nat Rev Genet*. 2007; 8, 413–423.
21. Dixon JR, Selvaraj S, Yue F, *et al*. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012; 485, 376–380.
22. de Wit E, Bouwman BAM, Zhu Y, *et al*. The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature*. 2013; 501, 227–231.
23. Krijger PHL, Di Stefano B, de Wit E, *et al*. Cell-of-origin-specific 3D genome structure acquired during somatic cell reprogramming. *Cell Stem Cell*. 2016; 18, 597–610.
24. Holwerda SJB, de Laat W. CTCF: the protein, the binding partners, the binding sites and their chromatin loops. *Philos Trans R Soc Lond B Biol Sci*. 2013; 368, 20120369.
25. Merkenschlager M, Nora EP. CTCF and cohesin in genome folding and transcriptional gene regulation. *Annu Rev Genomics Hum Genet*. 2016; 17, 17–43.
26. Mizuguchi T, Fudenberg G, Mehta S, *et al*. Cohesin-dependent globules and heterochromatin shape 3D genome architecture in *S. pombe*. *Nature*. 2014; 516, 432–435.
27. Brangwynne CP, Tompa P, Pappu RV. Polymer physics of intracellular phase transitions. *Nat Phys*. 2015; 11, 899–904.
28. Kampmann M. Facilitated diffusion in chromatin lattices: mechanistic diversity and regulatory potential. *Mol Microbiol*. 2005; 57, 889–899.
29. Bénichou O, Chevalier C, Meyer B, Voituriez R. Facilitated diffusion of proteins on chromatin. *Phys Rev Lett*. 2011; 106, 38102.
30. Erdel F, Müller-Ott K, Rippe K. Establishing epigenetic domains via chromatin-bound histone modifiers. *Ann N Y Acad Sci*. 2013; 1305, 29–43.
31. Buckley SM, Aranda-Orgilles B, Strikoudis A, *et al*. Regulation of pluripotency and cellular reprogramming by the ubiquitin-proteasome system. *Cell Stem Cell*. 2012; 11, 783–798.
32. Kim DH, Marinov GK, Pepke S, *et al*. Single-cell transcriptome analysis reveals dynamic changes in lncRNA expression during reprogramming. *Cell Stem Cell*. 2015; 16, 88–101.
33. Grand RS, Pichugina T, Gehlen LR, *et al*. Chromosome conformation maps in fission yeast reveal cell cycle dependent sub nuclear structure. *Nucleic Acids Res*. 2014; 42, 12585–12599.
34. Pichugina T, Sugawara T, Kaykov A, *et al*. A diffusion model for the coordination of DNA replication in *Schizosaccharomyces pombe*. *Sci Rep*. 2016; 6, 18757.
35. Dryden NH, Broome LR, Dudbridge F, *et al*. Unbiased analysis of potential targets of breast cancer susceptibility loci by capture Hi-C. *Genome Res*. 2014; 24, 1854–1868.
36. Jäger R, Migliorini G, Henrion M, *et al*. Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. *Nat Commun*. 2015; 6, 6178.
37. Mifsud B, Tavares-Cadete F, Young AN, *et al*. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet*. 2015; 47, 598–606.
38. Williams A, Spilianakis CG, Flavell RA. Interchromosomal association and gene regulation in trans. *Trends Genet*. 2010; 26, 188–197.
39. Felipe Barella L, ulio Cezar de Oliveira J, Cezar de Freitas Mathias P. Pancreatic islets and their roles in metabolic programming. *Nutrition*. 2014; 30, 373–379.
40. Vickers MH. Early life nutrition, epigenetics and programming of later life disease. *Nutrients*. 2014; 6, 2165–2178.
41. Jarick I, Vogel CIG, Scherag S, *et al*. Novel common copy number variation for early onset extreme obesity on chromosome 11q11 identified by a genome-wide analysis. *Hum Mol Genet*. 2011; 20, 840–852.
42. Comuzzie AG, Cole SA, Laston SL, *et al*. Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *PLoS One*. 2012; 7, e51954.
43. Fall T, Ingelsson E. Genome-wide association studies of obesity and metabolic syndrome. *Mol Cell Endocrinol*. 2014; 382, 740–757.
44. Sjögren M, Lyssenko V, Jonsson A, *et al*. The search for putative unifying genetic factors for components of the metabolic syndrome. *Diabetologia*. 2008; 51, 2242–2251.
45. Hara K, Fujita H, Johnson TA, *et al*. Genome-wide association study identifies three novel loci for type 2 diabetes. *Hum Mol Genet*. 2014; 23, 239–246.
46. Zeggini E, Scott LJ, Saxena R, *et al*. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet*. 2008; 40, 638–645.
47. Morris AP, Voight BF, Teslovich TM, *et al*. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet*. 2012; 44, 981–990.
48. Sladek R, Prokopenko I. Genome-wide association studies of type 2 diabetes. In *The Genetics of Type 2 Diabetes and Related Traits: Biology, Physiology and Translation* (ed. Florez CJ), 2016; pp. 13–61. Springer International Publishing: Cham.
49. Manolio TA, Collins FS, Cox NJ, *et al*. Finding the missing heritability of complex diseases. *Nature*. 2009; 461, 747–753.
50. Vattikuti S, Guo J, Chow CC. Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits. *PLoS Genet*. 2012; 8, e1002637.

https://doi.org/10.1017/S2040174417000678 Published online by Cambridge University Press

51. Farh KK, Marson A, Zhu J, *et al*. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*. 2015; 518, 337–343.
52. Schierding W, Cutfield WS, O'Sullivan JM. The missing story behind genome wide association studies: single nucleotide polymorphisms in gene deserts have a story to tell. *Front Genet*. 2014; 5, 39.
53. Marsman J, Horsfield JA. Long distance relationships: enhancer–promoter communication and dynamic gene transcription. *Biochim Biophys Acta Gene Regul Mech*. 2012; 1819, 1217–1227.
54. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature*. 2012; 489, 109–113.
55. Chen J, Tian W. Explaining the disease phenotype of intergenic SNP through predicted long range regulation. *Nucleic Acids Res*. 2016; 44, 8641–8654.
56. Schierding W, Antony J, Cutfield WS, *et al*. Intergenic GWAS SNPs are key components of the spatial and regulatory network for human growth. *Hum Mol Genet*. 2016; 25, 3372–3382.
57. Smemo S, Tena JJ, Kim K-H, *et al*. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*. 2014; 507, 371–375.
58. Claussnitzer M, Dankel SN, Kim K-H, *et al*. FTO obesity variant circuitry and adipocyte browning in humans. *N Engl J Med*. 2015; 373, 895–907.
59. Tolhuis B, Palstra RJ, Splinter E, *et al*. Looping and interaction between hypersensitive sites in the active β-globin locus. *Mol Cell*. 2002; 10, 1453–1465.
60. Drissen R, Palstra R-J, Gillemans N, *et al*. The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes Dev*. 2004; 18, 2485–2490.
61. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet*. 2015; 16, 197–212.
62. Naumova N, Smith EM, Zhan Y, Dekker J. Analysis of long-range chromatin interactions using chromosome conformation capture. *Methods*. 2012; 58, 192–203.
63. Zhao Z, Tavoosidana G, Sjölinder M, *et al*. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet*. 2006; 38, 1341–1347.
64. Rodley CDM, Bertels F, Jones B, O'Sullivan JM. Global identification of yeast chromosome interactions using genome conformation capture. *Fungal Genet Biol*. 2009; 46, 879–886.
65. Schierding W, O'Sullivan JM. Connecting SNPs in diabetes: a spatial analysis of meta-GWAS loci. *Front Endocrinol (Lausanne)*. 2015; 6, doi: 10.3389/fendo.2015.00102.
66. Dean A. In the loop: long range chromatin interactions and gene regulation. *Brief Funct Genomics*. 2011; 10, 3–10.
67. Harmston N, Lenhard B. Chromatin and epigenetic features of long-range gene regulation. *Nucleic Acids Res*. 2013; 41, 7185–7199.
68. Doss S. Cis-acting expression quantitative trait loci in mice. *Genome Res*. 2005; 15, 681–691.
69. Davis JR, Fresard L, Knowles DA, *et al*. An efficient multiple-testing adjustment for eQTL studies that accounts for linkage disequilibrium between variants. *Am J Hum Genet*. 2016; 98, 216–224.
70. Corradin O, Cohen AJ, Luppino JM, *et al*. Modeling disease risk through analysis of physical interactions between genetic variants within chromatin regulatory circuitry. *Nat Genet*. 2016; 48, 1313–1320.
71. Ong C-T, Corces VG. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet*. 2014; 15, 239–246.
72. Nora EP, Goloborodko A, Valton A-L, *et al*. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell*. 2017; 169, 930–944.e22.
73. Wang H, Maurano MT, Qu H, *et al*. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res*. 2012; 22, 1680–1688.
74. Maurano M, Wang H, John S, *et al*. Role of DNA methylation in modulating transcription factor occupancy. *Cell Rep*. 2015; 12, 1184–1195.
75. Banovich NE, Lan X, McVicker G, *et al*. Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels. *PLoS Genet*. 2014; 10, e1004663.
76. Flavahan WA, Drier Y, Liau BB, *et al*. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*. 2015; 529, 110–114.
77. Martin P, McGovern A, Orozco G, *et al*. Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci. *Nat Commun*. 2015; 6, 10069.
78. Dekker J. The three "C" s of chromosome conformation capture: controls, controls, controls. *Nat Methods*. 2006; 3, 17–21.
79. de Wit E, de Laat W. A decade of 3C technologies : insights into nuclear organization. *Genes Dev*. 2012; 26, 11–24.
80. Tak YG, Farnham PJ. Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenet Chromat*. 2015; 8, 57.
81. Kichaev G, Yang W-Y, Lindstrom S, *et al*. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet*. 2014; 10, e1004722.
82. Pasaniuc B, Price AL. Dissecting the genetics of complex traits using summary association statistics. *Nat Rev Genet*. 2016; 18, 117–127.
83. Huang Y, Cate SP, Battistuzzi C, *et al*. An association between a functional polymorphism in the monoamine oxidase a gene promoter, impulsive traits and early abuse experiences. *Neuropsychopharmacology*. 2004; 29, 1498–1505.
84. Yilmaz Z, Davis C, Loxton NJ, *et al*. Association between MC4R rs17782313 polymorphism and overeating behaviors. *Int J Obes*. 2015; 39, 114–120.
85. Rodley CDM, Grand RS, Gehlen LR, *et al*. Mitochondrial-nuclear DNA interactions contribute to the regulation of nuclear transcript levels as part of the inter-organelle communication system. *PLoS One*. 2012; 7, e30943.
86. Doynova MD, Berretta A, Jones MB, *et al*. Interactions between mitochondrial and nuclear DNA in mammalian cells are non-random. *Mitochondrion*. 2016; 30, 187–196.
87. Jacobson E, Perry JK, Long DS, *et al*. A potential role for genome structure in the translation of mechanical force during immune cell development. *Nucleus*. 2016; 7, 462–475.
88. Lamm E. The genome as a developmental organ. *J Physiol*. 2014; 592, 2283–2293.
89. Bard JBL. Waddington's legacy to developmental and theoretical biology. *Biol Theory*. 2008; 3, 188–197.

https://doi.org/10.1017/S2040174417000678 Published online by Cambridge University Press