ASIP OUTSTANDING INVESTIGATOR AWARD LECTURE

Biological Processes Discovered by High-Throughput Sequencing

Brian J. Reon* and Anindya Dutta*[†]

From the Departments of Pathology* and Biochemistry and Molecular Genetics,[†] University of Virginia School of Medicine, Charlottesville, Virginia

Accepted for publication
October 30, 2015.

Address correspondence to
Anindya Dutta, M.D., Ph.D.,
University of Virginia School of
Medicine, 1340 Jefferson Park
Ave, Jordan Hall, Room 1232,
Charlottesville, VA 22908-
0733. E-mail: ad8q@virginia.edu.

Advances in DNA and RNA sequencing technologies have completely transformed the field of genomics. High-throughput sequencing (HTS) is now a widely used and accessible technology that allows scientists to sequence an entire transcriptome or genome in a timely and cost-effective manner. Application of HTS techniques has led to many key discoveries, including the identification of long noncoding RNAs, microDNAs, a family of small extrachromosomal circular DNA species, and tRNA-derived fragments, which are a group of small non-miRNAs that are derived from tRNAs. Furthermore, public sequencing repositories provide unique opportunities for laboratories to parse large sequencing databases to identify proteins and noncoding RNAs at a scale that was not possible a decade ago. Herein, we review how HTS has led to the discovery of novel nucleic acid species and uncovered new biological processes during the course. (*Am J Pathol* 2016, 186: 722–732; <http://dx.doi.org/10.1016/j.ajpath.2015.10.033>)

High-throughput sequencing (HTS) technologies have revolutionized our understanding of biological systems and opened the path to investigate many unappreciated biological phenomena. When HTS first became available for widespread use, it cost approximately \$10 million to sequence a single human genome. However, advances in sequencing technologies during the past decade have made HTS much more cost-effective, and newer technologies boast the ability to sequence \$1000 genomes. As a consequence of the massive cost reductions, use of the technology has become ubiquitous. HTS is a versatile technology and is amenable to almost any scalable nucleic acid technique. Like any technique, HTS has its own drawbacks that can limit or even yield misleading conclusions, including GC read bias and batch effects.^{1,2} Although one must take caution when analyzing and comparing HTS data, these technical biases can, in large part, be corrected for.

HTS has been used to examine a wide variety of biological questions, including mapping the global three-dimensional organization of entire genomes and transcriptome of single cells.^{3–5} Furthermore, pan-genomic analyses have

led to critical insights into tumor biology and determinants of chemotherapeutic resistance.^{6–8} Throughout the years, much of the data produced by HTS has been deposited to public data repositories.⁹ The amount of publicly available data is staggering and has served as a rich resource for the scientific community to address many hypotheses that go beyond the capacity of any individual laboratory. In this review, we highlight three examples of exploiting HTS technology and publicly available data to identify novel nucleic acids and transcripts, which uncover novel biology.

Supported by NIH grants R01 CA60499, R01 CA166054, R01 GM084465, and P01 CA104106 Project 3.

Disclosures: None declared.

The ASIP Outstanding Investigator Award is given by the American Society for Investigative Pathology to a mid-career investigator with demonstrated excellence in research in experimental pathology at the time of the award. Anindya Dutta, recipient of the 2015 ASIP Outstanding Investigator Award, delivered a lecture entitled “High-Throughput Sequencing for Discovering New Biology” on March 28, 2015, at the annual meeting of the American Society for Investigative Pathology in Boston, MA.

Figure 1 microDNA detection and microDNA-mediated microdeletions. **A:** Rolling circle amplification of microDNAs. Inner black circle is a microDNA that is being amplified by rolling circle amplification. Each color represents an amplified copy of the microDNA, with the **arrowheads** denoting junction sites. The paired-end sequencing tag illustrates the mapping strategy, with the chimeric red and blue read mapping to a junction site, for high-throughput sequencing of microDNA libraries. **B:** Typical Sanger sequencing results from microDNA rolling circle amplification (RCA) products. Junction sites from the RCA products are highlighted. **C:** Model of microDNA production leading to a genomic microdeletion. **Arrows** represent 2- to 15-bp direct repeat, which, through possible repair mechanisms, can lead to the excision of a microDNA (genomic region between the direct repeats) and a corresponding microdeletion in the genome.

MicroDNA

Discovery and Attributes of MicroDNA

Our post-zygotic genomes were long thought to be stable; however, we now know that through errors in DNA replication and repair, there are considerable levels of genetic mosaicism between somatic cells.^{10,11} A lesser-appreciated source of somatic mosaicism and copy number variation is a ubiquitous pool of nonlinear extrachromosomal DNA, termed extrachromosomal circular DNA (eccDNA).¹² eccDNAs are mostly derived from genomic tandem repeats, including ribosomal RNA and satellite repeats, and in humans, range in size from a few kb to >20 kb in length.¹³ These circular DNAs constitute a significant pool of non-chromosomal genetic information. Recently, a novel subclass of eccDNAs, microDNAs, was discovered using a slightly modified eccDNA extraction protocol.¹⁴

MicroDNAs were first discovered by Sanger sequencing of cloned multiple displacement amplification products from eccDNAs (Figure 1, A and B). The sequenced microDNAs, derived from rolling circle amplification products, revealed a characteristic pattern of several hundred bp direct repeats, consistent with multiple displacement amplification products from a circular template. Interestingly, microDNAs are much smaller (<1 kb) in length than previously characterized eccDNAs.

Further characterization of microDNAs and other eccDNAs was hampered by the low throughput of traditional screening techniques, relying mainly on Sanger sequencing of cloned circular DNAs and Southern blots using probes against repetitive elements.¹² Recently, HTS of rolling circle amplification products from microDNAs was performed to get the first global genomic view of mammalian eccDNAs.¹⁴ Using a multifaceted computational algorithm, sequencing reads derived from microDNAs were identified, in part, by finding sequencing reads that span the junction of rolling circle

amplification products (Figure 1A). MicroDNAs range from 100 to 400 bp in length, with two major size peaks at approximately 200 and 400 bp. These size estimates were independently confirmed using electron microscopy, which also demonstrated that microDNAs are a heterogeneous mixture of both single- and double-stranded circles. Surprisingly, unlike previously described eccDNAs, microDNAs tend to derive from nonrepetitive regions of the genome. Most microDNAs have short, 2 to 15 bp, stretches of microhomologous direct repeats flanking the microDNA site of origin (Figure 1C). Furthermore, microDNAs are preferentially derived from genic regions, especially from exons and 5' untranslated regions (5'-UTRs), and also have a higher GC content than the genomic average.

Since their initial discovery in embryonic mouse brains, microDNAs have been observed in every biological system tested, including human cancer cell lines, a variety of other mouse tissues, and *Drosophila* S2 cells. Furthermore, HTS has revealed that the cellular pool of microDNAs is surprisingly complex, covering between 10,000 and 40,000 unique sites per mouse tissue, 40,000 to 80,000 sites in cancer cell lines, and up to 100,000 unique sites in chicken DT40 cells.¹⁵ Within a given species, there are common sites in the genome that produce microDNAs (microDNA hotspots) irrespective of cell/tissue lineage or developmental stage. MicroDNA hotspots have both high GC content and high gene density relative to the genomic average. The fact that microDNAs are continually produced from microDNA hotspots suggests that microDNAs are not produced as mere by-products of DNA metabolism or other stochastic events. Interestingly, there are many examples of microDNAs that are only found in a particular cell line or tissue. Indeed, a panel of five prostate and ovarian cancer cell lines could be separated into their respective groups purely on the basis of hierarchical clustering of unique microDNAs, suggesting that lineage-specific factors or chromatin state affect microDNA production. Many of these differences appear to

be linked to the transcriptional landscape of the cell, because microDNAs are heavily enriched in active promoters and RNA polymerase II binding sites. Furthermore, microDNAs are disenriched in laminin-associated domains, which are regions of the genome that interact with the nuclear lamina and are associated with a heterochromatic state.¹⁶ In addition to being associated with sites of active chromatin, microDNAs are enriched in genes with a high density of exons, further suggesting a connection to transcriptional processes.¹⁵

Possible Mechanisms of MicroDNA Biogenesis

Because of the fact that the genomic source of microDNAs is flanked by tandem sequence repeats, it has been suggested that they are produced as a by-product of intrachromosomal homologous recombination between direct repeats or slippage of DNA polymerase at the direct repeats during replication or repair.¹² As mentioned earlier, microDNAs have two major size peaks of 200 and 400 bp, which is approximately the expected size of DNA wrapped around mononucleosomes and dinucleosomes, respectively. In addition to GC richness, many microDNAs also display AA, AT, or TT dinucleotides at regular intervals, which is reminiscent of the pattern of genomic sequences that preferentially associate with nucleosomes.^{17,18} However, more work is needed to examine if there is any role of nucleosome deposition in the production of microDNA.

The clear association between sites of microDNA production and active transcription and high exon density makes it tempting to speculate on the role that transcription-associated DNA damage might have on microDNA biogenesis. R-loops are triple-stranded RNA/DNA hybrids that

consist of displaced single-stranded loops of DNA resulting from the base pairing of nascent RNA with template DNA during transcription.¹⁹ R-loops are known to lead to DNA damage and genomic instability.²⁰ Similar to microDNA, R-loops typically occur in G-rich DNA and commonly form near 5' and 3' UTRs.^{21,22} Splicing factor activity has also been shown to play an important role in minimizing the formation of R-loops.^{23,24} The striking similarities between regions that produce R-loops and microDNAs suggest that microDNAs might be produced from the displaced single-stranded DNA from unresolved R-loops (Figure 2A).

Although microDNA production is correlated with nucleosome and R-loops, our best understanding of how microDNAs are produced comes from measuring microDNA abundance and complexity from a variety of DNA repair knockout DT40 cell lines.¹⁵ All major DNA repair pathways, homologous recombination (BRCA1, BRCA2, CtIP, NBS1, and Rad54), nonhomologous end joining (Ku70, Lig4, and NBS1), and mismatch repair (MSH3), have been assessed for their ability to alter microDNA levels. Although no single DNA repair mutant completely abolished microDNA production, the mismatch repair mutant, MSH3^{-/-}, had an 80% reduction in the amount of microDNAs, with no changes in the size of the microDNAs. The mismatch repair mutant was also the only DNA repair mutant that had changes in the characteristics of the microDNAs (namely, an increase in the percentage of microDNAs arising from CpG islands).

One possibility is that microDNAs are produced during the repair of replication slippage events. Replication slippage can occur in regions of the genome with short direct repeats and produces a single-stranded DNA loop as a repair intermediate, which can be resolved by the mismatch repair

Figure 2 Generation of single-stranded (ss) or double-stranded (ds) microDNA. **A:** RNA polymerase transcribing across GC-rich DNA can lead to the formation of R-loops. Excision and ligation of the R-looped region releases ssDNA-RNA hybrid that may produce microDNA. **B: Left panel:** During DNA replication, replication slippage can occur at short direct repeats, which leads to the formation of a DNA loop. Depending on whether the looping occurs on the newly synthesized nascent DNA or the template strand, excision of the DNA loop (ss microDNA) could lead to a genomic deletion. **Right panel:** After a DNA break or stalling of the replication fork, small regions of microhomology on nascent strand DNA could facilitate its circularization, and the formation of ss microDNA with no reciprocal genomic deletion. **C:** During the repair of two dsDNA breaks, microhomology-mediated circularization of the intervening DNA sequence could lead to the production of a ds microDNA and a genomic microdeletion. Figure modified from Dillon et al,¹⁵ with permission from *Cell Reports*.

pathway.²⁵ Ligation of the looped DNA would lead to the formation of a single-stranded microDNA that could be converted to a double-stranded microDNA through some unknown mechanism (Figure 2B). It is also possible that microDNAs are formed by circularization of nascent DNA, bridged by microhomology, at stalled replication forks (Figure 2B). These examples might suggest that DNA replication is essential to produce microDNAs. However, microDNAs are also found at high abundance in nondividing tissues, such as adult mouse brains. This suggests that some non-replication-dependent repair process that also involves DNA synthesis mediates the production of some microDNAs. Another possibility is that in post-replicative cells, microDNAs are formed from two adjoining DNA ds breaks, with microhomology-mediated circularization of the intervening DNA (Figure 2C). Future mechanistic studies are needed to elucidate what contribution, if any, the previously mentioned repair pathways have toward producing the cellular pool of microDNA.

Implications of MicroDNA

MicroDNAs are a significant source of extrachromosomal genetic material, representing approximately 10 to 50 kb of additional information per cell. One of the biggest questions regarding microDNAs is as follows: Do microDNAs have a function or are they simply a by-product of nucleotide metabolism? Although the functionality of microDNA has yet to be demonstrated, several intriguing observations point to possible ways in which microDNA could affect cellular homeostasis. First, it is well known that many transcription factors and chromatin-modifying complexes bind at the 5' ends of genes and regulate their expression. Because many microDNAs originate from the 5' UTRs of genes, it is possible that transcription factors and other protein complexes would also bind their recognition sites on microDNAs. Indeed, several transcription factor binding motifs are enriched in microDNA (P. Kumar and A. Dutta, unpublished data). Recent work has shown that there is a balance between a limited pool of transcription factors and their competing binding sites, and perturbations to this homeostasis can affect transcription from certain genes.²⁶ By titrating away transcription factors, microDNAs could influence the balance of a transcriptional program, and genes with weaker transcription factor binding sites might not be activated.

Small DNA circles lacking a promoter have been shown to be transcribed both *in vitro* and in human cells, in a RNA polymerase III-dependent manner.²⁷ MicroDNAs are approximately the same size as the DNA circles that were transcribed in the previously mentioned study, which were approximately 120 bp. It is conceivable that microDNA could also serve as a template for promoterless transcription. Given the relatively limited number of microDNAs in an individual cell, transcription of microDNAs could serve as an amplification step, allowing for more robust effects.

Furthermore, transcripts from microDNAs originating from genic sites could act as sponges for miRNA targeting. Alternatively, because microDNAs are enriched from regions with multiple splicing events, another possibility is that transcripts from microDNAs could sponge splicing factors, a phenomenon that has already been shown to occur when ribosomal proteins are up-regulated.²⁸ Although more work is needed to tease out any functional role for microDNA, this exciting area of research could uncover new or unappreciated biology.

Extracellular nucleic acids were first discovered in human plasma in the late 1940s.²⁹ Since then, much emphasis has been placed on the potential of circulating nucleic acids in the serum or plasma to serve as biomarkers for various human pathologies.^{30,31} An interesting possibility is that microDNA could be used as a novel nucleic acid biomarker. As mentioned earlier, microDNA production is correlated with transcriptional activity and clustering of microDNAs alone is enough to correctly segregate tumor cell lines by tumor type. Furthermore, microDNA would presumably be more stable than circulating RNA and linear DNA, which are susceptible to degradation by serum exonucleases. It will be interesting to see if future studies are able to detect microDNA in human serum or if microDNA quantity or complexity is correlated with cancer or other pathologies.

tRFs

Discovery of tRFs

Since their discovery in the early 1990s, miRNAs have reformed how we view post-transcriptional regulation. miRNAs are a class of small noncoding RNAs, approximately 20 nucleotides (nt) in length, that primarily regulate protein levels through base pairing with a transcript's 3' UTR. Mediated by the RNA-induced silencing complex, miRNAs lead to the sequestration or degradation of the bound mRNAs.³² Although initial progress into cataloging an organism's miRNA repertoire was slow, microarrays and small RNA-seq greatly expanded our understanding of small RNA biology. Small RNA-seq generates tens of millions of reads that, among other things, are used to quantitate small RNA abundances. During the initial read mapping steps, many reads are discarded because they map to multiple genomic positions, or to abundant RNAs, such as tRNAs and rRNAs. Because of read censoring, whole classes of small noncoding RNAs have gone unmeasured and undiscovered. Recently, several groups independently examined the discarded pool of sequencing reads and discovered that many of the abundant non-miRNA reads mapped to tRNA loci, and represented a novel class of highly abundant small noncoding RNAs, termed tRNA-derived fragments (tRFs).^{33–35} Although surprising, the presence of tRNA fragments is certainly not unprecedented, because others have shown that during cellular

Figure 3 tRNA-derived fragment (tRF) biogenesis. **Left panel:** Precursor tRNAs are processed by RNase P and RNase Z to remove trailer and leader sequences, CCA-adding enzyme and removal of the tRNA intron (not shown) by tRNA splicing endonuclease, leading to the production of a mature tRNA (**center panel**). **Right panel:** In a Drosha- and Dicer-independent manner, unknown nucleases cleave the D- and T-loops, leading to the production of tRF-5s and tRF-3s, respectively. tRF-1s are by-products of 3'-trailer cleavage.

stress, tRNAs are cleaved in the anti-codon loop by angiogenin, forming tRNA halves.³⁶

Characteristics and Biogenesis of tRFs

The argument could be made that small fragments of tRNA present in small RNA-seq data are simply a reflection of the high abundance of tRNAs and smaller reads are just by-products of random tRNA degradation. However, if the small RNA-seq reads that mapped to tRNAs were background noise, one would expect to see read coverage over the entirety of the tRNA, which is not what is seen. tRFs are present as precisely defined fragments, originating from the 5' and 3' ends of the mature tRNA and the 3' trailer sequence of the premature tRNA, classified as tRF-5, tRF-3, and tRF-1, respectively. The tRF-5 class of tRFs is composed of three major size groups of 15, 22, and 32 nt, where the cleavage sites occur in the D-loop, D arm, and anti-codon arm of the tRNA, respectively. tRF-3s are primarily detected as 18 and 22 nt RNAs that arise from the T-loop of the mature tRNA. Conversely, tRF-1s range from 15 to 22 nt and begin at the cleavage site of the 3' trailer sequence and end at an RNA polymerase III transcription termination signal (Figure 3). These size classes are aggregate observations for all tRFs within a tRF family. However, each tRF family derived from a parental tRNA is mostly present as a specific size peak. For example, >90% of the tRF-5s derived from the tRNA GlyGCC are 22 nt, suggesting that tRFs are not the by-products of random exonucleolytic cleavage of tRNAs. Furthermore, individual tRF families from a specific tRNA

are not present at equal abundances, as would be expected for RNA being produced as random degradation products. Interestingly, tRFs are ubiquitous and present in every cancer cell line and mouse tissue that has been examined. More surprisingly, tRFs are evolutionarily conserved beyond humans and mice, and are present in *Drosophila*, yeast, and bacteria.³⁷

miRNAs are processed by the cleavage of primary miRNA transcripts by Drosha in the nucleus, forming a precursor miRNA, which is then cleaved in the cytoplasm by Dicer, forming a mature miRNA.³⁸ Although the levels of miRNAs are highly dependent on Drosha and Dicer, tRF production appears to be independent of these nucleases. Indeed, small RNA-seq from mouse embryonic stem cells with deletions in Dicer or DGCR8, a binding partner of Drosha in the microprocessor complex, had decreased levels of miRNAs but no decrease of tRFs. Although most tRFs are Dicer independent, previous work has shown that the biogenesis of individual tRFs is dependent on Dicer.³⁴ It is still not clear what proteins are involved in processing most tRFs, and whether the proteins that produce tRF-5s are the same for tRF-3s. Also, given that the tRFs from a common tRNA are not present at equal abundance, how does the cell decide which tRF to produce?

Function of tRFs

As the field of small RNA biology has progressed, more focus has been placed on designing computational and experimental methods to discover the targets of small RNAs and their protein-binding partners. Recently, several

experimental methods have been developed that not only measure the small RNAs that interact with a specific protein, but also identify the bound target RNAs.³⁹ One of the methods, photoactivatable ribonucleoside-enhanced cross-linking and immunoprecipitation (PAR-CLIP), is a technique that uses a photoactivatable nucleoside, 4-thiouridine, to label RNAs in the cell to allow for efficient cross-linking of RNAs to proteins after exposure to UV light. Proteins of interest can be pulled down, and the associated RNAs can be sequenced. This technique has been used to identify the RNA interactome of multiple RNA-binding proteins, including AGOs 1, 2, 3, and 4.³⁹

Although the initial PAR-CLIP studies were mainly focused on miRNAs, groups have reanalyzed the PAR-CLIP data to determine whether tRFs are associated with the AGO proteins.^{37,40} Interestingly, tRF-5s and tRF-3s associate with AGO 1, 3, and 4, but not AGO 2. Also, tRF-1s rarely associate with any of the AGO proteins, suggesting that the interactions with tRF-5s and tRF-3s are specific. In an independent study, using RNA immunoprecipitation, a separate group also found that a tRF-3, CU1276, preferentially interacted with AGO 1, 3, and 4 compared with AGO 2.⁴¹

During cDNA synthesis, a cross-linked 4-thiouridine can erroneously be reverse transcribed as a guanidine.³⁹ Because the C/T mutation only occurs at cross-linked sites, the mutation frequency can be used to discern true associations, as well as the structural organization of the RNA and protein complex. Using the C/T mutation frequency, tRFs have been found to interact with AGO proteins in a similar manner to miRNAs (eg, the sites with the highest C/T mutation frequency were the same and spares the 5' seed sequence used to base pair with the target), and with a similar cloning frequency to miRNAs. PAR-CLIP has the added benefit that the identity of targets can be inferred by locating the sites of highest C/T mutation frequency and the surrounding 40 nt within the target RNA, termed cross-linked-centered regions. miRNA seeds (first 6 to 8 nt of a miRNA) are most commonly located 1 to 2 nt directly downstream of the major cross-linked position in cross-linked centered regions.^{39,42} Surprisingly, tRF seeds have similar matching profiles as miRNAs, further suggesting that tRFs are functional and interact with mRNAs.³⁷

A separate technique, cross-linking, ligation, and sequencing of hybrids (CLASH), is able to capture small RNAs and their cellular targets by forming intermolecular RNA ligation products in purified ribonucleoprotein complexes.⁴³ HTS of CLASH products has identified tens of thousands of miRNA-mRNA chimeras,⁴⁴ and more recently, tRFs were found to form chimeras in AGO1 CLASH data.³⁷ In fact, more tRF-3-mRNA chimeras were present in AGO1 CLASH than miRNAs, even though miRNAs were more abundant.³⁷

Knockdown of tRF-1001, derived from tRNA Ser-TGA, causes cells to delay passage through G₂, and this function of tRF-1001 was dependent on the sequence of the tRF.³³ A

tRF-3 derived from tRNA Gly-GCC was decreased in diffuse large B-cell lymphoma, and repressed the replication and repair factor RPA1.⁴¹ A tRF-5 derived from tRNA Glu-CTC is induced by respiratory syncytial virus to suppress the cellular gene and antiviral protein APOE2.⁴⁵ Other groups have reported that tRFs act in canonical miRNA pathways. A tRF-3, CU1276, derived from the tRNA-GlyGCC, targets the 3' UTR of RPA1 in an AGO-dependent manner, resulting in the decrease of RPA1 protein levels. In addition, when CU1276 is expressed in B-cell lymphoma cell lines, cell growth is suppressed.⁴¹ Despite these examples, the functional consequence of tRFs interacting with mRNAs is largely unknown, but these recent observations of tRF-associated targets in PAR-CLIP and CLASH data strongly suggest that tRFs are playing a functional role in the cell.

Future Considerations

Although much of tRF biology remains unknown, several groups have begun to find unique and unexpected roles for tRFs. One study found that several tRFs, originating from the anti-codon loop, were induced in breast cancer cell lines after hypoxia. Furthermore, some of these tRFs interact with YBX1, an oncogenic RNA-binding protein, preventing YBX1 from binding and stabilizing other oncogenic mRNAs, such as AKT1 and EIF4G1, resulting in their down-regulation and a decrease in cell growth and invasiveness.⁴⁶ In *Haloflex volcanii*, it has been shown that a stress-induced tRF-5 binds to the ribosome, and through interactions with the small ribosomal subunit, interferes with peptidyl transferase activity, resulting in a decrease in protein synthesis.⁴⁷

These studies highlight the diverse nature of tRF functions and, as the field of tRF biology continues to grow, more activities will surely be discovered. It is important to take caution when studying tRFs, however, because some of the traditional methods used to study miRNAs, such as siRNAs and shRNAs, may not be suitable for studying tRFs. siRNAs and shRNAs with tRF sequences will be loaded in AGO2 containing RNA-induced silencing complex, and because most tRFs do not preferentially interact with AGO2 compared with other AGO proteins, could lead to phenotypes that do not reflect the actual tRF biology. As new work continues to expand our understanding of tRFs, new tools are needed to help researchers find biologically relevant tRFs to study and identify possible targets of tRFs. One resource we have generated is a tRF database, analogous to miRBase for miRNAs, called tRFdb.^{48,49} tRFdb provides expression data for tRF-5s, tRF-3s, and tRF-1s from more than 100 RNA-seq libraries spanning eight different species, ranging from humans and mice to bacteria. This resource can serve as a starting point for scientists looking for tRFs possibly involved in various pathologies and normal biology.

Given the nonrandom manner in which tRFs are produced, it is possible that tRFs, similar to other RNA species, could serve as novel clinical biomarkers. For example, recent work has shown that a specific tRF-3 is strongly expressed in normal germinal center B cells but is nearly undetectable in B-cell lymphomas derived from germinal centers.⁴¹ More systematic analysis is needed to determine whether other tRFs may serve as biomarkers.

Many more questions still remain about tRF biology. tRF-1s do not appear to associate with AGO proteins, and yet tRF-1001 was among the first to have demonstrated biological function. How do these tRFs act? Furthermore, tRNAs are known to be heavily modified post-transcriptionally, and tRF-5s and tRF-3s most likely share their parental tRNA modifications.⁵⁰ It has not yet been addressed whether tRFs are modified, and if tRF modifications are present, how does it affect tRF function, such as changing target-binding specificity or the kinetics of tRF activity. Also, because Drosha and Dicer do not process tRFs, how are tRFs properly loaded on the AGOs 1, 3, and 4? Although a myriad of questions remain, it is clear that tRFs represent a unique and abundant pool of small RNAs, which might play a significant regulatory role in normal and diseased cells.

lncRNAs

Discovery and Function of lncRNA

The advent of HTS transformed basic science research and, in particular, truly revolutionized the noncoding RNA (ncRNA) field. Through the ENCODE (Encyclopedia of DNA Elements) project and other global sequencing initiatives, it became clear that up to 76% of the mappable human genome is transcribed and a large number of these transcripts do not code for proteins.^{51–53} Although the exact number of ncRNAs is debated, these sequencing efforts have shown that ncRNAs are ubiquitously expressed, with numbers rivaling those of coding RNAs. As the examples of ncRNAs increased, they began to be subdivided into distinct categories. One subcategory is long ncRNAs (lncRNAs), somewhat arbitrarily defined as RNA species >200 bp, which do not code for proteins.⁵⁴ lncRNAs can be further subdivided on the basis of their relative position to neighboring genes. lncRNAs that are transcribed from the opposite strand of a protein-coding gene are called antisense lncRNAs, lncRNAs located within the introns of protein-coding genes are termed intronic lncRNAs, and lncRNAs that lie in regions of the genome between genes are called long, intervening, ncRNAs. As the number of lncRNAs began to increase, several skeptics pointed out that a large fraction of the ncRNA transcripts identified in high-throughput data sets were not evolutionarily conserved and were likely transcriptional noise, having no functional role in cells.⁵⁵ This has since been disproved, because many lncRNAs share the same active chromatin marks as

protein-coding genes, are evolutionarily conserved more than introns, but less than protein-coding gene exons, and, in many instances, have been found to play critical roles in normal and disease physiology.^{56–58} Initially, HTS helped the discovery and cataloging of lncRNAs; now, with the plethora of publically available sequencing data, scientists can rapidly identify lncRNAs and other genes of interest by making novel comparisons between pre-existing data sets. One example from our laboratory was the identification of a lncRNA that is decreased during prostate cancer progression, DRAIC (down-regulated RNA in cancer, inhibitor of cell invasion and migration), which we will discuss below.⁵⁹ We have identified other functional lncRNAs, such as MyoD upstream noncoding as a myogenic differentiating factor,⁶⁰ Alu containing p21 transcriptional repressor as a promoter of cell proliferation,⁶¹ and H19 as another promyogenic factor that acts as a primary transcript for promyogenic miRNAs miR-675-5p and miR-675-3p.⁶²

Similar to proteins, lncRNAs have modular domains that are capable of independently interacting with proteins through electrostatic interactions and nucleic acids through base pairing. Many lncRNAs have been shown to regulate gene expression in *cis* or in *trans*, through the recruitment of chromatin-modifying complexes. HOTAIR (HOX transcript antisense RNA) is a lncRNA that is overexpressed in aggressive breast cancer and functions by recruiting the polycomb repressive complex 2 to target gene loci, driving the tumor cells to a less differentiated state.⁶³ Although many of the characterized lncRNAs function in the nucleus, there are many examples of lncRNAs that function in diverse cellular processes in the cytoplasm. Uchl1-as is an antisense lncRNA that partly overlaps with the protein-coding gene Uchl1. Under stress conditions, Uchl1-as base pairs with a complementary region on Uchl1, which enhances translation of the Uchl1 mRNA through an embedded inverted SINEB2 element on Uchl1-as.⁶⁴ Other cytoplasmic lncRNAs have been found to act as miRNA sponges and sequester miRNA from targeting protein-coding genes.⁶⁵

Discovery and Function of DRAIC

Prostate cancer is the most common cause of male cancer in the developed world. Although the 10-year survival for all prostate cancer is >80%, advanced metastatic disease results in >30,000 deaths per year. In patients requiring systemic treatment, androgen-deprivation therapy typically results in a rapid clinical response, although most tumors will eventually recur, becoming castration-resistant prostate cancer. It is now appreciated that a sizable fraction of tumors that no longer respond to androgen-deprivation therapy remain reliant on androgen receptor signaling.⁶⁶ Although much work has been directed toward understanding prostate cancer pathogenesis, the drivers and suppressors of prostate cancer progression are not fully understood. In addition to canonical cancer signaling

pathways, several lncRNAs have been shown to be involved in prostate cancer progression, including SChLAP1 (second chromosome locus associated with prostate-1), PCAT29 (prostate cancer associated transcript 29), NEAT1 (nuclear paraspeckle assembly transcript 1), and DRAIC.^{59,67–69} DRAIC was discovered by identifying differentially expressed genes between LNCap and C4-2B cells, androgen-sensitive and castration-resistant cell lines, respectively.

We also screened for differentially expressed genes that followed the same androgen dependence trend in LNCap cells treated with an androgen analog, R1881. DRAIC was found to be repressed by R1881 and low in C4-2B cells. In addition, DRAIC is repressed by R1881 in a dose- and time-dependent manner. Furthermore, DRAIC is much lower in castration-resistant prostate cancer cell lines compared with androgen-dependent cell lines. As one might expect, there are many androgen-dependent androgen receptor binding sites near DRAIC. Two other transcription factors, FOXA1 and NKX3-1, have been identified that have binding sites flanking the DRAIC locus and positively regulate its expression. In primary patient samples, DRAIC was also found to be down-regulated in patients 22 weeks after androgen-deprivation therapy, suggesting that the suppression of DRAIC expression could be an earlier indicator or

facilitator of castration-resistant prostate cancer. Consistent with this theory, in a cohort of advanced prostate cancer patients, patients with lower expression of DRAIC have decreased disease-free survival. Notably, this trend extends beyond prostate cancer, and DRAIC down-regulation is strongly associated with overall and disease-free survival in a large variety of cancers, including renal clear-cell carcinoma and lower-grade gliomas.⁵⁹

Interestingly, a decrease in DRAIC expression changes cell morphology to a mesenchymal shape, resembling a classic epithelial-to-mesenchymal transition. Consistent with these observations, even though knockdown of DRAIC slightly decreases cell growth, knockdown of DRAIC causes a dramatic increase in prostate cancer's migratory and invasive abilities, consistent with the decrease of DRAIC being associated with poor clinical outcomes (Figure 4). One particularly interesting observation is that the expression patterns of DRAIC are similar to those of an adjacent lncRNA, PCAT29, which has previously been shown to repress cancer cell migration and invasion *in vitro* and *in vivo*.⁶⁸ Similar to DRAIC, PCAT29 has binding sites for androgen receptor, FOXA1, and NKX3-1 in its promoter and surrounding area. The only reported difference between the two lncRNAs is that DRAIC is predominately located in the cytoplasm, whereas PCAT29 is mostly found in the

Figure 4 Regulation and action of down-regulated RNA in cancer, inhibitor of cell invasion and migration (DRAIC)/PCAT29. In androgen-dependent (AD) cells, high levels of FOXA1 and NKX3-1 override the repressive action of the androgen receptor (AR), leading to the high expression of both DRAIC and PCAT29. As prostate tumors progress to become castration resistant (CR), FOXA1 and NKX3-1 expression is lost and activated AR pathways are able to strongly suppress the expression of DRAIC and PCAT29, leading to an increase in a tumor's invasiveness and a decrease in patient disease-free survival. Figure adapted from Sakurai et al,⁵⁹ with permission from *Molecular Cancer Research*. ADT, androgen deprivation therapy; Chr., chromosome; lncRNA, long noncoding RNA.

nucleus. The DRAIC/PCAT29 regulon was one of the first examples of neighboring, jointly regulated lncRNAs affecting tumor pathogenesis in a concerted pathway. One of the main questions remaining about the DRAIC/PCAT29 nexus is in what ways do they mechanistically affect a cell's migratory potential. In general, some degree of mechanistic insight is gleaned from a lncRNA's subcellular localization. So, it will be interesting to see if cytoplasmic DRAIC and nuclear PCAT29 function in similar pathways.

Concluding Remarks

Since the inception of HTS, progress in the field of genomics has been staggering, and HTS has been a key driving force for discoveries in many biological processes. We hope that this review illustrates the power of HTS to not only measure a wanted experimental outcome but also to uncover unique biological processes. Indeed, whether it is using unique HTS methods to detect novel nucleic acid species, such as microDNAs, or using pre-existing HTS data to uncover novel RNAs, HTS has and will continue to be a powerful and versatile technique that expands our understanding of biology.

References

1. Minoche AE, Dohm JC, Himmelbauer H: Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol* 2011, 12:R112
2. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, Geman D, Baggerly K, Irizarry RA: Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* 2010, 11:733–739
3. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J: Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 2009, 326:289–293
4. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA: Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 2015, 161:1202–1214
5. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW: Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 2015, 161:1187–1201
6. Gonzalez Bosquet J, Marchion DC, Chon H, Lancaster JM, Chanock S: Analysis of chemotherapeutic response in ovarian cancers using publicly available high-throughput data. *Cancer Res* 2014, 74:3902–3912
7. Hoadley KA, Yau C, Wolf DM, Cherniack AD, Tamborero D, Ng S, Leiserson MDM, Niu B, McLellan MD, Uzunangelov V, Zhang J, Kandoth C, Akbani R, Shen H, Omberg L, Chu A, Margolin AA, van't Veer LJ, Lopez-Bigas N, Laird PW, Raphael BJ, Ding L, Robertson AG, Byers LA, Mills GB, Weinstein JN, Van Waes C, Chen Z, Collisson EA, Benz CC, Perou CM, Stuart JM: Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* 2014, 158:929–944
8. The Cancer Genome Atlas Research Network: Integrated genomic characterization of papillary thyroid carcinoma. *Cell* 2015, 159:676–690
9. Wheeler DL, Church DM, Edgar R, Federhen S, Helmberg W, Madden TL, Pontius JU, Schuler GD, Schriml LM, Sequeira E, Suzek TO, Tatusova TA, Wagner L: Database resources of the National Center for Biotechnology Information: update. *Nucleic Acids Res* 2004, 32:D35–D40
10. Campbell IM, Shaw CA, Stankiewicz P, Lupski JR: Somatic mosaicism: implications for disease and transmission genetics. *Trends Genet* 2015, 31:382–392
11. McConnell MJ, Lindberg MR, Brennand KJ, Piper JC, Voet T, Cowing-Zitron C, Shumilina S, Lasken RS, Vermeesch JR, Hall IM, Gage FH: Mosaic copy number variation in human neurons. *Science* 2013, 342:632–637
12. Cohen S, Segal D: Extrachromosomal circular DNA in eukaryotes: possible involvement in the plasticity of tandem repeats. *Cytogenet Genome Res* 2009, 124:327–338
13. Cohen S, Agmon N, Sobol O, Segal D: Extrachromosomal circles of satellite repeats and 5S ribosomal DNA in human cells. *Mob DNA* 2010, 1:11
14. Shibata Y, Kumar P, Layer R, Willcox S, Gagan JR, Griffith JD, Dutta A: Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues. *Science* 2012, 336:82–86
15. Dillon LW, Kumar P, Shibata Y, Wang YH, Willcox S, Griffith JD, Pommier Y, Takeda S, Dutta A: Production of extrachromosomal microDNAs is linked to mismatch repair pathways and transcriptional activity. *Cell Rep* 2015, 11:1749–1759
16. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W, van Steensel B: Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* 2008, 453:948–951
17. Segal E, Widom J: What controls nucleosome positions? *Trends Genet* 2009, 25:335–343
18. Segal E, Fondufe-Mittendorf Y, Chen L, Thåström A, Field Y, Moore IK, Wang JP, Widom J: A genomic code for nucleosome positioning. *Nature* 2006, 442:772–778
19. Skourti-Stathaki K, Proudfoot NJ: A double-edged sword: r loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev* 2014, 28:1384–1396
20. Aguilera A: The connection between transcription and genomic instability. *EMBO J* 2002, 21:195–201
21. Ginno PA, Lim YW, Lott PL, Korf I: GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome Res* 2013, 23:1590–1600
22. Ginno PA, Lott PL, Christensen HC, Korf I, Chédin F: R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell* 2012, 45:814–825
23. Li X, Manley JL: Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* 2005, 122:365–378
24. Aguilera A, García-Muse T: R loops: from transcription byproducts to threats to genome stability. *Mol Cell* 2012, 46:115–124
25. Schofield MJ, Hsieh P: DNA mismatch repair: molecular mechanisms and biological function. *Annu Rev Microbiol* 2003, 57:579–608
26. Brewster RC, Weinert FM, Garcia HG, Song D, Rydenfelt M, Phillips R: The transcription factor titration effect dictates level of gene expression. *Cell* 2014, 156:1312–1323
27. Seidl CI, Lama L, Ryan K: Circularized synthetic oligodeoxynucleotides serve as promoterless RNA polymerase III templates for small RNA generation in human cells. *Nucleic Acids Res* 2013, 41:2552–2564
28. Munding EM, Shiue L, Katzman S, Donohue J, Ares M: Competition between Pre-mRNAs for the splicing machinery drives global regulation of splicing. *Mol Cell* 2013, 51:338–348
29. Mandel P, Métais P: Les acides nucléiques du plasma sanguin l'homme [undetermined language]. *C R Seances Soc Biol Fil* 1948, 142:241–243

30. Lo YM: Circulating nucleic acids in plasma and serum: an overview. *Ann N Y Acad Sci* 2001, 945:1–7
31. Butt AN, Swaminathan R: Overview of circulating nucleic acids in plasma/serum: update on potential prognostic and diagnostic value in diseases excluding fetal medicine and oncology. *Ann N Y Acad Sci* 2008, 1137:236–242
32. Ameres SL, Zamore PD: Diversifying microRNA sequence and function. *Nat Rev Mol Cell Biol* 2013, 14:475–488
33. Lee YS, Shibata Y, Malhotra A, Dutta A: A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev* 2009, 23: 2639–2649
34. Haussecker D, Huang Y, Lau A, Parameswaran P, Fire AZ, Kay MA: Human tRNA-derived small RNAs in the global regulation of RNA silencing. *RNA* 2010, 16:673–695
35. Cole C, Sobala A, Lu C, Thatcher SR, Bowman A, Brown JWS, Green PJ, Barton GJ, Hutvagner G: Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. *RNA* 2009, 15:2147–2160
36. Fu H, Feng J, Liu Q, Sun F, Tie Y, Zhu J, Xing R, Sun Z, Zheng X: Stress induces tRNA cleavage by angiogenin in mammalian cells. *FEBS Lett* 2009, 583:437–442
37. Kumar P, Anaya J, Mudunuri SB, Dutta A: Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets. *BMC Biol* 2014, 12:78
38. Bartel DP: MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004, 116:281–297
39. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jungkamp AC, Munschauer M, Ulrich A, Wardle GS, Dewell S, Zavolan M, Tuschl T: Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 2010, 141:129–141
40. Kishore S, Gruber AR, Jedlinski DJ, Syed AP, Jorjani H, Zavolan M: Insights into snoRNA biogenesis and processing from PAR-CLIP of snoRNA core proteins and small RNA sequencing. *Genome Biol* 2013, 14:R45
41. Maute RL, Schneider C, Sumazin P, Holmes A, Califano A, Basso K, Dalla-Favera R: tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. *Proc Natl Acad Sci U S A* 2013, 110: 1404–1409
42. Bartel DP: MicroRNAs: target recognition and regulatory functions. *Cell* 2009, 136:215–233
43. Kudla G, Granneman S, Hahn D, Beggs JD, Tollervey D: Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. *Proc Natl Acad Sci U S A* 2011, 108: 10010–10015
44. Helwak A, Kudla G, Dudnakova T, Tollervey D: Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* 2013, 153:654–665
45. Deng J, Ptashkin RN, Chen Y, Cheng Z, Liu G, Phan T, Deng X, Zhou J, Lee I, Lee YS, Bao X: Respiratory syncytial virus utilizes a tRNA fragment to suppress antiviral responses through a novel targeting mechanism. *Mol Ther* 2015, 23:1622–1629
46. Goodarzi H, Liu X, Nguyen HCB, Zhang S, Fish L, Tavazoie SF: Endogenous tRNA-derived fragments suppress breast cancer progression via YBX1 displacement. *Cell* 2015, 161:790–802
47. Gebetsberger J, Zywicki M, Künzi A, Polacek N: tRNA-derived fragments target the ribosome and function as regulatory non-coding RNA in *Haloflex volcanii*. *Archaea* 2012, 2012:10–12
48. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ: miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* 2006, 34:D140–D144
49. Kumar P, Mudunuri SB, Anaya J, Dutta A: tRFdb: a database for transfer RNA fragments. *Nucleic Acids Res* 2014, 43: D141–D145
50. Towns WL, Begley TJ: Transfer RNA methyltransferases and their corresponding modifications in budding yeast and humans: activities, predilections, and potential roles in human health. *DNA Cell Biol* 2012, 31:434–454
51. ENCODE Project Consortium: The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 2004, 306:636–640
52. Kapranov P, Willingham AT, Gingeras TR: Genome-wide transcription and the implications for genomic organization. *Nat Rev Genet* 2007, 8:413–423
53. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J, Veeravalli L, Ruan Y, Lassmann T, Carninci P, Brown JB, Lipovich L, Gonzalez JM, Thomas M, Davis CA, Shiekhattar R, Gingeras TR, Hubbard TJ, Notredame C, Harrow J, Guigó R: The GENCODE v7 catalogue of human long non-coding RNAs: analysis of their structure, evolution and expression. *Genome Res* 2012, 22: 1775–1789
54. Rinn JL, Chang HY: Genome regulation by long noncoding RNAs. *Annu Rev Biochem* 2012, 81:145–166
55. Struhl K: Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat Struct Mol Biol* 2007, 14:103–105
56. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES: Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 2009, 458:223–227
57. Sauvageau M, Goff LA, Lodato S, Bonev B, Groff AF, Gerhardinger C, Sanchez-Gomez DB, Hacisuleyman E, Li E, Spence M, Liapis SC, Mallard W, Morse M, Swerdel MR, D'Eccles MF, Moore JC, Lai V, Gong G, Yancopoulos GD, Frendewey D, Kellis M, Hart RP, Valenzuela DM, Arlotta P, Rinn JL: Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *Elife* 2013, 2:e01749
58. Lin N, Chang KY, Li Z, Gates K, Rana Z, Dang J, Zhang D, Han T, Yang CS, Cunningham TJ, Head S, Duester G, Dong PDS, Rana TM: An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. *Mol Cell* 2014, 53: 1005–1019
59. Sakurai K, Reon BJ, Anaya J, Dutta A: The lincRNA DRAIC/PCAT29 locus constitutes a tumor-suppressive nexus. *Mol Cancer Res* 2015, 13:828–838
60. Mueller AC, Cichewicz MA, Dey BK, Layer R, Reon BJ, Gagan JR, Dutta A: MUNC, a long noncoding RNA that facilitates the function of MyoD in skeletal myogenesis. *Mol Cell Biol* 2015, 35:498–513
61. Negishi M, Wongpalee SP, Sarkar S, Park J, Lee KY, Shibata Y, Reon BJ, Abounader R, Suzuki Y, Sugano S, Dutta A: A new lincRNA, APTR, associates with and represses the CDKN1A/p21 promoter by recruiting polycomb proteins. *PLoS One* 2014, 9: e95216
62. Dey BK, Pfeifer K, Dutta A: The H19 long noncoding RNA gives rise to microRNAs miR-675-3p and miR-675-5p to promote skeletal muscle differentiation and regeneration. *Genes Dev* 2014, 28: 491–501
63. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai M-C, Hung T, Argani P, Rinn JL, Wang Y, Brzoska P, Kong B, Li R, West RB, van de Vijver MJ, Sukumar S, Chang HY: Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 2010, 464:1071–1076
64. Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S, Fedele S, Pesce E, Ferrer I, Collavin L, Santoro C, Forrest AR, Carninci P, Biffo S, Stupka E, Gustincich S: Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* 2012, 491:454–457
65. Cesana M, Cacchiarelli D, Legnini I, Santini T, Sthandier O, Chinappi M, Tramontano A, Bozzoni I: A long noncoding RNA

- controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* 2011, 147:358–369
66. Bluemn EG, Nelson PS: The androgen/androgen receptor axis in prostate cancer. *Curr Opin Oncol* 2012, 24:251–257
67. Prensner JR, Iyer MK, Sahu A, Asangani IA, Cao Q, Patel L, Vergara IA, Davicioni E, Erho N, Ghadessi M, Jenkins RB, Triche TJ, Malik R, Bedenis R, McGregor N, Ma T, Chen W, Han S, Jing X, Cao X, Wang X, Chandler B, Yan W, Siddiqui J, Kunju LP, Dhanasekaran SM, Pienta KJ, Feng FY, Chinnaiyan AM: The long noncoding RNA SCHLAPI promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nat Genet* 2013, 45:1392–1398
68. Malik R, Patel L, Prensner JR, Shi Y, Iyer M, Subramaniyan S, Carley A, Niknafs YS, Sahu A, Han S, Ma T, Liu M, Asangani I, Jing X, Cao X, Dhanasekaran SM, Robinson D, Feng FY, Chinnaiyan AM: The lncRNA PCAT29 inhibits oncogenic phenotypes in prostate cancer. *Mol Cancer Res* 2014, 12:1081–1087
69. Chakravarty D, Sboner A, Nair SS, Giannopoulou E, Li R, Hennig S, Mosquera JM, Pauwels J, Park K, Kossai M, MacDonald TY, Fontugne J, Erho N, Vergara IA, Ghadessi M, Davicioni E, Jenkins RB, Palanisamy N, Chen Z, Nakagawa S, Hirose T, Bander NH, Beltran H, Fox AH, Elemento O, Rubin MA: The oestrogen receptor alpha-regulated lncRNA NEAT1 is a critical modulator of prostate cancer. *Nat Commun* 2014, 5:5383