Gene Technology
ISSN: 2329-6682
Wan and Wang, Gene Technology2013, 2:1
DOI: 10.4172/2329-6682.1000103
Short Communication Open Access

Decipher the Mystery of Long Non-Coding RNAs - Novel CRISPR/Cas System for Genome-Editing

Guohui Wan* and Min Wang

Department of Cancer Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

With the advent of high throughput RNA sequencing technology and higher density tiling arrays, the transcriptional activities across the human genome have been investigated and led to the observation of widespread transcription of the genome [1,2]. While only less than 2% of the mammalian genome codes for proteins, tens of thousands of genomic sites are pervasively transcribed and produce non-coding RNAs, including microRNAs, Piwi-interacting RNAs, small interfering RNAs, small nucleolar RNAs and long non-coding RNAs (lncRNAs). LncRNAs were originally defined as transcribed RNA molecules longer than 200 nucleotides and with little protein-coding potential. Recently, emerging evidences have shown that lncRNAs play important roles in a variety of biological processes including development, differentiation, metabolism, genome regulation and cancer progression [3-5]. Nonetheless, due to the heterogeneity and low abundance, in most cases, most lncRNAs have no genetic evidence to support their *in vivo* function.

Gene-targeted knockout technology has provided a powerful tool for elucidating the function of lncRNA genes *in vivo*, making the connection from mouse to mechanism. For example, *Malat1*, known as metastasis associated lung adenocarcinoma transcript 1, is among the most abundant and highly conserved lncRNAs. *Malat1* was identified as an oncogene that promoted tumorigenesis and found to regulate pre-mRNA splicing in nuclear speckles and promote E2F1 target gene expression during cell cycle progression in vitro [6-9]. Recently, *Malat1* knockout models have shown that loss of *Malat1 in vivo* has no apparent phenotypes and compatible with formation of nuclear speckles [10,11], but alters the transcription of *Malat1* neighboring genes and metastasis-associated genes [12,13]. These loss-of-function models further support the critical function of *Malat1* as a regulator of gene expression governing hallmarks of lung cancer metastasis. Along with more and more differentially expressed lncRNAs have been unraveled and screened during various physiological processes and disease, understanding the functions of these lncRNA genes require lose-of-function screen by deleting or modifying genes, and followed by studying the resulting phenotypes. More recently, a novel powerful genome-editing tool - the type II prokaryotic CRISPR/Cas system has been successfully employed for genome engineering, thus holding great potential for deciphering the function of lncRNAs *in vivo*.

The clustered, regularly interspaced, short palindromic repeats (CRISPR)/CRISPR associated (Cas) were originally identified as components of the bacterial adaptive immune response to invading viral and plasmid sequences [14,15]. This CRISPR/Cas system relies on integration of foreign DNA fragments into CRISPR loci. After transcription and processing, these inserts produce a short CRISPR RNA (crRNA), which recognizes a complementary stretch of nucleotides (the protospacer) within foreign DNA and then anneals to a trans-activating CRISPR RNA (tracrRNA), forming ribonucleoprotein complexes with the Cas9 nuclease and generating site-specific double strands breaks in the foreign DNA.

Generation of a targeted genomic double strands break has been considered as the rate-limiting step in the development of gene targeting technology for genome engineering. By combining crRNA and tracrRNA into a single chimeric guide RNA (gRNA) and expressing it alongside Cas9 endonuclease, several groups have shown that they can engineer the CRISPR/Cas system to function with custom gRNA to efficiently induce double strands break in specific genes in human cells [16-18]. For example, Mali et al. [16] obtained targeting rates of 10 to 25% in 293T cells, 13 to 8% in K562 cells, and 2 to 4% in induced pluripotent stem cells. Cong et al. [17] applied the CRISPR/Cas system to mediate genomic cleavage as efficiently as a pair of TALE nucleases (TALEN) targeting the same EMX1 element. Cho et al. [18] showed that CRISPR/Cas system can target two genomic sites and induce indels with frequencies of up to 33%. These three independent groups showed the easy programmability and wide applicability of the CRISPR/Cas technology, establishing the potential of the CRISPR/Cas system for high-throughput applications.

LncRNAs are marked with their abundance in the genome. It is clear that the majority of disease-associated SNPs are located within non-coding regions and impacts lncRNA expression [19]. Kumar et al. [20] showed that the associated of SNPs with expression levels of lncRNAs are also associated with complex traits and diseases. Studying the epistatic lncRNAs through CRISPR/Cas system may allow for dissection of phenotypes that are normally masked by compensatory mechanisms. Accordingly it requires the simultaneous introduction of multiple gRNAs into human cells to achieve multiplex lncRNA genes editing of multiple targeted loci. Subsequent studies further support that CRISPR/Cas system can efficiently mediate the simultaneous disruption of five genes in mouse embryonic stem cells, allowing the one-step generation of animals carrying mutations in multiple genes or reporter and conditional alleles [21,22]. Furthermore, the CRISPR/Cas system also provides the probability for studies of complex genetic disease in a Petri dish using human embryonic stem cells and induced pluripotent stem cells [23].

The CRISPR/Cas system may yet have limitations although it undoubtedly possesses great potential for genome editing. For instance, target site selection may be limited by requirement for a dubbed protospace adjacent motif (PAM) sequence. Two current Cas9 nucleases are derived from *Streptococcus pyogenes* and *Streptococcus thermophiles* require PAMs of 5'-NGG-3' and 5'-NNAGAAW-3' respectively, which may constraint the selection of target sequences [14]. Another important limitation of the CRISPR/Cas system is its potential off-target effects [16-18]. Although each base within the guide RNA sequence contributes to overall specificity, multiple mismatches between the guide RNA and its complementary target DNA sequence can be tolerated depending on the quantity, position and base identify

***Corresponding author:** Guohui Wan, Department of Cancer Biology, The University of Texas MD Anderson Cancer Center, 7777 Knight Road, Houston, Texas 77054, USA, E-mail: gwan@mdanderson.org

Received September 19, 2013; **Accepted** October 01, 2013; **Published** October 03, 2013

Citation: Wan G, Wang M (2013) Decipher the Mystery of Long Non-Coding RNAs - Novel CRISPR/Cas System for Genome-Editing. Gene Technology 2: 103. doi: 10.4172/2329-6682.1000103

of mismatches. However, Ran et al recently developed a strategy that combines the D10A mutant nickase version of Cas9 from *Streptococcus pyogenes (SpCas9n)* with a pair of offset gRNAs complementary to opposite strands of the target sites to introduce targeted double strand breaks [24]. This double nicking strategy reduces off-target activity by 50 to 1,500-fold in cell lines and facilitates gene knockout in mouse zygotes without sacrificing on-target cleavage efficiency, thus extensively improving the specificity of Cas9-mediated genome editing.

In the last decade, the rapid discovery of lncRNAs by high-throughput technologies has advanced our knowledge of transcriptome complexity. LncRNAs have been shown as a novel layer of regulation of diverse physiological activities. Dysregulation of lncRNAs and lncRNA expression-associated SNPs can lead to numerous diseases including cancers. It is inevitable to dissect into lncRNA structure and physiological function by *in vivo* knockout phenotypes. Timely, the novel CRISPR/Cas technology for genome-editing is moving so fast that it undoubtedly allows the accelerated study of lncRNAs *in vivo in* the near future.

References

1. Bertone P, Stolc V, Royce TE, Rozowsky JS, Urban AE, et al. (2004) Global identification of human transcribed sequences with genome tiling arrays. Science 306: 2242-2246.
2. ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, et al. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447: 799-816.
3. Prensner JR, Chinnaiyan AM (2011) The emergence of lncRNAs in cancer biology. Cancer Discov 1: 391-407.
4. Rinn JL, Chang HY (2012) Genome regulation by long noncoding RNAs. Annu Rev Biochem 81: 145-166.
5. Wan G, Liu Y, Han C, Zhang X, Lu X (2013) Noncoding RNAs in DNA Repair and Genome Integrity. Antioxid Redox Signal.
6. Li L, Feng T, Lian Y, Zhang G, Garen A, et al. (2009) Role of human noncoding RNAs in the control of tumorigenesis. Proc Natl Acad Sci U S A 106: 12956-12961.
7. Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, et al. (2010) The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol Cell 39: 925-938.
8. Bernard D, Prasanth KV, Tripathi V, Colasse S, Nakamura T, et al. (2010) A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. EMBO J 29: 3082-3093.
9. Yang L, Lin C, Liu W, Zhang J, Ohgi KA, et al. (2011) ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. Cell 147: 773-788.
10. Nakagawa S, Ip JY, Shioi G, Tripathi V, Zong X, et al. (2012) Malat1 is not an essential component of nuclear speckles in mice. RNA 18: 1487-1499.
11. Eißmann M, Gutschner T, Hämmerle M, Günther S, Caudron-Herger M, et al. (2012) Loss of the abundant nuclear non-coding RNA MALAT1 is compatible with life and development. RNA Biol 9: 1076-1087.
12. Zhang B, Arun G, Mao YS, Lazar Z, Hung G, et al. (2012) The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. Cell Rep 2: 111-123.
13. Gutschner T, Hämmerle M, Eissmann M, Hsu J, Kim Y, et al. (2013) The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. Cancer Res 73: 1180-1189.
14. Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. Science 327: 167-170.
15. Wiedenheft B, Sternberg SH, Doudna JA (2012) RNA-guided genetic silencing systems in bacteria and archaea. Nature 482: 331-338.
16. Mali P, Yang L, Esvelt KM, Aach J, Guell M, et al. (2013) RNA-guided human genome engineering via Cas9. Science 339: 823-826.
17. Cong L, Ran FA, Cox D, Lin S, Barretto R, et al. (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339: 819-823.
18. Cho SW, Kim S, Kim JM, Kim JS (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol 31: 230-232.
19. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, et al. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A 106: 9362-9367.
20. Kumar V, Westra HJ, Karjalainen J, Zhernakova DV, Esko T, et al. (2013) Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. PLoS Genet 9: e1003201.
21. Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, et al. (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell 153: 910-918.
22. Yang H, Wang H, Shivalila CS, Cheng AW, Shi L, et al. (2013) One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. Cell 154: 1370-1379.
23. Ding Q, Regan SN, Xia Y, Oostrom LA, Cowan CA, et al. (2013) Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. Cell Stem Cell 12: 393-394.
24. Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, et al. (2013) Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. Cell 154: 1380-1389.