npg

REVIEW

Junk DNA and the long non-coding RNA twist in cancer genetics

H Ling[1], K Vincent[1], M Pichler[1], R Fodde[2], I Berindan-Neagoe[1,3,4], FJ Slack[5] and GA Calin[1]

The central dogma of molecular biology states that the flow of genetic information moves from DNA to RNA to protein. However, in the last decade this dogma has been challenged by new findings on non-coding RNAs (ncRNAs) such as microRNAs (miRNAs). More recently, long non-coding RNAs (lncRNAs) have attracted much attention due to their large number and biological significance. Many lncRNAs have been identified as mapping to regulatory elements including gene promoters and enhancers, ultraconserved regions and intergenic regions of protein-coding genes. Yet, the biological function and molecular mechanisms of lncRNA in human diseases in general and cancer in particular remain largely unknown. Data from the literature suggest that lncRNA, often via interaction with proteins, functions in specific genomic loci or use their own transcription loci for regulatory activity. In this review, we summarize recent findings supporting the importance of DNA loci in lncRNA function and the underlying molecular mechanisms via *cis* or *trans* regulation, and discuss their implications in cancer. In addition, we use the 8q24 genomic locus, a region containing interactive SNPs, DNA regulatory elements and lncRNAs, as an example to illustrate how single-nucleotide polymorphism (SNP) located within lncRNAs may be functionally associated with the individual's susceptibility to cancer.

INTRODUCTION

The Encyclopedia of DNA Elements (ENCODE) project has revealed that at least 75% of the human genome is transcribed into RNAs, while protein-coding genes comprise only 3% of the human genome.[1] Because of a long-held protein-centered bias, many of the genomic regions that are transcribed into non-coding RNAs (ncRNAs) had been viewed as 'junk' in the genome, and the associated transcription had been regarded as transcriptional 'noise' lacking biological meaning.[2] The last decade has witnessed an explosive expansion in the understanding of biological function and clinical significance of ncRNA transcripts, exemplified by the large number of published reports linking microRNAs (miRNAs) and various human diseases including cancer.[3] With the advancement of sequencing technology and bioinformatics, other types of short or long ncRNAs, such as endogenous small interfering RNAs, PIWI-interacting RNAs, small nucleolar RNAs, natural antisense transcripts (NATs), circular RNAs, long intergenic ncRNAs (lincRNAs), enhancer ncRNAs and transcribed ultraconserved regions (T-UCRs), have been characterized and classified.[4,5] Among these ncRNAs, long ncRNAs (lncRNAs), defined as being at least 200 nucleotides in length, have received much attention due to their abundant presence in the human genome, as well as their tissue-specific expression patterns and functional relevance in complex physiological and pathological processes.[6] Distinct from the short miRNAs, the length of lncRNAs allows them to fold into more complex three-dimensional structures, likely to determine specific interactions of lncRNA with biomolecule partners such as transcription factors, histones or other chromatin-modifying proteins. Consequently, alterations in lncRNA expression levels could affect a broad spectrum of genes via their protein partners and as such cause profound phenotypic changes.[7] LncRNAs could also have sequence-specific interactions with DNA or RNA in the forms of duplex or triplex structures[8,9] and create complex regulatory networks composing of DNA, RNA and proteins.

The mapping of several lncRNAs to regulatory genomic regions such as promoters and enhancers[1] indicates a possible involvement of these noncoding transcripts in gene regulation. In addition, genome-wide association studies (GWAS) revealed that < 10% of the disease-related single-nucleotide polymorphisms (SNPs) are in exons of protein-coding genes, whereas nearly half of the disease-associated SNPs are outside protein-coding genes.[10] Although lncRNA function remains largely unknown, recent studies have clearly demonstrated the functional importance of lncRNAs in embryonic development,[11] cell differentiation[12] and various human diseases including cancer.[5,13,14] Mechanistically, lncRNAs that are transcribed from regulatory elements or cancer-associated genomic regions may cooperate with their genomic DNA elements to fine-tune the complex biological activities necessary for precise regulation. This might be of particular relevance in the regulation of complex biological activities that do not obey to 'binary switch' (on and off) regulation, but are rather regulated in a subtler, dosage-dependent manner.

The topic of lncRNA has been covered in several excellent in-depth review papers.[13,15–20] Here, we focus on the interplay between DNA and lncRNA in the human genome and the relevance of these interactions in human cancer. We introduce

[1]Department of Experimental Therapeutics, MD Anderson Cancer Center, University of Texas, Houston, TX, USA; [2]Department of Pathology, Erasmus MC Cancer Institute, University Medical Center Rotterdam, Rotterdam, The Netherlands; [3]Department of Immunology and Research Center for Functional Genomics, Biomedicine and Translational Medicine University of Medicine and Pharmacy 'I. Hatieganu', Cluj-Napoca, Romania; [4]Department of Functional Genomics, The Oncology Institute Ion Chiricuta, Cluj-Napoca, Romania and [5]Department of Pathology, Beth Israel Deaconess Medical Center, Harvard medical School, Boston, MA, USA. Correspondence: Dr G Calin, Department of Experimental Therapeutics, The University of Texas, MD Anderson Cancer Center, 1881 East Road, Unit 1950, APT 1125, Houston, TX 77030, USA.
E-mail: gcalin@mdanderson.org
Received 15 October 2014; revised 3 December 2014; accepted 4 December 2014; published online 26 January 2015

various types of lncRNAs from regulatory genomic elements, summarize recently identified molecular mechanisms of DNA–RNA interaction in the context of cancer and discuss the clinical relevance of the findings.

THE MISSING CULPRIT GENES

In many non-hypothesis-driven studies, large-scale genotyping from population-based samples is used to evaluate disease gene associations. Among these, GWAS studies provided valuable information as to the genetic variants in cancer risk, disease diagnosis, prognosis and treatment response.[21] However, the molecular mechanisms underlying such links remain largely undefined, owing to the fact that many of these genetic variants (43%) are located in gene 'desert' regions that lack protein-coding genes[13] (summarized in Table 1).

Similarly, non-GWAS studies also point to the same observations that in many cases protein-coding genes are not the culprits responsible for disease phenotypes. This notion can be exemplified and supported by the role played by miRNA-15a/16-1 in chronic lymphocytic leukemia (CLL).[22] A recurring pattern of 13q14.3 deletions was observed in CLL indicative of the presence of a tumor suppressor in this region. However, the protein-coding genes identified from this genomic region did not fulfill this tumor-suppressing function.[22] Instead, two miRNAs were identified and subsequently proven by multiple studies to underlie the etiology of CLL.[22] Following this initial finding, several studies identified numerous miRNAs involved in a broad spectrum of human malignancies. Nevertheless, miRNAs and protein-coding genes are not the only determining factors of disease phenotype. Other DNA regulatory elements may have an important role in causing morbid phenotypes by altering gene transcription modalities. Moreover, other types of ncRNAs are transcribed from cancer-associated genomic regions and participate in cancer pathogenesis.

'JUNK DNA' ENCODES FOR LNCRNAS

The protein-centered dogma had viewed genomic regions not coding for proteins as 'junk' DNA. We now understand that many lncRNAs are transcribed from 'junk' regions, and even those encompassing transposons, pseudogenes and simple repeats represent important functional regulators with biological relevance.[23,24] For the convenience of this review, we subdivided lncRNAs into several categories based on their genomic locus relative to protein-coding genes or their unique structural features. However, these classifications are not exclusive, and this grouping does not have any bearing on their biological activity or functional mechanisms.

Promoter-associated lncRNAs

Gene promoters interact with transcription factors and RNA polymerases to activate transcription.[25] The recent identification of ncRNA transcripts located within the promoter region of several genes[26,27] has clearly indicated that more complex regulatory mechanisms should be envisaged. A tiling microarray aimed at the study of ncRNAs mapping in the proximity of the transcription start site of 56 cell-cycle-related genes revealed extensive transcription activity in the gene promoter region without protein-coding feature. Among these lncRNA, the non-spliced 1.5-kb ncRNA PANDA transcribed from 5 kb upstream of the CDKN1A transcription start site was proven to function in the DNA damage response.[28] Interestingly, while CDKN1A mediates cell cycle arrest, PANDA promotes cell survival in response to DNA damage by preventing the transcription factor NF-YA from binding specific promoters of apoptosis-inhibiting genes.[28] This indicates that following DNA damage response, both cell cycle arrest and anti-apoptotic genes (and possibly genes with other functions) can be induced from the same locus, and a complex network will determine the biological phenotypes. In another study, a promoter-associated lncRNA complementary to the rRNA gene promoter binds to rRNA gene to form a lncRNA-DNA triplex. This RNA-DNA triplex prevents the binding of transcription termination factor 1 to the rRNA gene and recruits DNMT3b to silence it.[8]

Enhancer ncRNAs

Enhancers are defined as DNA elements that, independently of their proximity or orientation with respect to the gene transcription site, are able to enhance gene expression levels.[29] Notably, many active enhancer regions are transcribed into lncRNAs.[30] In mouse neurons, out of the 12 000 neuronal activity-regulated enhancers defined by p300/CBP occupation and histone H3-Lysine 4 mono-methylation (H3K4Me1), 2000 were found to bi-directionally express lncRNAs, termed 'enhancer RNAs' or eRNAs, that are predominantly non-polyadenylated.[31] Positive association of eRNA expression at neuronal enhancers with the levels of nearby protein-coding genes suggests that eRNA may regulate mRNA synthesis.[31] Next to eRNAs, polyadenylated 'enhancer-like ncRNAs' were identified from genomic enhancer regions and shown by RNA interference to activate neighboring protein-coding genes in cis.[32] In addition, several T-lymphocyte-specific enhancers are bound by RNA polymerase II and general transcription factors, and express both polyadenylated and nonpolyadenylated lncRNAs.[33]

Evf2 lncRNA represents yet another example of enhancer RNA that regulates gene expression of the Dlx cluster through interaction with the transcription factor Dlx2.[34] HOTTIP, an lncRNA expressed from the distal tip of the HoxA locus, drives expression

LncRNA	Cancer type	LncRNA function	Gene locus	SNP	Risk allele	GWAS
PCGEM1	Prostate	Oncogenic[118,119]	2q32	rs6434568	C > A	Prostate[120]
PCGEM1	Prostate	Oncogenic[118,119]	2q32	rs16834898	A > C	Prostate[120]
HULC	Hepatocellular	Oncogenic[121]	6p24	rs7763881	A > C	Hepatocellular[79]
CCAT2	Colorectal	Oncogenic[41]	8q24	rs6983267	G > T	Prostate, coloretal[95,122,123]
PCAT1	Prostate	Oncogenic[88]	8q24	rs1026411	C > T?	Breast[90]
PCAT1	Prostate	Oncogenic[88]	8q24	rs12543663	A > C?	Prostate[91]
PRNCR1	Prostate	Oncogenic[36]	8q24	rs1456315	?	Prostate[92]
PRNCR1	Prostate	Oncogenic[36]	8q24	rs7463708	?	Prostate[92]
ANRIL	Neurofibromas, Esophageal	Oncogenic[73,124]	9p21	rs2151280	T > C	Neurofibromas[74]
H19	Hepatocellular	Oncogenic or suppressive[75,76]	11p15	rs2839698	C > T	Bladder[77]
H19	Hepatocellular	Oncogenic or suppressive[75,76]	11p15	rs2107425	C > T	Breast cancer[78]
MALAT1	Lung	Oncogenic[125]	11q13	rs619586	A > G	Hepatocellular[79]

Table 1. Examples of cancer-associated lncRNAs containing cancer predisposition SNPs

Abbreviations: GWAS, genome-wide association studies; lncRNA, long non-coding RNA; SNP, single-nucleotide polymorphism.

of several *HoxA* genes.[35] Using an engineered reporter plasmid, it was elegantly shown that HOTTIP activates *HoxA* genes by *cis* regulation.[35] Recently, two lncRNAs highly expressed in aggressive prostate cancers, PRNCR1 and PCGEM1, were found to enhance transcription of ~ 2000 androgen receptor-responsive genes by binding to the androgen receptor.[36] This study expanded the functional mechanisms of enhancer RNAs by demonstrating a sophisticated underlying mechanism of *trans* regulation.

T-UCRs

Untraconserved regions (UCRs) refer to a subset of conserved genome sequences longer than 200 bp that are conserved with 100% identity between orthologous regions of the human, rat and mouse genomes. Although a high degree of genomic conservation usually indicates functional relevance, more than half of the 481 ultraconserved regions described by Bejerano *et al*.[37] have no protein-coding potential. Microarray analysis showed that 93% of the UCRs have transcriptional activity in at least one tissue, and consequently are referred to as T-UCR. T-UCR profiling in a panel of 133 human leukemia and carcinoma samples and 40 corresponding normal tissues identified specific signatures associated with each cancer type. For instance, uc.349A and uc.352, both mapping to the familial CLL-associated fragile chromosomal region 13q21.33-q22.2, are differentially expressed in normal versus malignant B-CLL CD5-postitive cells.[38] Following the initial report, several studies have reported the importance of the role played by T-UCRs in cancer. For instance, uc.338, a T-UCR whose expression is markedly increased in human hepatocellular carcinoma compared with noncancerous adjacent tissues, promotes anchorage-dependent and anchorage-independent cell proliferation.[39] Studies from our group showed that uc.475, a hypoxia-induced noncoding ultraconserved transcript, enhances cell proliferation specifically under hypoxic conditions.[40] In addition, we identified a novel lncRNA, named CCAT2, transcribed from a highly conserved 'gene-desert' region, and encompassing the cancer-associated SNP rs6983267. We showed that CCAT2 is an oncogenic lncRNA promoting chromosomal instability and colorectal cancer metastasis.[41] More recently, the uc.283+A T-UCR was shown to interfere with miRNA processing by binding to primary miRNA-195 (pri-miR-195) via sequence complementarity.[42] Despite these findings, the biological activities and functional mechanisms of the majority of T-UCRs still remain largely unexplored. It should be noted that for functional attribution of T-UCRs in human diseases, precise gene annotation is the key, and this requires rigorous analysis determining sense or antisense orientation of ncRNA, for instance, by northern blotting, strand-specific PCR and deep sequencing.

NATs

NATs are endogenous RNA molecules that are partially or fully complementary to protein-coding transcripts. According to their genomic origin, NATs can be separated into *cis*-NATs, which are transcribed from the same genomic loci as their sense transcripts but from the opposite DNA strand, and *trans*-NATs, which are transcribed from genomic regions that are distinct from those encoding their sense counterpart.[43,44] Although generally expressed at relatively low level compared with the sense transcripts, NATs have been shown to effectively regulate expression level of their protein-coding targets.[45] Systematic global transcriptome analysis suggested that ~ 70% of transcripts have antisense partners, and that perturbation of antisense RNA can alter the expression of the sense gene.[46] NATs activate or inactivate sense gene transcription by mechanisms including epigenetic modifications.[45] ANRIL is a NAT transcribed from the INK4A-INK4B gene-cluster locus encoding for the tumor suppressor genes *CDKN2A* and *CDKN2B*.[47] Through interaction with CBX7, a component of polycomb repressive complex 1 (PRC1) able to

recognize H3K27me3-repressive marks, ANRIL recruits the protein complex to its locus for sustained repression of the INK4A-INK4B gene cluster.[48] NATs also affect gene expression through post-transcriptional regulation such as splicing. During epithelial–mesenchymal transition, a NAT at the ZEB2 locus is transcriptionally activated. This ZEB2 NAT inhibits splicing of an internal ribosome entry site-containing intron, and positively regulates ZEB2 protein expression.[49] The regulation of sense transcript by NATs provides a natural way of improving or reducing protein expression.

LincRNAs

Initially identified using histone marker signatures associated with RNA polymerase II, lincRNAs have received much attention because of their lack of overlap with protein-coding genes. Therefore, their effect can be characterized without ambiguity in the attribution of biological functions.[19] HOTAIR is among the first lincRNA that was functionally and mechanistically elucidated.[50] Transcribed from a *HOXC* gene cluster, HOTAIR controls gene expression via a trans-effect, that is, affecting transcription on chromosomes other than the one producing the gene.[50] This was achieved by interaction of HOTAIR with polycomb repressive complex 2 (PRC2) and LSD1, which promotes repressive histone marks (such as H3K27me3) to silence the HOXD locus.[51] LincRNA-p21, a polyadenylated RNA transcribed from the upstream opposite strand to p21, is induced by DNA damage and acts as a downstream regulator of the p53 transcriptional response.[52] LincRNA-p21 physically associates with hnRNP K through its 5′ end and represses p53-responsive apoptotic genes.[52]

THE DNA-RNA TWIST IN CANCER GENETICS

The elucidation of the mechanisms underlying lncRNA function falls far behind the discovery pace of new lncRNAs. Although lncRNAs could be easily classified into different types according to their genomic locus or other features, this classification does not shed light on the mechanisms. Instead, lncRNAs from different classes might possibly share similar molecular mechanisms. Generally, the mode of action of lncRNAs can be classified into *cis* and *trans* regulation, depending on whether the lncRNA regulates neighboring genes on the same chromosomal regions where they are located or distant genes on other chromosomes, respectively (See Figure 1). In both cases, lncRNAs need to interact directly or indirectly with genomic DNA elements, in most cases with assistance of proteins, to perform specific biological functions. In addition, SNP variants inside a lncRNA sequence may not only affect the function of the DNA element, but also affect the primary sequence, and possibly the higher-order structure, and consequently the activities of the lncRNA.

Cis regulation within the genomic context

LncRNAs have several unique properties as *cis*-acting molecules.[53] First, lncRNAs are in close proximity, when compared with proteins, to their genomic locus during transcription and are thus able to direct locus- and allele-specific regulation. Second, the length of lncRNAs gives an advantage to bind with multiple epigenetic complexes and work as initiators or mediators in genomic looping feats necessary for active chromatin of gene transcription. Third, the length of lncRNAs makes it possible to function during transcription, and immediately after transcriptional termination the degradation signals might prevent diffused action at other genomic sites. Many lncRNAs mediate local functions in *cis*, interacting with chromatin-modifying proteins to regulate their neighboring genes. These include several previously mentioned enhancer RNAs and NATs. For instance, HOTTIP recruits WD repeat domain 5 (WDR5)/mixed lineage leukemia (MLL) complex to drive the H3K4M3 signature and gene transcription of
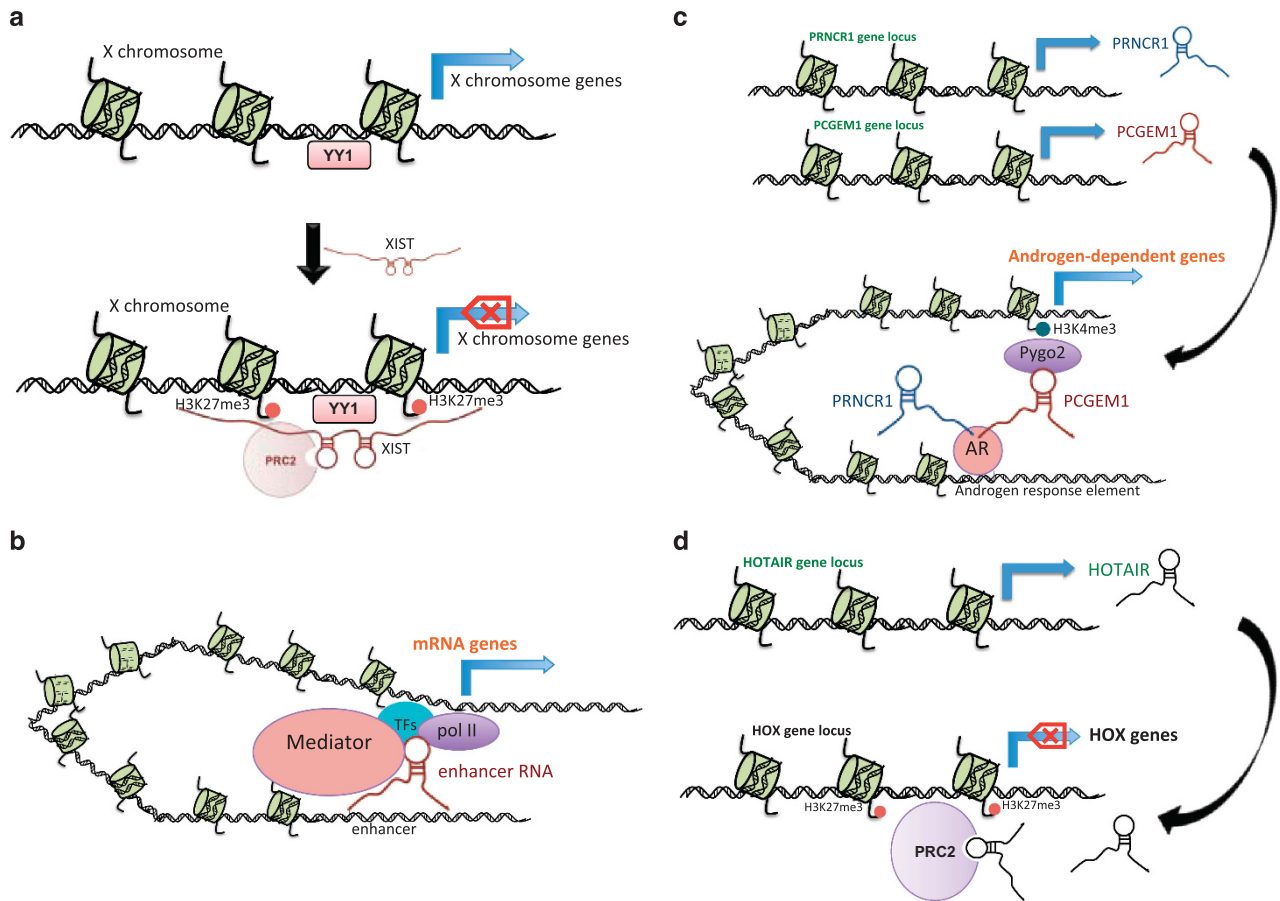
Figure 1. LncRNA functioning mechanisms via DNA–RNA interaction in *cis* or *trans*. (**a**) XIST loads onto the its own genomic locus via YY1 and recruits PCR2 complex to maintain repressive chromatin marked by H3K27me3 on the same X chromosome. (**b**) Enhancer RNAs transcribed from enhancer region maintain enhancer-promoter looping by recruiting mediators and transcription factors, and enhance transcription of neighboring mRNA genes. (**c**) PRNCR1, transcribed from 8q24, and PCGEM1, produced from 2q32, bind to androgen receptor (AR) to promote the chromatin status H3K4me3 and activate the AR-regulated genes located distant from their genomic loci. (**d**) HOTAIR recruits PRC2 and loads onto distant genomic loci to initiate repressive chromatin marked by H3K27me3 and block HOX gene transcription.

HoxA distal genes.[35] Chromosomal looping facilitates HOTTIP to act on its target genes.[35] This mechanism was elegantly demonstrated with a luciferase reporter artificially tethered with HOTTIP.[35] The lncRNA Mistral employs a similar mechanism of MLL interaction to recruit to and activate the *Hoxa6* and *Hoxa7* genes.[54] The lncRNA ecCEBPA uses a different mechanism, by binding to DNMT1 to prevent methylation of the *CEBPA* gene.[55]

The *cis* regulation could also elicit broader epigenetic changes, as in the cases of Xist, an lncRNA silencing an entire female X chromosome, and of several other lncRNAs regulating gene imprinting. Xist is transcribed exclusively from the inactive X chromosome in females, and tethered to the X inactivation center by the transcription factor Yin Yang 1 (YY1).[56] Xist RNA coats the X chromosome and serves as a scaffold for recruitment of silencing factors such as PRC2.[57] Interestingly, a repeated motif named 'Repeat A' within the Xist RNA encompassing a stem-loop structure was shown to be responsible for the recruitment of the PRC2 complex to the inactive X chromosome.[58] As an example of regulating gene imprinting, the lncRNA Air, transcribed from the paternal allele, recruits G9a to methylate H3K9 residues over an adjacent 300-kb genomic region, thus silencing the expression of distantly located genes including *Igf2r*, *Slc22a2* and *Slc22a3* on the paternal chromosome.[59]

LncRNAs not only regulate protein-coding genes, but can also activate neighboring lncRNAs. An example of this is the regulation of Xist by Tsix, a lncRNA transcribed in the antisense orientation in

relation to Xist from the activate X chromosome.[60] Tsix recruits PRC2 and methyltransferase DNMT3A to the Xist promoter, thus maintaining a repressive chromatin domain for long-term silencing of the *Xist* gene.[61] In addition, Tsix and Xist can form RNA duplex structures, which are subsequently subjected to RNA interference into small regulatory RNAs.[62]

Although it has been shown that the *in cis* mechanism employs genomic looping to exert a regulatory effect, whether lncRNAs are necessary to maintain the loop still remains to be determined. Lai et al.[63] demonstrated that knockdown of either lncRNAs or mediator (coactivator complex bridging regulatory information from enhancers to the promoter) abolished the chromatin interactions, supporting a participation of both the mediator and the lncRNA in looping enhancer–promoter interactions. Further, the lncRNA–mediator interaction regulates the kinase activity of the mediator protein, and subsequently promotes phosphorylation of serine 10 on histone H3, a chromatin mark for transcriptional activation.[63] However, the role of lncRNA in maintaining chromatin looping was not observed in other studies. For instance, depletion of HOTTIP did not disrupt looping chromatin architecture, as determined by high-throughput chromosome conformation capture.[35] A recent study similarly suggests that chromatin looping linking p53-binding sites and their targets does not depend on the lncRNAs transcribed from the p53-binding sites.[64]

Trans regulation at distant genomic loci

The property of interaction with proteins such as transcription factors or chromatin modifiers suggests the possibility of *trans* regulation by lncRNAs able to act outside the genomic locus they map to. About 20% of all lincRNAs have PRC2 as an interaction partner to regulate gene expression, thus suggesting widespread *trans*-regulated chromatin remodeling, as previously characterized for HOTAIR.[65] Similarly, a cross-linking immunoprecipitation followed by sequencing (CLIP-seq) study of RNAs associated with the SFRS1 splicing factor identified more than 6000 spliced ncRNAs.[66] Although not yet experimentally proven, it can be envisioned that a single ncRNA could affect a wide range of genes regulated by SFRS1. A more recent study showing regulation of androgen receptor-responsive genes by PRNCR1 and PCGEM1 also represents a *trans* mechanism through which more than 2000 genes are regulated by lncRNAs.[36]

While it is clear that lncRNAs target proteins to exert their *in trans* effects, the factors determining the RNA-protein interaction are not well-defined. Interestingly, several studies suggest that the secondary structure, instead of the primary lncRNA sequence, dictates a specific interaction. For instance, the tumor suppressor function of the MEG3 lncRNA was maintained by the conservation of the secondary structure, though not in its primary sequence.[67,68] In addition, repetitive sequences were found to contribute to the interaction with protein partners. In the case of Xist, although the *cis* regulatory mechanism is well established, it still provides an example to explain the importance of higher order structures in RNA–protein interaction. A cluster of nine repetitive elements within Xist was found to form stem-loop structures essential for the interaction with PRC1 and for H3K27 trimethylation, while another region encompassing repetitive elements was shown to bind to YY1 through a stem-loop structure tethering Xist onto the X chromosome.[56,69] Studies on short interspersed elements that are derived from transposons have also showed that repetitive sequences are the recognition domains for RNA polymerase II binding, and that such interactions leads to repression of mammalian heat-shock genes.[70,71]

Another puzzling question relative to the mechanisms underlying *in trans* regulation is how the lncRNAs recognize specific genomic loci. One possibility is that the primary or secondary structure of lncRNAs defines their preferred interaction with certain genomic regions. Using a technique named chromatin isolation by RNA purification (ChIRP), in combination with deep sequencing of genomic binding sites, an enriched binding motif was identified for HOTAIR.[9] The exact structure responsible for such RNA–DNA interaction remains to be determined. Notably, a promoter-associated lncRNA forms a triplex with the transcription termination factor 1 binding site, and subsequently recruits DNMT3b to silence *rRNA* gene.[8] The specific recognition of genomic loci could also be achieved by the relay of protein partners, as illustrated by the activation of androgen receptors-responsive genes by PRNCR1 and PCGEM1 via interaction with androgen receptor.[36]

Linking SNPs, lncRNAs and cancer

The fact that ~90% of disease-associated SNPs are in genomic regions not coding for proteins[10] suggests that these 'gene-poor' regions may represent a 'gold mine' towards the identification and characterization of novel lncRNAs. To facilitate such an effort, a lincSNP database has been established to link lncRNAs with disease-related SNPs.[72] Although the association does not necessarily mean a causal relationship between specific lncRNAs and disease phenotypes, the possibility of finding long-sought lncRNA culprits is a very attractive one. In addition, a disease predisposition SNP may flag the existence of regulatory element of a gene whose function is only weakly affected by the SNP variant(s). These 'disease predisposing' SNPs could be located upstream, within, or downstream of the lncRNAs. Here, we only review the cancer-related lncRNAs that also encompass cancer-risk SNPs (see Table 1). ANRIL was found to be a hotspot for risk locus for gliomas and basal cell carcinomas in GWAS.[73] The rs2151280 SNP variants located within the *ANRIL* gene were significantly associated with susceptibility to neurofibromas.[74] Moreover, the T allele of rs2151280 was correlated with lower ANRIL levels, suggesting that this SNP variant could affect ANRIL expression.[74] The rs2839698 and rs2107425 SNPs located within H19, a lncRNA with both oncogenic[75] and tumor suppressive activity,[76] were reported to be associated with bladder cancer risk.[77] Rs2107425 is also found to confer increased breast cancer risk in a different study.[78] HULC, a lncRNA involved in hepatocellular carcinoma, encompasses the rs7763881 SNP that determines susceptibility to hepatocellular carcinoma in HBV patients.[79] Similarly, this group also identified that the rs619586 variants, located within the *MALAT1* gene, were associated with hepatocellular carcinoma risk though with marginal significance.[79]

A TWISTED 8Q24 GENOMIC REGION

The 8q24 genomic region is frequently altered by amplification, deletion, viral integration or translocation in many types of human cancers.[80] A large-scale study identified the 8q24 region as the most frequently (14%) amplified region among inhuman cancers.[81] In addition, GWAS point to 8q24 as a hotspot for cancer-associated SNPs owing to the density, strength, as well as the high allele frequency of these SNPs.[82] However, the 2 Mb SNP-rich 8q24 region has nevertheless been considered a 'gene desert' largely because of the absence of functionally annotated genes with the only notable exception of the *MYC* proto-oncogene.[83] Several 8q24 loci have demonstrated enhancer activity, and it has been proposed that these enhancer activities might regulate *MYC* expression through looping with its promoter.[84] Recently, several reports revealed that lncRNAs including CCAT1,[85] CCAT2,[41] CARLo-5,[86] PVT1,[87] PCAT1,[88] and PRNCR1 [36] are transcribed from these regions (Figure 2). Among these, CCAT2, PCAT1 and PRNCR1 encompass the cancer predisposition SNPs (Table 1).[41,89–92] Several of these lncRNAs (for example, CCAT1 and CCAT2) regulate MYC expression,[41,85] while the rs6983267 SNP that resides within the *CCAT2* gene shows allele-specific effect on the lncRNA CARLo-5 expression levels.[86] Recently, MYC copy-number gains were found to depend on PVT1 in mice with chromosome engineering.[87]

The *CCAT2* gene is located in a very special region: first, this genomic region has shown enhancer activity affected by the SNP variants.[93,94] Second, the rs6983267 SNP it encompasses is one of the most consistently identified predisposition SNPs in multiple types of cancer including colorectal cancer, prostate cancer, ovarian cancer, head and neck cancer and inflammatory breast cancer.[95,96] Third, its genomic sequence is highly conserved among mammals, supporting a functional role for this element.[41] Deletion of the 8q24 region encompassing the rs6983267 was found to reduce intestinal tumor multiplicity in $Apc^{Min/+}$ mice.[97] However, the genetic deletion removes not only the DNA enhancer elements, but also the CCAT2 gene, thus allowing for different explanations for the observed phenotypic changes. Our study showing MYC regulation via knockdown approaches suggest that CCAT2 could independently regulate *MYC* transcription. Analysis of colorectal cancer samples showed a correlation of MYC and CCAT2 at the transcriptional level, further providing experimental support for the causal relationship. Most interestingly, overexpression of CCAT2 transforms a chromosomal stable cell line with near-diploid status into a chromosomally unstable one, with a marked increase in polyploidy. This is well in agreement with the high CCAT2 expression levels found in microsatellite stable tumors, often characterized by aneuploidy, when compared with the near-diploid MSI-High colon tumors.[41]
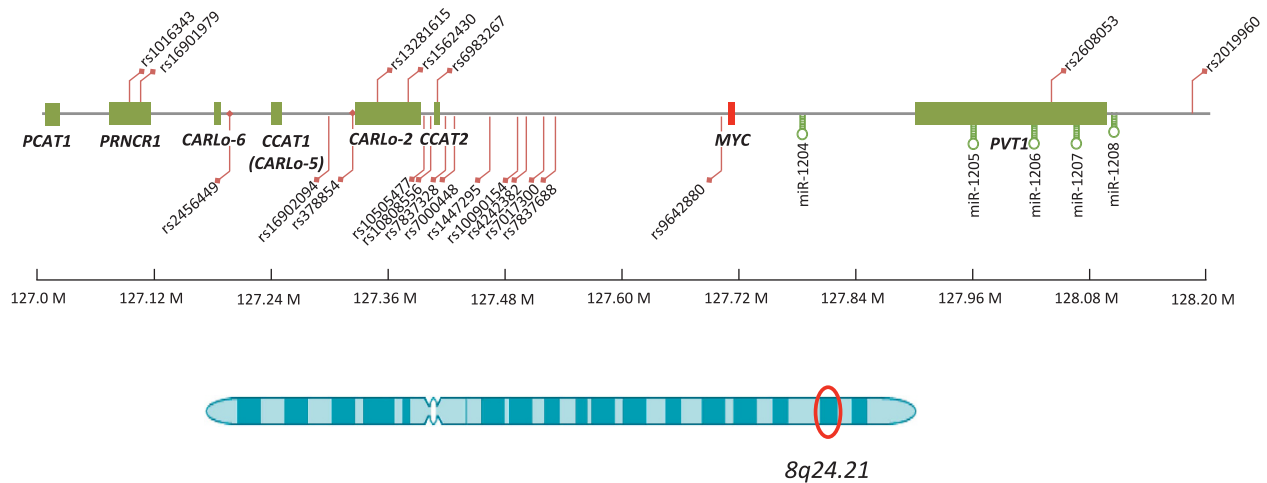
Figure 2. LncRNA and cancer predisposition SNPs on 8q24 genomic region. The 8q24.21 genomic region contains multiple lncRNA genes, located either upstream or downstream of the proto-oncogene *MYC*. Most of them have shown functional involvement in cancer, and some regulate *MYC* expression levels. The same region also features multiple cancer predisposition SNPs, either within or outside of the noncoding gene, suggesting a complex regulation network linking SNPs, lncRNAs and *MYC*.

Although we proved the oncogenic nature of CCAT2 in promoting chromosomal instability and colorectal cancer, whether the rs6983267 SNP variants affect CCAT2 function still remains to be further elucidated. From this perspective, we reported a significant positive correlation between CCAT2 and *MYC* expression in GG samples but not in TT samples of CRCs.[41]

As *MYC* and its regulatory networks have been proposed as one of the most important drivers in colon cancer development (as implicated by the large-scale TCGA project),[98] we hypothesize that a complex regulatory network containing DNA elements (enhancers) and RNA transcripts (lncRNAs) for the *MYC* gene is active in the 8q24 region and acts to fine tune the expression and function of this critical gene. The concept of super enhancers, defined as large clusters of transcriptional enhancers driving gene expression, has also recently surfaced, and points to *MYC* regulation in the 8q24 region as a typical example.[99]

It is also possible that lncRNAs may have fundamental biological effects, independent of MYC transcription, and that these factors together initiate or promote cancer pathogenesis. A genome-wide association approach identified that 75% of the disease-associated SNPs affect expression of lncRNA, but not that of neighboring protein-coding genes.[100] In addition, such effects are tissue-dependent, reflecting regulation of a complex trait.[100] As we learned from PCAT1 and CCAT2, lncRNAs transcribed from the 8q24 locus may affect double-stranded DNA break repair[101] and chromosome instability,[41] which consequently exert a broader biological effect in promoting cancer pathogenesis.

THE CLINICAL RELEVANCE OF THE DNA-RNA TWIST IN CANCER
Many lincRNAs such as ANRIL,[48] HOTAIR,[102] PCAT-1,[88] PRNCR1,[36] PCGEM1,[36] CCAT2,[41] and MALAT1[103] have been shown to associate with human cancer. Recently, XIST, a lncRNA for X-chromosome inactivation, was also shown to suppress hematologic cancer.[104] The abnormal expression profile and functional importance of lncRNAs in cancer suggest translation potential of this knowledge into clinical applications for the cancer patients.

LncRNAs are generally more tissue-specific than protein-coding genes and thus may be more specifically associated with certain cancer subtypes.[6] This tissue-specific expression pattern can possibly enhance the utility of lncRNAs as biomarkers for the early diagnosis of localized cancers from different body fluids, for the detection of cancer metastasis, the prediction of clinical

outcome and/or to reveal the origin of metastatic cancers. For instance, increased MALAT1 expression levels predict metastasis and poor survival in early-stage NSCLC.[105] Likewise, elevated HOTAIR levels are associated with poor prognosis in several cancer types including breast,[102] liver,[106] colorectal,[107] gastrointestinal[108] and pancreatic[109] cancers. A mouse study demonstrated that HOTAIR initiate breast cancer metastases.[102,103] Also, CCAT2 levels in primary tumors showed an inverse correlation with metastasis-free survival of breast cancer patients.[41,110] Furthermore, a bioinformatics study identified 120 individual lncRNAs that are significantly associated with progression-free survival in prostate cancer.[111]

An ideal lncRNA biomarker requires robust detection in plasma and other biofluids such as urine. Although lncRNA stability in such environments remains largely unknown, several studies have suggested the potential of lncRNAs as biomarkers. MALAT1 fragment levels in patient plasma were found to significantly differentiate human subjects with or without prostate cancer.[112] The specific association of PCA3 with prostate cancer has been developed into a FDA-approved commercial Progensa PCA3 assay aiding for the recommendation of repeated prostate biopsies.[113] The finding of lncRNA germline and somatic mutations in leukemia and colorectal cancer[114] suggests that a combined strategy of genotyping the DNA sequence and measurement of lncRNA expression levels may strengthen the disease connection.

The DNA–RNA coordination in determining a specific activity indicates that disruption of either one component could have functional consequences. LncRNAs may represent ideal therapeutic targets. Another attractive feature of lncRNA therapeutics is the capacity to increase protein output in a more natural way, for instance, by targeting NATs. The effect of *cis*-acting NATs may be more focused on a local gene, and potentially such therapy has less off-target effects. Here the clear understanding of the mechanism of lncRNA within its genomic context is the key for such therapeutic development.

CONCLUSION
'One man's junk is another man's treasure'. The recent advances in lncRNA research have revealed transcriptional treasures from the once derided 'junk' DNA regions. Although currently only a small fraction of the lncRNAs have been functionally characterized, we believe that the reservoir of functional lncRNAs will quickly

expand as the result of many emerging technologies for high-throughput screening and functional validation. For instance, studies on protein interaction coupled with the transcriptome data can be greatly facilitated by photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP);[115] genomic occupation sites of lncRNAs can be profiled by ChIRP and subsequent DNA sequencing;[9] functional motifs within RNA can be detected by RNA–mechanically induced trapping of molecular interactions (RNA-MITOMI);[116] RNA movement can be traced by live imaging using engineered fluorescent RNAs.[117] However, because of the extremely large number of lncRNAs in the human genome, it may be more practical to first focus on the disease-associated lncRNAs suggested by other studies such as expression analysis and GWAS findings. These disease-related SNPs can be useful marks to flag functioning lncRNAs. In addition, lncRNAs identified in such regions, either functionally affected or altered in their expression levels by specific SNP variants, may be the culprits underlying the mechanisms of disease predisposition. Elucidation of such mechanisms needs a detailed understanding of lncRNA structure, structure-function relationship and a suitable experimental system to distinguish the subtle differences.

Owing to tissue-specific expression patterns and site-specific action of lncRNAs, drugs targeting lncRNAs could achieve more selective therapeutic effect than conventional drugs. In addition, the allele-specific regulatory mechanisms of lncRNAs may be exploited for precise control of gene expression, presumably with fewer side effects. Synthetic oligonucleotides with high affinity and specificity, such as those with locked nucleic acid modifications, allow for targeted regulation of lncRNA expression. Small molecule chemical compounds showing specificity towards a lncRNA could also be tested as candidates to interrupt lncRNA–protein interaction, or interfere with the lncRNA loading onto its target genomic regions.

The regulatory scheme in human cells is complicated, and it is rare that a single molecule can explain an entire disease phenotype. It can be envisioned that in a specific genomic locus there are intertwined transcripts of many kinds, including protein-coding genes, overlapping intronic and noncoding RNAs in the sense or antisense orientation relative to the protein-coding genes, further complicated by the various isoforms caused by alternative splicing. Thus, a loss or gain of a genomic region, as frequently seen in cancer, will not only affect DNA regulatory elements, but also affect the transcription landscape. This concept can be further expanded to include regulatory circuitry at several genomic loci containing both coding and non-coding genes with reciprocal interactions and feedback loops to determine a disease phenotype. Hence, it is of critical importance to consider the genetic context, including gene locus, neighboring genes, chromatin status and target genomic regions, for a comprehensive functional annotation or therapeutic manipulations in the battle against cancer.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1 Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A et al. Landscape of transcription in human cells. Nature 2012; 489: 101–108.
2 Comings DE. The structure and function of chromatin. Adv Hum Genet 1972; 3: 237–431.
3 Calin GA, Croce CM. MicroRNA signatures in human cancers. Nat Rev Cancer 2006; 6: 857–866.
4 Ling H, Fabbri M, Calin GA. MicroRNAs and other non-coding RNAs as targets for anticancer drug development. Nat Rev Drug Discov 2013; 12: 847–865.
5 Esteller M. Non-coding RNAs in human disease. Nat Rev Genet 2011; 12: 861–874.
6 Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 2012; 22: 1775–1789.
7 Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. Mol Cell 2011; 43: 904–914.
8 Schmitz KM, Mayer C, Postepska A, Grummt I. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Gene Dev 2010; 24: 2264–2269.
9 Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell 2011; 44: 667–678.
10 Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci USA 2009; 106: 9362–9367.
11 Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell 2011; 147: 1537–1550.
12 Fatica A, Bozzoni I. Long non-coding RNAs: new players in cell differentiation and development. Nat Rev Genet 2014; 15: 7–21.
13 Batista PJ, Chang HY. Long noncoding RNAs: cellular address codes in development and disease. Cell 2013; 152: 1298–1307.
14 Wapinski O, Chang HY. Long noncoding RNAs and human disease. Trends Cell Biol 2011; 21: 354–361.
15 Mercer TR, Mattick JS. Structure and function of long noncoding RNAs in epigenetic regulation. Nat Struct Mol Biol 2013; 20: 300–307.
16 Geisler S, Coller J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. Nat Rev Mol Cell Biol 2013; 14: 699–712.
17 Yang L, Froberg JE, Lee JT. Long noncoding RNAs: fresh perspectives into the RNA world. Trends Biochem Sci 2014; 39: 35–43.
18 Cech TR, Steitz JA. The noncoding RNA revolution-trashing old rules to forge new ones. Cell 2014; 157: 77–94.
19 Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. Cell 2013; 154: 26–46.
20 Lee JT, Bartolomei MS. X-inactivation, imprinting, and long noncoding RNAs in health and disease. Cell 2013; 152: 1308–1323.
21 Kim HS, Minna JD, White MA. GWAS meets TCGA to illuminate mechanisms of cancer predisposition. Cell 2013; 152: 387–389.
22 Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E et al. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci USA 2002; 99: 15524–15529.
23 Gong C, Maquat LE. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. Nature 2011; 470: 284–288.
24 Mariner PD, Walters RD, Espinoza CA, Drullinger LF, Wagner SD, Kugel JF et al. Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. Mol Cell 2008; 29: 499–509.
25 Levine M, Tjian R. Transcription regulation and animal diversity. Nature 2003; 424: 147–151.
26 Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 2008; 322: 1845–1848.
27 Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA et al. Divergent transcription from active promoters. Science 2008; 322: 1849–1851.
28 Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD et al. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nat Genet 2011; 43: 621–629.
29 Blackwood EM, Kadonaga JT. Going the distance: a current view of enhancer action. Science 1998; 281: 60–63.
30 Orom UA, Shiekhattar R. Long noncoding RNAs usher in a new era in the biology of enhancers. Cell 2013; 154: 1190–1193.

31 Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J *et al*. Widespread transcription at neuronal activity-regulated enhancers. *Nature* 2010; **465**: 182–187.

32 Orom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G *et al*. Long noncoding RNAs with enhancer-like function in human cells. *Cell* 2010; **143**: 46–58.

33 Koch F, Fenouil R, Gut M, Cauchy P, Albert TK, Zacarias-Cabeza J *et al*. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat Struct Mol Biol* 2011; **18**: 956–963.

34 Feng J, Bi C, Clark BS, Mady R, Shah P, Kohtz JD. The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Gene Dev* 2006; **20**: 1470–1484.

35 Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y *et al*. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 2011; **472**: 120–124.

36 Yang L, Lin C, Jin C, Yang JC, Tanasa B, Li W *et al*. lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. *Nature* 2013; **500**: 598–602.

37 Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS *et al*. Ultraconserved elements in the human genome. *Science* 2004; **304**: 1321–1325.

38 Calin GA, Liu CG, Ferracin M, Hyslop T, Spizzo R, Sevignani C *et al*. Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell* 2007; **12**: 215–229.

39 Braconi C, Valeri N, Kogure T, Gasparini P, Huang N, Nuovo GJ *et al*. Expression and functional role of a transcribed noncoding RNA with an ultraconserved element in hepatocellular carcinoma. *Proc Natl Acad Sci USA* 2011; **108**: 786–791.

40 Ferdin J, Nishida N, Wu X, Nicoloso MS, Shah MY, Devlin C *et al*. HINCUTs in cancer: hypoxia-induced noncoding ultraconserved transcripts. *Cell Death Differ* 2013; **20**: 1675–1687.

41 Ling H, Spizzo R, Atlasi Y, Nicoloso M, Shimizu M, Redis RS *et al*. CCAT2, a novel noncoding RNA mapping to 8q24, underlies metastatic progression and chromosomal instability in colon cancer. *Genome Res* 2013; **23**: 1446–1461.

42 Liz J, Portela A, Soler M, Gomez A, Ling H, Michlewski G *et al*. Regulation of pri-miRNA processing by a long noncoding RNA transcribed from an ultraconserved region. *Mol Cell* 2014; **55**: 138–147.

43 Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M *et al*. Antisense transcription in the mammalian transcriptome. *Science* 2005; **309**: 1564–1566.

44 Wang XJ, Gaasterland T, Chua NH. Genome-wide prediction and identification of cis-natural antisense transcripts in Arabidopsis thaliana. *Genome Biol* 2005; **6**: R30.

45 Faghihi MA, Wahlestedt C. Regulatory roles of natural antisense transcripts. *Nat Rev Mol Cell Biol* 2009; **10**: 637–643.

46 Yu W, Gius D, Onyango P, Muldoon-Jacobs K, Karp J, Feinberg AP *et al*. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* 2008; **451**: 202–206.

47 Pasmant E, Laurendeau I, Heron D, Vidaud M, Vidaud D, Bieche I. Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. *Cancer Res* 2007; **67**: 3963–3969.

48 Yap KL, Li S, Munoz-Cabello AM, Raguz S, Zeng L, Mujtaba S *et al*. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol Cell* 2010; **38**: 662–674.

49 Beltran M, Puig I, Pena C, Garcia JM, Alvarez AB, Pena R *et al*. A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. *Gene Dev* 2008; **22**: 756–769.

50 Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA *et al*. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 2007; **129**: 1311–1323.

51 Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F *et al*. Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 2010; **329**: 689–693.

52 Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D *et al*. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* 2010; **142**: 409–419.

53 Lee JT. Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Gene Dev* 2009; **23**: 1831–1842.

54 Bertani S, Sauer S, Bolotin E, Sauer F. The noncoding RNA Mistral activates Hoxa6 and Hoxa7 expression and stem cell differentiation by recruiting MLL1 to chromatin. *Mol Cell* 2011; **43**: 1040–1046.

55 Di Ruscio A, Ebralidze AK, Benoukraf T, Amabile G, Goff LA, Terragni J *et al*. DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* 2013; **503**: 371–376.

56 Jeon Y, Lee JT. YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* 2011; **146**: 119–133.

57 Simon MD, Pinter SF, Fang R, Sarma K, Rutenberg-Schoenberg M, Bowman SK *et al*. High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature* 2013; **504**: 465–469.

58 Maenner S, Blaud M, Fouillen L, Savoye A, Marchand V, Dubois A *et al*. 2-D structure of the A region of Xist RNA and its implication for PRC2 association. *PLoS Biol* 2010; **8**: e1000276.

59 Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R *et al*. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* 2008; **322**: 1717–1720.

60 Lee JT, Davidow LS, Warshawsky D. Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat Genet* 1999; **21**: 400–404.

61 Sun BK, Deaton AM, Lee JT. A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. *Mol Cell* 2006; **21**: 617–628.

62 Ogawa Y, Sun BK, Lee JT. Intersection of the RNA interference and X-inactivation pathways. *Science* 2008; **320**: 1336–1341.

63 Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA *et al*. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* 2013; **494**: 497–501.

64 Melo CA, Drost J, Wijchers PJ, van de Werken H, de Wit E, Oude Vrielink JA *et al*. eRNAs are required for p53-dependent enhancer activity and gene transcription. *Mol Cell* 2013; **49**: 524–535.

65 Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D *et al*. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci USA* 2009; **106**: 11667–11672.

66 Sanford JR, Wang X, Mort M, Vanduyn N, Cooper DN, Mooney SD *et al*. Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res* 2009; **19**: 381–394.

67 Zhang X, Rice K, Wang Y, Chen W, Zhong Y, Nakayama Y *et al*. Maternally expressed gene 3 (MEG3) noncoding ribonucleic acid: isoform structure, expression, and functions. *Endocrinology* 2010; **151**: 939–947.

68 Novikova IV, Hennelly SP, Sanbonmatsu KY. Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Res* 2012; **40**: 5034–5051.

69 Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA, Wang H *et al*. Role of histone H3 lysine 27 methylation in X inactivation. *Science* 2003; **300**: 131–135.

70 Allen TA, Von Kaenel S, Goodrich JA, Kugel JF. The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. *Nat Struct Mol Biol* 2004; **11**: 816–821.

71 Yakovchuk P, Goodrich JA, Kugel JF. B2 RNA and Alu RNA repress transcription by disrupting contacts between RNA polymerase II and promoter DNA within assembled complexes. *Proc Natl Acad Sci USA* 2009; **106**: 5569–5574.

72 Ning S, Zhao Z, Ye J, Wang P, Zhi H, Li R *et al*. LincSNP: a database of linking disease-associated SNPs to human large intergenic non-coding RNAs. *BMC Bioinformatics* 2014; **15**: 152.

73 Pasmant E, Sabbagh A, Vidaud M, Bieche I. ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. *FASEB J* 2011; **25**: 444–448.

74 Pasmant E, Sabbagh A, Masliah-Planchon J, Ortonne N, Laurendeau I, Melin L *et al*. Role of noncoding RNA ANRIL in genesis of plexiform neurofibromas in neurofibromatosis type 1. *J Natil Cancer Inst* 2011; **103**: 1713–1722.

75 Matouk IJ, DeGroot N, Mezan S, Ayesh S, Abu-lail R, Hochberg A *et al*. The H19 non-coding RNA is essential for human tumor growth. *PloS one* 2007; **2**: e845.

76 Zhang L, Yang F, Yuan JH, Yuan SX, Zhou WP, Huo XS *et al*. Epigenetic activation of the MiR-200 family contributes to H19-mediated metastasis suppression in hepatocellular carcinoma. *Carcinogenesis* 2013; **34**: 577–586.

77 Verhaegh GW, Verkleij L, Vermeulen SH, den Heijer M, Witjes JA, Kiemeney LA. Polymorphisms in the H19 gene and the risk of bladder cancer. *Eur Urol* 2008; **54**: 1118–1126.

78 Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG *et al*. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 2007; **447**: 1087–1093.

79 Liu Y, Pan S, Liu L, Zhai X, Liu J, Wen J *et al*. A genetic variant in long non-coding RNA HULC contributes to risk of HBV-related hepatocellular carcinoma in a Chinese population. *PloS one* 2012; **7**: e35145.

80 Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F *et al*. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 1992; **258**: 818–821.

81 Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J *et al*. The landscape of somatic copy-number alteration across human cancers. *Nature* 2010; **463**: 899–905.

82 Sur I, Tuupanen S, Whitington T, Aaltonen LA, Taipale J. Lessons from functional analysis of genome-wide association studies. *Cancer Res* 2013; **73**: 4180–4184.

83 Huppi K, Pitt JJ, Wahlberg BM, Caplen NJ. The 8q24 gene desert: an oasis of non-coding transcriptional activity. *Front Genet* 2012; **3**: 69.

84 Jia L, Landan G, Pomerantz M, Jaschek R, Herman P, Reich D et al. Functional enhancers at the gene-poor 8q24 cancer-linked locus. *PLoS Genet* 2009; **5**: e1000597.

85 Xiang JF, Yin QF, Chen T, Zhang Y, Zhang XO, Wu Z et al. Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res* 2014; **24**: 513–531.

86 Kim T, Cui R, Jeon YJ, Lee JH, Lee JH, Sim H et al. Long-range interaction and correlation between MYC enhancer and oncogenic long noncoding RNA CARLo-5. *Proc Natl Acad Sci USA* 2014; **111**: 4173–4178.

87 Tseng YY, Moriarity BS, Gong W, Akiyama R, Tiwari A, Kawakami H et al. PVT1 dependence in cancer with MYC copy-number increase. *Nature* 2014; **512**: 82–86.

88 Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC et al. Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat Biotechnol* 2011; **29**: 742–749.

89 Li L, Sun R, Liang Y, Pan X, Li Z, Bai P et al. Association between polymorphisms in long non-coding RNA PRNCR1 in 8q24 and risk of colorectal cancer. *J Exp Clin Cancer Res* 2013; **32**: 104.

90 Gaudet MM, Kirchhoff T, Green T, Vijai J, Korn JM, Guiducci C et al. Common genetic variants and modification of penetrance of BRCA2-associated breast cancer. *PLoS Genet* 2010; **6**: e1001183.

91 Eeles RA, Kote-Jarai Z, Giles GG, Olama AA, Guy M, Jugurnauth SK et al. Multiple newly identified loci associated with prostate cancer susceptibility. *Nat Genet* 2008; **40**: 316–321.

92 Chung S, Nakagawa H, Uemura M, Piao L, Ashikawa K, Hosono N et al. Association of a novel long non-coding RNA in 8q24 with prostate cancer susceptibility. *Cancer Sci* 2011; **102**: 245–252.

93 Tuupanen S, Turunen M, Lehtonen R, Hallikas O, Vanharanta S, Kivioja T et al. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat Genet* 2009; **41**: 885–890.

94 Pomerantz MM, Ahmadiyeh N, Jia L, Herman P, Verzi MP, Doddapaneni H et al. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat Genet* 2009; **41**: 882–884.

95 Ghoussaini M, Song H, Koessler T, Al Olama AA, Kote-Jarai Z, Driver KE et al. Multiple loci with different cancer specificities within the 8q24 gene desert. *J Natl Cancer Inst* 2008; **100**: 962–966.

96 Bertucci F, Lagarde A, Ferrari A, Finetti P, Charafe-Jauffret E, Van Laere S et al. 8q24 Cancer risk allele associated with major metastatic risk in inflammatory breast cancer. *PLoS one* 2012; **7**: e37943.

97 Sur IK, Hallikas O, Vaharautio A, Yan J, Turunen M, Enge M et al. Mice lacking a Myc enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. *Science* 2012; **338**: 1360–1363.

98 Cancer Genome Atlas N, Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012; **487**: 330–337.

99 Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA et al. Super-enhancers in the control of cell identity and disease. *Cell* 2013; **155**: 934–947.

100 Kumar V, Westra HJ, Karjalainen J, Zhernakova DV, Esko T, Hrdlickova B et al. Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. *PLoS Genet* 2013; **9**: e1003201.

101 Prensner JR, Chen W, Iyer MK, Cao Q, Ma T, Han S et al. PCAT-1, a long noncoding RNA, regulates BRCA2 and controls homologous recombination in cancer. *Cancer Res* 2014; **74**: 1651–1660.

102 Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 2010; **464**: 1071–1076.

103 Gutschner T, Diederichs S. The hallmarks of cancer: a long non-coding RNA point of view. *RNA Biol* 2012; **9**: 703–719.

104 Yildirim E, Kirby JE, Brown DE, Mercier FE, Sadreyev RI, Scadden DT et al. Xist RNA is a potent suppressor of hematologic cancer in mice. *Cell* 2013; **152**: 727–742.

105 Ji P, Diederichs S, Wang W, Boing S, Metzger R, Schneider PM et al. MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* 2003; **22**: 8031–8041.

106 Yang Z, Zhou L, Wu LM, Lai MC, Xie HY, Zhang F et al. Overexpression of long non-coding RNA HOTAIR predicts tumor recurrence in hepatocellular carcinoma patients following liver transplantation. *Ann Surgical Oncol* 2011; **18**: 1243–1250.

107 Kogo R, Shimamura T, Mimori K, Kawahara K, Imoto S, Sudo T et al. Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res* 2011; **71**: 6320–6326.

108 Niinuma T, Suzuki H, Nojima M, Nosho K, Yamamoto H, Takamaru H et al. Upregulation of miR-196a and HOTAIR drive malignant character in gastro-intestinal stromal tumors. *Cancer Res* 2012; **72**: 1126–1136.

109 Kim K, Jutooru I, Chadalapaka G, Johnson G, Frank J, Burghardt R et al. HOTAIR is a negative prognostic factor and exhibits pro-oncogenic activity in pancreatic cancer. *Oncogene* 2013; **32**: 1616–1625.

110 Redis RS, Sieuwerts AM, Look MP, Tudoran O, Ivan C, Spizzo R et al. CCAT2, a novel long non-coding RNA in breast cancer: expression study and clinical correlations. *Oncotarget* 2013; **4**: 1748–1762.

111 Du Z, Fei T, Verhaak RG, Su Z, Zhang Y, Brown M et al. Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. *Nat Struct Mol Biol* 2013; **20**: 908–913.

112 Ren S, Wang F, Shen J, Sun Y, Xu W, Lu J et al. Long non-coding RNA metastasis associated in lung adenocarcinoma transcript 1 derived miniRNA as a novel plasma-based biomarker for diagnosing prostate cancer. *Eur J Cancer* 2013; **49**: 2949–2959.

113 Rittenhouse H, Blase A, Shamel B, Schalken J, Groskopf J. The long and winding road to FDA approval of a novel prostate cancer test: our story. *Clin Chem* 2013; **59**: 32–34.

114 Wojcik SE, Rossi S, Shimizu M, Nicoloso MS, Cimmino A, Alder H et al. Non-codingRNA sequence variations in human chronic lymphocytic leukemia and colorectal cancer. *Carcinogenesis* 2010; **31**: 208–215.

115 Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 2010; **141**: 129–141.

116 Martin L, Meier M, Lyons SM, Sit RV, Marzluff WF, Quake SR et al. Systematic reconstruction of RNA functional motifs with high-throughput microfluidics. *Nat Methods* 2012; **9**: 1192–1194.

117 Paige JS, Wu KY, Jaffrey SR. RNA mimics of green fluorescent protein. *Science* 2011; **333**: 642–646.

118 Srikantan V, Zou Z, Petrovics G, Xu L, Augustus M, Davis L et al. PCGEM1, a prostate-specific gene, is overexpressed in prostate cancer. *Proc Natl Acad Sci USA* 2000; **97**: 12216–12221.

119 Fu X, Ravindranath L, Tran N, Petrovics G, Srivastava S. Regulation of apoptosis by a prostate-specific and prostate cancer-associated noncoding gene, PCGEM1. *DNA Cell Biol* 2006; **25**: 135–141.

120 Xue Y, Wang M, Kang M, Wang Q, Wu B, Chu H et al. Association between lncrna PCGEM1 polymorphisms and prostate cancer risk. *Prost Cancer Prost Dis* 2013; **16**: 139–144, S131.

121 Du Y, Kong G, You X, Zhang S, Zhang T, Gao Y et al. Elevation of highly up-regulated in liver cancer (HULC) by hepatitis B virus X protein promotes hepatoma cell proliferation via down-regulating p18. *J Biol Chem* 2012; **287**: 26302–26311.

122 Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet* 2007; **39**: 984–988.

123 Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet* 2007; **39**: 645–649.

124 Chen D, Zhang Z, Mao C, Zhou Y, Yu L, Yin Y et al. ANRIL inhibits p15(INK4b) through the TGFbeta1 signaling pathway in human esophageal squamous cell carcinoma. *Cell Immunol* 2014; **289**: 91–96.

125 Gutschner T, Hammerle M, Eissmann M, Hsu J, Kim Y, Hung G et al. The non-coding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res* 2013; **73**: 1180–1189.