

Expression Specificity of Disease-Associated lncRNAs: Toward Personalized Medicine

Quan Nguyen and Piero Carninci

Abstract Long noncoding RNAs (lncRNAs) perform diverse regulatory functions in transcription, translation, chromatin modification, and cellular organization. Misregulation of lncRNAs is found linked to various human diseases. Compared to protein-coding RNAs, lncRNAs are more specific to organs, tissues, cell types, developmental stages, and disease conditions, making them promising candidates as diagnostic and prognostic biomarkers and as gene therapy targets. The functional annotation of mammalian genome (FANTOM) consortium utilizes cap analysis of gene expression (CAGE) method to quantify genome-wide activities of promoters and enhancers of coding and noncoding RNAs across a large collection of human and mouse tissues, cell types, diseases, and time-courses. The project discovered widespread transcription of major lncRNA classes, including lncRNAs derived from enhancers, bidirectional promoters, antisense lncRNAs, and repetitive elements. Results from FANTOM project enable assessment of lncRNA expression specificity across tissue and disease conditions, based on differential promoter and enhancer usage. More than 85 % of disease-related SNPs are within noncoding regions and are strikingly overrepresented in enhancer and promoter regions, suggestive of the importance of lncRNA loci at these SNP harboring regions to human diseases. In this chapter, we discuss lncRNA expression specificity, review diverse functions of disease-associated lncRNAs, and present perspectives on their potential therapeutic applications for personalized medicine. The future development of lncRNA applications relies on technologies to identify and validate their functions, structures, and mechanisms. Comprehensive understanding of genome-wide interaction networks of lncRNAs with proteins, chromatin, and other RNAs in regulating cellular processes will allow personalized medicine to use lncRNAs as highly specific biomarkers in diagnosis, prognosis, and therapeutic targets.

Q. Nguyen · P. Carninci (✉)

Division of Genomic Technologies, RIKEN Yokohama Campus, RIKEN Center for Life Science Technologies, 1-7-22 Suehiro-Cho, Tsurumi-Ku, Yokohama City, Kanagawa 230-0045, Japan
e-mail: carninci@riken.jp

Current Topics in Microbiology and Immunology (2016) 394: 237–258

DOI 10.1007/82_2015_464

© Springer International Publishing Switzerland 2015

Published Online: 30 August 2015

Abbreviations

CAGE	Cap analysis of gene expression
ChIP	Chromatin immunoprecipitation
ENCODE	Encyclopedia of DNA elements
eRNAs	Enhancer RNAs
FANTOM	Functional annotation of mammalian genome
GWAS	Genome-wide association study
lncRNAs	Long noncoding RNAs
miRNAs	MicroRNAs
TSS	Transcription start sites
RNA-seq	RNA sequencing

Contents

1	Introduction	239
1.1	Widespread Transcription of LncRNA and LncRNA Annotation	239
1.2	Diverse Functions and Mechanisms of Action of LncRNAs	240
1.3	LncRNA Expression Specificity	241
2	LncRNAs Derived from Enhancers, Promoters, and Repetitive Elements—Insights from FANTOM Promoter Analysis	241
2.1	A Promoter-Centric Approach to Identifying and Characterizing LncRNA Transcription	241
2.2	LncRNA Expression Specificity: Insights from Quantified Human Promoterome	243
2.3	Cell-Specific Enhancer Usage	244
2.4	LncRNAs Derived from Repetitive Elements	245
3	LncRNAs and Human Diseases	246
3.1	LncRNAs with Strong Links to Human Diseases	246
3.2	Known LncRNAs from Enhancers and Repetitive Elements with Significant Association to Human Diseases	248
3.3	LncRNAs as Biomarkers for Diagnosis, Prognosis and as Targets for Gene Therapy	249
3.4	Strategies for Perturbation of Disease-Associated LncRNAs	250
4	Perspectives	251
4.1	Studies of LncRNA Structure and Their Interactions with RNA, DNA, and Proteins	251
4.2	Personalized Medicine	252
5	Conclusion	253
	References	254

1 Introduction

1.1 *Widespread Transcription of LncRNA and LncRNA Annotation*

A decade ago, genome-wide expression studies in mammals, using high-density tiling array or massive sequencing of full-length complementary DNAs, identified widespread transcription of tens of thousands of long non-protein-coding RNAs (lncRNAs), which are commonly defined as RNAs longer than 200 nucleotides, and have no coding potential for long peptides (Carninci et al. 2005; Bertone et al. 2004). With a rapid increase in sequencing depth and in the number of tissues, cell types, and organisms being sequenced, the list of lncRNAs is steadily growing. The ENCODE project (ENCyclopedia Of DNA Elements) estimated that 80.4 % of the human genome participates in at least one biochemical event in at least one cell type (Bernstein et al. 2012) and at least 74.7 % is transcribed (Djebali et al. 2012). Assembled transcripts from over 4.5 billion uniquely mapped reads in an RNA-seq data set of 23 human tissues under multiple conditions, with additional annotation information of known spliced expressed sequenced tags (ESTs), cDNAs and genes, could be mapped to 85.2 % of the genome (Hangauer et al. 2013). A recent ab initio assembly of over 43 Tb of sequence from 7256 RNA-seq libraries shows that there are at least two times more lncRNAs (68 %) than coding RNAs and the majority of these lncRNAs (79 %) were so far unannotated (Iyer et al. 2015).

The annotation of reproducible lncRNA loci (also called lncRNA genes in GENCODE) is rapidly growing to a comparable number to that of coding genes. In GENCODE annotation, the number of lncRNA loci increased from 9277 in 2012 (Derrien et al. 2012) to 15,900 lncRNA loci (version 22, October 2014). Current continuous attempts to annotate lncRNAs have produced several large systemic databases of lncRNAs such as lncRNAtor (Park et al. 2014), LNCipedia (Volders et al. 2015), and NONCODE (Xie et al. 2013). The numbers of lncRNA transcripts from these databases differ, largely subjective to criteria used for defining lncRNAs, sequencing depth, and diversity of sequenced samples. For example, while there are 56,018 human lncRNA loci in NONCODE version 4.0, the lncRNAtor version 1.0 contains 14,051 lncRNA gene units. Current unpublished analysis of RNA-seq and CAGE data in FANTOM5 suggests transcription of 45–50,000 lncRNA loci in the human genome (Unpublished results).

The debate on whether or not lncRNAs are the products of transcription noise is moving toward how lncRNAs produce functions (Pennisi 2014). Many lncRNAs possess key features distinct from transcription noise, including active transcription regulation as indicated in high frequency and conservation of transcription factor binding sites in lncRNA promoters, high precision of transcription in splicing processes, a similar range of half-lives, higher sequence conservation than random intergenic regions, lncRNA expression profiles associated with profiles of mRNA, and increasing evidence of specific lncRNA biological functions especially those related to diseases (Derrien et al. 2012; Pennisi 2014; Necsulea et al. 2014; Guttman

et al. 2009; Mercer et al. 2008). However, current knowledge of lncRNA functions remains limited to approximately 287 functionally characterized lncRNA loci, as documented in the lncRNADB database (Quek et al. 2014).

1.2 Diverse Functions and Mechanisms of Action of LncRNAs

LncRNAs utilize diverse mechanisms of action by interacting with most types of binding partners, including RNA, DNA, and proteins. These interactions may vary according to cell types, tissues, and organs. Comprehensive reviews on lncRNA functions and mechanisms are available elsewhere (Morris and Mattick 2014; Li and Chang 2014). In this section, we exemplify potential links between regulatory functions of lncRNAs in transcription, translation, chromatin modification, and cellular organization to diseases.

For transcriptional regulation, lncRNAs act either on *cis* or *trans* by forming complex with proteins to regulate expression, often by controlling chromatin states (Rinn and Chang 2012). The X-inactive-specific transcript (Xist) recruits Polycomb repressive complex (PRC2) to inactivate X chromosome, and deletion of Xist resulted in X chromosome reactivation, promoting hematologic cancer in mice (Yildirim et al. 2013). Linc1992 THRIL [TNF α and Heterogeneous nuclear ribonucleoprotein L (hnRNPL) Related Immunoregulatory LincRNA] binds to hnRNPL and TNF α promoter to upregulate TNF α transcription (Li et al. 2014). LncRNA CCAT1-L (Colon cancer-associated transcript) is abundant in colorectal cancer cell lines or patients' mucosa samples, whereas it is undetectable or lowly expressed in other cell types or control samples (Xiang et al. 2014). CCAT1-L regulates long-range chromatin looping between MYC (v-myc avian myelocytomatosis viral oncogene homolog) promoter and enhancers to increase MYC expression and stimulates tumorigenesis (Xiang et al. 2014). LncRNA can also regulate expression at translational level. The antisense lncRNA of the ubiquitin carboxyl terminal hydrolase gene (Uchl1) is exported to the cytoplasm and facilitates the binding of the sense Uchl1mRNA to active polysomes for more efficient translation (Carrieri et al. 2012). The Uchl1 protein is a potential therapeutic target for treatment of Parkinson's disease (Liu et al. 2002). Several lncRNAs are known as essential organizing factors of the nucleus, for example, NEAT1 (Nuclear-enriched abundant transcript 1), and Xist and Malat1 (metastasis-associated lung adenocarcinoma transcript 1) (Hirose et al. 2014; Rinn and Guttman 2014; Mao et al. 2011). H19 lncRNA and miR-675 are derived from a same genomic locus and both contribute to gastric cancer, but by two different pathways to suppress two different genes, namely ISM1 (isthmin 1, angiogenesis inhibitor) and CALN1 (calneuron 1) (Li et al. 2014). LncRNAs also act as regulators of signaling pathways. For example, downregulation of the lncRNA low expression in tumor (LET) inversely affects expression and stability of genes in hypoxia signaling network, which may contribute to hepatocellular carcinoma (HCC) metastasis (Yang et al. 2013). Leukemia-induced noncoding activator RNA (LUNAR) enhances

IGF1R (insulin-like growth receptor factor 1) expression, contributing to maintenance of the IGF1 pathway and promoting T cell acute lymphoblastic leukemia (T-ALL) growth (Trimarchi et al. 2014).

1.3 LncRNA Expression Specificity

lncRNAs are highly specific to cell type, organs, and species. An RNA-seq study on 15 cell lines showed that 29 % of all lncRNAs were transcribed in only one cell type, and that only 10 % expressed in all cell lines, whereas 53 % of protein-coding mRNAs were constitutively transcribed in all cell lines (Djebali et al. 2012). The majority of lncRNAs are expressed at low levels spanning five to six orders of magnitude, from 10^{-2} to 10^3 Reads Per Kilobase per Million mapped reads (rpkm) for non-polyadenylated lncRNAs or 10^{-2} to 10^4 rpkm for polyadenylated lncRNAs (Djebali et al. 2012). lncRNAs are organ-specific, with the highest number of lncRNAs distributed in testes (55 % for young lncRNAs and 46 % for old lncRNAs), followed by neural and liver tissues (Necsulea et al. 2014). In previous studies, in the mouse, lncRNAs were also abundantly detected during early development and organogenesis (Carninci et al. 2003), which cannot be extensively sampled in human. In the GENCODEv7 database, at least 30 % of lncRNAs are found transcribed only within primate lineage (Derrien et al. 2012). A de novo study using RNA-sequencing data of 11 tetrapod species showed that in a combined data set containing 13,533 lncRNAs, commonly detected in at least three out of 11 species, 81 % were primate-specific (Necsulea et al. 2014). It is evident that the level of sequence conservation for lncRNA primary sequences is low (Guttman et al. 2009; Yue et al. 2014; Fort et al. 2014). For example, only approximately 15 % of human lncRNAs have homologs in mouse (Yue et al. 2014). Interestingly, in human, significant evidence of purifying selection for SNPs within lncRNAs was obtained, suggestive of lineage-specific functions of human lncRNAs (Ward and Kellis 2012). Although still in its infancy, lncRNAs structural studies produce emerging evidence for a high level of secondary structural conservation of lncRNA functional domains, such as those in MEG3, SRA1, and HOTAIR (Mercer and Mattick 2013).

2 LncRNAs Derived from Enhancers, Promoters, and Repetitive Elements—Insights from FANTOM Promoter Analysis

2.1 A Promoter-Centric Approach to Identifying and Characterizing LncRNA Transcription

In the FANTOM project, CAGE sequencing of 5' capped RNAs at transcription start sites (TSS) was used to generate expression data for activity analysis of

enhancers and promoters of both coding and noncoding RNAs (Takahashi et al. 2012; Kanamori-Katayama et al. 2011). CAGE differential expression analysis at single nucleotide level enables discovery of alternative promoter usage in different tissues and developmental stages (Carninci et al. 2006; Haberer et al. 2014). Quantitative expression of promoter regions from CAGE data allows ab initio motif activity response analysis of transcription factors, allowing construction of transcriptional network (Suzuki et al. 2009; Akalin et al. 2009). An example of such study is the comprehensive regulatory circuitry in differentiation and growth arrest of human monocytic cell line using CAGE time-course data (Suzuki et al. 2009).

Promoter-centric analysis of capped RNAs has identified and characterized expression of major classes of lncRNAs (Fig. 1). FANTOM3 discovered

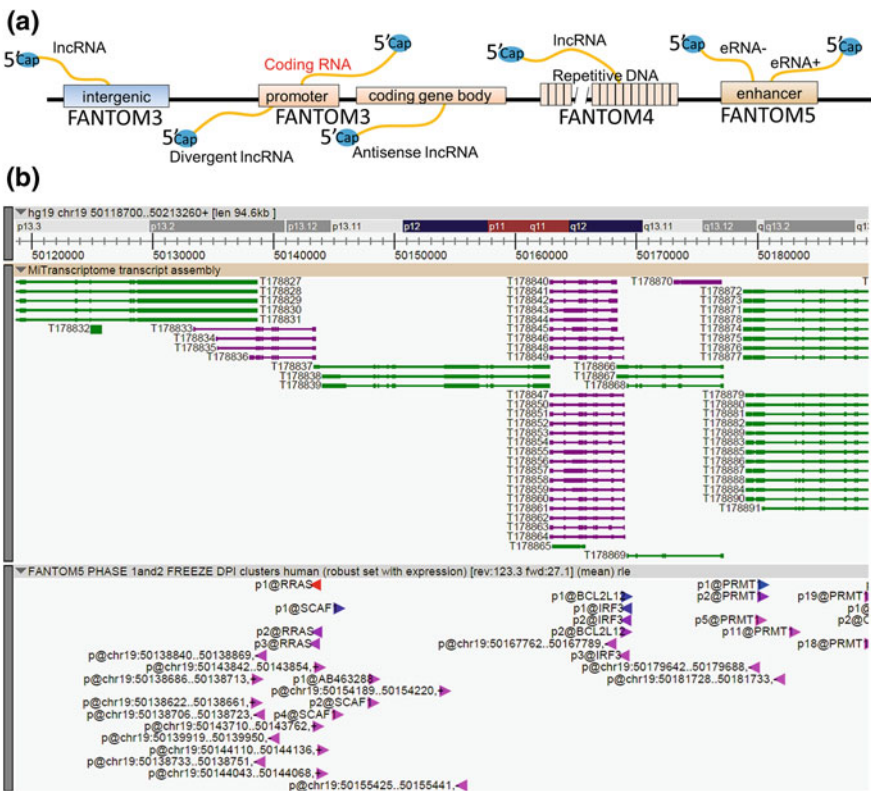


Fig. 1 a CAGE technology applied in FANTOM project (<http://fantom.gsc.riken.jp/5/>) leads to detection of several large classes of lncRNA, including lncRNAs derived from promoters, enhancers, repetitive elements, and antisense RNAs. b CAGE sequencing quantifies directional transcription of coding/noncoding genes at different promoters. CAGE data were from FANTOM5 human pooled robust promoter data set (<http://fantom.gsc.riken.jp/5/data/>). CAGE clusters are shown as *directional arrows*, with *names* and *colors* representing different promoter clustering levels. Reference transcripts (green bars for sense and purple bar for antisense transcripts) were from MiTranscriptome database (Iyer et al. 2015) using ZENBU for visualization (Severin et al. 2014)

genome-wide transcription events of lncRNAs (Carninci et al. 2005), especially the widespread production of antisense RNAs (Katayama et al. 2005). Following this success, FANTOM4 revealed at least 6–30 % of the capped transcripts detected in a sequencing library were transcribed from repetitive regions in human and mouse genome (Faulkner et al. 2009). Further, FANTOM5 comprehensively studied transcripts derived from regulatory elements, resulting in the most comprehensive promoter atlas and enhancer atlas (Forrest et al. 2014; Andersson et al. 2014). Moreover, from investigating enhancer and promoter activities of 19 human and 14 mouse time-course differentiation samples, FANTOM5 found a temporally coordinated transcription pattern of enhancers, promoters, and transcription factors (Arner et al. 2015). In this cascade of time-course transcription under different stimuli, enhancer RNAs were first transcribed, followed by transcription of regulatory genes, then by other responsive genes.

2.2 *LncRNA Expression Specificity: Insights from Quantified Human Promoterome*

lncRNAs are highly developmental and tissue-specific and are not readily detected in many biological cell lines and tissues. From sequencing of 975 human samples including 573 primary cell samples, 152 postmortem tissues, and 250 cancer cell lines, the FANTOM5 project produced a comprehensive human cell-type-specific promoterome atlas, covering activity of 185,000 promoters, which represent promoters of 91–94 % coding and noncoding known genes (Forrest et al. 2014). Across this large collection of tissues and cell types, 20 % of promoters were found expressed in more than 50 % of all samples (ubiquitously expressed), while 80 % were considered cell-type-specific (expressed in fewer than 50 % of all sequenced samples). The large FANTOM5 promoter expression data set with a diverse collection of samples from various diseases enables sample ontology enrichment analysis (SOEA), which associates promoters to disease ontology terms (Forrest et al. 2014). For each CAGE promoter, SOEA tests the overrepresentation of disease ontology terms in a ranked list of samples based on expression of the promoter. Applying this approach for all 127,645 human CAGE peaks revealed that a large proportion of transcribed RNAs in human are enriched in immune system, especially in monocytes and bone marrows (Forrest et al. 2014). A similar sample set enrichment analysis (SSEA) approach was performed at transcript level using data from 7256 RNA-seq libraries identified 7942 lncRNAs stringently associated with cancer or cell lineage specificity or both (Iyer et al. 2015).

Through quantitative promoter analysis of different tissues in mouse and human, Carninci et al. (2006) first revealed that alternative promoter usage is common and that differential promoter usage is tissue-specific. More and more studies have shown relevance of this phenomenon to diseases and development. Alternative transcription start sites were found associated with colorectal tumors (Thorsen et al. 2011),

or in macrophage responses to activating reagents (Carninci et al. 2006), or in early zebra fish embryonic development (Haberle et al. 2014). At genome-wide level, the promoterome atlas shows that approximately 80 % of human promoters are cell-type-specific, while only 6 % can be considered as housekeeping promoters (Forrest et al. 2014). Moreover, 51 % of human promoters showed changing activities over time during time-course differentiation, including stem cells, progenitors, differentiated cells, and cells under different stimulating conditions (Arner et al. 2015). Similarly, 13 % of expressed enhancers displayed changing activities during the time-course.

2.3 *Cell-Specific Enhancer Usage*

Defining active enhancers by a double-CAGE-peak pattern appears to be more accurate than traditional enhancer-detection approaches by ChIP sequencing (Chromatin immunoprecipitation) and by sequencing of DNase I hypersensitive sites (DHSs) (Andersson et al. 2014). This approach quantifies enhancer transcription level, and thereby enabling assessment of cell-specific enhancer usage and their activity. In vitro enhancer validation assay showed that to identify cell-specific enhancers in monocytes, B cells, and T cells, the method based on enhancer CAGE transcription activity was more consistent than using chromatin accessibility information obtained from sequencing DHSs or ChIP sequencing (Andersson et al. 2014). CAGE data also allow correlation of enhancer activities to promoter activities for target genes of enhancers. Applying this method, FANTOM5 produced a comprehensive collection of ~44,000 active enhancers, which displayed bidirectional transcription across a diversity of human tissues and cell states (Anderson et al. 2014). In general, lncRNAs derived from active enhancers are non-polyadenylated (90 %), not spliced (95 %), short, not overlapping downstream mRNAs or lncRNAs, and unstable (Andersson et al. 2014; Core et al. 2014). Further, binding activity of putative transcription factors to domains within enhancers and promoters can be assessed by motif response activities analysis (MARA) (Suzuki et al. 2009).

A genome-wide binding profiling showed that enhancer RNAs act to increase chromatin accessibility by transcriptional regulatory complexes to defined genomic regulatory regions, contributing to cell-type-specific transcriptional regulation (Mousavi et al. 2013). The lncRNA CCAT1-L, which is transcribed from a super-enhancer cluster located at 515 kb upstream of the MYC gene, regulates long-range binding of enhancers to MYC promoters, thereby upregulating MYC transcription in colorectal cancer (Xiang et al. 2014). The transcription repressors, Rev-Erb nuclear receptors, suppress macrophage gene expression by downregulating lncRNA expression of distal enhancers, which are macrophage lineage determining factors (Lam et al. 2013).

2.4 *lncRNAs Derived from Repetitive Elements*

FANTOM4 project identified 250,000 retrotransposon-derived transcription start sites, accounting for 6–30 % of capped transcripts in a cell (Faulkner et al. 2009). Depending on estimation approaches, approximately 30–50 % of human DNA is constituted by repetitive elements, which consist of short interspersed noncoding element (SINEs), long interspersed noncoding element (LINEs), long terminal repeat elements (LTRs), and other less common types of repetitive elements. The ENCODE project found 18 % of CAGE-defined TSS overlapping repetitive regions. Shannon entropy analysis of expression uniformity, a measure of tissue specificity (Schug et al. 2005), showed that transcripts from repetitive regions were more narrowly expressed than those from genic regions, suggestive of higher cell-line specificity for repetitive RNA (Djebali et al. 2012). The higher cell specificity was also found in transposable element (TE) containing lncRNAs (Kelley and Rinn 2012). Transcription from LTR has recently been reported to produce functional lncRNAs. Knockdown of a subset of these lncRNAs reduced expression of multiple gene markers of pluripotency (Fort et al. 2014). TE composition analysis of 9241 lncRNA found that 83 % of these lncRNAs contain transposable elements, occupying 41.9 % total length of lncRNA regions (Kelley and Rinn 2012). Retrotransposon TSS within protein-coding genes can drive alternative transcription initiation of these genes (Kelley and Rinn 2012). In addition, 35 % of retrotransposon-associated TSS are tissue-specific, two times higher than that for other TSS types (17 %) (Faulkner et al. 2009).

Transposable elements (TEs) commonly found in lncRNAs may act as binding domains of lncRNAs, a theory called repeat insertion domain of lncRNA (RIDL) hypothesis (Johnson and Guigo 2014). For example, a nuclear-enriched lncRNA (antisense Uchl1) containing an embedded inverted SINEB2 repeat accelerates protein translation of the sense protein-coding gene Uchl1, which is associated with neurodegenerative diseases (Carrieri et al. 2012). The SINE B2 domain was shown to be an essential functional domain of the AS-Uchl1 (Carrieri et al. 2012). lncRNAs containing TEs are involved in a wide range of cellular functions. A hybrid lncRNA derived from integration of LINE1 (Long interspersed element 1) and the X gene of an integrated Hepatitis B virus (HBV) was detected in 23.3 % of HBV-associated hepatocellular carcinoma tumors, and correlated with poorer survival, possibly acting via Wnt/ β -catenin signaling pathway (Shukla et al. 2013; Lau et al. 2014). Another example of a functional repetitive RNA is the telomeric RNA, known as TERRA (Telomeric Repeat containing RNAs). Transcription of repetitive regions at telomere ends produces lncRNA TERRA, which is essential in telomere length regulation, telomere recombination, and telomere end damage repair (Azzalin and Lingner 2014; Yu et al. 2014). An example on interaction of repeat containing lncRNAs with DNA for transcription regulation is the C₀T-1 RNA. LINE1 DNA, which comprises approximately 17 % of human genome, is widely transcribed to generate C₀T-1 stable lncRNA (Hall et al. 2014). C₀T-1 RNAs is transcribed from euchromatin regions tightly associate with chromatin in *cis*,

preventing chromosome condensation, a function similar to the Xist lncRNA (Hall et al. 2014). Many more potential functions of different types of TE containing RNAs remain to be explored.

3 LncRNAs and Human Diseases

3.1 *LncRNAs with Strong Links to Human Diseases*

By curating data from above 500 publications, a database of experimentally verified lncRNA-related diseases shortlisted 321 lncRNAs, which are associated with 221 diseases (<http://cmbi.bjmu.edu.cn/lncnadisease>) (Chen et al. 2013). Among various types of diseases, most commonly, lncRNAs are found related to cancer. At least six lncRNAs have been shown to be involved in prostate carcinogenesis, three of which are highly prostate-specific, including prostate cancer antigen 3 (PCA3), prostate cancer gene expression marker 1 (PCGEM1), and prostate cancer-associated lncRNA transcript 1 (PCAT1) (Walsh et al. 2014). The prostate cancer-associated lncRNA transcript-1 (PCAT-1) represses BRCA2 tumor suppressor by post-transcriptional repression of its 3' UTR in a similar way to microRNA-like or competitive-endogenous RNAs, but not by epigenetic (Prensner et al. 2014). HOX antisense intergenic RNA (HOTAIR) lncRNA is involved in several types of cancers including gastric adenocarcinoma, colorectal cancer, and breast cancer. HOTAIR recruits PRC2 to specific loci for trimethylation of Histone 3 Lysine 27 (H3K27me3) and represses a series of genes (Gupta et al. 2010). Human colorectal cancer-specific CCAT1-L lncRNA promotes long-range chromatin looping between MYC promoter and its enhancers (Xiang et al. 2014). More cancer-associated lncRNAs are being discovered. A recent large-scale *ab initio* transcriptome analysis of 27 cancer types in different tissues and organs found 7942 lncRNAs statistically associated with cancer and/or lineage in human (Iyer et al. 2015).

Besides cancer, lncRNAs are found associated with a range of other disease types, in which two top categories are cardiovascular diseases and neurodegenerative diseases (Chen et al. 2013) (Fig. 2). For example, the antisense APOA1-AS represses the sense APOA1 mRNA, resulting in reduction of the Apolipoprotein A-1 protein, a main component of high-density lipoprotein (HDL) in plasma (Halley et al. 2014). In addition, lncRNAs appear to be important for cognitive functions as brain is the second most common organ expressing the highest number of lncRNAs and many lncRNAs are specific for mammals and primates (Necsulea et al. 2014; Morris and Mattick 2014). The upregulation of the natural antisense lncRNA BACE1-AS (β -secretase-1) increases BACE1 stability and thus maintaining high level of BACE1 enzyme, which may lead to pathophysiology in

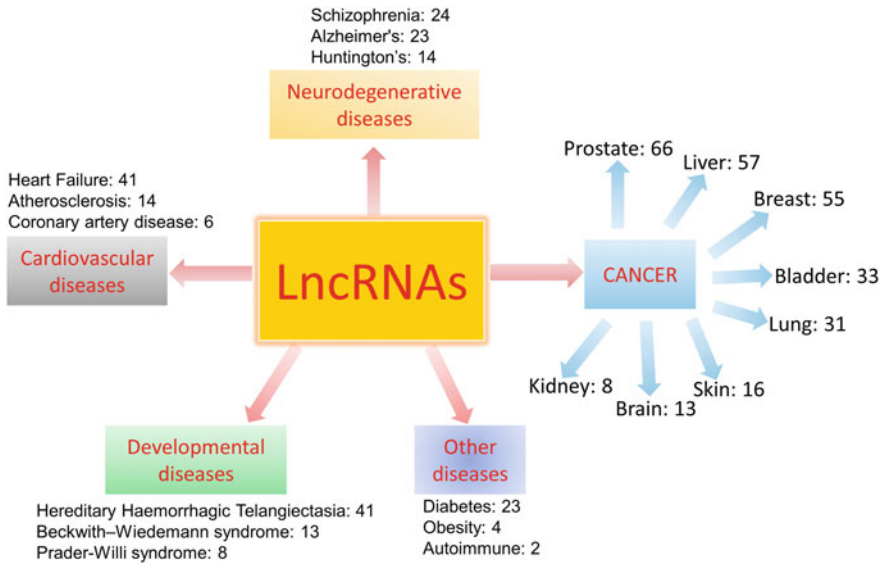


Fig. 2 Various types of diseases found associated with lncRNAs. The number shown for each type of disease is the number of lncRNAs found associated with the disease by experimental evidence on interactions, epigenetics, mutation, expression, and genomic location. The figure was produced by the authors using statistics from the lncRNA Disease database (updated 2014 June 14th) (Chen et al. 2013)

Alzheimer’s disease (Faghihi et al. 2008). ANRIL is involved in type-2 diabetes and coronary artery diseases. The lncRNA ANRIL can cross talk with microRNAs at epigenetic levels. ANRIL binds to PRC2 and epigenetically represses miR-99a/miR-449a, thereby controlling mTOR and CDK6/E2F1 pathways (Zhang et al. 2014). lncRNAs are also involved in immune response processes, as in the case of a lncRNA overlapping 3’UTR region of the interleukin-7 receptor α -subunit gene (lnc-IL7R), which when being repressed reduced trimethylation of H3K27 at proximal promoter regions of inflammatory mediators, diminishing LPS-induced inflammatory responses (Cui et al. 2014). lncRNAs bind to STAT3 in the cytoplasm and promote STAT3 phosphorylation, which is essential for dendritic cell differentiation and T cell activation (Wang et al. 2014). For development and growth diseases, the lncRNA IPW (imprinted gene in the Prader-Willi syndrome region), which is normally transcribed from a paternal allele on chromosome 15, interacts with G9A methyltransferase to maintain H3K9me3 state at the DLK1-DIO3 region on chromosome 14 to repress maternally expressed genes (MEGs) (Stelzer et al. 2014). The aberrant upregulation of MEGs may contribute to Prader-Willi phenotypes.

3.2 Known LncRNAs from Enhancers and Repetitive Elements with Significant Association to Human Diseases

Exons of lncRNAs contain two times higher the number of disease-associated SNPs than those in exons of coding RNAs (Iyer et al. 2015). Notably, lncRNAs are on average longer than coding RNAs, so the higher number of SNPs in lncRNAs may be partly attributed to the larger sizes. Among 301 known cancer-linked SNPs, 88 % are at introns or intergenic regions (Cheetham et al. 2013). The rs944289 SNP is in noncoding region and is linked to the downregulation of a 3.2 kb downstream thyroid-specific lncRNA PCTSC3 (papillary thyroid carcinoma susceptibility candidate 3) which is a possible tumor suppressor (Jendrzewski et al. 2012). Andersson et al. (2014) found significantly more disease-related SNPs in promoters and enhancers than in exon regions or in random sequence of a wide range of cell types and diseases. As a proof of concept, the authors used in vitro luciferase assay to show that reduced enhancer activities due to two SNPs within enhancers are associated with diabetes and Crohn's diseases (Andersson et al. 2014). The cancer-associated variant, rs6983267, regulates expression of an adjacent lncRNA CARLo-5 (cancer-associated region long noncoding RNAs) via long-range interaction between MYC enhancer and CARLo-5 promoter, which correlates with increased cancer susceptibility (Kim et al. 2014).

A noticeable number of lncRNAs derived from or containing repetitive elements are found involved in pluripotency and immune responses. A lncRNA chimera human-viral transcript derived from an integrated genomic region of Hepatitis B virus gene X to a LINE1 site enhances Hepatocellular carcinoma tumor proliferation via Wnt/ β -catenin signaling pathway and promotes metastasis via epithelial to mesenchymal transition (Lau et al. 2014). Transcription of transposable elements, especially those originated from endogenous retroviruses (ERV), is a part of pluripotency regulation network (Kunarso et al. 2010). Long terminal repeat derived transcripts, particularly those belong to endogenous retrovirus families, are found enriched in human, mouse embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs) compared to mouse epithelial fibroblasts (MEFs), human fetal dermal fibroblast (HDF-f), B lymphocytes, and T lymphocytes (Fort et al. 2014). Interestingly, a primate-specific endogenous retrovirus (HERVH) binds naive-pluripotency transcription factor LBP9 to drive transcription of hESC-specific alternative and chimeric transcripts, with over 10 % being lncRNAs, to regulate pluripotency (Wang et al. 2014).

Repetitive lncRNA such as TERRA RNA at telomere ends is required for recruiting telomerase complex and is needed for telomerase protection (de Silanes et al. 2014; Porro et al. 2014). A novel response class, consisting of lncRNAs and small ncRNAs, named as DDRNAs, is needed for site-specific DNA repair and may act in recruitment of DNA damage repair complexes (Francia et al. 2012). Overexpression of PCAT1 lncRNA impairs DNA damage repairs (Prensner et al.

2014). RNA can form hybrids with complementary DNA, which then act as template for homologous recombination and DNA damage repair (Keskin et al. 2014).

3.3 *lncRNAs as Biomarkers for Diagnosis, Prognosis and as Targets for Gene Therapy*

Screening of lncRNAs for potential therapeutic targets is being developed. Most lncRNA loci have been identified and genome-wide lncRNA differential expression analysis starts to reveal hundreds of potential candidates. For example, an RNA-sequencing expression analysis of a noninvasive lung cancer cell line (CL1-0) and a more metastatic prone sub-clone (CL1-5) identified 111 lung cancer-associated lncRNAs, which include candidates with experimental evidence support such as the lung cancer metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) and the lncRNA *SCAL1* (*smoke and cancer-associated lncRNA-1*) (Thai et al. 2013).

Several lncRNAs have been shown as promising biomarkers in diagnosis and prognosis. Overexpression of lncRNA PCAT-1 by 50 % or more may be a prognostic indicator for colorectal cancer progression (Ge et al. 2013). HOTAIR is a predictor of tumor metastasis and survival in breast cancer progression. Approximately 125-fold overexpression of HOTAIR was found in more than one-third of all primary tumors studied (Gupta et al. 2010). An analysis from public database for 2255 patients suggests that HOTAIR expression level is strongly correlated with hazard ratio for esophageal squamous cell carcinomas and colorectal cancer (Deng et al. 2014). Applying CAGE for 50 matched human hepatocarcinoma liver samples, Hashimoto et al. (manuscript submitted) found 43 LTR-derived lncRNAs strongly upregulated in more than 50 % of cancer tissues. Kaczowski et al. (manuscript submitted) compared FANTOM CAGE expression in 216 different cancer cell lines with corresponding primary cell lines and identified a core set of pan-cancer biomarkers, including enhancer RNAs and RNAs from repetitive elements.

lncRNAs have been shown as promising therapeutic targets. Knockdown of the lncRNA-JADE represses histone H4 acetylation in DNA damage response pathway and reduces breast tumor growth in vivo in mice (Wan et al. 2013). In a test for a potential therapeutic intervention to Angelman syndrome, antisense oligonucleotides were successfully applied to knockdown UBE3A-ATS transcripts, allowing the expression of paternal Ube3a in neuron both in vitro and in vivo, which in turn recovers UBE3A ligase protein expression and mediates some cognitive deficits in Angelman mouse model (Meng et al. 2014). Intratumoral injection of a plasmid carrying a toxin produced under the control of the lncRNA H19 promoter was applied to reduce tumor size in bladder, ovarian, and pancreatic cancer. In a clinical trial phase 2 for diphtheria toxin-ABC-819, 33 % complete ablation of bladder cancer tumor and 66 % prevention of new tumors in the first 3 months were reported (Gofrit et al. 2014).

3.4 *Strategies for Perturbation of Disease-Associated LncRNAs*

For selection of lncRNA perturbation technologies, a collection of lncRNA knock-down options are available. These include both traditional reagents such as antisense oligonucleotides (ASOs), small interfering RNAs (siRNAs), short hairpin RNAs (shRNAs), new classes of inhibitory molecules such as AntagoNAT oligonucleotides (single-stranded gapmer LNAs-modified with phosphorothioate backbone) to knockdown natural antisense transcripts (NATs), and precise genome-editing nuclease technologies, most commonly including the use of chimeric nucleases Transcription Activator Like Effector Nucleases (TALENs), Zinc Finger Nucleases (ZFNs) and Clustered Regulatory Interspaced Short Palindromic Repeat (CRISPR)/Cas-based RNA guided DNA nucleases (Gaj et al. 2013; Takahashi and Carninci 2014; Kim and Kim 2014). Use of ASOs was reported to produce effective knock-down of lncRNA *in vivo* by 50–80 % for MALAT1 in human and mouse (Gutschner et al. 2013). In human and monkey liver cell lines and in intravenous injection experiments on African green monkeys, downregulation of APOA1-AS by ASOs enhanced expression in APO (Apolipoprotein) gene cluster, including APOA1, APOC3, and APOA4 (Halley et al. 2014). ANRIL (antisense noncoding RNA in the INK4 locus) recruited PCR2 complex to specifically repress mir99a and mir449a in gastric cancer, while siRNA knockdown of ANRIL decreased expression of mRNA targets of these two miRNAs (Zhang et al. 2014). Knockdown of the low abundance antisense brain-derived neurotrophic factor (BDNF-AS) by SiRNA increased BDNF mRNA and BDNF protein levels in hippocampal neurospheres (Modarresi et al. 2012). Retroviral transduction could stably overexpress HOTAIR to several 100-fold in human breast cancer cell lines (Gupta et al. 2010). Application of AntagoNATs for transiently upregulating expression of sense protein-coding genes in a locus-specific manner opens a new pharmacological strategy to expression perturbation (Modarresi et al. 2012). Knockout of MALAT1 by 1000-fold in human lung tumor cells was achieved using zinc finger nucleases, creating an efficient loss-of-function model (Gutschner et al. 2013). Overexpression of lncRNA CCAT1-L by 15- to 30-fold using TALENs showed that upregulation of CCAT1-L lncRNA enhanced MYC expression (Xiang et al. 2014). Efficient deletion of a large 23-kb fragment within the lncRNA *Rian* was achieved by CRISPR/Cas9 system in mice (Han et al. 2014).

Selection of lncRNA knockdown targets needs to consider molecular mechanism of actions of the targets. For example, the targets can be different if the lncRNA directly participates in regulation, or if their transcription process is needed to generate chromatin context for regulating transcription of other genes (Bassett et al. 2014; Latos et al. 2012). A number of lncRNAs form complexes with epigenetic factors, which repress cancer suppressor genes by modifying chromatin state. For these targets, repressing RNA–protein complexes such as HOTAIR-PRC2, ANRIL-CBX7 (chromobox homolog 7), PCAT-1-PRC2, and H19-EZH2 (enhancer of zeste homolog 2) may specifically reactivate cancer suppressor genes (Fatemi et al. 2014). The interdependence of the lncRNA and protein in these complexes suggests

that the endogenous levels of proteins and lncRNAs may decide target selection, i.e., it may be more effective to use small molecule inhibitors to inactivate the protein in the case that the lncRNA is abundant and vice versa (Gupta et al. 2010). On the other direction, for more direct tumor suppressing lncRNAs such as TUG1 (taurine upregulated gene 1) (Zhang et al. 2014) and PINT (p53 induced transcript) (Marin-Bejar et al. 2013), direct upregulation of those lncRNAs may reduce tumor growth. A database for putative coding genes affected by lncRNA knockdown or overexpression, for instance the lncRNA2Target database, is useful for lncRNA target selection (Jiang et al. 2014). In addition, selection of lncRNAs should take into account half-lives of targeted lncRNAs, for which the current understanding is still limited. Most lncRNAs produced from bidirectionally balanced transcription are suppressed post-transcriptionally by ribo-nucleolytic RNA exosomes complex (Andersson et al. 2014). In a microarray stability assay for 823 lncRNAs in mouse Neuro-2a cells, Clark et al. (2012) showed that lncRNA half-lives are similar in range of coding RNA, and over 6 % of lncRNAs are highly stable (>12 h).

Furthermore, selecting knockdown regions within lncRNA targets should consider structure and functional domains. lncRNAs commonly comprise two exons, fewer compared to mRNAs (Derrien et al. 2012), but exons of lncRNAs can be large in size, requiring careful selection and combination of targeted knockdown regions. However, structure of lncRNAs is still poorly understood, due to lack of high-throughput biophysical and biochemical tools for RNA structure analysis (Mortimer et al. 2014). Similarly, functional domains of lncRNAs are still not well studied due to technological constraints. Although repeat insertion domains present in most lncRNAs may act as binding domains of many lncRNAs to proteins and DNA, these repetitive domains may be more challenging to be specifically and effectively knockdowned (Kelley and Rinn 2012; Johnson and Guigo 2014). Recent development of domain-specific chromatin isolation by RNA purification (dChIRP) technology enables investigation of binding sites of single RNA domain in RNA–RNA, RNA–DNA, and RNA–protein interactions (Quinn et al. 2014). Such technology will provide useful parameters for selecting knockdown regions within lncRNAs. Another important challenge for developing lncRNA generic therapy is about specificity, efficiency, and immunogenicity of gene-delivery strategies. In-depth discussion about this challenge can be found elsewhere (Takahashi and Carninci 2014).

4 Perspectives

4.1 *Studies of LncRNA Structure and Their Interactions with RNA, DNA, and Proteins*

Recent advances in high-throughput experimental structure sequencing methods, e.g., structure-seq, in combination with computational modeling starts to produce rich information of secondary in vivo structure of tens of thousands transcripts

(Wan et al. 2014; Ding et al. 2014). The information may aid the design of knock-down targets to avoid stably folded RNA regions. However, these techniques are constrained by low resolution, which is usually not sufficient to predict functional domains of lncRNAs. In contrast, recent development of domain-specific chromatin isolation by RNA purification (dChIRP) enables thorough study of a functional domain of lncRNAs in pair-wise interactions of RNA–RNA, RNA–DNA, and RNA–protein, yet the throughput is low (Quinn et al. 2014). More advanced combination of computational biology and experimental approaches will increase resolution and throughput of RNA structure, which will advance lncRNA functional studies. More specialized tools are being rapidly developed for (1) RNA–protein interactions such as cross-linking immunoprecipitation CLIP-seq, with various protocols such as PAR-CLIP, HITS-CLIP, and iCLIP (Sugimoto et al. 2012; Hafner et al. 2010), for (2) chromosome organization by chromosome conformation captures (3C, 4C, 5C, Hi-C, Chia-PET, and 6C) (de Wit and de Laat 2012), for (3) RNA–DNA interaction by capture hybridization analysis of targets (CHART) (Simon et al. 2011) or chromatin isolation by RNA purification (ChIRP) (Chu et al. 2011) and a modified version of ChIRP to study domain-specific RNA–DNA interaction (dChIRP) (Quinn et al. 2014), for (4) RNA–RNA interaction by cross-linking ligation and sequencing of hybrids (CLASH) (Helwak and Tollervy 2014) or by RNA antisense purification (RAP-RNA) (Engreitz et al. 2014), and in situ labeling technologies and imaging (Chakraborty et al. 2012). Remarkably, application of dChIRP can decipher the lncRNA architecture and functions at domain-specific level and can detect pair-wise RNA–RNA, RNA–protein, and RNA–DNA interactions (Quinn et al. 2014). Combination of high resolution mapping of RNA–chromatin interaction sites using RNA antisense purification (RAP) with chromosome conformation capture (Hi-C) and computational modeling could revealed the mechanism of the lncRNA Xist to spread along the X chromosome by utilizing three-dimensional conformation of the genome (Engreitz et al. 2013).

4.2 *Personalized Medicine*

With the great magnitude of expression specificity, lncRNAs may be molecules of choice for future personalized medicine. Large-scale studies in FANTOM projects establish that tissue and disease specificity are important characters of lncRNAs. More insights on potential RNA therapies are discussed elsewhere (Takahashi and Carninci 2014). Promoter design can play a vital role in optimization of non-viral gene expression therapy to reduce inflammatory responses, to increase tissue specificity, and to increase expression levels (Pringle et al. 2012; Hyde et al. 2008). For example, complete removal of CpG dinucleotides in enhancer/promoter regions of non-viral expression vector administered in cystic fibrosis treatment resulted in stronger and longer expression of transgenes with undetectable inflammatory responses (Hyde et al. 2008). For this type of application, the use of FANTOM5-rich database of promoter usage with tissue specificity, promoter

structure, and promoter activities can help increase efficacy and specificity of therapeutic vectors (Forrest et al. 2014).

lncRNAs will provide additional options for gene therapies. The consensus number of distinct molecular targets of FDA-approved drugs (Food and Drug Administration) until 2006 was as low as 324, in which 266 (or 82.1 %) targets are human-genome-derived proteins (Overington et al. 2006). The total estimated number of druggable coding genes in a human genome is limited to approximately 2000–3000 genes (Russ and Lampel 2005). Expansion of the potentially druggable targets may need to include lncRNAs. Importantly, since disease-associated SNPs present more frequently in transcribed regions encompassing enhancers, promoters, and lncRNAs, the interpretation of genome-wide association studies (GWAS) should take into account these regulatory elements (Gong et al. 2014). Primate-specific SNPs found in lncRNA exons carry significantly higher selective constraint than those in intergenic regions (Necsulea et al. 2014). For therapeutic application of lncRNAs to be approved in clinical settings, it is likely that the effects of lncRNA perturbation should be characterized at regulatory network level. While functions and mechanisms of lncRNA are still poorly characterized, caution has to be taken on their application. The Progenesa™ PCA3 urine test (Gen-Probe Inc., San Diego, CA, USA) using lncRNA prostate cancer antigen 3 (PCA3) as a marker for prostate cancer was approved by FDA. However, a recent assessment by the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group found that the current data are insufficient to support the clinical validity of PCA3 test for diagnosis and management of prostate cancer, unless further supporting evidence is available (Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group 2014). Better understanding of possible effects of lncRNA perturbation on cellular processes and availability of more clinical data will foster lncRNA therapeutic applications.

5 Conclusion

lncRNAs play important regulatory roles, and misregulation of lncRNAs is found associated with various diseases. Genome-wide sequencing of capped RNAs in FANTOM project enables promoter-centric analysis of transcription contributed to discovery of major lncRNA classes such as those derived from bidirectional promoters, enhancers, repetitive regions, and antisense RNAs. Moreover, by quantifying activities of regulatory sequences for lncRNAs, including promoters and enhancers, high level of lncRNA expression specificity can be found between individuals, organs, tissues, and cell types. From genome-wide CAGE sequencing of multitude of systematically classified primary cell samples, tissues and cell lines within FANTOM project, differential promoter and enhancer usage of lncRNAs can be linked to disease ontology terms (Forrest et al. 2014; Andersson et al. 2014). The lncRNA expression specificity in relation to human diseases makes lncRNAs promising candidates for biomarkers in diagnosis and prognosis, and for targets in

therapeutic treatments. This is important because lncRNA candidates add more options to the current limited number of druggable genes as well as limited use of disease-related SNPs from GWAS studies. Interestingly, a majority of SNPs lie within lncRNA, promoter, and enhancer regions, which open possibility to make use of information from adjacent lncRNAs and regulatory genomic regions such as promoters and enhancers to better link SNPs to diseases. Combined use of lncRNAs, coding genes, and SNPs may bring personalized medicine closer to clinical applications in the near future. Some challenges to be solved require high-throughput technologies for studying structure and interaction network, and technologies for effective perturbation of lncRNA expression.

Acknowledgements We especially thank FANTOM consortium for generating unprecedented amount of data for promoter-centric analysis in thousands of samples. We apologize for not being able to mention all important, related work from colleagues. This work was supported by a research grant from the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) to the RIKEN Center for Life Science Technologies and the Human Frontier Science Program to P.C.

References

- Akalin A et al (2009) Transcriptional features of genomic regulatory blocks. *Genome Biol* 10(4): R38
- Andersson R et al (2014a) An atlas of active enhancers across human cell types and tissues. *Nature* 507(7493):455–461
- Andersson R et al (2014b) Nuclear stability and transcriptional directionality separate functionally distinct RNA species. *Nat Commun* 5:5336
- Arner E et al (2015) Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* 347(6225):1010–1014
- Azzalin CM, Lingner J (2014) Telomere functions grounding on TERRA firma. *Trends in Cell Biol* 25(1):29–36
- Bassett AR et al (2014) Considerations when investigating lncRNA function in vivo. In: Weigel D (ed) vol 3
- Bernstein BE et al (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57–74
- Bertone P et al (2004) Global identification of human transcribed sequences with genome tiling arrays. *Science* 306(5705):2242–2246
- Carninci P et al (2003) Targeting a complex transcriptome: the construction of the mouse full-length cDNA encyclopedia. *Genome Res* 13(6b):1273–1289
- Carninci P et al (2005) The transcriptional landscape of the mammalian genome. *Science* 309(5740):1559–1563
- Carninci P et al (2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* 38(6):626–635
- Carrieri C et al (2012) Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* 491(7424):454–457
- Chakraborty D et al (2012) Combined RNAi and localization for functionally dissecting long noncoding RNAs. *Nat Methods* 9(4):360–362
- Cheatham SW et al (2013) Long noncoding RNAs and the genetics of cancer. *Br J Cancer* 108(12):2419–2425

- Chen G et al (2013) LncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Res* 41(Database issue):D983–D986
- Chu C et al (2011) Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol Cell* 44(4):667–678
- Clark MB et al (2012) Genome-wide analysis of long noncoding RNA stability. *Genome Res* 22(5):885–898
- Core LJ et al (2014) Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet* 46(12):1311–1320
- Cui H et al (2014) The human long noncoding RNA lnc-IL7R regulates the inflammatory response. *Eur J Immunol* 44(7):2085
- de Silanes IL et al (2014) Identification of TERRA locus unveils a telomere protection role through association to nearly all chromosomes. *Nat Commun* 5
- de Wit E, de Laat W (2012) A decade of 3C technologies: insights into nuclear organization. *Genes Dev* 26(1):11–24
- Deng Q et al (2014) Prognostic value of long non-coding RNA HOTAIR in various cancers. *PLoS ONE* 9(10):e110059
- Derrien T et al (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 22(9):1775–1789
- Ding Y et al (2014) In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* 505(7485):696–700
- Djebali S et al (2012) Landscape of transcription in human cells. *Nature* 489(7414):101–108
- Engreitz JM et al (2013) The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* 341(6147):1237973
- Engreitz JM et al (2014) RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent pre-mRNAs and chromatin sites. *Cell* 159(1):188–199
- Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group (2014) Recommendations from the EGAPP Working Group: does PCA3 testing for the diagnosis and management of prostate cancer improve patient health outcomes? *Genet Med* 16(4):338–346
- Faghihi MA et al (2008) Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of β -secretase expression. *Nature Med* 14(7):723–730
- Fatemi RP, Velmeshev D, Faghihi MA (2014) De-repressing lncRNA-targeted genes to upregulate gene expression: focus on small molecule therapeutics. *Mol Ther Nucleic Acids* 3:e196
- Faulkner GJ et al (2009) The regulated retrotransposon transcriptome of mammalian cells. *Nat Genet* 41(5):563–571
- Forrest AR et al (2014) A promoter-level mammalian expression atlas. *Nature* 507(7493):462–470
- Fort A et al (2014) Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nat Genet* 46:558–566
- Francia S et al (2012) Site-specific DICER and DROSHA RNA products control the DNA-damage response. *Nature* 488(7410):231–235
- Gaj T, Gersbach CA, Barbas CF 3rd (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol* 31(7):397–405
- Ge X et al (2013) Overexpression of long noncoding RNA PCAT-1 is a novel biomarker of poor prognosis in patients with colorectal cancer. *Med Oncol* 30(2):588
- Gofrit ON et al (2014) DNA based therapy with diphtheria toxin-A BC-819: a phase 2b marker lesion trial in patients with intermediate risk nonmuscle invasive bladder cancer. *J Urol* 191(6):1697–1702
- Gong J et al (2014) lncRNASNP: a database of SNPs in lncRNAs and their potential functions in human and mouse. *Nucleic Acids Res* 43(Database issue):D181–186
- Gupta RA et al (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464(7291):1071–1076
- Gutschner T et al (2013) The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res* 73(3):1180–1189
- Guttman M et al (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458(7235):223–227

- Haberle V et al (2014) Two independent transcription initiation codes overlap on vertebrate core promoters. *Nature* 507(7492):381–385
- Hafner M et al (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141(1):129–141
- Hall LL et al (2014) Stable COT-1 repeat RNA is abundant and is associated with euchromatic interphase chromosomes. *Cell* 156(5):907–919
- Halley P et al (2014) Regulation of the apolipoprotein gene cluster by a long noncoding RNA. *Cell Rep* 6(1):222–230
- Han J et al (2014) Efficient in vivo deletion of a large imprinted lncRNA by CRISPR/Cas9. *RNA Biol* 11(7):829–835
- Hangauer MJ, Vaughn IW, McManus MT (2013) Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. *PLoS Genet* 9(6):e1003569
- Helwak A, Tollervey D (2014) Mapping the miRNA interactome by cross-linking ligation and sequencing of hybrids (CLASH). *Nat Protocols* 9(3):711–728
- Hirose T et al (2014) NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. *Mol Biol Cell* 25(1):169–183
- Hyde SC et al (2008) CpG-free plasmids confer reduced inflammation and sustained pulmonary gene expression. *Nat Biotechnol* 26(5):549–551
- Iyer MK et al (2015) The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet* 47(3):199–208
- Jendrzewski J et al (2012) The polymorphism rs944289 predisposes to papillary thyroid carcinoma through a large intergenic noncoding RNA gene of tumor suppressor type. *Proc Natl Acad Sci USA* 109(22):8646–8651
- Jiang Q et al (2014) LncRNA2Target: a database for differentially expressed genes after lncRNA knockdown or overexpression. *Nucleic Acids Res* 43(Database issue):D193–196
- Johnson R, Guigo R (2014) The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs. *RNA* 20(7):959–976
- Kanamori-Katayama M et al (2011) Unamplified cap analysis of gene expression on a single-molecule sequencer. *Genome Res* 21(7):1150–1159
- Katayama S et al (2005) Antisense transcription in the mammalian transcriptome. *Science* 309(5740):1564–1566
- Kelley D, Rinn J (2012) Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol* 13(11):R107
- Keskin H et al (2014) Transcript-RNA-templated DNA recombination and repair. *Nature* 515(7527):436–439
- Kim H, Kim J-S (2014) A guide to genome engineering with programmable nucleases. *Nat Rev Genet* 15(5):321–334
- Kim T et al (2014) Long-range interaction and correlation between MYC enhancer and oncogenic long noncoding RNA CARLo-5. *Proc Natl Acad Sci USA* 111(11):4173–4178
- Kunarso G et al (2010) Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet* 42(7):631–634
- Lam MT et al (2013) Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature* 498(7455):511–515
- Latos PA et al (2012) Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science* 338(6113):1469–1472
- Lau C-C et al (2014) Viral-human chimeric transcript predisposes risk to liver cancer development and progression. *Cancer Cell* 25(3):335–349
- Li L, Chang HY (2014) Physiological roles of long noncoding RNAs: insight from knockout mice. *Trends Cell Biol* 24(10):594–602
- Li Z et al (2014a) The long noncoding RNA THRIL regulates TNF α expression through its interaction with hnRNPL. *Proc Natl Acad Sci USA* 111(3):1002–1007
- Li H et al (2014b) Overexpression of lncRNA H19 enhances carcinogenesis and metastasis of gastric cancer. *Oncotarget* 5(8):2318–2329

- Liu Y et al (2002) The UCH-L1 gene encodes two opposing enzymatic activities that affect alpha-synuclein degradation and Parkinson's disease susceptibility. *Cell* 111(2):209–218
- Mao YS et al (2011) Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat Cell Biol* 13(1):95–101
- Marin-Bejar O et al (2013) Pint lincRNA connects the p53 pathway with epigenetic silencing by the polycomb repressive complex 2. *Genome Biol* 14(9):R104
- Meng L et al (2014) Towards a therapy for Angelman syndrome by targeting a long non-coding RNA. *Nature* 518(7539):409–412
- Mercer TR, Mattick JS (2013) Structure and function of long noncoding RNAs in epigenetic regulation. *Nat Struct Mol Biol* 20(3):300–307
- Mercer TR et al (2008) Specific expression of long noncoding RNAs in the mouse brain. *Proc Natl Acad Sci USA* 105(2):716–721
- Modarresi F et al (2012) Inhibition of natural antisense transcripts in vivo results in gene-specific transcriptional upregulation. *Nat Biotechnol* 30(5):453–459
- Morris KV, Mattick JS (2014) The rise of regulatory RNA. *Nat Rev Genet* 15(6):423–437
- Mortimer SA, Kidwell MA, Doudna JA (2014) Insights into RNA structure and function from genome-wide studies. *Nat Rev Genet* 15:469
- Mousavi K et al (2013) eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. *Mol Cell* 51(5):606–617
- Necsulea A et al (2014) The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* 505(7485):635–640
- Overington JP, Al-Lazikani B, Hopkins AL (2006) How many drug targets are there? *Nat Rev Drug Discov* 5(12):993–996
- Park C et al (2014) lncRNAtor: a comprehensive resource for functional investigation of long non-coding RNAs. *Bioinformatics* 30(17):2480–2485
- Pennisi E (2014) Lengthy RNAs earn respect as cellular players. *Science* 344(6188):1072
- Porro A et al (2014) Functional characterization of the TERRA transcriptome at damaged telomeres. *Nat Commun* 5
- Prensner JR et al (2014) PCAT-1, a long noncoding RNA, regulates BRCA2 and controls homologous recombination in cancer. *Cancer Res* 74(6):1651–1660
- Pringle IA et al (2012) Rapid identification of novel functional promoters for gene therapy. *J Mol Med (Berl)* 90(12):1487–1496
- Quek XC et al (2014) lncRNADB v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res* 43(Database issue):D168–173
- Quinn JJ et al (2014) Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. *Nat Biotechnol* 32(9):933–940
- Rinn JL, Chang HY (2012) Genome regulation by long noncoding RNAs. *Annu Rev Biochem* 81:145–166
- Rinn J, Guttman M (2014) RNA and dynamic nuclear organization. *Science* 345(6202):1240–1241
- Russ AP, Lampel S (2005) The druggable genome: an update. *Drug Discovery Today* 10(23–24):1607–1610
- Schug J et al (2005) Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biol* 6(4):R33
- Severin J, Lizio M, Harshbarger J (2014) Interactive visualization and analysis of large-scale sequencing datasets using ZENBU. *Nat Biotechnol* 32(3):217–219
- Shukla R et al (2013) Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell* 153(1):101–111
- Simon MD et al (2011) The genomic binding sites of a noncoding RNA. *Proc Natl Acad Sci USA* 108(51):20497–20502
- Stelzer Y et al (2014) The noncoding RNA IPW regulates the imprinted DLK1-DIO3 locus in an induced pluripotent stem cell model of Prader-Willi syndrome. *Nat Genet* 46(6):551–557
- Sugimoto Y et al (2012) Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biol* 13(8):R67

- Suzuki H et al (2009) The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nat Genet* 41(5):553–562
- Takahashi H, Carninci P (2014) Widespread genome transcription: new possibilities for RNA therapies. *Biochem Biophys Res Commun* 452(2):294–301
- Takahashi H et al (2012) 5' end-centered expression profiling using cap-analysis gene expression and next-generation sequencing. *Nat Protoc* 7(3):542–561
- Thai P et al (2013) Characterization of a novel long noncoding RNA, SCAL1, induced by cigarette smoke and elevated in lung cancer cell lines. *Am J Respir Cell Mol Biol* 49(2):204–211
- Thorsen K et al (2011) Tumor-specific usage of alternative transcription start sites in colorectal cancer identified by genome-wide exon array analysis. *BMC Genom* 12(1):505
- Trimarchi T et al (2014) Genome-wide mapping and characterization of notch-regulated long noncoding RNAs in acute leukemia. *Cell* 158(3):593–606
- Volders PJ et al (2015) An update on LNCipedia: a database for annotated human lncRNA sequences. *Nucleic Acids Res* 43(Database issue):D174–D180
- Walsh AL et al (2014) Long noncoding RNAs and prostate carcinogenesis: the missing 'linc'? *Trends Mol Med* 20(8):428–436
- Wan G et al (2013) A novel non-coding RNA lncRNA-JADE connects DNA damage signalling to histone H4 acetylation. *EMBO J* 32(21):2833–2847
- Wan Y et al (2014) Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* 505(7485):706–709
- Wang J et al (2014) Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* 516:405–409
- Wang P et al (2014) The STAT3-binding long noncoding RNA lnc-DC controls human dendritic cell differentiation. *Science* 344(6181):310–313
- Ward LD, Kellis M (2012) Evidence of abundant purifying selection in humans for recently acquired regulatory functions. *Science* 337(6102):1675–1678
- Xiang JF et al (2014) Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res* 24(5):513–531
- Xie C et al (2013) NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic Acids Res* 42(Database issue):D98–D103
- Yang F et al (2013) Repression of the long noncoding RNA-LET by histone deacetylase 3 contributes to hypoxia-mediated metastasis. *Mol Cell* 49(6):1083–1096
- Yildirim E et al (2013) Xist RNA is a potent suppressor of hematologic cancer in mice. *Cell* 152(4):727–742
- Yu TY, Kao YW, Lin JJ (2014) Telomeric transcripts stimulate telomere recombination to suppress senescence in cells lacking telomerase. *Proc Natl Acad Sci USA* 111(9):3377–3382
- Yue F et al (2014) A comparative encyclopedia of DNA elements in the mouse genome. *Nature* 515(7527):355–364
- Zhang EB et al (2014a) Long noncoding RNA ANRIL indicates a poor prognosis of gastric cancer and promotes tumor growth by epigenetically silencing of miR-99a/miR-449a. *Oncotarget* 5(8):2276–2292
- Zhang EB et al (2014b) P53-regulated long non-coding RNA TUG1 affects cell proliferation in human non-small cell lung cancer, partly through epigenetically regulating HOXB7 expression. *Cell Death Dis* 5:e1243