CellPress

Review

Extraordinary Cancer Epigenomics: Thinking Outside the Classical Coding and Promoter Box

Matthew Murtha[1] and Manel Esteller[1,2,3,*]

The advent of functional genomics powered by high-throughput sequencing has given us a new appreciation of the genomic sequences that lie outside the canonical promoter–coding sequence box. These regions harbor distant regulatory elements, enhancers, super-enhancers, insulators, alternative promoters, and sequences that transcribe as noncoding RNAs (ncRNAs) such as miRNAs and long ncRNAs. These functional genomics studies have also enabled a clearer understanding of the role of the 3D structure of the genome in epigenetic regulation. Here we review the impact that epigenetic changes, and specifically DNA methylation, have on these extraordinary sequences in driving cancer progression.

Functional Genomics: Beyond the Gene

The human genome encodes all of the intrinsic components of human life, as well as containing instructions for the appropriate timing of expression and execution. Dysregulation of these processes results in pathologies such as cancer, a leading cause of death in modern societies. In part to rise to the clinical challenges created by cancer, the sequence of the human genome has been under intense investigation for some time. The advent of massively parallel high-throughput (HTP) sequencing has led to considerable progress towards unlocking the underlying codes embedded in genomic sequences. There are two areas of functional genomics currently under intense study: (i) functional analysis of noncoding regions of the genome; and (ii) characterization of epigenetic modifications that govern genomic regulation. This review highlights studies at the intersection of these trends and its relevance to in cancer.

Genomics has revealed that much of the genome comprises noncoding DNA that does not result in the production of functional proteins. Although it was first considered 'junk' DNA [1], functional genomics now shows that much of this noncoding DNA is important, conserved, and fulfills biological functions. Consortia such as the Encyclopedia of DNA Elements (ENCODE)[i] have found that the vast majority of the genome (~80%) is involved in biochemical events such as associations with binding proteins or RNAs [2]. Genome-wide association studies (GWASs) use large cohorts of patient genomic samples to assign risk of developing a particular disease to genetic variants. These variants are typically at SNPs and are correlated with disease traits by linkage disequilibrium; that is, the variant is observed in patients with the disease more often than would be expected if the variant were randomly associated with the disease [3,4]. Integration of these studies with the functional annotations in ENCODE shows that 88% of SNPs associated with a risk of cancer are at positions that lie outside the coding regions of genes or their promoters, suggesting severe consequences of altered functions of the noncoding genome [2].

Trends

There is a new appreciation for the noncoding regions of the human genome as their functions in gene regulation are delineated through the use of functional genomics.

The cataloging of epigenetic marks and their impact on cellular function is proceeding rapidly.

The intersection of epigenetics and functional genomics reveals the myriad ways by which the expression of the human genome is regulated.

Understanding how epigenetic marks on the noncoding genome can promote tumor progression can lead to new therapeutic strategies.

[1]Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), Avinguda Gran Via de L'Hospitalet 199–203, L'Hospitalet del Llobregat, Barcelona, Catalonia, Spain
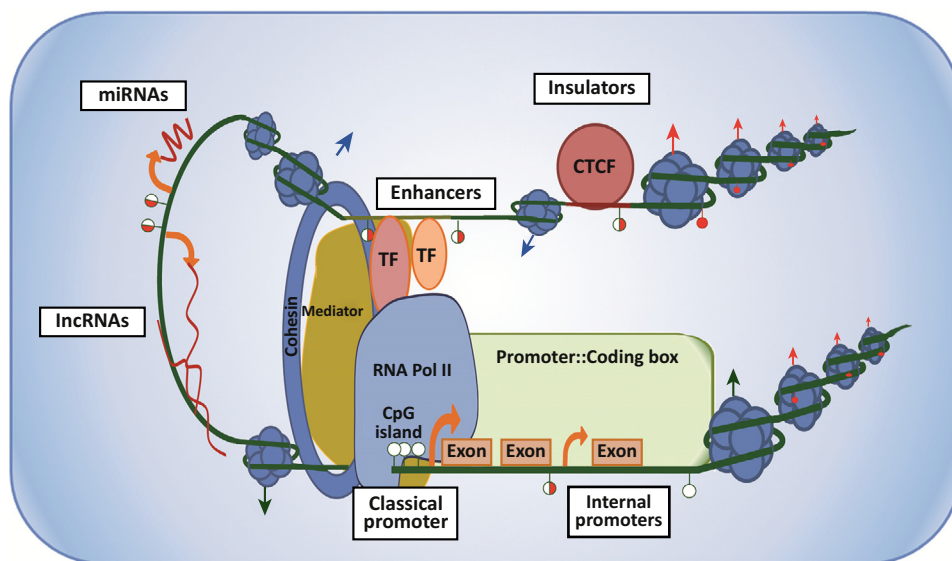[2]Department of Physiological Sciences II, School of Medicine, University of Barcelona, Barcelona, Catalonia, Spain
[3]Institucio Catalana de Recerca i Estudis Avançats, Barcelona, Catalonia, Spain

*Correspondence: mesteller@idibell.cat (M. Esteller).

Key Figure

Epigenetic Regulation of Non-Protein-Coding Genome Elements

Trends in Cancer

Figure 1. Epigenetic changes in cancer affect many non-protein-coding regions of the genome. Depicted are some of these elements outside the 'promoter–coding' box: miRNAs, long noncoding RNAs (lncRNAs), enhancers, insulators, and internal promoters. Enhancers are brought into contact with promoter elements by the actions of mediator and cohesin, which are recruited by transcription factors (TFs). CTCF, CCCTC-binding factor; red circles, DNA methylation at cytosine residues; white circles, unmethylated DNA; half-red circles, differentially methylated cytosines that contribute to cancer progression; blue triangles, histone modification H3K27Ac (enhancer mark); green triangles, H3K4me3 (active chromatin); red triangles, H3K27me3 (marks repressive chromatin).

Epigenetics refers to inheritable changes in gene expression seen in the absence of changes to the DNA sequence. Epigenetic regulation of gene expression often proceeds through modifications to the lysine tails of histones and methylation of cytosine bases [5–7]. HTP sequencing has provided a boom in studies delineating the impact of epigenetic modifications on the regulation of the genome and disease risk. These studies are leading to a new appreciation of the myriad ways in which the genome is regulated. Here we review how the epigenome interacts with the less well-studied elements of the genome – the noncoding regions – to drive cancer progression (Figure 1, Key Figure). While histone modifications are a major component of epigenetic regulation [5–7], we focus on the role of DNA methylation of the noncoding genome.

Enhancers

Enhancers are an important class of noncoding genomic elements, dysfunction of which clearly contributes to cancer. Analysis of ENCODE and FANTOM5 annotations reveals a stronger association between disease-associated SNPs and enhancer sequences than coding sequences [8][ii]. Novel strategies to identify active enhancers are detailed in Box 1.

Since risk loci identified by GWASs consistently map to noncoding regions, variants in enhancers are increasingly recognized as responsible for variation in complex traits. GWASs can be used to correlate disease-associated SNPs with changes in gene expression, yielding expression quantitative trait loci (eQTLs) [9–12]. These analyses have identified many meaningful

CellPress

Box 1. Methods to Identify Enhancers

Identifying the genomic location of enhancers is an important goal for functional genomics studies. Chromatin immunoprecipitation sequencing (ChIP-seq) studies of transcription factor binding and histone modifications are typically used *in lieu* of functional data to infer the activity of *cis*-regulatory modules. The drawback to using ChIP-seq data is that binding does not perfectly correlate with function; a recent analysis of Encyclopedia of DNA Elements (ENCODE)-predicted enhancers found only ~30% with genuine enhancer/CRM activity [131]. A further drawback to these approaches is that they identify large segments of the genome (typically greater than 1 kb) that are not amenable to motif analysis.

Several approaches have been developed to overcome these obstacles. *Cis*-regulatory elements by sequencing takes advantage of barcoded transcripts to measure the activity of putative enhancers via modified RNA-seq [132,133]. Functional identification of regulatory elements within accessible chromatin (FIREWACh) and site-specific integration fluorescence-activated cell sorting followed by sequencing (SIF-seq) are two functional approaches utilizing detection of fluorescent reporters that were developed simultaneously [134,135]. Both approaches take advantage of transgenic approaches, FIREWACh via lentivirus insertion into the genome and SIF-seq via homologous recombination of the *HPRT* gene in the mouse genome. Last, there is the self-transcribed active regulatory-region sequencing approach, which takes advantage of enhancers' ability to direct their own transcription [136].

---

relationships between SNPs, gene expression, and disease risk that have advanced our understanding of a wide variety of diseases [12–15]. Extending this approach to epigenetic modifications such as DNA methylation can help clarify how epigenetics and the noncoding genome interact in cancer. Sites of differential methylation that correlate with gene expression changes are called methylation eQTLs (meQTLs) and represent a promising approach to the study of cancer epigenetics and its impact on gene expression [16–18].

Aiming to determine whether enhancer methylation links cancer susceptibility loci to cancer-driving mechanisms, Aran and Hellman analyzed 390 estrogen receptor-positive breast tumors from The Cancer Genome Atlas (TCGA)[iii] and found that methylation sites and enhancers characterized the level of intertumor expression variation better than promoter methylation [19]. Thus, while enhancer sites contain both sequence and methylation polymorphisms that define enhancer activity, the expression status of the target gene is better explained by methylation status than by polymorphic sequence alone.

Analysis of 3649 primary human tumors created a meQTL catalog of DNA methylation associations for 21% of interrogated cancer risk polymorphisms [18]. This study linked risk alleles to genes previously characterized as having known roles in cancer in addition to as yet unidentified cancer genes. The association between breast cancer risk allele rs2380205 and the *FBXO18* gene is of particular note. FBXO18 is an F-box protein helicase that actively participates in the formation of double-strand breaks and activation of tumor protein-dependent apoptosis following DNA replication stress [20]. FBXO18-deficient cells therefore have a propensity for increased cell survival, since they exhibit impaired inactivation of the cytotoxic stress-induced cascade [21,22]. These observations suggest a causal link between FBXO18 and breast cancer, making this an interesting target for future research.

Genome-wide binding studies of mediator and cohesin components revealed that these factors concentrate at a subset of regulatory elements that group genomically to form what are termed 'super-enhancers' [23]. Super-enhancers are associated with key developmental regulatory genes and their functions are modified in cancer [24–26]. For example, a somatic heterozygous insertion creates a binding site for the transcription factor (TF) MYB, which leads to the formation of a super-enhancer upstream of the TAL1 oncogene that contributes to the pathology of T cell acute lymphoblastic leukemia (T-ALL) [27]. Induction of this super-enhancer is evident from the accumulation of H327KAc histone marks and binding of major leukemogenic transcription components RUNX1, GATA-3, and TAL1 itself. This was one of the first studies to provide a genetic mechanism for oncogenic super-enhancers in malignant cells.

CellPress

Another example comes from neuroblastoma, where the rs2168101 G>T SNP is located within a super-enhancer that spans the first intron of the gene LMO1 [28]. The G allele creates a binding site for GATA3 that drives overexpression of LMO1 compared with the T allele, which lacks GATA3 binding. This can create a predisposition to development of neuroblastoma through modulation of super-enhancer activity [28].

The relationship between DNA methylation and super-enhancers can also be measured using functional genomics data [29,30]. For example, Heyn and colleagues analyzed the methylation status of over 5000 super-enhancers in 78 normal tissue samples, 714 primary tumors, and 24 metastatic samples. They found that tumors undergo a shift in super-enhancer DNA methylation profile compared with healthy controls [30]. Loss of methylation at super-enhancers consistently coincided with increased gene expression, while gain of methylation correlated with lower levels of expression.

Methylation can affect enhancer activity in several ways, including potential structural roles that regulate nucleosome occupancy and modifications. One study found that DNA methylation of super-enhancers was necessary to maintain appropriate levels of H3K27Ac histone modification [31]. Comparing wild-type colon cells and cancerous cell lines deficient in DNA methylation, Lay et al. found that a small but significant proportion of enhancers were affected by the loss of DNA methylation through changes in nucleosome positioning and histone modifications [31].

The methylation status of enhancers is, then, of the utmost importance. In addition to new methods to detect enhancer sequences (described in Box 1) there are also new methods to determine their methylation status. Besides whole-genome bisulfite sequencing, there are now microarray platforms designed specifically to include enhancer sequences, which will be powerful tools in future research [32].

Internal Promoters

Epigenetic changes can also redirect the transcriptional network. By comparing methylation levels between primary tumors and metastases, Visoso et al. found a hypomethylation event that reactivates a cryptic transcript of TBC1D16, a Rab GTPase-activating protein [33]. The novel short isoform of TBC1D16 exacerbates melanoma growth and metastasis by targeting RAB5C and regulating EGFR and confers poor clinical outcome while showing greater sensitivity to BRAF and MEK inhibitors than cells lacking the short transcript. This is an example of the contribution of epigenetics to metastasis and to the prediction of drug response by unearthing cryptic transcription start sites (TSSs).

Similar studies suggest that epigenetic modifications impart a switch in promoter usage of DCKL1, a gene that specifically marks colon/pancreatic cancers in mice and is expressed by human adenocarcinomas (hCRCs) [34]. Downregulation of DCKL1 results in loss of cancer stem cells (CSCs) and inhibits the growth of xenograft tumors; however, DCKL1 expression is observed in hCRCs. This expression is the result of a majority of hCRCs that use an alternative promoter in intron V to express short transcripts of DCLK1. This is in contrast to normal cells, which use the canonical alpha promoter, which becomes methylated and repressed in cancer. Analysis of a cohort of 92 hCRC patients revealed that high expression of the shorter alternative promoter-initiated transcripts correlates with worse survival rates, demonstrating the clinical consequences of epigenetically regulated promoter usage.

A similar dynamic of epigenetic modifications regulates the isoform usage of the oncogene p53 in response to DNA damage in a cell-line model of carcinogenesis [35]. However, the prevalence of this effect in the cancer patient population at large remains unclear. The relationship between

---

Box 2. Methods to Determine 3D Genomic Interactions

State-of-the-art biochemical techniques for delineating 3D genomic interactions now all take advantage of high-throughput (HTP) sequencing.

DamID is an approach used to catalog genomic sequences that interact with lamina-associated domains of the inner face of the nuclear membrane [137–139]. In DamID, bacterial adenine methyltransferase is fused to a protein of interest and allowed to interact with physically proximal DNA. Sequences containing methylated products are enriched by Dam-specific restriction enzymes, then sequenced and mapped back to the reference genome thus demarcating genomic regions that interact with the protein of interest.

Another group of widely used techniques revolve around chromosome conformation capture (3C) methods coupled to HTP sequencing [140]. In principle, these methods quantify the relative spatial proximity between individual genomic loci through the digestion and re-ligation of crosslinked chromatin. These techniques are broadly divided into those that interrogate the whole genome (e.g., Hi-C and its derivatives DNase Hi-C and single-cell Hi-C) [46,49,141–144] and those that analyze interactions of targeted loci (e.g., 3C, circularized chromatin conformation capture, chromatin interaction analysis by paired-end tag sequencing) [145–148].

---

promoter methylation and usage of cryptic internal promoters to generate oncogenic isoforms requires further investigation.

The Genome 3D Structure

The human genome is packaged into the nucleus in a 3D structure with distinct properties and characteristics for each cell type. It is now apparent that this 3D organization impacts gene regulation [36,37]. Methods for determining the 3D structure of the genome are reviewed in Box 2. One of the most important consequences of this 3D organization is the direct interaction of enhancers with their target promoters through the formation of chromatin loops, which allows them to operate at a great genomic distance from the TSS.

The factors that allow DNA looping and interactions between promoters and cognate enhancers are rapidly being elucidated. For example, Kagey *et al.* found that genes encoding mediator and cohesin complexes have a severe impact on gene regulation [38]. Their studies show that mediator, a transcriptional coactivator, forms a protein complex with cohesin to loop the DNA between enhancers and promoters. Further, Nipbl, a cohesin-loading factor, is associated with these complexes, providing a mechanism by which the cohesin can be loaded onto DNA to form and maintain chromatin loops. Both the mediator and cohesin genes have known roles in cancer [39–43].

Since 2009, functional genomics and bioinformatics approaches have been used to generate whole-genome contact probability maps in human and other organisms. These studies confirmed the presence of chromosome territories (CTs) – the preference of interphase chromosomes to occupy distinct spatial territories and self-associate rather than mix homogenously in the nucleus [44–46]. They have also demonstrated the existence of 'topology-associated domains' (TADs) – large, megabase-sized local chromatin-interaction domains where regions within the domain self-interact rather than interacting with loci of other domains [45–49]. TADs are defined and constrained by the binding of CCCTC-binding factor (CTCF), with inverted binding sites marking either side of the domain [49–51]. CTCF acts as an 'insulator' inhibiting the spread of histone modifications and disrupting aberrant communication between enhancers and promoters [52,53].

Insulators

Identifying dysregulation of TADs by epigenetic mechanisms is an important field for future investigation [54]. A remarkable example was recently published by Flavahan *et al.* in a study that investigated the impact of changes in the methylation status of CTCF sites in the methylator

CellPress

subtype of glioblastoma [55]. Gliomas with a putative CpG island methylator phenotype (G-CIMP) are linked to isocitrate dehydrogenase (IDH) gain-of-function mutations [56–59]. Mutant IDHs produce 2-hydroxyglutarate, an oncometabolite that interferes with the function of iron-dependent hydroxylases [60]. IDH mutants have severely altered methylomes, which is likely to be due to 2-hydroxyglutarate-mediated inhibition of TET family 5′-methylcytosine hydroxylases, which catalyze a key step in the removal of DNA methylation [57,61,62]. Analysis of the methylomes of IDH-mutant tumor cells showed hypermethylated CTCF-binding sites, which compromise the binding of CTCF [55]. As a result, TAD integrity is lost, which results in aberrant gene activation. Specifically, this study showed that loss of CTCF binding allowed a constitutively active enhancer that i normally sequestered in a separate TAD to interact and activate *PDGFRA*, a prominent glioma oncogene, thus providing a potential mechanism by which disruption of the methylome drives the progression of G-CIMP gliomas.

This example reveals how epigenetic changes such as DNA methylation at CTCF sites can disrupt the 3D architecture of the genome and lead to aberrant activation of oncogenes and cancer progression.

Noncoding RNAs

Noncoding RNAs (ncRNAs) are a major class of biomolecules with low protein-coding potential that are involved in a diverse array of cellular processes. These RNAs are typically named based on their size [e.g., microRNAs, long ncRNAs (lncRNAs)], function [ribosomal RNAs, small nucleolar RNAs (snoRNAs), PIWI-interacting RNAs (piRNAs)], or genomic origin (transcribed ultraconserved regions). Their expression is regulated through several mechanisms including epigenetic modifications. It is becoming increasingly evident their functions are disrupted in cancer and they may provide therapeutic targets.

Head-to-Head Antisense Transcripts

An example of a ncRNA altering expression of a cancer-related gene is the head-to-head antisense transcript of the VIM gene promoting VIM expression. VIM is a member of the intermediate filament family and is commonly deregulated in cancer [63]. Its antisense transcript forms an R-loop structure that, when disrupted by H1 ribonuclease or antisense knockdown, results in VIM downregulation [63]. This is the first example of a head-to-head antisense transcript mediating enhancement of transcription through an R-loop mechanism. ncRNAs are subject to various regulatory mechanisms, including inactivation associated with nearby CpG island methylation. For example, Ferreira *et al.* found cancer-specific hypermethylation of the snoRNAs SNORD123, U70C, and ACA59B in a colorectal cancer cell line [64]. snoRNAs have a wide variety of cellular functions such as chemical modification of RNA, pre-RNA processing, and control of alternative splicing. Analysis of a comprehensive collection of normal tissues, cancer cell lines, and primary tumors demonstrated that hypermethylation of snoRNAs is a common feature in many tumors, particularly leukemia [64], suggesting an important role for their epigenetic inactivation in cancer. A similar analysis found that piRNAs are also down-regulated in cancer [65].

miRNAs

miRNAs (small RNAs of 18–25 nucleotides) repress expression of proteins by binding seed sequences in mRNAs, inhibiting translation and reducing mRNA stability. Dysfunctional miRNAs are implicated in various cancers [66–68] and in metastasis [69]. An integrated analysis of TCGA data revealed a master miRNA regulatory network that governs the ovarian cancer mesenchymal subtype, which has poor survival [70]. As with protein-coding genes, the expression of miRNAs is subject to numerous regulatory mechanisms, including epigenetic control. There are several noteworthy cases where aberrant epigenetic signaling disrupts miRNA function leading to cancer.

CellPress

One of the first such was the discovery that miR-124 contributes to pathogenesis through its role in diminishing levels of CDK6 [71]. Analysis of miRNAs in the colorectal cancer cell line HCT-116 found that mIR-24 acts to represses CDK6, which affects the phosphorylation status of Rb protein, a downstream effector of CDK6. Further analysis found that miR-124 is epigenetically silenced in at least ten tumor types, such as breast, gastric, liver, and hematopoietic malignancies [72].

One major group of miRNAs dysregulated in cancer is the miR-200 family. miR-200 RNAs were first described as involved in the induction of epithelial-to-mesenchymal transition (EMT) via downregulation of E-cadherin and increased ZEB proteins [73–77]. Specifically, the CpG islands of miR-200ba429 and miR-200c141 are unmethylated in cell lines with epithelial features and the epithelia of normal colon mucosa crypts. Conversely, they are hypermethylated in cells with mesenchymal features. This epigenetic state can be changed by forcing EMT or mesenchymal-to-epithelial transition, demonstrating that miR-200 epigenetic silencing is not a static event but rather a dynamic process that reflects the epithelial or mesenchymal phenotype [73]. Later, it was shown that loss of miR-200 repression of ETS1 under hypoxic conditions results in an angiogenic response in cancer cells [78]. The mechanisms by which the miR-200 family is regulated include not only DNA methylation [79] but also DNA looping and read-through of the upstream gene *PTPN6* [80].

Another well-studied example is the Let-7 family of miRNAs, which target HRAS and HMGA2 along with having roles in the regulation of proliferation and as cell cycle regulators. Let-7 was originally identified in *Caenorhabditis elegans* in mutagenesis screens for factors governing developmental timing [81]. Let-7 members are downregulated epigenetically in several cancers and their overexpression results in inhibited growth and transformation of cancer cell lines and tumor xenografts [82,83]. This is due to the fact that Let-7 targets key components of cell cycle progression such as k-Ras, cyclin D1, Cdc34, Hmga2, E2f2, and Lin28 [83,84].

Other examples of miRNA silenced by epigenetic modifications are hsa-miR-9-1 and hsa-miR-9-3, which are hypermethylated in clear cell renal cell carcinoma (CCRCC). Comparisons between metastatic and nonmetastatic tumors revealed that methylation of hsa-miR-9-3 is significantly associated with an increased risk of recurrence and a significant decrease in recurrence-free survival time [84]. Aberrant expression of miR-9 has also been observed in acute myeloid leukemia (AML) [85].

miR-34a resides near a CpG island within the gene *EF570049* and contains a p53-binding site [86]. In 2007, miR-34 was shown to be part of the p53 tumor-suppressor network [87]. It plays key roles in chronic lymphocytic leukemia (CLL), with aberrant epigenetic regulation and downregulation in CLL patients with p53 mutations [87–91].

These examples are summarized in Table 1.

Table 1. Selected Epigenetically Deregulated miRNAs in Cancer

miRNA family	Tumor	Expression in cancer	Target gene	Refs
Let-7	Lung, pancreatic, breast	Typically downregulated	*RAS*	[82–84]
miR-124	Various	Silenced	*CDK6*	[71]
miR-200	Various	Silenced	*ZEB1,2, EMT*	[73–75]
miR-34	Lung, melanoma, CLL	Silenced	*CDK6, p53*	[69,87–91]
miR-9	CCRCC, AML	Silenced	*FoxO1, FoxO3*	[85,86]

CellPress

lncRNAs

Another class of elements within the noncoding genome is the lncRNAs. lncRNA is a class of RNAs that are typically more than 200 bp in length, are polyadenylated, and are spliced. Like protein-coding genes, their promoters are characterized by trimethylation of histone H3 at lysine 4 and are transcribed by RNA polymerase II [92]. lncRNAs are involved in several normal biological processes, such as regulating gene expression, signal transduction, and the stability of mRNAs [93–95]. In cancer, lncRNAs can operate both as tumor suppressors and as oncogenes [96].

lncRNAs are fully involved in epigenetic regulation of gene expression through epigenetic self-regulation and through interactions with epigenetic modifiers such as the PRC2 and SWI/SNF complexes [97]. Up to 20% of all lncRNAs have been implicated in PRC2 binding, with their binding specificity still a matter of debate [98–101].

One of the best-studied lncRNAs with oncogenic potential is HOX transcript antisense RNA (HOTAIR). Overexpression of HOTAIR is correlated with many types of cancer, including colorectal, hepatocellular, and aggressive breast cancer [102–104]. Overexpressed HOTAIR associates with the PRC2 complex and redirects PRC2 binding genome wide to a pattern resembling PRC2 binding in embryonic fibroblasts. This results in increased cancer invasiveness and metastasis [105].

An interesting example of the intersection of lncRNAs, epigenetics, and cancer is at the *Igf2/H19* gene locus. One of the first gene loci to be identified as imprinted, Igf2 is expressed from the paternal and H19 from the maternal allele [106–108]. H19 is a 2.3-kb, capped, spliced, and polyadenylated lncRNA that is highly expressed during embryogenesis and then strongly downregulated after birth in all tissues except muscle. H19 is likely to function through two primary mechanisms: first, H19 produces miR-675, which promotes gastric, colorectal, and glioma cancers [109–112]; and second, the H19 lncRNA harbors multiple binding sites for the Let-7 family of miRNAs and acts as a 'molecular sponge' limiting active Let-7 miRNAs by competitively binding and sequestering them [113]. H19 knockdown results in the same phenotype as Let-7 overexpression. This mechanism of action is part of the competitive endogenous RNA (ceRNA) hypothesis, which proposes that transcripts with shared miRNA-binding sites compete for post-transcriptional control of mRNAs and has been suggested as a unifying function of lncRNAs [114]. While there is good evidence to support this ceRNA model of lncRNA modulation of miRNA behavior, there is some experimental evidence questioning its validity [115].

The roles of lncRNAs in gene regulation are constantly being expanded. For example, the lncRNA LED is suppressed in human tumors by DNA methylation. Strongly activated enhancers produce ncRNA transcripts termed enhancer RNAs (eRNAs). A study mapping p53-responsive eRNAs found that LED binds p53-bound enhancers and suppression of LED leads to lower activation of the enhancers and its target gene. p53 is a canonical tumor suppressor and the expression of a key target, the potent cell-cycle inhibitor and tumor suppressor CDKN1A, is sensitive to LED levels [116]. This work provides a novel regulatory interaction between enhancer function and lncRNAs and is a promising avenue for future studies.

Transcribed ultraconserved regions represent an interesting subclass of lncRNAs. They are a subset of genomic elements that are absolutely conserved between orthologous regions of the human, rat, and mouse genomes. These sequences can be located in both intra- and intergenic regions. In a screen for elements whose expression is aberrantly silenced in cancer, it was found that several T-UCRs, such as Uc.160+, Uc.283+A, and Uc.346+, are silenced in response to CpG island hypermethylation in cancer cells [117]. Further studies show that the Uc.283+A

transcript controls pri-miRNA processing. This regulation required complementarity between an ultraconserved sequence in Uc.283+A and the lower stem region of pri-miR-195. This complementarity prevents pri-miRNA cleavage by Drosha and suggests a model whereby lower-stem strand invasion by Uc.283+A impairs microprocessor recognition. This work reveals a regulatory network dependent on ncRNA–ncRNA interactions involving several ncRNA classes important in cancer.

Concluding Remarks

Advances in functional genomics and the biostatistics of large data sets allow us to draw correlations between genomic variants known to be associated with disease and DNA elements in the noncoding genome. Further, these advances help us identify the functions of these risk elements, illuminating the myriad ways by which expression of the human genome is regulated. These advances are furthered by our continuing investigation into the characterization of epigenetic modifications and their molecular mechanisms.

Many questions remain, however (see Outstanding Questions). While epigenetics and distally acting regulatory elements provide a link between risk alleles and disease-related genes, many steps in the mechanistic pathways remain to be determined, as does their relevance to the disease state. Cell identity is ultimately determined by the interactions of the repertoire of genes expressed. There are many challenges to understanding the full network regulating gene expression, as there are multiple layers of regulation (e.g., transcriptional, post-transcriptional, post-translational) and multiple feedback loops that can augment these regulatory layers. Each cell type (e.g., liver cells, neurons) therefore reflects its own transcriptional network equilibrium. Understanding cancer as a diseased cell state thus necessitates identifying how the oncogenic lesion disrupts normal gene expression and function. Deconvoluting the impact of cell state from current functional genomics data sets is of paramount importance to our understanding of how the various regulatory layers interconnect to define cell identity.

Epigenetic modification has long been understood as a key barrier to cell identity transitions [118] and is thus vital to understanding the transition from normal to cancerous tissue. The future of this avenue of research will be the enhanced resolution afforded by single-cell epigenomics [119,120]. Tumors are now understood to be heterogeneous both transcriptionally and epigenetically, with some cells operating as 'cancer stem cells' or 'tumor-propagating cells' [121]. Single-cell epigenomics will allow us to understand the epigenetic differences imparting stem-like properties to cells.

Another avenue requiring further investigation is retroelements – repetitive sequences that comprise a significant portion of the mammalian genome [122]. A large number of sequences are characterized by long terminal repeats (LTRs) derived from the remnants of ancestral infectious retroviruses that have invaded the germ line at some point during evolution [123]. These sequences are called 'endogenous retroviruses' (ERVs) and generally feature mutations or recombination events that prohibit them from generating infectious particles. ERVs have been linked to the development of cancer through several mechanisms [124]. Because they rarely make infectious particles, ERVs do not contribute to oncogenesis through insertional mutagenesis in the way that infectious exogenous retroviruses are known to do [125]. Rather, production of ERV env genes, which encode envelope proteins, can cause cell fusion with profound implications in cancer [126,127]. Further, LTR sequences from ERVs can drive aberrant expression of oncogenes [128]. Given that ERV genes are silenced through epigenetic mechanisms, it is likely that there are oncogenic alterations in their epigenetic regulation that can lead to cancer, and these represent an enticing avenue for future research.

Outstanding Questions

How much of the noncoding genome is subject to active epigenetic regulation?

What epigenetic marks are most descriptive of the underlying cancer-regulatory network? Can these marks be utilized in personalized medicine?

What are the roles of retroelements in normal tissue and how do they contribute to oncogenesis?

What is the order of events among epigenetic changes, gene expression changes, and cancer phenotype?

Delineating causation among epigenetic changes, gene expression changes, and the cancer phenotype is also critical for future targeted therapies. For example, consider the case of IDH-mutant G-CIMP gliomas, which feature a hypermethylated genome and activated PDGF due to methylation of CTCF-binding sites. PDGF activation confers a growth advantage on the IDH-mutant cells in gliomasphere assays, yet patients harboring IDH mutations tend to survive longer than those with non-mutant IDH [55,57]. Does this represent a protective epigenome that can compensate for the activation of *PDGF*? Or is PDGF activation a weaker driver of tumorigenesis and lethality than other, non-IDH1 glioma oncogenes? Delineating the relative impact of an aberrant epigenome versus the expression of known oncogenes will clarify the pathways that lead to cancer.

Last, translating these findings to the clinic will require an investment in personalized medicine and epigenomic profiling of patient tumor samples. Epigenetic marks at non-promoter/gene elements can shed light on the cell type that the tumor originated from [129,130]. As epigenetic profiling becomes more accessible it is increasingly likely that we will be able to leverage single-cell profiling to accurately determine both the cell type of origin and the precise oncogenic lesion, allowing highly specific and effective therapies to be administered. Thinking outside the classical coding and promoter box by studying the epigenome of normal tissue and cancer cells will lead to better therapies for the public health crisis that is cancer.

Resources

i www.genome.gov/10005107/encode-project/

ii fantom.gsc.riken.jp/5/

iii tcga-data.nci.nih.gov

References

1. Ling, H. *et al.* (2015) Junk DNA and the long non-coding RNA twist in cancer genetics. *Oncogene* 34, 5003–5011

2. ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74

3. Easton, D.F. *et al.* (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 447, 1087–1093

4. Hoffmann, T.J. and Witte, J.S. (2015) Strategies for imputing and analyzing rare variants in association studies. *Trends Genet.* 31, 556–563

5. Audia, J.E. and Campbell, R.M. (2016) Histone modifications and cancer. *Cold Spring Harb. Perspect. Biol.* 8, a019521

6. Heyn, H. and Esteller, M. (2012) DNA methylation profiling in the clinic: applications and challenges. *Nat. Rev. Genet.* 13, 679–692

7. Voigt, P. and Reinberg, D. (2011) Histone tails: ideal motifs for probing epigenetics through chemical biology approaches. *Chembiochem* 12, 236–252

8. Andersson, R. *et al.* (2014) An atlas of active enhancers across human cell types and tissues. *Nature* 507, 455–461

9. Dimas, A.S. *et al.* (2009) Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* 325, 1246–1250

10. Nica, A.C. *et al.* (2011) The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet.* 7, e1002003

11. Majewski, J. and Pastinen, T. (2011) The study of eQTL variations by RNA-seq: from SNPs to phenotypes. *Trends Genet.* 27, 72–79

12. Li, Q. *et al.* (2013) Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. *Cell* 152, 633–641

13. Fairfax, B.P. *et al.* (2012) Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat. Genet.* 44, 502–510

14. Fairfax, B.P. *et al.* (2014) Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* 343, 1246949

15. Zhang, X. *et al.* (2015) Identification of common genetic variants controlling transcript isoform variation in human whole blood. *Nat. Genet.* 47, 345–352

16. Aran, D. *et al.* (2013) DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. *Genome Biol.* 14, R21

17. Aran, D. and Hellman, A. (2014) Unmasking risk loci: DNA methylation illuminates the biology of cancer predisposition: analyzing DNA methylation of transcriptional enhancers reveals missed regulatory links between cancer risk loci and genes. *Bioessays* 36, 184–190

18. Heyn, H. *et al.* (2014) Linkage of DNA methylation quantitative trait loci to human cancer risk. *Cell Rep.* 7, 331–338

19. Aran, D. and Hellman, A. (2013) DNA methylation of transcriptional enhancers and cancer predisposition. *Cell* 154, 11–13

20. Kim, J. *et al.* (2002) The novel human DNA helicase hFBH1 is an F-box protein. *J. Biol. Chem.* 277, 24530–24537

21. Jeong, Y-T. *et al.* (2013) FBH1 promotes DNA double-strand breakage and apoptosis in response to DNA replication stress. *J. Cell Biol.* 200, 141–149

22. Fugger, K. *et al.* (2013) FBH1 co-operates with MUS81 in inducing DNA double-strand breaks and cell death following replication stress. *Nat. Commun.* 4, 1423

23. Whyte, W.A. *et al.* (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153, 307–319

24. Lovén, J. *et al.* (2013) Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* 153, 320–334

25. Niederriter, A.R. *et al.* (2015) Super enhancers in cancers, complex disease, and developmental disorders. *Genes (Basel)* 6, 1183–1200

26. Pott, S. and Lieb, J.D. (2015) What are super-enhancers? *Nat. Genet.* 47, 8–12

27. Mansour, M.R. *et al.* (2014) Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* 346, 1373–1377

28. Oldridge, D.A. *et al.* (2015) Genetic predisposition to neuroblastoma mediated by a LMO1 super-enhancer polymorphism. *Nature* 528, 418–421

29. Charlet, J. *et al.* (2016) Bivalent regions of cytosine methylation and H3K27 acetylation suggest an active role for DNA methylation at enhancers. *Mol. Cell* 62, 422–431

30. Heyn, H. *et al.* (2016) Epigenomic analysis detects aberrant super-enhancer DNA methylation in human cancer. *Genome Biol.* 17, 11

31. Lay, F.D. *et al.* (2015) The role of DNA methylation in directing the functional organization of the cancer epigenome. *Genome Res.* 25, 467–477

32. Moran, S. *et al.* (2016) Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics* 8, 389–399

33. Vizoso, M. *et al.* (2015) Epigenetic activation of a cryptic *TBC1D16* transcript enhances melanoma progression by targeting EGFR. *Nat. Med.* 21, 741–750

34. O'Connell, M.R. *et al.* (2015) Epigenetic changes and alternate promoter usage by human colon cancers for expressing DCLK1-isoforms: clinical implications. *Sci. Rep.* 5, 14983

35. Blackburn, J. *et al.* (2015) Damage-inducible intragenic demethylation of the human *TP53* tumor suppressor gene is associated with transcription from an alternative intronic promoter. *Mol. Carcinog.* Published online December 16, 2015. http://dx.doi.org/10.1002/mc.22441

36. Homouz, D. and Kudlicki, A.S. (2013) The 3D organization of the yeast genome correlates with co-expression and reflects functional relations between genes. *PLoS One* 8, e54699

37. Dekker, J. and Mirny, L. (2016) The 3D genome as moderator of chromosomal communication. *Cell* 164, 1110–1121

38. Kagey, M.H. *et al.* (2010) Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467, 430–435

39. Mäkinen, N. *et al.* (2011) *MED12*, the mediator complex subunit 12 gene, is mutated at high frequency in uterine leiomyomas. *Science* 334, 252–255

40. Je, E.M. *et al.* (2012) Mutational analysis of *MED12* exon 2 in uterine leiomyoma and other common tumors. *Int. J. Cancer* 131, E1044–E1077

41. Nagasawa, S. *et al.* (2015) *MED12* exon 2 mutations in phyllodes tumors of the breast. *Cancer Med.* 4, 1117–1121

42. Singh, V.P. and Gerton, J.L. (2015) Cohesin and human disease: lessons from mouse models. *Curr. Opin. Cell Biol.* 37, 9–17

43. Watrin, E. *et al.* (2016) Gene regulation and chromatin organization: relevance of cohesin mutations to human disease. *Curr. Opin. Genet. Dev.* 37, 59–66

44. Branco, M.R. and Pombo, A. (2006) Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol.* 4, e138

45. Lieberman-Aiden, E. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293

46. Dixon, J.R. *et al.* (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380

47. Sexton, T. *et al.* (2007) Gene regulation through nuclear organization. *Nat. Struct. Mol. Biol.* 14, 1049–1055

48. Bickmore, W.A. and van Steensel, B. (2013) Genome architecture: domain organization of interphase chromosomes. *Cell* 152, 1270–1284

49. Rao, S.S.P. *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680

50. Tang, Z. *et al.* (2015) CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell* 163, 1611–1627

51. Guo, Y. *et al.* (2015) CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* 162, 900–910

52. Narendra, V. *et al.* (2015) Transcription. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* 347, 1017–1021

53. Hark, A.T. *et al.* (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus. *Nature* 405, 486–489

54. Fudenberg, G. *et al.* (2011) High order chromatin architecture shapes the landscape of chromosomal alterations in cancer. *Nat. Biotechnol.* 29, 1109–1113

55. Flavahan, W.A. *et al.* (2016) Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* 529, 110–114

56. Noushmehr, H. *et al.* (2010) Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* 17, 510–522

57. Turcan, S. *et al.* (2012) IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature* 483, 479–483

58. Brennan, C.W. *et al.* (2013) The somatic genomic landscape of glioblastoma. *Cell* 155, 462–477

59. Ceccarelli, M. *et al.* (2016) Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell* 164, 550–563

60. Dang, L. *et al.* (2009) Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* 462, 739–744

61. Lu, C. *et al.* (2012) IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* 483, 474–478

62. Rasmussen, K.D. and Helin, K. (2016) Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev.* 30, 733–750

63. Boque-Sastre, R. *et al.* (2015) Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. *Proc. Natl. Acad. Sci. U. S. A.* 112, 5785–5790

64. Ferreira, H.J. *et al.* (2012) CpG island hypermethylation-associated silencing of small nucleolar RNAs in human cancer. *RNA Biol.* 9, 881–890

65. Ferreira, H.J. *et al.* (2014) Epigenetic loss of the PIWI/piRNA machinery in human testicular tumorigenesis. *Epigenetics* 9, 113–118

66. Di Leva, G. *et al.* (2014) MicroRNAs in cancer. *Annu. Rev. Pathol.* 9, 287–314

67. Lin, S. and Gregory, R.I. (2015) MicroRNA biogenesis pathways in cancer. *Nat. Rev. Cancer* 15, 321–333

68. Rupaimoole, R. *et al.* (2016) miRNA deregulation in cancer cells and the tumor microenvironment. *Cancer Discov.* 6, 235–246

69. Lujambio, A. *et al.* (2008) A microRNA DNA methylation signature for human cancer metastasis. *Proc. Natl. Acad. Sci. U. S. A.* 105, 13556–13561

70. Yang, D. *et al.* (2013) Integrated analyses identify a master microRNA regulatory network for the mesenchymal subtype in serous ovarian cancer. *Cancer Cell* 23, 186–199

71. Lujambio, A. *et al.* (2007) Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. *Cancer Res.* 67, 1424–1429

72. Baer, C. *et al.* (2013) Genome-wide epigenetic regulation of miRNAs in cancer. *Cancer Res.* 73, 473–477

73. Davalos, V. *et al.* (2012) Dynamic epigenetic regulation of the microRNA-200 family mediates epithelial and mesenchymal transitions in human tumorigenesis. *Oncogene* 31, 2062–2074

74. Gregory, P.A. *et al.* (2008) The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat. Cell Biol.* 10, 593–601

75. Park, S-M. *et al.* (2008) The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes Dev.* 22, 894–907

76. Liu, X. *et al.* (2016) MicroRNA-200 family profile: a promising ancillary tool for accurate cancer diagnosis. *Am. J. Ther.* 23, e388–e397

77. Perdigão-Henriques, R. *et al.* (2016) miR-200 promotes the mesenchymal to epithelial transition by suppressing multiple members of the Zeb2 and Snail1 transcriptional repressor complexes. *Oncogene* 35, 158–172

78. Chan, Y.C. *et al.* (2011) miR-200b targets Ets-1 and is down-regulated by hypoxia to induce angiogenic response of endothelial cells. *J. Biol. Chem.* 286, 2047–2056

Trends in Cancer

CellPress

79. Ning, X. *et al.* (2015) DNMT1 and EZH2 mediated methylation silences the microRNA-200b/a/429 gene and promotes tumor progression. *Cancer Lett.* 359, 198–205

80. Batista, L. *et al.* (2016) Regulation of miR-200c/141 expression by intergenic DNA-looping and transcriptional read-through. *Nat. Commun.* 7, 8959

81. Reinhart, B.J. *et al.* (2000) The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 403, 901–906

82. Brueckner, B. *et al.* (2007) The human *let-7a-3* locus contains an epigenetically regulated microRNA gene with oncogenic function. *Cancer Res.* 67, 1419–1423

83. Wang, T. *et al.* (2015) Aberrant regulation of the LIN28A/LIN28B and Let-7 loop in human malignant tumors and its effects on the hallmarks of cancer. *Mol. Cancer* 14, 125

84. Hildebrandt, M.A.T. *et al.* (2010) Hsa-miR-9 methylation status is associated with cancer development and metastatic recurrence in patients with clear cell renal cell carcinoma. *Oncogene* 29, 5724–5728

85. Park, Y.R. *et al.* (2016) MicroRNA-9 suppresses cell migration and invasion through downregulation of TM4SF1 in colorectal cancer. *Int. J. Oncol.* 48, 2135–2143

86. Baer, C. *et al.* (2012) Extensive promoter DNA hypermethylation and hypomethylation is associated with aberrant microRNA expression in chronic lymphocytic leukemia. *Cancer Res.* 72, 3775–3785

87. He, L. *et al.* (2007) A microRNA component of the p53 tumour suppressor network. *Nature* 447, 1130–1134

88. Mraz, M. *et al.* (2009) miR-34a, miR-29c and miR-17-5p are downregulated in CLL patients with TP53 abnormalities. *Leukemia* 23, 1159–1163

89. Asslaber, D. *et al.* (2010) MicroRNA-34a expression correlates with MDM2 SNP309 polymorphism and treatment-free survival in chronic lymphocytic leukemia. *Blood* 115, 4191–4197

90. Fabbri, M. *et al.* (2011) Association of a microRNA/TP53 feedback circuitry with pathogenesis and outcome of B-cell chronic lymphocytic leukemia. *JAMA* 305, 59–67

91. Musilova, K. and Mraz, M. (2015) MicroRNAs in B-cell lymphomas: how a complex biology gets more complex. *Leukemia* 29, 1004–1017

92. Guttman, M. *et al.* (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458, 223–227

93. Guttman, M. *et al.* (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 477, 295–300

94. Rinn, J.L. *et al.* (2007) Functional demarcation of active and silent chromatin domains in human *HOX* loci by noncoding RNAs. *Cell* 129, 1311–1323

95. Gong, C. and Maquat, L.E. (2011) lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. *Nature* 470, 284–288

96. Sahu, A. *et al.* (2015) Long noncoding RNAs in cancer: from function to translation. *Trends Cancer* 1, 93–109

97. Lee, J.T. (2012) Epigenetic regulation by long noncoding RNAs. *Science* 338, 1435–1439

98. Khalil, A.M. *et al.* (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 106, 11667–11672

99. Davidovich, C. *et al.* (2013) Promiscuous RNA binding by Polycomb repressive complex 2. *Nat. Struct. Mol. Biol.* 20, 1250–1257

100. Cifuentes-Rojas, C. *et al.* (2014) Regulatory interactions between RNA and Polycomb repressive complex 2. *Mol. Cell* 55, 171–185

101. Kretz, M. and Meister, G. (2014) RNA binding of PRC2: promiscuous or well ordered? *Mol. Cell* 55, 157–158

102. Loewen, G. *et al.* (2014) Functions of lncRNA HOTAIR in lung cancer. *J. Hematol. Oncol.* 7, 90

103. Niinuma, T. *et al.* (2012) Upregulation of miR-196a and HOTAIR drive malignant character in gastrointestinal stromal tumors. *Cancer Res.* 72, 1126–1136

104. Hajjari, M. and Salavaty, A. (2015) HOTAIR: an oncogenic long non-coding RNA in different cancers. *Cancer Biol. Med.* 12, 1–9

105. Gupta, R.A. *et al.* (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464, 1071–1076

106. Bartolomei, M.S. *et al.* (1991) Parental imprinting of the mouse *H19* gene. *Nature* 351, 153–155

107. DeChiara, T.M. *et al.* (1991) Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* 64, 849–859

108. Barlow, D.P. *et al.* (1991) The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the *Tme* locus. *Nature* 349, 84–87

109. Tsang, W.P. *et al.* (2010) Oncofetal H19-derived miR-675 regulates tumor suppressor RB in human colorectal cancer. *Carcinogenesis* 31, 350–358

110. Keniry, A. *et al.* (2012) The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. *Nat. Cell Biol.* 14, 659–665

111. Shi, Y. *et al.* (2014) Long non-coding RNA H19 promotes glioma cell invasion by deriving miR-675. *PLoS One* 9, e86295

112. Zhuang, M. *et al.* (2014) The long non-coding RNA H19-derived miR-675 modulates human gastric cancer cell proliferation by targeting tumor suppressor RUNX1. *Biochem. Biophys. Res. Commun.* 448, 315–322

113. Kallen, A.N. *et al.* (2013) The imprinted H19 lncRNA antagonizes Let-7 microRNAs. *Mol. Cell* 52, 101–112

114. Thomson, D.W. and Dinger, M.E. (2016) Endogenous microRNA sponges: evidence and controversy. *Nat. Rev. Genet.* 17, 272–283

115. Denzler, R. *et al.* (2014) Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance. *Mol. Cell* 54, 766–776

116. Léveillé, N. *et al.* (2015) Genome-wide profiling of p53-regulated enhancer RNAs uncovers a subset of enhancers controlled by a lncRNA. *Nat. Commun.* 6, 6520

117. Lujambio, A. *et al.* (2010) CpG island hypermethylation-associated silencing of non-coding RNAs transcribed from ultraconserved regions in human cancer. *Oncogene* 29, 6390–6401

118. Waddington, C.H. (1956) Genetic assimilation of the bithorax phenotype. *Evolution* 10, 1

119. Schwartzman, O. and Tanay, A. (2015) Single-cell epigenomics: techniques and emerging applications. *Nat. Rev. Genet.* 16, 716–726

120. Clark, S.J. *et al.* (2016) Single-cell epigenomics: powerful new methods for understanding gene regulation and cell identity. *Genome Biol.* 17, 72

121. Kreso, A. and Dick, J.E. (2014) Evolution of the cancer stem cell model. *Cell Stem Cell* 14, 275–291

122. Lander, E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921

123. Kassiotis, G. (2014) Endogenous retroviruses and the development of cancer. *J. Immunol.* 192, 1343–1349

124. Chiappinelli, K.B. *et al.* (2015) Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cell* 162, 974–986

125. Lee, E. *et al.* (2012) Landscape of somatic retrotransposition in human cancers. *Science* 337, 967–971

126. Bjerregaard, B. *et al.* (2006) Syncytin is involved in breast cancer–endothelial cell fusions. *Cell. Mol. Life Sci.* 63, 1906–1911

127. Duelli, D. and Lazebnik, Y. (2007) Cell-to-cell fusion as a link between viruses and cancer. *Nat. Rev. Cancer* 7, 968–976

128. Cohen, C.J. *et al.* (2009) Endogenous retroviral LTRs as promoters for human genes: a critical assessment. *Gene* 448, 105–114

129. Vaskova, E.A. *et al.* (2013) "Epigenetic memory" phenomenon in induced pluripotent stem cells. *Acta Naturae* 5, 15–21

130. Campos, E.I. *et al.* (2014) Epigenetic inheritance: histone bookmarks across generations. *Trends Cell Biol.* 24, 664–674

131. Kwasnieski, J.C. *et al.* (2014) High-throughput functional testing of ENCODE segmentation predictions. *Genome* 24, 1595–1602

Trends in Cancer

CellPress

132. Kwasnieski, J.C. *et al.* (2012) Complex effects of nucleotide variants in a mammalian *cis*-regulatory element. *Proc. Natl. Acad. Sci. U. S. A.* 109, 19498–19503

133. White, M.A. *et al.* (2013) Massively parallel *in vivo* enhancer assay reveals that highly local features determine the *cis*-regulatory function of ChIP-seq peaks. *Proc. Natl. Acad. Sci. U. S. A.* 110, 11952–11957

134. Murtha, M. *et al.* (2014) FIREWACh: high-throughput functional detection of transcriptional regulatory modules in mammalian cells. *Nat. Methods* 11, 559–565

135. Dickel, D.E. *et al.* (2014) Function-based identification of mammalian enhancers using site-specific integration. *Nat. Methods* 11, 566–571

136. Arnold, C.D. *et al.* (2013) Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* 339, 1074–1077

137. van Steensel, B. and Henikoff, S. (2000) Identification of *in vivo* DNA targets of chromatin proteins using tethered Dam methyltransferase. *Nat. Biotechnol.* 18, 424–428

138. van Steensel, B. *et al.* (2001) Chromatin profiling using targeted DNA adenine methyltransferase. *Nat. Genet.* 27, 304–308

139. Wu, F. and Yao, J. (2013) Spatial compartmentalization at the nuclear periphery characterized by genome-wide mapping. *BMC Genomics* 14, 591

140. Dekker, J. *et al.* (2002) Capturing chromosome conformation. *Science* 295, 1306–1311

141. Ma, W. *et al.* (2015) Fine-scale chromatin interaction maps reveal the *cis*-regulatory landscape of human lincRNA genes. *Nat. Methods* 12, 71–78

142. Kind, J. *et al.* (2015) Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* 163, 134–147

143. Nagano, T. *et al.* (2013) Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502, 59–64

144. Nagano, T. *et al.* (2015) Single-cell Hi-C for genome-wide detection of chromatin interactions that occur simultaneously in a single cell. *Nat. Protoc.* 10, 1986–2003

145. de Wit, E. and de Laat, W. (2012) A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* 26, 11–24

146. Zhang, J. *et al.* (2012) ChIA-PET analysis of transcriptional chromatin interactions. *Methods* 58, 289–299

147. Li, G. *et al.* (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* 148, 84–98

148. Raviram, R. *et al.* (2016) 4C-ker: a method to reproducibly identify genome-wide interactions captured by 4C-seq experiments. *PLoS Comput. Biol.* 12, e1004780