



King's Research Portal

DOI:

[10.3389/fcell.2022.832348](https://doi.org/10.3389/fcell.2022.832348)

Document Version

Publisher's PDF, also known as Version of record

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Cain, J., Montibus, B., & Oakey, R. (2022). Intragenic CpG islands and their impact on gene regulation. *Frontiers in Cell and Developmental Biology*, 10, Article 832348. <https://doi.org/10.3389/fcell.2022.832348>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Intragenic CpG Islands and Their Impact on Gene Regulation

James A. Cain[†], Bertille Montibus[†] and Rebecca J. Oakey^{*†}

Department of Medical and Molecular Genetics, King's College London, Guy's Hospital, London, United Kingdom

OPEN ACCESS

Edited by:

Robert Feil,
UMR5535 Institut de Génétique
Moléculaire de Montpellier (IGMM),
France

Reviewed by:

Maxim Van Cleef Greenberg,
UMR7592 Institut Jacques Monod
(IJM), France

*Correspondence:

Rebecca J. Oakey
rebecca.oakey@kcl.ac.uk

†ORCID ID:

James A. Cain,
orcid.org/0000-0001-8141-0666
Bertille Montibus,
orcid.org/0000-0002-6895-3954
Rebecca J. Oakey,
orcid.org/0000-0003-2706-8139

Specialty section:

This article was submitted to
Developmental Epigenetics,
a section of the journal
Frontiers in Cell and Developmental
Biology

Received: 09 December 2021

Accepted: 20 January 2022

Published: 11 February 2022

Citation:

Cain JA, Montibus B and Oakey RJ
(2022) Intragenic CpG Islands and
Their Impact on Gene Regulation.
Front. Cell Dev. Biol. 10:832348.
doi: 10.3389/fcell.2022.832348

The mammalian genome is depleted in CG dinucleotides, except at protected regions where they cluster as CpG islands (CGIs). CGIs are gene regulatory hubs and serve as transcription initiation sites and are as expected, associated with gene promoters. Advances in genomic annotations demonstrate that a quarter of CGIs are found within genes. Such intragenic regions are repressive environments, so it is surprising that CGIs reside here and even more surprising that some resist repression and are transcriptionally active within a gene. Hence, intragenic CGI positioning within genes is not arbitrary and is instead, selected for. As a wealth of recent studies demonstrate, intragenic CGIs are embedded within genes and consequently, influence 'host' gene mRNA isoform length and expand transcriptome diversity.

Keywords: polyadenylation, epigenetics, DNA methylation, orphan CpG-Islands, CpG island (CGI), alternative polyadenylation (APA), mRNA processing

INTRODUCTION

Gene regulation is a prerequisite of life, the seemingly simple decision of whether to express a gene or not is present in nearly all organisms. Regulatory elements are sequence specific motifs in the mammalian genome that coordinate gene expression. One fundamental class of regulatory element are the CpG islands (CGIs). CGIs are regions of the genome that are enriched for cytosine and guanine dinucleotides (CpGs) and have been defined bioinformatically as having a GC content over 50%, an observed CpG ratio compared to the whole genome (Obs/Exp) of over 0.6 and a length of over 200 bps (Gardiner-Garden and Frommer, 1987) (**Figure 1A**). CpG's in isolation are modified with DNA methylation, the addition of a methyl group onto cytosine, which is a heritable epigenetic mark. However, when CpGs congregate into islands they are generally protected from DNA methylation (Bird, 1978; Bird et al., 1979; Bird et al., 1985).

Instances where CGIs are DNA methylated have been correlated with transcriptional silencing when they are in close proximity to transcription start sites (TSS) of genes (Deaton and Bird, 2011). Using bioinformatic criteria, most CGIs are indeed found at TSSs, but this does not take into account their biochemical potential to exhibit or lack DNA methylation (Larsen et al., 1992; Takai and Jones, 2002; Saxonov et al., 2006). To overcome this, methods that specifically enrich for DNA fragments containing CGIs, with and without DNA methylation, were developed and combined with next generation sequencing to biochemically detect CGI coordinates (Illingworth et al., 2008; Blackledge et al., 2012) (**Figure 1B**). These studies found that only half of CGIs in mouse and human genomes are associated with TSS of genes and the rest are unannotated "orphans", that are located either distal to genes (intergenic) or within genes themselves (intragenic) (**Figure 1C**). Not only are these "orphan" CGIs more likely to be DNA methylated (**Figure 1C**), but they are also more likely to demonstrate DNA methylation and transcriptional signatures in a tissue-specific manner (Illingworth et al., 2010; Deaton et al.,

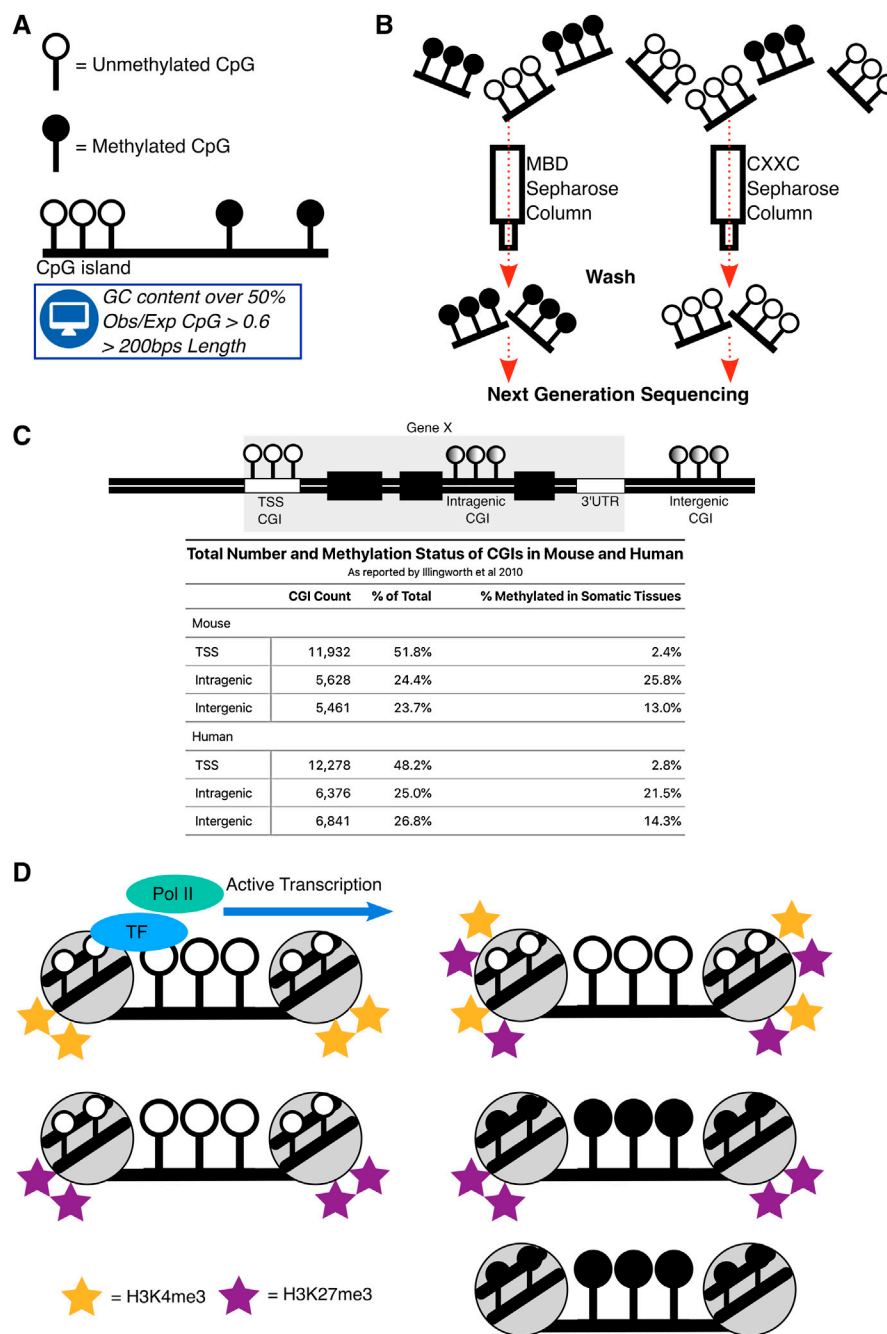


FIGURE 1 | Definitions and states of CpG islands. **(A)** Depiction of the typical CpGs in the mammalian genome which are DNA methylated in isolation but are devoid of this mark in the CGI context. CGIs were originally defined bioinformatically. **(B)** Schematic demonstrating CGIs locations were biochemically determined. MBD and CXXC proteins fixed to a Sepharose column allowed purification of DNA methylated and unmethylated CGIs in mammals. **(C)** Representation of CGIs across mammalian genomes and a table, summarising the proportion and methylation status of CGIs as reported by Illingworth et al. (2010) in mouse and human across TSS, Intragenic and Intergenic regions. Total CGI numbers in mouse = 23,021, human = 25,495 **(D)** Summary of the main states of CGIs across the genome. Their active state is associated with binding of transcription factors (TF) and subsequent RNA Polymerase II (Pol II) binding. Repressed states of CGIs are through combinations of DNA methylation, H3K4me3 and H3K27me3.

2011). Intragenic CGIs (iCGIs) are particularly striking as they are embedded within genes across mammalian species and are often not considered by most analyses, which use the standard bioinformatic definition of CGIs.

iCGIs can impact gene expression in a multitude of ways either by being transcriptionally active themselves or through interactions with biological processes in close vicinity. Biochemical methods discovered that a quarter of all CGIs are

within genes, existing as iCGIs. Recent studies have now tied iCGIs to multiple functions (Maunakea et al., 2010; Jeziorska et al., 2017; Amante et al., 2020). Comprehensive reviews discuss CGIs more broadly and in the context of development and disease (Deaton and Bird, 2011; Greenberg and Bourc'his, 2019). This minireview aims to update the current knowledge of CGIs and capture their repertoire of functions outside of canonical TSSs.

CGIS ARE PROMOTERS INDEPENDENT OF GENOMIC POSITION

Chromatin, the complex of DNA and histone proteins which forms chromosomes, can exist in an open or closed configuration indicative of active or inactive gene expression. The state of chromatin across the mammalian genome is studied through analysis of histone tail modifications, marking the histones that DNA is wrapped around. Over 100 histone modifications exist, some are well understood and some remain enigmatic, without a known biological function (Zhao and Garcia, 2015). Still, histone modifications are correlated to states of chromatin and are invaluable markers when studying gene regulation. CGIs overlap with >70% of canonical TSSs in the human genome and are typically associated with promoters (Saxonov et al., 2006), where they can exhibit multiple states, referred to here as 'CGI states'. These states can be categorised depending on their histone marks.

One state is bivalency, where CGIs are transcriptionally repressed, devoid of DNA methylation and exhibit both active histone 3 lysine 4 trimethylation (H3K4me3) and repressive histone 3 lysine 27 trimethylation (H3K27me3) modifications. Bivalency has been proposed to poise CGI promoters for activation (Bernstein et al., 2006; Voigt et al., 2013), but has more recently been suggested to protect CGIs from DNA methylation whilst simultaneously keeping them transcriptionally inactive (Maupetit-Méhouas et al., 2016; Kumar and Jothi, 2020; Shah et al., 2021). The majority of promoter associated CGIs across the human genome exhibit bivalency (Court et al., 2019). This state likely arises due to the sequence composition of CGIs rather than their location, as CGIs experimentally introduced into the β -globin locus in mouse embryonic stem cells also displayed bivalency (Krebs et al., 2014; Wachter et al., 2014). The shift from bivalent CGIs to active CGIs is initiated through the binding of transcription factors, leading to the removal of H3K27me3, whilst maintaining the H3K4me3 mark. Removal of H3K4me3 and maintenance of H3K27me3 at the CGI is repressive, otherwise known as polycomb-only mediated repression, and is observed at a minority of promoter CGIs in somatic tissues (Mikkelsen et al., 2007; Farcas et al., 2012; Court et al., 2019; Blackledge et al., 2020).

A more stable form of repression at CGIs is through DNA methylation. In somatic tissues, DNA methylation represses promoter CGIs at the inactivated X chromosome (Augui et al., 2011; Galupa and Heard, 2018), germline genes (Velasco et al., 2010; Dahlet et al., 2021; Mochizuki et al., 2021), imprinted genes (Barlow and Bartolomei, 2014), and some lineage-committed genes (Dahlet et al., 2020). Whilst H3K27me3 and DNA

methylation are both repressive, they are typically mutually exclusive at CGIs (Brinkman et al., 2012; Statham et al., 2012). Chromatin immunoprecipitation experiments indicate that H3K27me3 and DNA methylation can co-exist at some imprinted genes (Maupetit-Méhouas et al., 2016). CGIs can therefore exhibit multiple states of chromatin which are indicative of their transcriptional potential (Blackledge and Klose, 2011) (Figure 1D).

iCGIs are more likely to be DNA methylated (Figure 1C) and those lacking DNA methylation can exhibit bivalent chromatin signatures and when transcriptionally active, show transcription factor binding and the promoter mark, H3K4me3 (Lee et al., 2017; Amante et al., 2020; Choi et al., 2020). iCGIs can therefore exist in the same 'states' as promoter CGIs, albeit in different proportions. The iCGI states themselves are regulated in a tissue-specific manner, and with crosstalk from the gene that 'hosts' them. This can lead to both, consequences on the iCGI itself and their corresponding host gene.

CONSEQUENCES OF BEING AN INTRAGENIC CGI WITHIN A GENE

The location of iCGIs within a gene is a turbulent place for a promoter region because active transcription results in the silencing of DNA which has been transcribed through. At first, this sounds paradoxical, but it has been identified at various loci that transcription through a gene promoter can silence it. This phenomenon was first demonstrated at the α -globin locus in a case of α -thalassemia where the *LUC7L* gene is juxtaposed upstream of *HBA2*. Here, *LUC7L* transcription extends through the *HBA2* promoter CGI which is subsequently DNA methylated and silenced (Tufarelli et al., 2003). This can be observed naturally at regions of genes that contain clusters of overlapping genes, such as at the imprinted loci, *Gnas* and *Igfr2* and likely, at *Kcnq1* too. At the *Gnas* locus, incoming transcription from upstream *Nesp* removes H3K4me3 at the *Gnas* CGI and establishes DNA methylation and silencing (Chotalia et al., 2009; Williamson et al., 2011). Transcription of the *Airn* long non-coding RNA (lncRNA) through the *Igfr2* promoter silences *Igfr2* (Latos et al., 2012; Santoro et al., 2013). Similarly, the *Kcnq1* locus contains an overlapping transcript, *Kcnq1ot1*, that overlaps with the *Kcnq1* CGI promoter. Silencing of *Kcnq1* is correlated to transcription of the overlapping *Kcnq1ot1*, suggesting transcription itself is causing gene repression (Golding et al., 2011). Genome-wide analysis now highlights that this repression is through interactions between the transcribing RNA Polymerase II and the deposition of the elongation associated histone mark, H3K36me3. This in turn recruits DNMT3B, to deposit *de novo* intragenic DNA methylation (Baubec et al., 2015; Neri et al., 2017; Dahlet et al., 2020) (Figure 2A).

This may indicate that tissue-specific patterns of DNA methylation at iCGIs are a by-product of transcription through the gene itself, where iCGI function as a promoter is silenced when the host gene is transcriptionally active. Whilst iCGIs hosted within an active gene are generally

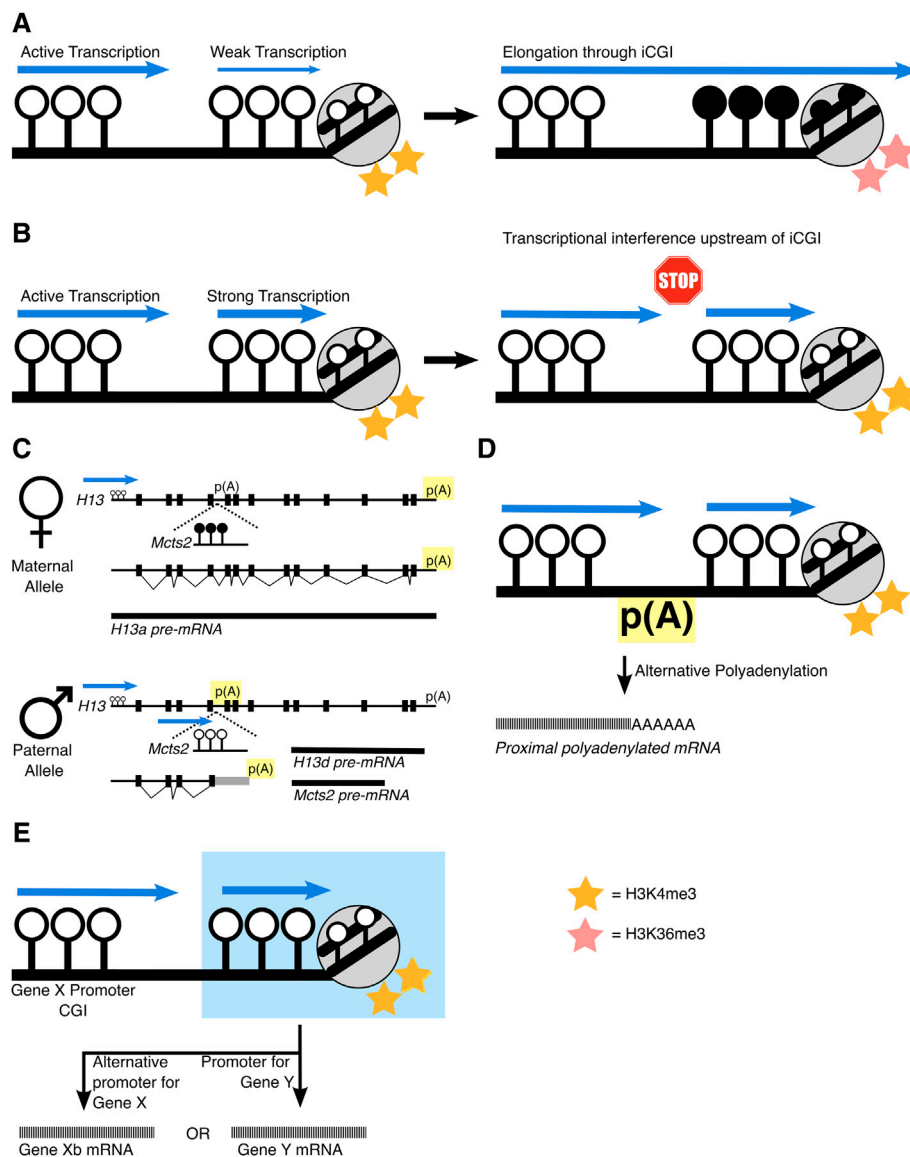


FIGURE 2 | Schematics of how iCGIs impact gene regulation mechanisms. **(A)** Transcription through a ‘weak’ iCGI can silence it, depositing H3K36me3 and DNA methylation at the iCGI. **(B)** However, if the iCGI exhibits strong transcriptional activity, it can lead to transcriptional interference. This can result in events akin to those at the **(C)** H13/Mcts2 locus, that exhibits allele-specific PAS usage. Usage of the PAS is highlighted in yellow. **(D)** Similar mechanisms have been found other iCGIs. Alternatively, and in some cases, simultaneously, **(E)** the iCGI can act as a promoter itself, highlighted in blue, for either the host gene itself (gene X) or for a different ‘nested’ gene (gene Y).

silenced, subsets of iCGIs that show more RNA Polymerase II binding are protected from this silencing and maintain their H3K4me3 promoter status (Jeziorska et al., 2017). This indicates that iCGIs can resist the silencing power of transcription, but only if they are ‘strong’ enough to do so (Figures 2A,B).

What are the factors that dictate CGI strength? There is speculation that long CGIs may exhibit more sites for RNA Polymerase II binding (Elango and Yi, 2011) and a higher CpG density is correlated to enhanced transcription factor binding (Hartl et al., 2019). Given that iCGIs are generally shorter than promoter CGIs and less CpG dense, this may

explain why a subset of iCGIs are silenced. But, despite this, subsets of iCGIs escape transcriptional silencing and this can have a series of effects on the host gene itself.

CONSEQUENCES ON THE GENE FOR HOSTING AN ACTIVE INTRAGENIC CGI

Polyadenylation and splicing are co-transcriptional processes that can generate a diversity of mature mRNA isoforms from a single gene. Briefly, regulation of splicing and polyadenylation can control which exons of the pre-mRNA are utilised and when

the pre-mRNA should be terminated. Alternative regulation of either of these processes impact the function of the mature mRNA (Proudfoot, 2011; Lee and Rio, 2015) and both recruit large protein machineries that regulate these processes co-transcriptionally (Lee and Rio, 2015; Tian and Manley, 2016; Gruber and Zavolan, 2019). As such, it seems plausible that iCGI activity can impact splicing and polyadenylation when they are co-occurring in close proximity.

Coincidentally, there are a wealth of studies linking active iCGIs to alternative polyadenylation (APA) events, specifically intronic APA (iAPA), which can alter the protein coding sequence of mRNA transcripts as it is terminated prematurely. This was first demonstrated at the imprinted *Mcts2/H13* locus (Wood et al., 2008). Here, *H13* isoforms are alternatively polyadenylated depending on the parental origin of DNA methylation at the iCGI within *H13*'s fifth intron. This iCGI is a promoter for a nested gene, *Mcts2* and when active (paternal allele), polyadenylation of *H13* occurs within intronic regions. However, when *Mcts2* and its iCGI promoter are silenced (maternal allele), polyadenylation occurs at the 3'UTR of *H13* (Figure 2C). This mechanism of APA is the same at the imprinted *Nap115/Herc3* locus. Here, an iCGI is a promoter for *Nap115* and its parental origin is correlated with the polyadenylation site choice of the host gene, *Herc3* (Cowley et al., 2012).

Outside of the imprinted context, two recent studies which perturbed DNA methylation showed similar results at iCGIs. Knockout of DNA methyltransferases (*DNMT1* & *DNMT3B*) in cancer cells increased initiating RNA Polymerase II at the iCGI which was correlated with the usage of proximal polyadenylation sites of two host genes (Nanavaty et al., 2020). Similar polyadenylation site usage was also found when DNA methylation was perturbed at the iCGI within the *NFATc1* locus, resulting in alternative *NFATc1* isoforms. These locus specific effects have been detected genome-wide by a recent bioinformatic screen, emphasising that iCGI activity leads to premature transcription termination upstream of the iCGI, most likely through APA (Amante et al., 2020) (Figures 2B,D).

These findings demonstrate that a transcriptionally active iCGI can influence alternative polyadenylation and highlight the ways in which iCGIs can shape the transcriptome. Mechanistically, this is likely due to RNA polymerase II prematurely stopping because of meeting another initiating polymerase at the iCGI, otherwise known as transcriptional interference (TI) (Shearwin et al., 2005) (Figure 2B). Here, the polyadenylation machinery selects the nearest site to avoid the production of an unstable mRNA transcript. It is still undetermined whether iCGI activity influences APA serendipitously through TI, or if this is a direct mechanism to regulate pre-mRNA termination.

An active iCGI can also influence isoform choice more directly, by acting as an alternative promoter for the host gene (Figure 2D). The *SHANK3* gene for example, contains an iCGI which is differentially methylated between hippocampus and cortex astrocytes. In hippocampus astrocytes, the iCGI is active and devoid of DNA methylation where it serves as an alternative promoter for *SHANK3*, transcribing a shorter mRNA transcript.

Whereas in cortex astrocytes, when the iCGI is silenced through DNA methylation and instead, the canonical full length *SHANK3* isoform is transcribed (Maunakea et al., 2010).

CGI FUNCTION AS ENHANCER REGIONS

Recent work suggests that CGIs may have another regulatory role as enhancers. Enhancers are *cis*-regulatory 50-150bp DNA sequences that are characterised by enriched transcription factor binding sites, H3K4me1 and H3K27ac histone modifications and when active, regions of bidirectional transcription produce enhancer RNAs (eRNAs) (Kim et al., 2010; Santa et al., 2010; Li et al., 2016). eRNAs confer *cis*-regulatory effects by recruiting transcriptional machinery to target genes to induce gene activation (Arner et al., 2015), otherwise referred to as enhancer looping. Intragenic enhancers can interfere with host gene expression through transcriptional interference (Onodera et al., 2012; Cinghu et al., 2017), similar to active iCGIs. Bioinformatic analyses show that iCGIs themselves exhibit enhancer histone modifications, are actively transcribed to eRNAs, are conserved across mammalian species (Bell and Vertino, 2017) and show greater transcription factor binding (Steinhaus et al., 2020). Such signatures have been identified at an iCGI within *Kdm6b*, which exhibits H3K4me1 and loops to the promoter CGI to enhance *Kdm6b* expression (Montibus et al., 2021). Given that transcription of enhancer regions is required to deposit enhancer histone marks, it is unclear how CGIs are initially defined as enhancer regions (Kaikkonen et al., 2013).

Enhancer signatures are also found at the other type of 'orphan' CGI, intergenic CGIs. A recent study has challenged the idea that these CGIs directly serve as enhancers, and instead, boost proximal enhancer function (Pachano et al., 2021). Here, intergenic CGIs augment enhancers' ability to amplify only target genes that contain a CGI promoter themselves. As enhancers loop to promoter CGIs, the unmethylated intergenic CGIs that are within 3 kb of the proximal enhancer bring along machinery for efficient promoter CGI transcription. These intergenic CGIs also serve to protect transcription factor binding sites (TFBS) within the proximal enhancer from repressive DNA methylation (Pachano et al., 2021). The relationship between intergenic CGIs and proximal enhancers may be reciprocal, as the TFBS within the enhancer can assist recruitment of machinery to the intergenic CGI itself.

CONCLUSIONS, THE RELEVANCE OF INTRAGENIC CGIS IN BIOLOGY

CGIs are regions where transcription can initiate. Whilst most CGIs are localised to annotated TSSs, many can be found intragenically. In some cases, the iCGI is silenced; in others, active iCGIs impact pre-mRNA processing and promote or contain enhancer function.

iCGIs are more prone to DNA methylation during embryonic development and adult development compared to their TSS CGI

counterparts (Illingworth et al., 2010; Auclair et al., 2014), implying that regulation of iCGIs is crucial for tissue specific programming. For example, iCGIs are specifically expressed in brain tissues and their host genes function in brain-specific biological processes (Amante et al., 2020). In this case, iCGIs may function as TSSs for novel transcripts or result in APA of the host gene and therefore, expand the transcriptome during developmental processes such as neurogenesis. iCGIs are conserved across mammalian species (Illingworth et al., 2010), suggesting they are maintained and necessary for proper gene regulation. It is still unclear how the multitude of functionalities of iCGIs are specified, i.e., how does an iCGI know to serve as an alternative promoter or, to disrupt host gene polyadenylation.

Similarly, DNA methylation of iCGIs prevents spurious intragenic transcription (Neri et al., 2017; Dahlet et al., 2020). Blocking spurious intragenic transcriptional activity is a method to ensure productive elongation by RNA polymerase II. DNA hypomethylation is widespread in cancer cells and extends to intragenic regions (Ehrlich, 2002; Hon et al., 2012; Kulis et al., 2012), hinting that intragenic transcription may be widespread in cancer (Kulis et al., 2013). The *RB1* gene, for example, contains an imprinted iCGI where its DNA methylation is inversely correlated to expression of the full length *RB1* transcript (Kamber et al., 2009; Kulis et al., 2012). The region which contains the iCGI is commonly deleted in cases of chronic lymphocytic leukaemia, implying that iCGIs may be disrupted in cancer cells. Despite this, intragenic DNA hypomethylation in cancer is mainly outside of promoter CGIs and they are paradoxically, hypermethylated instead (Kulis et al., 2013; Court et al., 2019). It is currently unknown if the cancer signature of hypomethylation extends to iCGIs, or if they are hypermethylated like promoter CGIs and if this is functionally relevant.

Given their distinct regulation and that many are protected from DNA methylation it is reasonable to suggest that iCGIs are

required in mammalian biology. A clear challenge that has limited our understanding of iCGIs is their overlap with genomic annotations. Conventional short-read sequencing technologies present a challenge when trying to distinguish whether signals or reads stem from the host gene or the iCGI. The arrival of long-read sequencing technologies and the eventual decline in cost of such methods will allow these reads to be distinguished and aid understanding of iCGIs (Logsdon et al., 2020). This will further be enhanced by studying the methylation of iCGIs in more contexts, which will be possible when methods such as whole-genome bisulphite sequencing (WGBS) are more cost effective. Technology in its current state can also aid understanding of iCGIs, with greater reporting of genomic locations of CGIs in genome-wide analyses of DNA methylation, which are currently skewed to canonical TSSs.

AUTHOR CONTRIBUTIONS

JC wrote and composed the paper, BM and RO provided supervision and contributed significantly to the final version of the paper.

FUNDING

JC is supported by the UK Medical Research Council MR/N013700/1 and King's College London and is a member of the MRC Doctoral Training Partnership in Biomedical Sciences.

ACKNOWLEDGMENTS

We are extremely grateful to Hannah Mischo and Lisa Dressler for their insightful and useful comments on the review.

REFERENCES

- Amante, S. M., Montibus, B., Cowley, M., Barkas, N., Setiadi, J., Saadeh, H., et al. (2020). Transcription of Intragenic CpG Islands Influences Spatiotemporal Host Gene Pre-mRNA Processing. *Nucleic Acids Res.* 48, 8349–8359. doi:10.1093/nar/gkaa556
- Arner, E., Daub, C. O., Vitting-Seerup, K., Andersson, R., Lilje, B., Drabløs, F., et al. (2015). Transcribed Enhancers lead Waves of Coordinated Transcription in Transitioning Mammalian Cells. *Science* 347, 1010–1014. doi:10.1126/science.1259418
- Auclair, G., Guibert, S., Bender, A., and Weber, M. (2014). Ontogeny of CpG Island Methylation and Specificity of DNMT3 Methyltransferases during Embryonic Development in the Mouse. *Genome Biol.* 15, 545. doi:10.1186/s13059-014-0545-5
- Augui, S., Nora, E. P., and Heard, E. (2011). Regulation of X-Chromosome Inactivation by the X-Inactivation centre. *Nat. Rev. Genet.* 12, 429–442. doi:10.1038/nrg2987
- Barlow, D. P., and Bartolomei, M. S. (2014). Genomic Imprinting in Mammals. *Cold Spring Harbor Perspect. Biol.* 6, a018382. doi:10.1101/cshperspect.a018382
- Baubec, T., Colombo, D. F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A. R., et al. (2015). Genomic Profiling of DNA Methyltransferases Reveals a Role for DNMT3B in Genic Methylation. *Nature* 520, 243–247. doi:10.1038/nature14176
- Bell, J. S. K., and Vertino, P. M. (2017). Orphan CpG Islands Define a Novel Class of Highly Active Enhancers. *Epigenetics* 12, 449–464. doi:10.1080/15592294.2017.1297910
- Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., et al. (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* 125, 315–326. doi:10.1016/j.cell.2006.02.041
- Bird, A. P., Taggart, M. H., and Smith, B. A. (1979). Methylated and Unmethylated DNA Compartments in the Sea Urchin Genome. *Cell* 17, 889–901. doi:10.1016/0092-8674(79)90329-5
- Bird, A., Taggart, M., Frommer, M., Miller, O. J., and Macleod, D. (1985). A Fraction of the Mouse Genome that Is Derived from Islands of Nonmethylated, CpG-Rich DNA. *Cell* 40, 91–99. doi:10.1016/0092-8674(85)90312-5
- Bird, A. P. (1978). Use of Restriction Enzymes to Study Eukaryotic DNA Methylation. *J. Mol. Biol.* 118, 49–60. doi:10.1016/0022-2836(78)90243-7
- Blackledge, N. P., and Klose, R. (2011). CpG Island Chromatin. *Epigenetics* 6, 147–152. doi:10.4161/epi.6.2.13640
- Blackledge, N. P., Long, H. K., Zhou, J. C., Kriaucionis, S., Patient, R., and Klose, R. J. (2012). Bio-CAP: a Versatile and Highly Sensitive Technique to Purify and Characterise Regions of Non-methylated DNA. *Nucleic Acids Res.* 40, e32. doi:10.1093/nar/gkr1207
- Blackledge, N. P., Fursova, N. A., Kelley, J. R., Huseyin, M. K., Feldmann, A., and Klose, R. J. (2020). PRC1 Catalytic Activity Is Central to Polycomb System Function. *Mol. Cell* 77, 857–874. doi:10.1016/j.molcel.2019.12.001

- Brinkman, A. B., Gu, H., Bartels, S. J. J., Zhang, Y., Matarese, F., Simmer, F., et al. (2012). Sequential ChIP-Bisulfite Sequencing Enables Direct Genome-Scale Investigation of Chromatin and DNA Methylation Cross-Talk. *Genome Res.* 22, 1128–1138. doi:10.1101/gr.133728.111
- Choi, W.-Y., Hwang, J.-H., Cho, A.-N., Lee, A. J., Lee, J., Jung, I., et al. (2020). DNA Methylation of Intragenic CpG Islands Are Required for Differentiation from iPSC to NPC. *Stem Cell Rev. Rep.* 16, 1316–1327. doi:10.1007/s12015-020-10041-6
- Chotalia, M., Smallwood, S. A., Ruf, N., Dawson, C., Lucifero, D., Frontera, M., et al. (2009). Transcription Is Required for Establishment of Germline Methylation marks at Imprinted Genes. *Genes Dev.* 23, 105–117. doi:10.1101/gad.495809
- Cinghu, S., Yang, P., Kosak, J. P., Conway, A. E., Kumar, D., Oldfield, A. J., et al. (2017). Intragenic Enhancers Attenuate Host Gene Expression. *Mol. Cell* 68, 104–117. doi:10.1016/j.molcel.2017.09.010
- Court, F., Le Boiteux, E., Fogli, A., Müller-Barthélémy, M., Vauers-Barrière, C., Chautard, E., et al. (2019). Transcriptional Alterations in Glioma Result Primarily from DNA Methylation-independent Mechanisms. *Genome Res.* 29, 1605–1621. doi:10.1101/gr.249219.119
- Cowley, M., Wood, A. J., Böhm, S., Schulz, R., and Oakey, R. J. (2012). Epigenetic Control of Alternative mRNA Processing at the Imprinted Herc3/Nap115 Locus. *Nucleic Acids Res.* 40, 8917–8926. doi:10.1093/nar/gks654
- Dahlet, T., Argüeso Lleida, A., Al Adhami, H., Dumas, M., Bender, A., Ngondo, R. P., et al. (2020). Genome-wide Analysis in the Mouse Embryo Reveals the Importance of DNA Methylation for Transcription Integrity. *Nat. Commun.* 11, 3153. doi:10.1038/s41467-020-16919-w
- Dahlet, T., Truss, M., Frede, U., Al Adhami, H., Bardet, A. F., Dumas, M., et al. (2021). E2F6 Initiates Stable Epigenetic Silencing of Germline Genes during Embryonic Development. *Nat. Commun.* 12, 3582. doi:10.1038/s41467-021-23596-w
- Deaton, A. M., and Bird, A. (2011). CpG Islands and the Regulation of Transcription. *Genes Dev.* 25, 1010–1022. doi:10.1101/gad.2037511
- Deaton, A. M., Webb, S., Kerr, A. R. W., Illingworth, R. S., Guy, J., Andrews, R., et al. (2011). Cell Type-specific DNA Methylation at Intragenic CpG Islands in the Immune System. *Genome Res.* 21, 1074–1086. doi:10.1101/gr.118703.110
- De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B. K., et al. (2010). A Large Fraction of Extragenic RNA Pol II Transcription Sites Overlap Enhancers. *PLoS Biol.* 8, e1000384. doi:10.1371/journal.pbio.1000384
- Ehrlich, M. (2002). DNA Methylation in Cancer: Too Much, but Also Too Little. *Oncogene* 21, 5400–5413. doi:10.1038/sj.onc.1205651
- Elango, N., and Yi, S. V. (2011). Functional Relevance of CpG Island Length for Regulation of Gene Expression. *Genetics* 187, 1077–1083. doi:10.1534/genetics.110.126094
- Farcas, A. M., Blackledge, N. P., Sudbery, I., Long, H. K., McGouran, J. F., Rose, N. R., et al. (2012). KDM2B Links the Polycomb Repressive Complex 1 (PRC1) to Recognition of CpG Islands. *eLife* 1, e00205. doi:10.7554/eLife.00205
- Galupa, R., and Heard, E. (2018). X-chromosome Inactivation: A Crossroads between Chromosome Architecture and Gene Regulation. *Annu. Rev. Genet.* 52, 535–566. doi:10.1146/annurev-genet-120116-024611
- Gardiner-Garden, M., and Frommer, M. (1987). CpG Islands in Vertebrate Genomes. *J. Mol. Biol.* 196, 261–282. doi:10.1016/0022-2836(87)90689-9
- Golding, M. C., Magri, L. S., Zhang, L., Lalone, S. A., Higgins, M. J., and Mann, M. R. W. (2011). Depletion of Kcnq1ot1 Non-coding RNA Does Not Affect Imprinting Maintenance in Stem Cells. *Development* 138, 3667–3678. doi:10.1242/dev.057778
- Greenberg, M. V. C., and Bourc'his, D. (2019). The Diverse Roles of DNA Methylation in Mammalian Development and Disease. *Nat. Rev. Mol. Cell Biol.* 20, 590–607. doi:10.1038/s41580-019-0159-6
- Gruber, A. J., and Zavolan, M. (2019). Alternative Cleavage and Polyadenylation in Health and Disease. *Nat. Rev. Genet.* 20, 599–614. doi:10.1038/s41576-019-0145-z
- Hartl, D., Krebs, A. R., Grand, R. S., Baubec, T., Isbel, L., Wirbelauer, C., et al. (2019). CG Dinucleotides Enhance Promoter Activity Independent of DNA Methylation. *Genome Res.* 29, 554–563. doi:10.1101/gr.241653.118
- Hon, G. C., Hawkins, R. D., Caballero, O. L., Lo, C., Lister, R., Pelizzola, M., et al. (2012). Global DNA Hypomethylation Coupled to Repressive Chromatin Domain Formation and Gene Silencing in Breast Cancer. *Genome Res.* 22, 246–258. doi:10.1101/gr.125872.111
- Illingworth, R., Kerr, A., DeSousa, D., Jørgensen, H., Ellis, P., Stalker, J., et al. (2008). A Novel CpG Island Set Identifies Tissue-specific Methylation at Developmental Gene Loci. *PLoS Biol.* 6, e22. doi:10.1371/journal.pbio.0060022
- Illingworth, R. S., Gruenewald-Schneider, U., Webb, S., Kerr, A. R. W., James, K. D., Turner, D. J., et al. (2010). Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome. *PLoS Genet.* 6, e1001134. doi:10.1371/journal.pgen.1001134
- Jeziorska, D. M., Murray, R. J. S., De Gobbi, M., Gaentzsch, R., Garrick, D., Ayyub, H., et al. (2017). DNA Methylation of Intragenic CpG Islands Depends on Their Transcriptional Activity during Differentiation and Disease. *Proc. Natl. Acad. Sci. USA* 114, E7526–E7535. doi:10.1073/pnas.1703087114
- Kaikkonen, M. U., Spann, N. J., Heinz, S., Romanoski, C. E., Allison, K. A., Stender, J. D., et al. (2013). Remodeling of the Enhancer Landscape during Macrophage Activation Is Coupled to Enhancer Transcription. *Mol. Cell* 51, 310–325. doi:10.1016/j.molcel.2013.07.010
- Kanber, D., Berulava, T., Ammerpohl, O., Mitter, D., Richter, J., Siebert, R., et al. (2009). The Human Retinoblastoma Gene Is Imprinted. *PLoS Genet.* 5, e1000790. doi:10.1371/journal.pgen.1000790
- Kim, T.-K., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., et al. (2010). Widespread Transcription at Neuronal Activity-Regulated Enhancers. *Nature* 465, 182–187. doi:10.1038/nature09033
- Krebs, A. R., Dessus-Babus, S., Burger, L., and Schübeler, D. (2014). High-throughput Engineering of a Mammalian Genome Reveals Building Principles of Methylation States at CG Rich Regions. *eLife* 3, e04094. doi:10.7554/eLife.04094
- Kulis, M., Heath, S., Bibikova, M., Queirós, A. C., Navarro, A., Clot, G., et al. (2012). Epigenomic Analysis Detects Widespread Gene-Body DNA Hypomethylation in Chronic Lymphocytic Leukemia. *Nat. Genet.* 44, 1236–1242. doi:10.1038/ng.2443
- Kulis, M., Queirós, A. C., Beekman, R., and Martín-Subero, J. I. (2013). Intragenic DNA Methylation in Transcriptional Regulation, normal Differentiation and Cancer. *Biochim. Biophys. Acta Gene Regul. Mech.* 1829, 1161–1174. doi:10.1016/j.bbaggm.2013.08.001
- Kumar, D., and Jothi, R. (2020). Bivalent Chromatin Protects Reversibly Repressed Genes from Irreversible Silencing. *bioRxiv*. doi:10.1101/2020.12.02.406751
- Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. (1992). CpG Islands as Gene Markers in the Human Genome. *Genomics* 13, 1095–1107. doi:10.1016/0888-7543(92)90024-M
- Latos, P. A., Pauler, F. M., Koerner, M. V., Şenergin, H. B., Hudson, Q. J., Stocsits, R. R., et al. (2012). Airn Transcriptional Overlap, but Not its lncRNA Products, Induces Imprinted Igf2r Silencing. *Science* 338, 1469–1472. doi:10.1126/science.1228110
- Lee, Y., and Rio, D. C. (2015). Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu. Rev. Biochem.* 84, 291–323. doi:10.1146/annurev-biochem-060614-034316
- Lee, S.-M., Lee, J., Noh, K.-M., Choi, W.-Y., Jeon, S., Oh, G. T., et al. (2017). Intragenic CpG Islands Play Important Roles in Bivalent Chromatin Assembly of Developmental Genes. *Proc. Natl. Acad. Sci. USA* 114, E1885–E1894. doi:10.1073/pnas.1613300114
- Li, W., Notani, D., and Rosenfeld, M. G. (2016). Enhancers as Non-coding RNA Transcription Units: Recent Insights and Future Perspectives. *Nat. Rev. Genet.* 17, 207–223. doi:10.1038/nrg.2016.4
- Logsdon, G. A., Vollger, M. R., and Eichler, E. E. (2020). Long-read Human Genome Sequencing and its Applications. *Nat. Rev. Genet.* 21, 597–614. doi:10.1038/s41576-020-0236-x
- Maunakea, A. K., Nagarajan, R. P., Bilenky, M., Ballinger, T. J., D'Souza, C., Fouse, S. D., et al. (2010). Conserved Role of Intragenic DNA Methylation in Regulating Alternative Promoters. *Nature* 466, 253–257. doi:10.1038/nature09165
- Maupetit-Méhouas, S., Montibus, B., Nury, D., Tayama, C., Wassef, M., Kota, S. K., et al. (2016). Imprinting Control Regions (ICRs) Are Marked by Mono-Allelic Bivalent Chromatin when Transcriptionally Inactive. *Nucleic Acids Res.* 44, 621–635. doi:10.1093/nar/gkv960
- Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., et al. (2007). Genome-wide Maps of Chromatin State in Pluripotent and Lineage-Committed Cells. *Nature* 448, 553–560. doi:10.1038/nature06008
- Mochizuki, K., Sharif, J., Shirane, K., Uranishi, K., Bogutz, A. B., Janssen, S. M., et al. (2021). Repression of Germline Genes by PRC1.6 and SETDB1 in the Early

- Embryo Precedes DNA Methylation-Mediated Silencing. *Nat. Commun.* 12, 7020. doi:10.1038/s41467-021-27345-x
- Montibus, B., Cercy, J., Bouschet, T., Charras, A., Maupetit-Méhouas, S., Nury, D., et al. (2021). TET3 Controls the Expression of the H3K27me3 Demethylase Kdm6b during Neural Commitment. *Cell. Mol. Life Sci.* 78, 757–768. doi:10.1007/s00018-020-03541-8
- Nanavaty, V., Abrash, E. W., Hong, C., Park, S., Fink, E. E., Li, Z., et al. (2020). DNA Methylation Regulates Alternative Polyadenylation via CTCF and the Cohesin Complex. *Mol. Cell* 78, 752–764. doi:10.1016/j.molcel.2020.03.024
- Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., et al. (2017). Intragenic DNA Methylation Prevents Spurious Transcription Initiation. *Nature* 543, 72–77. doi:10.1038/nature21373
- Onodera, C. S., Underwood, J. G., Katzman, S., Jacobs, F., Greenberg, D., Salama, S. R., et al. (2012). Gene Isoform Specificity through Enhancer-Associated Antisense Transcription. *PLOS ONE* 7, e43511. doi:10.1371/journal.pone.0043511
- Pachano, T., Sánchez-Gaya, V., Ealo, T., Mariner-Fauli, M., Bleckwehl, T., Asenjo, H. G., et al. (2021). Orphan CpG Islands Amplify Poised Enhancer Regulatory Activity and Determine Target Gene Responsiveness. *Nat. Genet.* 53, 1036–1049. doi:10.1038/s41588-021-00888-x
- Proudfoot, N. J. (2011). Ending the Message: Poly(A) Signals Then and Now. *Genes Dev.* 25, 1770–1782. doi:10.1101/gad.172684.11
- Santoro, F., Mayer, D., Klement, R. M., Warczok, K. E., Stukalov, A., Barlow, D. P., et al. (2013). Imprinted Igf2r Silencing Depends on Continuous Airn lncRNA Expression and Is Not Restricted to a Developmental Window. *Development* 140, 1184–1195. doi:10.1242/dev.088849
- Saxonov, S., Berg, P., and Brutlag, D. L. (2006). A Genome-wide Analysis of CpG Dinucleotides in the Human Genome Distinguishes Two Distinct Classes of Promoters. *Proc. Natl. Acad. Sci.* 103, 1412–1417. doi:10.1073/pnas.0510310103
- Shah, R. N., Grzybowski, A. T., Elias, J., Chen, Z., Hattori, T., Lechner, C. C., et al. (2021). Re-evaluating the Role of Nucleosomal Bivalency in Early Development. *bioRxiv*. doi:10.1101/2021.09.09.458948
- Shearwin, K., Callen, B., and Egan, J. (2005). Transcriptional Interference - a Crash Course. *Trends Genet.* 21, 339–345. doi:10.1016/j.tig.2005.04.009
- Statham, A. L., Robinson, M. D., Song, J. Z., Coolen, M. W., Stirzaker, C., and Clark, S. J. (2012). Bisulfite Sequencing of Chromatin Immunoprecipitated DNA (BisChIP-Seq) Directly Informs Methylation Status of Histone-Modified DNA. *Genome Res.* 22, 1120–1127. doi:10.1101/gr.132076.111
- Steinhaus, R., Gonzalez, T., Seelow, D., and Robinson, P. N. (2020). Pervasive and CpG-dependent Promoter-like Characteristics of Transcribed Enhancers. *Nucleic Acids Res.* 48, 5306–5317. doi:10.1093/nar/gkaa223
- Takai, D., and Jones, P. A. (2002). Comprehensive Analysis of CpG Islands in Human Chromosomes 21 and 22. *Proc. Natl. Acad. Sci. U S A* 99, 3740–3745. doi:10.1073/pnas.052410099
- Tian, B., and Manley, J. L. (2016). Alternative Polyadenylation of mRNA Precursors. *Nat. Rev. Mol. Cell Biol.* 18, 18–30. doi:10.1038/nrm.2016.116
- Tufarelli, C., Stanley, J. A. S., Garrick, D., Sharpe, J. A., Ayyub, H., Wood, W. G., et al. (2003). Transcription of Antisense RNA Leading to Gene Silencing and Methylation as a Novel Cause of Human Genetic Disease. *Nat. Genet.* 34, 157–165. doi:10.1038/ng1157
- Velasco, G., Hube, F., Rollin, J., Neuillet, D., Philippe, C., Bouzinba-Segard, H., et al. (2010). Dnmt3b Recruitment through E2F6 Transcriptional Repressor Mediates Germ-Line Gene Silencing in Murine Somatic Tissues. *Proc. Natl. Acad. Sci.* 107, 9281–9286. doi:10.1073/pnas.1000473107
- Voigt, P., Tee, W.-W., and Reinberg, D. (2013). A Double Take on Bivalent Promoters. *Genes Dev.* 27, 1318–1338. doi:10.1101/gad.219626.113
- Wachter, E., Quante, T., Merusi, C., Arczewska, A., Stewart, F., Webb, S., et al. (2014). Synthetic CpG Islands Reveal DNA Sequence Determinants of Chromatin Structure. *eLife* 3, e03397. doi:10.7554/eLife.03397
- Williamson, C. M., Ball, S. T., Dawson, C., Mehta, S., Beechey, C. V., Fray, M., et al. (2011). Uncoupling Antisense-Mediated Silencing and DNA Methylation in the Imprinted Gnas Cluster. *PLoS Genet.* 7, e1001347. doi:10.1371/journal.pgen.1001347
- Wood, A. J., Schulz, R., Woodfine, K., Koltowska, K., Beechey, C. V., Peters, J., et al. (2008). Regulation of Alternative Polyadenylation by Genomic Imprinting. *Genes Dev.* 22, 1141–1146. doi:10.1101/gad.473408
- Zhao, Y., and Garcia, B. A. (2015). Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harb. Perspect. Biol.* 7, a025064. doi:10.1101/cshperspect.a025064

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Cain, Montibus and Oakey. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.