

Non-coding RNA: It's Not Junk

C. Richard Boland¹

Published online: 7 March 2017

© Springer Science+Business Media New York 2017

Since the three-dimensional structure of DNA was discovered by Watson and Crick in 1953, progress in the understanding of the arrangement of the human genome occurred slowly at first. After the initial breakthrough in deducing the double-helical structure of DNA, it took 11 years to recognize and decipher the triplet code for amino acids and another 15 years to recognize that eukaryotic genes are interrupted by noncoding introns separating the exons. Although it gradually became apparent that protein-coding genes are like occasional charms on a very long bracelet, the reason why such long stretches of noncoding existed was unknown.

In the 1980s, progress accelerated. The identification and isolation of the first human genes by positional cloning, a tedious and challenging task, led DNA scientists to call for the sequencing of the entire human genome, the world's largest collaborative scientific project to date. The Human Genome Project (HGP) led to the first draft of our living instructions in 2001 [1], a task 99% completed by 2004 [2]. Although this monumental project enabled genome visualization at a very granular level, understanding the functional implications of its complexity has remained elusive.

It came as a major surprise even to those working on the HGP that only ~1.5% of the human genome encodes ~21,000 distinct protein-coding genes [3]. So, what is the rest of our DNA doing? Is it simply packaging material, like wrapping paper and styrofoam? Some initially referred to the noncoding DNA as “junk,” a concept that triggered skepticism by many observers. Many of the noncoding

sequences are repeated transposable (i.e., moveable) elements that facilitate genomic rearrangements, and are of evolutionary importance. Our genome is also packed with many types of tandemly repeated DNA sequences: ~45% consists of repetitive elements such as long terminal repeats (LTRs), long and short interspersed nuclear elements (called LINES and SINES, including as many as one million *Alu* repeats), and perhaps another 25% is made up of shorter tandem repeats such as satellites, minisatellites, and microsatellites [4]. Although, as stated, 98.5% of the human genome consists of non-protein-coding DNA sequences, most of the genome is transcribed into RNA—if at low level [5]. Our genome also encodes tens of thousands of long non-coding RNAs (lncRNAs), which are non-protein-coding transcripts of >200 nucleotides expressed at tenfold lower abundance than protein-encoding messenger RNAs (mRNAs), but appear to be functionally different than shorter RNA species such as microRNAs (miRNAs—we have a few thousand of these which regulate the transcription of mRNAs), short interfering RNAs (siRNAs, which are 20–25 bases long and inhibit gene expression), Piwi-interacting RNAs (piRNAs, 26–31 bases long and probably involved in gene silencing), and small nucleolar RNAs (snoRNAs which interact with proteins and other RNAs but with incompletely understood functions).

The existence of so many RNA species provokes the natural question: “What is all of this stuff doing?” From a structural perspective, it appears to be rather messy. In spite of this perception, the scientific community is gradually recognizing that many or all of these expressed RNAs may substantially regulate cellular functions and behavior. Importantly, some proportion of these transcripts regulate how much and what form of each gene is transcribed and translated into protein. Moreover, there are distinct patterns

✉ C. Richard Boland
crboland@ucsd.edu

¹ GI Division, UCSD School of Medicine, Gilman Drive,
San Diego, CA 92093, USA

of non-coding RNA expression associated with numerous disease states, cancer in particular.

One of the consequences of the near-universal transcription of our genome by the cell is that it might be exploitable for diagnostic or therapeutic purposes. Since cancer is a genetic disease (i.e., nearly all cancers are driven some number of somatic alterations in DNA), investigators have looked for telltale evidence of mutated DNA in the blood, stool, urine, and other tissues or body fluids of cancer patients. Each cell has only two DNA copies of each gene, limiting the ability to detect abnormal DNA from cancers. Nevertheless, when DNA is transcribed into RNA, the number of copies may be amplified, with some RNA released by the cell. Consequently, many translational researchers have observed telltale RNA species in a variety of clinical settings. Some of these RNAs are surprisingly stable in the blood and even the stool.

There are at least 2000 non-coding microRNAs that are important regulators of the stability and expression of intracellular mRNAs. They have promoters, start sites, and may be silenced by methylation, just like protein-coding genes. MicroRNAs can be packaged and released from tumor cells, and circulate in a relatively stable state as exosomes. Although their function as export products is still incompletely understood, specific patterns of microRNA expression have been linked to many different cancers. Consequently, microRNAs are increasingly being explored as biomarkers in the diagnosis of cancer, all the more attractive due to the potentially large number of copies, likely more promising than DNA biomarkers [6].

It follows naturally that other non-coding RNAs such as the lncRNAs should be investigated for specific cancer-associated patterns of expression and examined for possible diagnostic purposes. In this issue of *Digestive Diseases and Sciences*, Li et al. [7] reported their experience with lncRNAs in colorectal cancer (CRC). They extracted total RNA from four early-stage CRCs, profiling the RNA using a commercially available microarray that included 77,103 lncRNAs and 18,853 mRNAs. They compared these profiles with profiles from five samples of hyperplastic or inflammatory polyps. Although these were not ideal controls, the choice of tissue reflected the constraints the investigators faced from their institutional review boards, which prevented sampling of normal colon in these patients. They identified 3296 lncRNAs and 2711 mRNAs that were differentially expressed between the CRCs and the non-neoplastic polyps. Some were upregulated and others were downregulated, and a small proportion of these were differentially expressed by a factor of ≥ 10 , specifically 85 lncRNAs and 194 mRNAs.

Using bioinformatics to search for systematic patterns of expression, they identified five candidate lncRNAs that

were significantly upregulated, and five that were significantly downregulated in CRC. Confirming the differential expression of the lncRNAs using quantitative PCR assays, they synthesized their results into a lncRNA/mRNA co-expression network in order to illustrate the connectivity among these dysregulated genetic elements. From a functional perspective, the upregulated mRNAs in these networks were from genes involved with a variety of cell behaviors fundamental to cancer (DNA replication, apoptosis, etc.), whereas the downregulated mRNAs were involved in differentiated cell functions. They correlated these results with DNA mismatch repair activity (proficient vs. defective), with metastatic lesions, and with other pertinent clinical features. Although a focus of the work was to test the hypothesis that there might be different patterns of lncRNA expression in CRCs contingent on their DNA mismatch repair status, no clear-cut correlations were found.

While this study of lncRNA in CRC does not provide any clinically actionable results, it does represent a launching point for the exploration of an area likely to yield future insights into cancer pathogenesis. What are the functional and clinical implications of differential lncRNA expression? Liu et al. [8] have reported that the exosomal lncRNA *CRNDE-h* is elevated in the serum of patients with CRC, which correlated with the degree of metastasis. Sun et al. [9] mined the Gene Expression Omnibus public database that used lncRNA microarrays on cases of CRC, identifying 15 dysregulated lncRNAs, which they validated using quantitative PCR on 84 CRC samples. Wang et al. reported that the lncRNA *NNT-AS1* is upregulated in CRC tissues, its abundance correlated with metastasis, and its inhibition in laboratory models decreased the malignant behavior of the cells. In like fashion, the lncRNA *NEATI* is upregulated in several gastrointestinal cancers [10]. Additionally, Damas et al. [11] have reported that the lncRNA *SNHG5* was significantly upregulated in CRC tissues (as were the lncRNAs *ncRAN* and *GAS5*), and intermediately so in adenomas. Furthermore, when *SNHG5* was inhibited in cultured cells, 150 genes were downregulated, including genes in the growth-promoting *STAT* pathway by blocking downstream signal transduction. There are numerous other examples in the literature in which lncRNAs function like oncogenes, or at least as cooperative drivers of malignant behavior.

Adding to the possibilities, Han et al. [12] recently showed that *CRNDE* stimulates CRC cell proliferation and chemoresistance by way of regulating *miR-181a-5p* and *Wnt-beta-catenin* signaling. Adding to this, Gao et al. [13] have revealed that *CRNDE* functions like a miRNA sponge, downregulating *miR-136*, conferring resistance to the chemotherapeutic agent oxaliplatin. Indeed, it is difficult to disentangle the cooperative functions of these non-

coding RNA species in the regulation of cell function. A current PubMed search of “lncRNA” and “cancer” yielded >3200 citations, with the most appearing in the past few years.

It would appear that the scientific community is just beginning to understand the complex interplay among RNA species. One particularly exciting advance is the use of clustered regularly interspaced short palindromic repeat (CRISPR)-mediated genome editing approaches to explore lncRNA function on a genome-wide scale. Liu et al. [14] have a genome-scale systematic approach for inhibiting lncRNA expression in cancer cell lines and induced pluripotent stem cells (iPSCs) in order to determine functional outcomes. They created a CRISPR-mediated interference library that enabled them to downregulate 16,401 different lncRNA loci in their cell models. Inhibition of 499 of the lncRNAs perturbed transcriptional networks and reduced the rate of cell growth. Most of the lncRNA loci were distant from any protein-coding gene or known enhancer element. Nearly 90% of the knockdowns were specific to just one of the cell types (either a cancer cell line or iPSC), indicating the specificity of the growth regulatory networks—and the challenge ahead of us to utilize this approach therapeutically.

It hence appears prudent to closely monitor the development of new technologies involving lncRNAs. These non-protein-coding RNA species are ripe for development as cancer biomarkers and as mediators of cellular behaviors ranging from metastasis to resistance to chemotherapeutic agents. This is just the beginning of the next phase of the understanding of the anatomy and physiology of the human genome.

Acknowledgments This study was supported in part by NIH Grant CA72851.

Compliance with ethical standards

Conflict of interest The author has no conflicts to report.

References

1. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409:860–921.
2. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*. 2004;431:931–945.
3. Lander ES. Initial impact of the sequencing of the human genome. *Nature*. 2011;470:187–197.
4. Casa V, Gabellini D. A repetitive elements perspective in Polycomb epigenetics. *Front Genet*. 2012;3:199.
5. Birney E, Stamatoyannopoulos JA, Dutta A, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007;447:799–816.
6. Shigeyasu K, Toden S, Zumwalt TJ, Okugawa Y, Goel A. Emerging role of MicroRNAs as liquid biopsy biomarkers in gastrointestinal cancers. *Clin Cancer Res*. 2017. doi:10.1158/1078-0432.CCR-16-1676.
7. Li Q, Li N, Lao Y, et al. Variable levels of long noncoding RNA expression in DNA mismatch repair proficient early-stage colon cancer. *Dig Dis Sci*. (Epub ahead of print). doi:10.1007/s10620-017-4465-6.
8. Liu T, Zhang X, Gao S, et al. Exosomal long noncoding RNA CRNDE-h as a novel serum-based biomarker for diagnosis and prognosis of colorectal cancer. *Oncotarget*. 2016;7:85551–85563.
9. Sun X, Hu Y, Zhang L, et al. Mining, validation, and clinical significance of colorectal cancer (CRC)-associated lncRNAs. *PLoS One*. 2016;11:e0164590.
10. Xiong DD, Feng ZB, Cen WL, et al. The clinical value of lncRNA NEAT1 in digestive system malignancies: a comprehensive investigation based on 57 microarray and RNA-seq datasets. *Oncotarget*. 2017. doi:10.18632/oncotarget.14756.
11. Damas ND, Marcatti M, Come C, et al. SNHG5 promotes colorectal cancer cell survival by counteracting STAU1-mediated mRNA destabilization. *Nat Commun*. 2016;7:13875.
12. Han P, Li JW, Zhang BM, et al. The lncRNA CRNDE promotes colorectal cancer cell proliferation and chemoresistance via miR-181a-5p-mediated regulation of Wnt/beta-catenin signaling. *Mol Cancer*. 2017;16:9.
13. Gao H, Song X, Kang T, et al. Long noncoding RNA CRNDE functions as a competing endogenous RNA to promote metastasis and oxaliplatin resistance by sponging miR-136 in colorectal cancer. *Onco Targets Ther*. 2017;10:205–216.
14. Liu SJ, Horlbeck MA, Cho SW, et al. CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science*. 2017;355. doi:10.1126/science.aah7111.