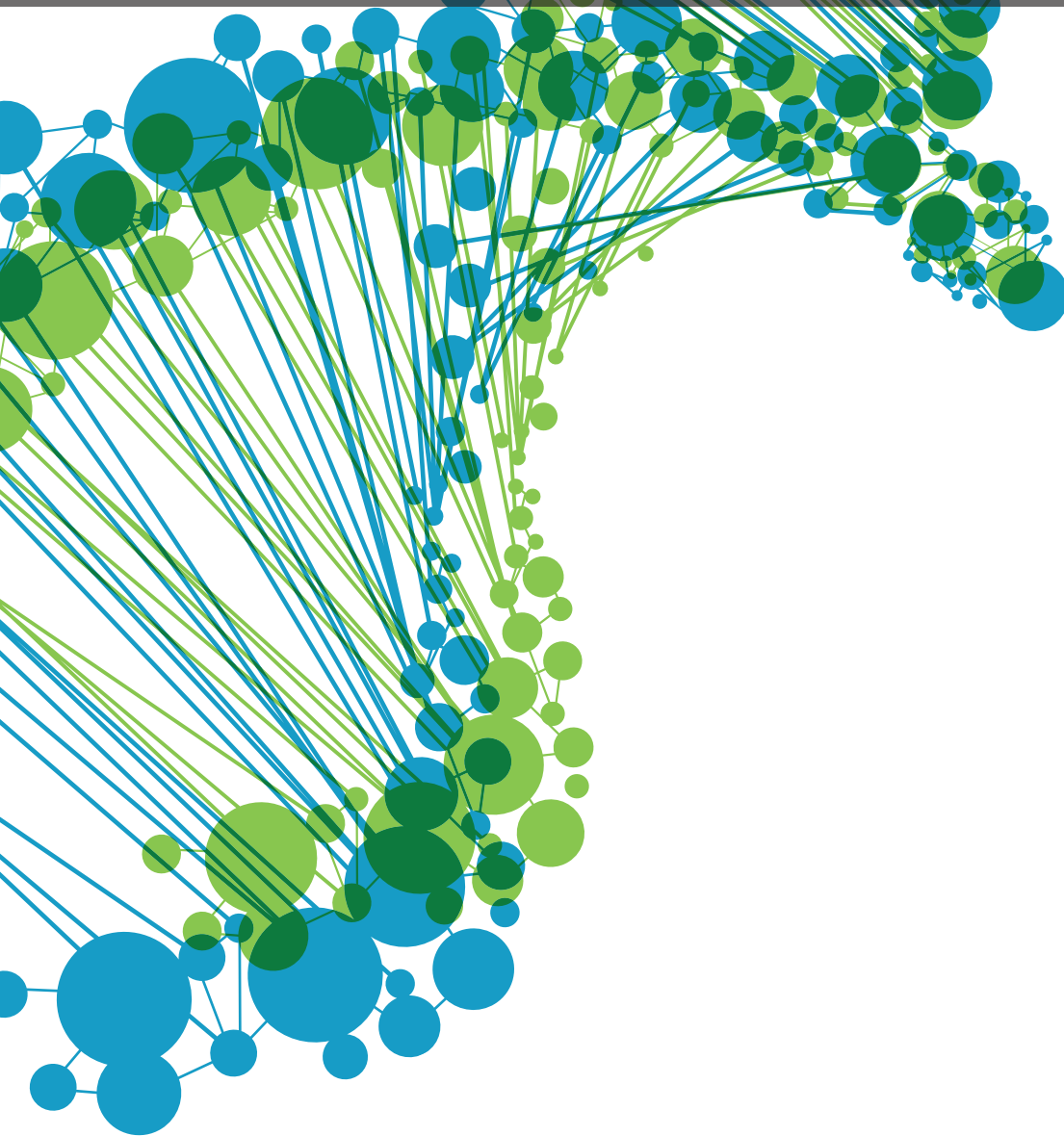# MOLECULAR FUNCTION AND REGULATION OF NON-CODING RNAs IN MULTIFACTORIAL DISEASES

EDITED BY : Mohammadreza Hajjari, Seyed Javad Mowla and Mohammad Ali Faghihi

frontiers Research Topics

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: **researchtopics@frontiersin.org**

MOLECULAR FUNCTION AND REGULATION OF NON-CODING RNAs IN MULTIFACTORIAL DISEASES

Topic Editors:
Mohammadreza Hajjari, Shahid Chamran University of Ahvaz, Iran
Seyed Javad Mowla, Tarbiat Modares University, Iran
Mohammad Ali Faghihi, University of Miami Miller School of Medicine, USA

Our understanding of the mechanisms underlying the development of multifactorial diseases such as diabetes, autism, Alzheimer's disease, and cancer has been greatly advanced. Non-coding RNAs (ncRNAs), generally including microRNAs and long non-coding RNAs, have recently been found to have potential roles in these diseases, and provide new opportunities for developing both specific biomarkers and therapeutic targets. However, the molecular function and regulation of these RNAs still remains challenging. Numerous studies are focusing on this field in order to fully appreciate the role and regulation of these molecules in human medicine and biology.

This e-book aims to bring together new findings on Non-coding RNAs in different complex diseases. It will highlight the characterization, roles, mechanism, and mode of action of these RNAs in complex disorders. We believe that the publications on this topic would be exponentially extended in future. The improved approaches at multiple levels may pave the way for designing and applying new biomarker and therapeutic targets for specific diseases based on these attractive molecules.

Table of Contents

frontiers
in Genetics

CrossMark

Editorial: Molecular Function and Regulation of Non-coding RNAs in Multifactorial Diseases

Mohammadreza Hajjari[1]*, Seyed Javad Mowla[2]* and Mohammad Ali Faghihi[3]*

[1] Department of Genetics, Shahid Chamran University of Ahvaz, Ahvaz, Iran, [2] Department of Molecular Genetics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran, [3] Department of Psychiatry and Behavioral Sciences, Center for Therapeutic Innovation, University of Miami Miller School of Medicine, Miami, FL, USA

The Editorial on the Research Topic

Molecular Function and Regulation of Non-coding RNAs in Multifactorial Diseases

Most of the parts in the human genomic DNA do not produce any proteins. They encode functional RNA molecules, which are not translated, and are thus named as non-coding RNAs (ncRNAs; Cammaerts et al.; Hajjari et al., 2014). Depending on the length and the function of these RNAs, they are categorized into different types such as microRNAs (miRNAs) and long non-coding RNAs (lncRNAs). Different studies have shown that these RNAs have crucial roles in cellular and molecular mechanisms. Considering their function, it is believed that the dysregulation of ncRNAs is involved in different diseases, especially such complex ones as cancers, neurological disorders, and cardiovascular diseases (Nouraee and Mowla; Merelo et al.; Hajjari et al., 2014). Accordingly, some researchers believe that these molecules can help with the diagnosis and the treatment of complex diseases. However, we have still a limited understanding of the functions of these molecules, and the different aspects of ncRNAs are still to be discovered in future. The above consideration was the motivation for the production of the topic entitled "Molecular function and regulation of non-coding RNAs in multi-factorial diseases." We believe that the articles in this issue and the attributed e-book can provide the interesting viewpoints for the researchers interested in this topic.

miRNAs are small RNAs with about 20–24 nucleotides length and function in post transcriptional regulation of gene expression. They usually destabilize and repress target RNAs via binding to $3'$UTRs, $5'$UTRs or the coding sequences of the transcripts (Lytle et al., 2007; Qin et al., 2010). miRNAs have been demonstrated to play major roles in a wide range of developmental processes as well as diseases. To understand the role of miRNA pathogenesis in different diseases such as cancers, two different strategies including the expression analysis and the genetic approaches are proposed (Reviewed in Cammaerts et al.). As such, the miRNAs are currently considered as new potential biomarkers for different diseases (Angelini and Emanueli). Furthermore, some reports have interestingly shown the plasma and saliva miRNAs as potentially sensitive and specific biomarkers (Lin et al.; Khoo et al., 2012). Nonetheless, there are still some concerns with the functionality of miRNAs that need to be addressed in future.

By modulating different targets or pathways, miRNAs have been quickly considered as potential therapeutic molecules for some diseases such as cancer (reviewed in Naidu et al., 2015) and cardiovascular ones (Nouraee and Mowla). miRNA-based therapeutics mainly focuses on modulation of the miRNA expression levels (Misso et al., 2014). MRX34, an agent mimicking miR-34 and developed by Mirna Therapeutics, has progressed into phase I clinical trials (identifier:

NCT01829971, Currently recruiting participants, Estimated study completion date: Dec 2016). The results may hopefully provide more insights for the researchers. In spite of different researches, there are still some challenges in designing a suitable carrier for targeting miRNAs into the desired cells. Bakhshinejad reports the studies proposing the nanocarriers as the potential tools to overcome delivery problems concerned with miRNA-based pharmaceutical tools (Bakhshinejad).

lncRNAs, the RNAs longer that 200 nucleotides, comprise the largest proportion of the human transcriptome. The biology of lncRNAs seems more complicated in comparison to miRNAs. These RNAs act through more different pathways and modes of action within the cell. They function at transcriptional and post-transcriptional levels, in the cytosol and nucleus through cis and trans-regulatory mechanisms (Hajjari et al., 2014; Angrand et al.). Different lncRNAs such as *HOTAIR, H19, MEG3, ANRIL, HULC*, and *XIST* are transcribed within the genome and play important roles in different cellular pathways (reviewed in Hajjari et al., 2014). Multiple lines of observation increasingly associate the mutations and the dysregulations of lncRNAs to some human diseases such as cancer (reviewed in Angrand et al.; Hajjari et al., 2014). On the current subject matter, Shahryari et al. have provided some evidence showing that some lncRNAs such as *SOX2OT* are also involved in pluoripotency (Shahryari et al.). Peschanscky et al. have also described the novel facets of *FMR4* lncRNA functionality and its relation to neurodevelopment (Peschansky et al.).

The growing evidence of the dysregulated lncRNAs not only introduces a new layer of complexity in the molecular mechanism of human diseases, but also opens up the opportunity to use lncRNAs as therapeutic targets and biomarkers. The prominent example of lncRNAs application in clinical practices is *PCA3* (Prostate cancer gene 3), an FDA-approved biomarker for the prostate cancer. According to this test, the *PSA* and *PCA3* RNA molecules are measured, and then the PCA3 Score (a ratio of *PCA3* RNA to *PSA* RNA) is calculated (http://www.fda.gov). Circulating lncRNAs have been also suggested as potential biomarkers for cancer and some other diseases. Exomal lncRNAs have been suggested as potential biomarkers to diagnose the malignant state of patients with prostate cancer (Işin et al.).

The rapidly growing list of ncRNAs holds promises that miRNAs and lncRNAs will become ever more important in illness management in future. The elucidation of the mechanisms by which lncRNAs act will help us to design suitable biomarkers or therapeutic agents for some diseases, especially cancer. The deciphering of the genetic networks and pathways regulated by the abnormally expressing lncRNAs, the characterization of miRNA-mRNA-lncRNA interaction, and the genetic/epigenetic mechanisms regulating the lncRNA expression provide more insights into the application of lncRNA research to human health. Considering the different molecular functions of ncRNAs, these RNAs can be the subject of a large diversity of further studies on complex diseases. We trust that the publications and the novelties on this subject will be exponentially extended in future.

AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct, and intellectual contribution to the work, and approved it for publication.

REFERENCES

Hajjari, M., Khoshnevisan, A., and Shin, Y. K. (2014). Molecular function and regulation of long non-coding RNAs: paradigms with potential roles in cancer. *Tumor Biol.* 35, 10645–10663. doi: 10.1007/s13277-014-2636-z

Khoo, S. K., Petillo, D., Kang, U. J., Resau, J. H., Berryhill, B., Linder, J., et al. (2012). Plasma-based circulating MicroRNA biomarkers for Parkinson's disease. *J. Parkinsons Dis.* 2, 321–331. doi: 10.3233/JPD-012144

Lytle, J. R., Yario, T. A., and Steitz, J. A. (2007). Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5′ UTR as in the 3′ UTR. *Proc. Natl. Acad. Sci. U.S.A.* 104, 9667–9672. doi: 10.1073/pnas.0703820104

Misso, G., Di Martino, M. T., De Rosa, G., Farooqi, A. A., Lombardi, A., Campani, V., et al. (2014). Mir-34: a new weapon against cancer? *Mol. Therapy Nucleic Acids* 3:e194. doi: 10.1038/mtna.2014.47

Naidu, S., Magee, P., Michela Garofalo, M. (2015). MiRNA-based therapeutic intervention of cancer. *J. Hematol. Oncol.* 8:68. doi: 10.1186/s13045-015-0162-0

Qin, W., Shi, Y., Zhao, B., Yao, C., Jin, L., Ma, J., et al. (2010). MiR-24 regulates apoptosis by targeting the open reading frame (ORF) region of FAF1 in cancer cells. *PLoS ONE* 5:e9429. doi: 10.1371/journal.pone.0009429

MicroRNAs as clinical biomarkers?

Timothy G. Angelini[1] and Costanza Emanueli[2, 3]*

[1] *Royal Surrey County Hospital, Guildford, UK,* [2] *Bristol Heart Institute, School of Clinical Sciences, University of Bristol, Bristol, UK,* [3] *National Heart and Lung Institute, Imperial College London, London, UK*

A clinical biomarker has been defined as "any cellular, biochemical, molecular, or genetic alterations by which a normal, abnormal, or simple biological process can be recognized or monitored" (Drucker and Krapfenbauer, 2003; Rahim et al., 2015). Biomarkers are widely employed throughout medicine on a day to day basis to diagnose, prognosticate, and predict outcomes of disease and illness or even as guides of treatment. They can either be used as a standalone, or more commonly, in conjunction with other test results. Moreover, biomarkers are often chosen as "surrogate endpoints" in clinical trials. Biomarkers must fulfill certain criteria in order to make it in to regular use in clinical practice. They must be specific to the condition that they are indicated in, sensitive, and must be practical in terms of accessibility of the sample, ease of testing method and provide information in which to guide clinical decisions. However, many biomarkers have unsatisfactory specificity or are not as sensitive as they are made out to be. In addition, biomarkers for some conditions are not yet available. Consequently, research in to novel biomarkers should be encouraged.

Biomarkers can be classified in to different classes, one being the sources where they measured from. In this respect, there are two distinct kinds of biomarker; intracellular, and extracellular (which can be broken down into three further groups by means of sampling; invasive, minimally invasive, and non-invasive). Examples of intracellular biomarkers include many of those used in oncology, such as BRCA1/2 (Drooger et al., 2015), and estrogen receptor and hormone receptor status in breast cancers (Yang et al., 2014), and BRAF genes in melanoma (Dar et al., 2015). For these tests, the cells from the suspected tumor must be invasively sampled, often by core sampling, or tissue scrapings. Extracellular biomarkers are often but not always linked with less invasive sampling, for example the withdrawal of a peripheral venous blood sample or urine collection. Examples are urine test for beta-human chorionic gonadotrophin (Chen et al., 2015) as an indicator of pregnancy or prostate specific antigen (PSA) (Ravery, 1999) in a suspected prostate cancer case.

An exciting, but not yet conclusive, new area of research with the potential to derive new clinical biomarkers has been into microRNAs (miRs). These short RNA molecules were discovered in 1993 (Lee et al., 1993) in *C. elegans* and found present in human samples in 2000 (van Rooij, 2011). From there, their love story with the biomedical community has seen a crescendo of enthusiasm and they have been highly regarded as both novel therapeutic targets, and possible intracellular and extracellular biomarkers. We might speculate that the reason behind this popularity is two-folds:

(1) The mode of action of miRs is relatively straightforward. MiRs do not contain any coding for protein production (and hence they are classified as noncoding RNAs–ncRNAs) and each miR is capable of post-translational regulation of the expression of a plethora of target messenger RNAs (mRNAs), which it recognizes through the semi complementary nucleotide base-pairing between its "seed sequence" (of just eight nucleotides) and one or more miR binding sites in the 3′-untranslated region of the mRNA targeted (Jackson and Standart, 2007). The biology of miR appears much simpler in comparison to, as an example, the long ncRNA that act via multiple, with still largely un-clarified mechanisms, at transcriptional and post-transcriptional levels, in the nucleus and the cytosol, and in *cis* and *trans* (Rinn and Chang, 2012). Moreover,

miRs exist in a limited number, thought to be around 2000 in humans. However, miRs variants are still being discovered through RNA-sequencing approaches.

(2) MiR are released by their producing cells in protected forms allowing them to remain for long periods in biological fluids. Moreover, miRs were initially believed to be tissue- and even cell-specific. This led to the assumption that what was found in the blood could sense the altered status of the cell types or tissues that were supposed to produce such miR. Now the concept of tissue/cell-specificity has been largely dismissed and it is becoming clear that at best we can talk of cell/tissue type enrichment for most, if not all, the known miRs. Hence, it is quite unrealistic to think that, under most clinical scenarios (with exceptions discussed later), the level of a single miR measured in the whole plasma or serum can be informative of a local slowly or relatively slowly evolving condition, such as, a developing cancer (Washam et al., 2013).

Circulating individual miRs, such as miR-21 (suggested to be useful for the detection of various carcinomas) (Wang et al., 2014; Wu et al., 2015), and circulating groups of miRs (such as a serum miR classifier encompassing miR-29a, miR-29c, miR-133a, miR-143, miR-145, miR-192, and miR-505, that has been proposed to detect hepatocellular carcinoma) (Lin et al., 2015) have been suggested as potential biomarkers that could be used in cancer detection and staging, and to follow-up already diagnosed cancer patients. However, in a recent article reporting the results of the comparison among 15 previous reports on potential new breast cancer biomarkers (Leinder et al., 2013) there was a scarce overlap between results: "Of the 143 circulating miRNAs reported to be differently regulated, 100 were supported by just 1 reference; 25 others had discordant results across publications and of the remaining 18 miRs, 8 had fold changes too low to be confirmed. Of the 10 concordant results, 9 were supported entirely by publications from the same institution and had authors in common" (Leinder et al., 2013; Witwer, 2015). This suggests that further efforts are needed before miR-based biomarkers can benefit cancer patients and that this could also apply to other disease conditions. Cancer has been the first translational area for miR work, closely followed by heart failure. Looking at both clinical scenarios, we find a typical example of unspecificity: circulating miR-21 has been proposed as a biomarker of both prostate cancer (Egidi et al., 2013) and myocardial fibrosis (in heart failure) (Thum et al., 2008). miR-21 is also the most expressed miR by vascular endothelial cells (Greco and Martelli, 2014), which are the cells directly lining the circulating blood and for this reason supposed to be the highest contributors to miRs circulating in the peripheral blood (Greco and Martelli, 2014; Witwer, 2015). A miR such as this, amongst others that are widely and highly expressed, might not be an ideal circulating candidate biomarker, suggesting that miRs that are usually under expressed, but upregulated under a particular condition could be better suited to be employed in a diagnostic test. For example, we recently found that miR-503 appears in the blood of diabetic patients at the last stage of critical limb ischemia, i.e., when they need an amputation (Caporali et al., 2011). It is possible

that circulating miR-503 could have a diagnostic/prognostic value when measured at the earlier stages of the disease, however, measuring miR-503 in a small leg muscle biopsy could provide more reliable information.

There are areas where we believe that circulating miRs show more promise and this is in the recognition of acute events, such as myocardial infarct (MI), as well as in the surgical setting, where time-restricted changes in miR expression have been reported consistently. For examples, the heart (and skeletal muscle)—enriched miR-1 has been noted to increase in patients after open heart surgery, after MI or transcoronary ablation of septal hypertrophy, an interventional procedure that mimics MI (D'Alessandra et al., 2010; Widera et al., 2011; Liebetrau et al., 2013; Nabialek et al., 2013). Diagnostic biomarkers are a key part in emergency service provision and the rapid diagnosis, and therefore treatment of patients with life threatening conditions. One of the most widely used in the emergency department are cardiac troponins (cTns: cTn-T and cTn-I) for the diagnosis of MI. CTns are used in conjunction with other investigations, such as electrocardiogram (ECG) changes, allowing, for example, to determine whether a patient has had a "STEMI" type MI (with ST segment Elevation by ECG). ECG cannot pick up non-STEMI cases and here laboratory biomarkers are highly important (Alpert et al., 2000). However, cTns are not always specific to MI, they can be raised in patients with other cardiac conditions and also after infection. In light of this, research into miRs as potentially better biomarkers has been carried out showing time-dependent increases in cardiac-enriched, ischemia-responsive miRs in the blood of MI patients (D'Alessandra et al., 2010; Nabialek et al., 2013). There has also been claim that miRs can help differentiating the diagnosis of a STEMI compared with other myocardial conditions, such as stable angina, non-STEMI, and Takotsubo cardiomyopathy (Nabialek et al., 2013; Ward et al., 2013; Jaguszewski et al., 2014). However, it is yet to be demonstrated that miRs can replace cTns as routine biomarkers used in the intensive coronary care unit, in depth investigations of specificity and sensitivity in different cohorts or patients are needed. Additionally, in the realm of emergency medicine, sensitivity and specificity of miRs response in terms of circulating changes are not the only issues, because the time necessary by this putative biomarker to appear elevated in the blood or another biological fluid is critical. Classically, cTns take a few hours to increase in the blood after an MI. MiR-1 has been proposed to go up earlier than cTns (Liebetrau et al., 2013). However, for these comparisons, the time to obtain the test results and the test reproducibility are big obstacles yet to be overcome. Differently from high sensitive cTns that today are measured by immune-enzymatic reactions allowing for results to be obtained in around 20 min, PCR-based miR analyses are still quite time-consuming. Alternative techniques for miR quantification have been proposed (Arata et al., 2012) but they are far away from being commonly employed by the scientific community, let alone the clinical diagnostic lab. Additionally, in a clinical laboratory staff cannot reason as in observational studies, where everybody is often quite satisfied with saying "miR-1 is increased in the blood of MI patients in comparisons to a control group." Diagnostic use

of miRs needs a different rigor and first of all a definition of what the "normal" threshold of miR concentration is, above which we can suspect in a patient who is experiencing a heart attack. Scientists working in the miR field are aware that this is not an easy task and by using PCR-based methods quantitative differences between different studies are common. In addition, the data normalization approaches are still debated and interference by heparin (used in interventional procedures) with the PCR reaction has been reported (Mayr et al., 2013), even if protocol to nullify the "heparin effect" are adopted. Alternative technologies can be developed, but this will require further investment, time and validation efforts (Arata et al., 2012).

In conclusion, We believe that miRs hold potential value as clinical biomarkers, but, their journey to the diagnostic lab is still long and needs improved approaches at multiple levels, starting with technical refinement in the miR concentration evaluation, the use of RNA-sequencing to possibly recognize new miRs that are better candidates (higher tissue/cell-specificity, lower expression under healthy conditions etc.) and the use of blood fractions potentially enriched in miRs (like exosomes and microparticles) for diagnostic tests. It is also possible that miR clusters have more specificity than single miRs. Moreover, miR could be associated to other biomarkers to improve the diagnostic power.

Funding and Acknowledgments

References

Alpert, J. S., Thygesen, K., Antman, E., and Bassand, J. P. (2000). Myocardial infarction redefined - a consensus document of the joint European Society of Cardiology/American College of Cardiology Committee for the redefinition of myocardial infarction. *J. Am. Coll. Cardiol.* 36, 959–969.

Arata, H., Komatsu, H., Hosokawa, K., and Maeda, M. (2012). Rapid and sensitive microRNA detection with laminar flow-assisted dendritic amplification on power-free microfluidic chip. *PLoS ONE* 7:e48329. doi: 10.1371/journal.pone.0048329

Caporali, A., Meloni, M., Vollenkle, C., Bonci, D., Sala-Newby, G. B., Addis, R., et al. (2011). Deregulation of microRNA-503 contributes to diabetes mellitus-induced impairment of endothelial function and reparative angiogenesis after limb ischemia. *Circulation* 123, 282–291. doi: 10.1161/CIRCULATIONAHA.110.952325

Chen, C. Y., Hwu, Y. M., Chen, C. P., and Chang, C. C. (2015). Quantative analysis of total Beta-subunit of human chorionic gonadotrophin concentration in urine reduction to assist in the diagnosis of ectopic pregnancy. *Int. J. Nanomed.* 10, 2475–2483. doi: 10.2147/IJN.S81201

D'Alessandra, Y., Devanna, P., Limana, F., Straino, S., Di Carlo, A., Brambilla, P. G., et al. (2010). Circulating microRNAs are new and sensitive biomarkers of myocardial infarction. *Eur. Heart J.* 31, 2765–2773. doi: 10.1093/eurheartj/ehq167

Dar, A. A., Nosrati, M., Bezrookove, V., de Semir, D., Majid, S., Thummala, S., et al. (2015). The role of BPTF in melanoma progression and in response to BRAF-targeted therapy. *J. Natl. Cancer Inst.* 107, djv034. doi: 10.1093/jnci/djv034

Drooger, J. C., Hooning, M. J., Seynaeve, C. M., Baaijens, M. H., Obdeijn, I. M., Sleijfer, S., et al. (2015). Diagnostic and therapeutic ionizing radiation and the risk of a first and second primary breast cancer, with special attention for BRCA1 and BRCA2 mutation carriers: a critical review of the literature. *Cancer Treat. Rev.* 41, 187–196. doi: 10.1016/j.ctrv.2014.12.002

Drucker, E., and Krapfenbauer, K. (2003). Pitfalls and limitations in translation from biomarker discovery to clinical utility in predictive and personalised medicine. *EPMA J.* 4, 7. doi: 10.1186/1878-5085-4-7

Egidi, M. G., Cochetti, G., Serva, M. R., Guelfi, G., Zampini, D., Mechelli, L., et al. (2013). Circulating microRNAs and kallikreins before and after radical prostatectomy: are ther really prostate cancer markers? *Biomed. Res. Int.* 2013, 241780. doi: 10.1155/2013/241780

Greco, S., and Martelli, F. (2014). MicroRNAs in hypoxia responses. *Antioxid. Redox Signal.* 21, 1164–1166. doi: 10.1089/ars.2014.6083

Jackson, R. J., and Standart, N. (2007). How do microRNAs regulate gene expression? *Sci. Signal.* 2007, re1. doi: 10.1126/stke.3672007re1

Jaguszewski, M., Osipova, J., Ghadri, J. R., Napp, L. C., Widera, C., Franke, J., et al. (2014). A signature of circulating microRNAs differentiates takotsubo cardiomyopathy from acute myocardial infarction. *Eur. Heart.* 35, 999–1006. doi: 10.1093/eurheartj/eht392

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75, 843–854.

Leinder, R. S., Li, L., and Thompson, C. L. (2013). Dampening enthusiasm for circulating microRNA in breast cancer. *PLoS ONE* 8:e57841. doi: 10.1371/journal.pone.0057841

Liebetrau, C., Mollmann, H., Dorr, O., Szardien, S., Troidl, C., Willmer, M., et al. (2013). Release kinetics of circulating muscle-enriched microRNAs in patients undergoing transcoronary ablation of septal hypertrophy. *J. Am. Coll. Cardiol.* 62, 992–998. doi: 10.1016/j.jacc.2013.05.025

Lin, X. J., Chong, Y., Guo, Z. W., Xie, C., Yang, X. J., Zhang, Q., et al. (2015). A serum microRNA classifier for early detection of hepatocellular carcinoma: a multicentre, retrospective, longitudinal biomarker identification study with a nested case-control study. *Lancet Oncol.* 16, 804–815. doi: 10.1016/S1470-2045(15)00048-0

Mayr, M., Lee, R., Kaudewitz, D., Zampetaki, A., and Channon, K. M. (2013). Effects of heparin on temporal microRNA profiles. *J. Am. Coll. Cardiol.* 63, 940–941. doi: 10.1016/j.jacc.2013.07.118

Nabialek, E., Wañha, W., Kula, D., Jadczyk, T., Krajewska, M., Kowalówka, A., et al. (2013). Circulating microRNAs (miR-423-5p, miR-208a and miR-1) in acute myocardial infarction and stable coronary heart disease. *Minerva Cardioangiol.* 61, 627–637.

Rahim, M. A., Rahim, Z. H., Ahmad, W. A., and Hashim, O. H. (2015). Can Saliva proteins be used to predict the onset of acute myocardial infarction among high risk patients? *Int. J. Med. Sci.* 12, 329–335. doi: 10.7150/ijms.11280

Ravery, V. (1999). The significance of recurrent PSA after radical prostatectomy: benign versus malignant sources. *Semin. Urol. Oncol.* 17, 127–129.

Rinn, J. L., and Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* 81, 145–166. doi: 10.1146/annurev-biochem-051410-092902

Thum, T., Gross, C., Fiedler, J., Fischer, T., Kissler, S., Bussen, M., et al. (2008). MicroRNA-21 contributes to myocardial disease by stimulating MAP kinase signalling in fibroblasts. *Nature* 456, 980–984. doi: 10.1038/nature07511

van Rooij, E. (2011). The art of microRNA research. *Circ. Res.* 108, 219–234. doi: 10.1161/CIRCRESAHA.110.227496

Wang, Y., Gao, X., Wei, F., Zhang, X., Yu, J., Zhao, H., et al. (2014). Diagnostic and prognostic value of circulating miR-21 for cancer: a systematic review and meta-analysis. *Gene* 533, 389–397. doi: 10.1016/j.gene.2013.09.038 533

Ward, J. A., Esa, N., Pidikiti, R., Freedman, J. E., Keaney, J. F., Tanriverdi, K., et al. (2013). Circulating cell and plasma microRNA profiles differ between non-ST-segment and ST-Segment-elevation myocardial infarction. *Fam. Med. Med. Sci. Res.* 2, 108. doi: 10.4172/2327-4972.1000108

Washam, C. L., Byrum, S. D., Leitzel, K., Ali, S. M., Tackett, A. J., Gaddy, D., et al. (2013). Identification of PTHrP(12-48) as a plasma biomarker associated with breast cancer bone metastatsis. *Can. Epidemiol. Biomarkers Prev.* 22, 972–983. doi: 10.1158/1055-9965

Widera, C., Gupta, S. K., Lorenzen, J. M., Bang, C., Bauersachs, J., Bethmann, K., et al. (2011). Diagnostic and prognostic impact of six circulating microRNAs in acute coronary syndrome. *J. Mol. Cell. Cardiol.* 51, 872–875. doi: 10.1016/j.yjmcc.2011.07.011

Witwer, K. W. (2015). Circulating MicroRNA biomarker studies: pitfalls and potential solutions. *Clin. Chem.* 61, 56–63. doi: 10.1373/clinchem.2014. 221341

Wu, K., Li, L., and Li, S. (2015). Circulating microRNA-21 as a biomarker for the detection of various carcinomas: an updated meta-analysis based on 36 studies. *Tumour Biol.* 36, 1973–1981. doi: 10.1007/s13277-014-2803-2

Yang, Y. F., Liao, Y. Y., Yang, M., Peng, N. F., Xie, S. R., and Xie, Y. F. (2014). Discordance in ER, PR and HER2 receptors between primary and recurrent/metastatic lesions and their impact on survival in breast cancer patients. *Med. Oncol.* 31, 214. doi: 10.1007/s12032-014-0214-2

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

miRNA therapeutics in cardiovascular diseases: promises and problems

Nazila Nouraee and Seyed J. Mowla*

Department of Molecular Genetics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran

microRNAs (miRNAs) are a novel class of non-coding RNAs which found their way into the clinic due to their fundamental roles in cellular processes such as differentiation, proliferation, and apoptosis. Recently, miRNAs have been known as micromodulators in cellular communications being involved in cell signaling and microenvironment remodeling. In this review, we will focus on the role of miRNAs in cardiovascular diseases (CVDs) and their reliability as diagnostic and therapeutic biomarkers in these conditions. CVDs comprise a variety of blood vessels and heart disorders with a high rate of morbidity and mortality worldwide. This necessitates introduction of novel molecular biomarkers for early detection, prevention, or treatment of these diseases. miRNAs, due to their stability, tissue-specific expression pattern and secretion to the corresponding body fluids, are attractive targets for cardiovascular-associated therapeutics. Explaining the challenges ahead of miRNA-based therapies, we will discuss the exosomes as delivery packages for miRNA drugs and promising novel strategies for the future of miRNA-based therapeutics. These approaches provide insights to the future of personalized medicine for the treatment of CVDs.

Keywords: miRNA, cardiovascular disease, cardio-miR, exosome, cell communication, delivery vehicle, secretory miRNA, miRNA therapeutics

Introduction

Among multifactorial diseases, cardiovascular diseases (CVDs) are significant due to their variable symptoms and high mortality rate accounting for one third of global deaths (Santulli, 2013). They include a wide range of disorders connected to blood vessels and heart ranging from coronary artery disease (CAD), pulmonary arterial hypertension (PAH), and congenital heart disease to deep vein thrombosis and cerebrovascular disease. Major CVD risk factors include family history, obesity, hypertension, diabetes mellitus, and hypercholesterolemia (Mendis et al., 2011). Since CVDs are hard-to-cure, several investigations have focused on different mechanisms underlying CVD in order to manage the symptoms. microRNAs have emerged as one of the most favorable molecular targets in this regard. microRNAs are a class of non-coding RNAs with a short length of 18–24 nucleotides. They mainly act as post-translational repressors of gene expression. By regulating the fundamental cellular mechanisms such as cell differentiation, proliferation, growth, and apoptosis, miRNAs have received enormous attention for therapeutic applications. Huge investigations and investments have been made to bring these molecules into the clinic. In this review, we will explain the novel diagnostic, therapeutic and prognostic approaches based on miRNAs in the field of CVDs (cardio-miRs) and the problems ahead of this research area; where we are and where we expect to be in the future.

miRNA Mechanisms of Action

Calin et al. (2002) were first to link miRNAs with cancer progression. Since then several studies have focused on the role of these small molecules in the pathogenesis of different human diseases. The primary findings emphasized on the post-transcriptional regulatory role of miRNAs through base pairing with the 3′ untranslated region (UTR) of their target mRNAs. This leads to the mRNA degradation or translation inhibition and in both cases the miRNA binding results in the suppression of their target mRNAs. This is the main mechanism reported for the miRNAs regulatory role; however, many variations have also been reported (Ha and Kim, 2014; Lin and Gregory, 2015). miRNA binding to the 5′ UTR of transcripts has also been demonstrated to be able to activate or suppress target genes (Lee et al., 2009). Some miRNAs can also bind to the open reading frame (ORF) of mRNA transcripts and repress translation. This mechanism was first reported by Tay et al. (2008). They demonstrated that miRNA binding to the coding region of pluripotency genes can regulate the embryonic stem cell differentiation. miRNA binding might also happen at the promoter of target gene which causes repression of gene translation. On the other hand, some miRNAs modulate their target expression by binding to RNA-binding proteins that regulate the expression of mRNA transcripts (Eiring et al., 2010). Salmena et al. (2011) proposed the competing endogenous RNA (ceRNA) hypothesis according to which, miRNAs, long non-coding RNAs (lncRNAs) and target mRNAs are in a finely tuned interaction through miRNA response elements (MREs) and their competition for target binding based on their total concentration in the cytoplasm defines a higher level of regulation. Recent studies have indicated another fascinating aspect of miRNA regulatory network. Fabbri et al. (2012) proved that miRNAs excreted from cancer cells, can directly bind the toll-like receptors (TLRs) at the surface of neighboring immune cells and activate the relevant signaling pathways in the recipient cells. The complexity of miRNA-mediated regulatory systems highlights the importance of these small molecules in the clinic and needs further proceeding technologies to get closer to clinical therapies.

miRNA Diagnostics in CVD

microRNAs show tissue-specific and time-dependent expression patterns turning them into a leading fact in using these small molecules as diagnostic and therapeutic targets (Lu et al., 2005). Similar to other developmental phenomena in embryogenesis, miRNAs play critical roles in cardiovascular development and also their expression profile changes according to different pathological conditions. Cardiac resynchronization therapy (CRT) after heart failure helps improve the heart arrhythmia and it also affects the myocardial miRNA expression. In responder patients, CRT affects cardiac processes including cardiac fibrosis, apoptosis, angiogenesis, and channel alterations and consequently alters the expression levels of miR-29 (implicated in cardiac fibrosis), miR-30, miR-92, and miR-145 (involved with cardiac angiogenesis),

miR-30 (modulated in cardiac apoptosis), and miR-26 (affected by modified ionic channel function; Sardu et al., 2014).

Recently circulatory miRNA expression profiling has provided strong molecular markers for detection of various diseases including CVD such as myocardial hyperthrophy, infarction, angiogenesis and fibrosis (Charan Reddy, 2014; Wronska et al., 2015). van Rooij et al. (2006) defined a miRNA signature for cardiac hypertrophy. They showed that miR-195 overexpression is sufficient to drive cardiac hypertrophy. miR-1 family, miR-133a/b and miR-208 are well-known for their implication in myogenesis both in skeletal muscles and cardiac development and their dysregulation has been detected in several CVDs including myocardial infarction, hypertrophy, and arrhythmias (Thum et al., 2008a). While miR-208 showed significant upregulation (pro-hypertrophic), miR-1 and miR-133a expression levels significantly decreased (anti-hypertrophic) in acute myocardial infarction (AMI) compared with normal individuals heart (Bostjancic et al., 2010). In CAD, miR-1 is upregulated in the left ventricular endocardium (Yang et al., 2007) and also plasma levels of miR-1, miR-133, and miR-208b have been reported to be elevated after AMI (Widera et al., 2011; Devaux et al., 2015). However, the circulatory miR-1 is not sensitive or specific enough to be known as an AMI specific biomarker since it is also affected by factors other than AMI. Moreover, by targeting Protein phosphatase 2A regulatory subunit B56 alpha (PP2A), miR-1 has been linked to the arrhythmia mainly through hyperphosphorylation of ryanodine receptor (RyR2; Terentyev et al., 2009). Following hypertension or cardiac infarction, cardiac hyperthrophy, and cardiac fibrosis are common diseases in which the excess amounts of extracellular matrix (ECM) proteins accumulate in cardiac tissue in order to adapt the system to the pathological conditions. Overexpression of miR-29 and miR-21 and downregulation of miR-133 and miR-30 have been linked to ECM remodeling and fibrosis by regulating several components of the ECM including collagen type I alpha 1 and 2 (Col1A1 and Col1A2) as targets of miR-29 (van Rooij et al., 2008), sprouty homolog 1 (SPRY1) as a target of miR-21 (Thum et al., 2008b) and connective tissue growth factor (CTGF) as target of miR-133 and miR-30c (Duisters et al., 2009). miR-133 is specifically expressed in cardiomyocytes while miR-30 is also detectable in cardiac fibroblasts as well as cardiomyocytes. These and other miRNAs (including miR-328) have been linked to atrial fibrillation which is mainly resulted from structural remodeling and fibrosis (Santulli et al., 2014a). **Table 1** summarizes a list of miRNAs involved in different CVDs (cardio-miRs).

miRNA-Bearing Exosomes as Communicators of CVD

Recent miRNA-based studies have focused on the role of exosomes as natural delivery vehicles for some proteins, mRNA, and miRNAs which facilitate communication between cells and their neighboring stroma. Exosomes are small vesicles (40–100 nm) originating from the plasma membrane or

TABLE 1 | List of cardio-miRs involved with different cardiovascular disorders.

miRNA name	Diseases name	Detection approach	Dysregulation	Reference
miR-208a/b, miR-499	AMI and myocardial injury	Microarray and real-time PCR	Upregulation in the plasma of patients	Corsten et al. (2010), Wang et al. (2010)
miR423-5p	Heart failure	Microarray and real-time PCR	Upregulation in the plasma of patients	Tijsen et al. (2010)
miR-328	Atrial fibrillation (adverse electrical remodeling)	Microarray and real-time PCR	Upregulation in left atrial samples (from dogs)	Lu et al. (2010)
miR-1	AMI	Real-time PCR	Upregulation in the patients' serum	Cheng et al. (2010)
miR-26a	Cardiovascular repair; acute coronary syndromes	Real-time PCR	Upregulation in the Patients' plasma	Icli et al. (2013)
miR-16, miR-27a, miR-101, and miR-150	Left ventricular contractility (LV)	Real-time PCR	Downregulation of miR-101 or miR-150 and upregulation of miR-16 or miR-27a correlate with higher risk of impaired LV contractility	Devaux et al. (2013)
miR-133a	AMI and coronary artery stenosis	Real-time PCR	Upregulation in the Patients' plasma	Wang et al. (2013)
miR-126, miR-17/miR-92a, miR-155	Coronary artery disease	Microarray and real-time PCR	Reduction in serum and plasma of patients	Fichtlscherer et al. (2010)
miR-126	Congestive heart failure;	Real-time PCR	Negative correlation with NYHA class; downregulation in patients; positive association with incident MI	Fukushima et al. (2011)
miR-126	Myocardial infarction (MI)	Real-time PCR	Upregulation in plasma samples of MI patients; positive association with MI incidence	Zampetaki et al. (2012)
miR-1, miR-134, miR-186, miR-208, miR-223, and miR-499	AMI and angina pectoris	Deep sequencing	Upregulation in serum samples of AP patients compared to AMI	Li et al. (2013)
miR-106b/25 cluster, miR-21/590-5p family, miR-126*, miR-451	Coronary artery disease	Microarray and real-time PCR	upregulation in plasma of patients with unstable angina	Ren et al. (2013)

AMI, acute myocardial infarction; AP, angina pectoris; NYHA, New York Heart Association.

multivesicular bodies (MVBs) and are present in almost all biological fluids. They shuttle between neighboring cells and transfer their cargoes which can be cellular components including proteins, mRNA, and non-coding RNAs. These cargoes can perform regulatory effects and control gene expression in the recipient cells (Pegtel et al., 2010). Exosomes are famous for being the communicators of microenvironment. Exosomal membrane proteins as well as other components are usually similar to their originating cell. Their membranes usually contain higher levels of sphingomyelin, cholesterol and phosphatidylserine compared to their originating cells. Their cargo also might represent the cellular composition of RNAs and proteins or they might show a separate profile. Methodologies including immunoblotting, affinity extraction into magnetic beads and flow cytometry have been used to identify the protein components of exosomes. Profiling of RNAs – miRNAs in particular – is usually performed by means of microarrays, qPCR-based arrays and most recently by next generation sequencing techniques which analyze the transcriptome – miRnome – of each sample.

Recent studies have proved targeted packaging of miRNAs and their biogenesis components including Dicer and AGO2 in exosomes (Melo et al., 2014). Exosome shuttling is implicated in a variety of disease including cancer, CVDs or viral infections. In CVD, exosomal transfer of miRNAs is a well-known mechanism through which, cells educate their adjacent environment in order to confront the pathological condition. In heart ischemia or fibrosis, paracrine or endocrine secretion of miRNA-bearing exosomes into cardiomyocytes or active fibroblasts of the heart have been demonstrated. This leads to trans-differentiation of fibroblasts into an active state with a different rate of growth factors secretion (van Rooij and Olson, 2009; Yoo et al., 2011). Bang et al. (2014) proved that following heart infarction, exosome-secretion from cardiomyocytes with a new miRNA profile, can reprogram the cardiac fibroblasts leading to cardiac hypertrophy. These mechanisms can introduce potential novel therapeutic targets. As well, recognizing tissue-specific CAF markers, the mechanisms underlying their reactivation and the miRNA signals in these processes, will potentially provide promising targets for CVD therapies. In addition to the fibroblasts in the microenvironment, exosomal transfer of signals is also used for cardiovascular protection after a disease. For example in heart ischemia, secretory exosomes with miRNA cargos from cardiomyocytes have been shown to activate and induce the bone marrow-derived stem cells. These latter cells release the second subset of exosomes that lead to myocardial regeneration or protection (Sahoo and Losordo, 2014). Several investigations have demonstrated the selective packaging of miRNAs in exosomes and their secretion to the stroma of cells by means of signal molecules. This is beyond shedding the vesicles by plasma membrane (Yang et al., 2011; Montecalvo et al., 2012; Stoorvogel, 2012). These disease-specific expression patterns of exosomal miRNAs – which can be non-invasively detected in

the body fluids – provides potential molecular markers and promising therapeutic targets for treatment of CVDs. Several circulating miRNAs have been linked to some CVDs. In patients with AMI and myocardial injury, higher levels of miR-208a have been detected in the plasma of patients and they showed higher specificity and sensitivity than conventional biomarkers such as cardiac troponin I (TnI) for diagnosis of the disease (Ji et al., 2009; Wang et al., 2010). Another potential biomarker for AMI is reported by Adachi et al. (2010) who detected higher plasma levels of miR-499 in patients with AMI compared with normal individuals. Using microarray analysis of exosomes obtained from cardiac progenitor cells, Gray et al. (2015) introduced 11 miRNAs significantly up-regulated in response to hypoxic conditions while miR-292 showed the highest variations in these exosomes. Moreover, exosome profiling from cardiac fibroblasts was shown to be enriched with miR-21-3p (miR-21*). These fibroblasts affect cardiomyocytes and secrete the mediators of cardiac hypertrophy (Bang et al., 2014). Knowing the mechanisms underlying these communications and recognizing potential strong biomarkers for diagnosis and treatment of CVDs will help us in development of future molecular therapeutics.

miRNA Therapeutics in CVD

In the past decade, outstanding researches in the field of miRNA drugs have changed the face of molecular medicine. miRNA-based therapeutics mainly focus on reinstating the miRNA expression levels. Two main approaches include overexpression of downregulated miRNAs and suppression of overexpressed ones. These approaches have been subject to different modifications in order to improve the efficiency of delivery and less off-target effects. Promising tools in this regard are the oligonucleotides that mimic the endogenous miRNA or suppress the mature miRNA by sequence complementarity. Addition of locked nucleic acids (LNAs) or 2′-O-methylation of the antisense oligonucleotides increases the binding specificity while cholesterol conjugation enhances the circulation time, serum stability, and cellular uptake. Expression vectors bearing the pre-miRNA sequences or tandem miRNA target sites (sponges) sound promising for *in vitro* overexpression or suppression of miRNAs, respectively. Another issue to overcome is the efficient delivery of each of the miRNA drugs. Naked oligonucleotides are less efficient due to their instability *in vitro* or *in vivo* which subject them to different nucleases. Lipid-based vehicles, viral systems, and cationic polymers are among the main delivery tools for miRNA-based therapeutics. Each of these strategies has its own challenges and still needs improvements to address problems such as cytotoxicity, immunogenicity, and low efficiency (van Rooij and Olson, 2012). Fiedler et al. (2011) compared the effect of different doses of cholesterol-based anti-miR-24 (antagomirs) on miR-24 expression in cardiomyocytes and cardiac endothelial cells. They showed a cell-type specific tendency or mechanism for antagomir uptake by these tissues. Cholesterol-based antagomirs for silencing miR-21 were also proved to inhibit cardiac fibrosis and dysfunction (Kumarswamy et al., 2012).

van Rooij et al. (2008) used anti-miR-29b oligonucleotides (cholesterol modified) after myocardial infarction and observed upregulation of ECM proteins leading to cardiac fibrosis. Also, Bonauer et al. (2009) demonstrated that intravenous injection of miR-92a antagomir, improved the function of damaged tissue in models of myocardial infarction. Inhibition of miR-92a results in neoangiogenic effects and functional recovery of ischemic tissues (Bonauer et al., 2009). Another promising miRNA in CVD therapeutics is miR-208 which is implicated in cardiac remodeling. LNA-modified anti-miR-208 oligonucleotides have successfully prevented pathologically associated cardiac remodeling during diastolic heart disease (van Rooij et al., 2007). A common issue in these approaches is *in vivo* instability and the low homing efficiency of oligonucleotides which results in modest changes in the expression levels of the target protein. Accordingly, improved stabilization or efficient delivery vehicles are required. Examples of such vehicles are adeno-associated viruses (AAVs). AAV9, a specific serotype of AAVs, has tropism for myocardiocytes and enriches in the heart (Bish et al., 2008). In another study Santulli et al. (2014b). proved that miR-126-3p down-regulation using an adeno-viral vector containing miR-126-3p target sites inhibits proliferative vascular smooth muscle cells and prevents restenosis in animal models. miR-126 is an endothelial specific miRNA that regulates vascular integrity and angiogenesis. These strategies provide an efficient tool for cardiac gene transfer and targeted delivery of miRNA-based therapeutics.

Due to their natural role in miRNA secretion and shuttling between different cells, exosomes are of great interest in miRNA therapeutics. Exosomes are flexible in size and cargo type and their non-synthetic nature potentiates them for more efficient and non-immunogenic delivery of cargo while they maintain the cargo integrity and stability. Exosomal membranes contain certain proteins which have binding affinity to specific receptors on the surface of recipient cells. So, they can selectively target cell types of interest and manipulating their miRNA components – as well as other molecular cargoes – will provide promising tools for the future of personalized medicine. In CVD, exosomes can be used as therapeutic agents, as protein delivery carriers or as gene therapy devices. Mesenchymal stem cell- derived exosomes have been investigated in the field of cardiac regenerative medicine especially in myocardial ischemia/reperfusion injuries (Lai et al., 2010). Martinez et al. (2006) have demonstrated that exosomes bearing differentiation signals can affect neovascularization and they are promising for treatment of angiogenic defects. Moreover, several exosomal miRNAs are implicated in angiogenesis and vascular repair. Zhang et al. (2010) showed that exosomes enriched with pre-miR-150 enhance endothelial cell migration. They transfected pre-miR-150/anti-miR-150 oligonucleotides to the THP-1 cell line and collected the conditioned media of these cells which was enriched with the exogenous oligonucleotides-bearing exosomes (Zhang et al., 2010). This conditioned medium affects the migration ability of endothelial cells. Exosome are novel promising elements for the future of CVD treatment due to their targeted delivery capacity and their microenvironment-dependent nature. The latter property triggers their activation in relation to the pathological microenvironment such as pH

or substrate concentration. However, further studies would be necessary to overcome obstacles such as engineering and purifying exosomes, cargo loading into them and optimizing their quality and characterization for targeted delivery.

Conclusion

Cardiovascular diseases are the one of the leading causes of death worldwide. Most of CVDs can be prevented by controlling behavioral and environmental risk factors and early diagnosis and management play an important role in this regard. In this review, we focused on miRNAs as small non-coding RNAs involved in a variety of key cellular processes. Several miRNA biomarkers have been introduced for different CV situations. miRNAs have emerged as attractive novel therapeutics and they have several advantages over other molecular therapeutics due to their small size, conserved sequences and their stability in the body fluids. First anti-cancer miRNA-based drug, MRX-34 (a liposome-based miR-34 mimic) developed by Mirna Therapeutics came to the clinic in 2013 for the treatment of hepatocellular

carcinoma (mirnarx.com; NCT01829971). Current achievements have portrayed a promising future for miRNA-based therapeutics although still several obstacles including their stability, renal clearance, off-target effects, inefficient endocytosis by target cells or the immunogenicity of delivery vehicles, need to be overcome. In CVD, some miRNA-based strategies have resulted in promising findings, including anti-miR-126 approaches that resulted in vascular smooth muscle cells and restenosis inhibition (Santulli et al., 2014b). Also, novel methodologies such as exosome-based delivery of miRNA drugs provide reliable evidence to overcome impediments such as inefficient, unspecific delivery, and immunogenic reactions. But these drugs still need investigations on different aspects of miRNA biology, their long term effects, subsequent biochemical and off-target effects and miRNA pathway analysis.

Acknowledgment

References

Adachi, T., Nakanishi, M., Otsuka, Y., Nishimura, K., Hirokawa, G., Goto, Y., et al. (2010). Plasma microRNA 499 as a biomarker of acute myocardial infarction. *Clin. Chem.* 56, 1183–1185. doi: 10.1373/clinchem.2010.144121

Bang, C., Batkai, S., Dangwal, S., Gupta, S., Foinquinos, A., Holzmann, A., et al. (2014). Cardiac fibroblast–derived microRNA passenger strand-enriched exosomes mediate cardiomyocyte hypertrophy. *J. Clin. Invest.* 124, 2136–2146. doi: 10.1172/JCI70577

Bish, L., Morine, K., Sleeper, M., Sanmiguel, J., Wu, D., Gao, G., et al. (2008). Adeno-associated virus (AAV) serotype 9 provides global cardiac gene transfer superior to AAV1, AAV6, AAV7, and AAV8 in the mouse and rat. *Hum. Gene Ther.* 19, 1359–1368. doi: 10.1089/hum.2008.123

Bonauer, A., Carmona, G., Iwasaki, M., Mione, M., Koyanagi, M., Fischer, A., et al. (2009). MicroRNA-92a controls angiogenesis and functional recovery of ischemic tissues in mice. *Science* 324, 1710–1713. doi: 10.1126/science.1174381

Bostjancic, E., Zidar, N., Stajer, D., and Glavac, D. (2010). MicroRNAs miR-1, miR-133a, miR-133b and miR-208 are dysregulated in human myocardial infarction. *Cardiology* 115, 163–169. doi: 10.1159/000268088

Calin, G., Dumitru, C., Shimizu, M., Bichi, R., Zupo, S., Noch, E., et al. (2002). Frequent deletions and down-regulation of microRNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. U.S.A.* 99, 15524–15529. doi: 10.1073/pnas.242606799

Charan Reddy, K. (2014). Regulatory noncoding RNAs in cardiovascular disease: shedding light on 'Dark Matter.' *J. Cardiovasc. Dis.* 3, 1–7.

Cheng, Y., Tan, N., Yang, J., Liu, X., Cao, X., He, P., et al. (2010). A translational study of circulating cell-free microRNA-1 in acute myocardial infarction. *Clin. Sci. (Lond.)* 119, 87–95. doi: 10.1042/CS20090645

Corsten, M., Dennert, R., Jochems, S., Kuznetsova, T., Devaux, Y., Hofstra, L., et al. (2010). Circulating MicroRNA-208b and MicroRNA-499 reflect myocardial damage in cardiovascular disease. *Circ. Cardiovasc. Genet.* 3, 499–506. doi: 10.1161/CIRCGENETICS.110.957415

Devaux, Y., Mueller, M., Haaf, P., Goretti, E., Twerenbold, R., Zangrando, J., et al. (2015). Diagnostic and prognostic value of circulating microRNAs in patients with acute chest pain. *J. Intern. Med.* 277, 260–271. doi: 10.1111/joim.12183

Devaux, Y., Vausort, M., Mccann, G., Kelly, D., Collignon, O., Ng, L., et al. (2013). A panel of 4 microRNAs facilitates the prediction of left ventricular contractility after acute myocardial infarction. *PLoS ONE* 8:e70644. doi: 10.1371/journal.pone.0070644

Duisters, R., Tijsen, A., Schroen, B., Leenders, J., Lentink, V., Van Der Made, I., et al. (2009). miR-133 and miR-30 regulate connective tissue growth factor:

implications for a role of microRNAs in myocardial matrix remodeling. *Circ. Res.* 104, 170–178. doi: 10.1161/CIRCRESAHA.108.182535

Eiring, A., Harb, J., Neviani, P., Garton, C., Oaks, J., Spizzo, R., et al. (2010). miR-328 functions as an RNA decoy to modulate hnRNP E2 regulation of mRNA translation in leukemic blasts. *Cell* 140, 652–665. doi: 10.1016/j.cell.2010.01.007

Fabbri, M., Paone, A., Calore, F., Galli, R., Gaudio, E., Santhanam, R., et al. (2012). MicroRNAs bind to Toll-like receptors to induce prometastatic inflammatory response. *PNAS* 109, E2110–E2116. doi: 10.1073/pnas.1209414109

Fichtlscherer, S., De Rosa, S., Fox, H., Schwietz, T., Fischer, A., Liebetrau, C., et al. (2010). Circulating microRNAs in patients with coronary artery disease. *Circ. Res.* 107, 677–684. doi: 10.1161/CIRCRESAHA.109.215566

Fiedler, J., Jazbutyte, V., Kirchmaier, B., Gupta, S., Lorenzen, J., Hartmann, D., et al. (2011). MicroRNA-24 regulates vascularity after myocardial infarction. *Circulation* 124, 720–730. doi: 10.1161/CIRCULATIONAHA.111.039008

Fukushima, Y., Nakanishi, M., Nonogi, H., Goto, Y., and Iwai, N. (2011). Assessment of plasma miRNAs in congestive heart failure. *Circ. J.* 75, 336–340. doi: 10.1253/circj.CJ-10-0457

Gray, W., French, K., Ghosh-Choudhary, S., Maxwell, J., Brown, M., Platt, M., et al. (2015). Identification of therapeutic covariant microRNA clusters in hypoxia-treated cardiac progenitor cell exosomes using systems biology. *Circ. Res.* 116, 255–263. doi: 10.1161/CIRCRESAHA.116.304360

Ha, M., and Kim, V. (2014). Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* 15, 509–524. doi: 10.1038/nrm3838

Icli, B., Wara, A., Moslehi, J., Sun, X., Plovie, E., Cahill, M., et al. (2013). MicroRNA-26a regulates pathological and physiological angiogenesis by targeting BMP/SMAD1 signaling. *Circ. Res.* 113, 1231–1241. doi: 10.1161/CIRCRESAHA.113.301780

Ji, X., Takahashi, R., Hiura, Y., Hirokawa, G., Fukushima, Y., and Iwai, N. (2009). Plasma miR-208 as a biomarker of myocardial injury. *Clin. Chem.* 55, 1944–1949. doi: 10.1373/clinchem.2009.125310

Kumarswamy, R., Volkmann, I., Jazbutyte, V., Dangwal, S., Park, D., and Thum, T. (2012). Transforming growth factor-β-induced endothelial-to-mesenchymal transition is partly mediated by microRNA-21. *Arterioscler. Thromb. Vasc. Biol.* 32, 361–369. doi: 10.1161/ATVBAHA.111.234286

Lai, R., Arslan, F., Lee, M., Sze, N., Choo, A., Chen, T., et al. (2010). Exosome secreted by MSC reduces myocardial ischemia/reperfusion injury. *Stem Cell Res.* 4, 214–222. doi: 10.1016/j.scr.2009.12.003

Lee, I., Ajay, S., Yook, J., Kim, H., Hong, S., Kim, N., et al. (2009). New class of microRNA targets containing simultaneous 5′-UTR and 3′-UTR interaction sites. *Genome Res.* 19, 1175–1183. doi: 10.1101/gr.089367.108

Li, C., Fang, Z., Jiang, T., Zhang, Q., Liu, C., Zhang, C., et al. (2013). Serum microRNAs profile from genome-wide serves as a fingerprint for diagnosis of acute myocardial infarction and angina pectoris. *BMC Med. Genomics* 6:16. doi: 10.1186/1755-8794-6-16

Lin, S., and Gregory, R. (2015). MicroRNA biogenesis pathways in cancer. *Nat. Rev. Cancer* 15, 321–333. doi: 10.1038/nrc3932

Lu, J., Getz, G., Miska, E., Alvarez-Saavedra, E., Lamb, J., Peck, D., et al. (2005). MicroRNA expression profiles classify human cancers. *Nature* 435, 834–838. doi: 10.1038/nature03702

Lu, Y., Zhang, Y., Wang, N., Pan, Z., Gao, X., Zhang, F., et al. (2010). MicroRNA-328 contributes to adverse electrical remodeling in atrial fibrillation. *Circulation* 122, 2378–2387. doi: 10.1161/CIRCULATIONAHA.110.958967

Martinez, M., Larbret, F., Zobairi, F., Coulombe, J., Debili, N., Vainchenker, W., et al. (2006). Transfer of differentiation signal by membrane microvesicles harboring hedgehog morphogens. *Blood* 108, 3012–3020. doi: 10.1182/blood-2006-04-019109

Melo, S., Sugimoto, H., O'connell, J., Kato, N., Villanueva, A., Vidal, A., et al. (2014). Cancer exosomes perform cell-independent microRNA biogenesis and promote tumorigenesis. *Cancer Cell* 26, 707–721. doi: 10.1016/j.ccell.2014.09.005

Mendis, S., Puska, P., and Norrving, B. (eds). (2011). *Global Atlas on Cardiovascular Disease Prevention and Control.* Geneva: World Health Organization in collaboration with the World Heart Federation and the World Stroke Organization.

Montecalvo, A., Larregina, A., Shufesky, W., Stolz, D., Sullivan, M., Karlsson, J., et al. (2012). Mechanism of transfer of functional microRNAs between mouse dendritic cells via exosomes. *Blood* 119, 756–766. doi: 10.1182/blood-2011-02-338004

Pegtel, D., Cosmopoulos, K., Thorley-Lawson, D., Van Eijndhoven, M., Hopmans, E., Lindenberg, J., et al. (2010). Functional delivery of viral miRNAs via exosomes. *Proc. Natl. Acad. Sci. U.S.A.* 107, 6328–6333. doi: 10.1073/pnas.0914843107

Ren, J., Zhang, J., Xu, N., Han, G., Geng, Q., Song, J., et al. (2013). Signature of circulating microRNAs as potential biomarkers in vulnerable coronary artery disease. *PLoS ONE* 8:e80738. doi: 10.1371/journal.pone.0080738

Sahoo, S., and Losordo, D. (2014). Exosomes and cardiac repair after myocardial infarction. *Circ. Res.* 114, 333–344. doi: 10.1161/CIRCRESAHA.114.300639

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P. (2011). A ceRNA hypothesis: the rosetta stone of a hidden rna language. *Cell* 146, 353–358. doi: 10.1016/j.cell.2011.07.014

Santulli, G. (2013). Epidemiology of cardiovascular disease in the 21st century: updated numbers and updated facts. *J. Cardiovasc. Dis.* 1, 1–2.

Santulli, G., Iaccarino, G., De Luca, N., Trimarco, B., and Condorelli, G. (2014a). Atrial fibrillation and microRNAs. *Front. Physiol.* 5:15. doi: 10.3389/fphys.2014.00015

Santulli, G., Wronska, A., Uryu, K., Diacovo, T., Gao, M., Marx, S., et al. (2014b). A selective microRNA-based strategy inhibits restenosis while preserving endothelial function. *J. Clin. Invest.* 124, 4102–4114. doi: 10.1172/JCI76069

Sardu, C., Marfella, R., Santulli, G., and Paolisso, G. (2014). Functional role of miRNA in cardiac resynchronization therapy. *Pharmacogenomics* 15, 1159–1168. doi: 10.2217/pgs.14.76

Stoorvogel, W. (2012). Functional transfer of microRNA by exosomes. *Blood* 119, 646–648. doi: 10.1182/blood-2011-11-389478

Tay, Y., Zhang, J., Thomson, A., Lim, B., and Rigoutsos, I. (2008). MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* 455, 1124–1128. doi: 10.1038/nature07299

Terentyev, D., Belevych, A., Terentyeva, R., Martin, M., Malana, G., Kuhn, D., et al. (2009). miR-1 overexpression enhances Ca(2+) release and promotes cardiac arrhythmogenesis by targeting PP2A regulatory subunit B56alpha and causing CaMKII-dependent hyperphosphorylation of RyR2. *Circ. Res.* 104, 514–521. doi: 10.1161/CIRCRESAHA.108.181651

Thum, T., Catalucci, D., and Bauersachs, J. (2008a). MicroRNAs: novel regulators in cardiac development and disease. *Cardiovasc. Res.* 79, 562–570. doi: 10.1093/cvr/cvn137

Thum, T., Gross, C., Fiedler, J., Fischer, T., Kissler, S., Bussen, M., et al. (2008b). MicroRNA-21 contributes to myocardial disease by stimulating MAP kinase signalling in fibroblasts. *Nature* 456, 980–984. doi: 10.1038/nature07511

Tijsen, A., Creemers, E., Moerland, P., De Windt, L., Van Der Wal, A., Kok, W., et al. (2010). MiR423-5p as a circulating biomarker for heart failure. *Circ. Res.* 106, 1035–1039. doi: 10.1161/CIRCRESAHA.110.218297

van Rooij, E., and Olson, E. (2009). Searching for MiR-acles in Cardiac Fibrosis. *Circ. Res.* 104, 138–140. doi: 10.1161/CIRCRESAHA.108.192492

van Rooij, E., and Olson, E. (2012). MicroRNA therapeutics for cardiovascular disease: opportunities and obstacles. *Nat. Rev. Drug Discov.* 11, 860–872. doi: 10.1038/nrd3864

van Rooij, E., Sutherland, L., Liu, N., Williams, A., Mcanally, J., Gerard, R., et al. (2006). A signature pattern of stress-responsive microRNAs that can evoke cardiac hypertrophy and heart failure. *Proc. Natl. Acad. Sci. U.S.A.* 103, 18255–18260. doi: 10.1073/pnas.0608791103

van Rooij, E., Sutherland, L., Qi, X., Richardson, J., Hill, J., and Olson, E. (2007). Control of stress-dependent cardiac growth and gene expression by a microRNA. *Science* 316, 575–579. doi: 10.1126/science.1139089

van Rooij, E., Sutherland, L., Thatcher, J., Dimaio, J., Naseem, R., Marshall, W., et al. (2008). Dysregulation of microRNAs after myocardial infarction reveals a role of miR-29 in cardiac fibrosis. *Proc. Natl. Acad. Sci. U.S.A.* 105, 13027–13032. doi: 10.1073/pnas.0805038105

Wang, F., Long, G., Zhao, C., Li, H., Chaugai, S., Wang, Y., et al. (2013). Plasma microRNA-133a is a new marker for both acute myocardial infarction and underlying coronary artery stenosis. *J. Trans. Med.* 11, 222. doi: 10.1186/1479-5876-11-222

Wang, G., Zhu, J., Zhang, J., Li, Q., Li, Y., He, J., et al. (2010). Circulating microRNA: a novel potential biomarker for early diagnosis of acute myocardial infarction in humans. *Eur. Heart J.* 31, 659–666. doi: 10.1093/eurheartj/ehq013

Widera, C., Gupta, S., Lorenzen, J., Bang, C., Bauersachs, J., Bethmann, K., et al. (2011). Diagnostic and prognostic impact of six circulating microRNAs in acute coronary syndrome. *J. Mol. Cell Cardiol.* 51, 872–875. doi: 10.1016/j.yjmcc.2011.07.011

Wronska, A., Kurkowska-Jastrzebska, I., and Santulli, G. (2015). Application of microRNAs in diagnosis and treatment of cardiovascular disease. *Acta Physiol. (Oxf.)* 213, 60–83. doi: 10.1111/apha.12416

Yang, B., Lin, H., Xiao, J., Lu, Y., Luo, X., Li, B., et al. (2007). The muscle-specific microRNA miR-1 regulates cardiac arrhythmogenic potential by targeting GJA1 and KCNJ2. *Nat. Med.* 13, 486–891. doi: 10.1038/nm1569

Yang, M., Chen, J., Su, F., Yu, B., Su, F., Lin, L., et al. (2011). Microvesicles secreted by macrophages shuttle invasion-potentiating microRNAs into breast cancer cells. *Mol. Cancer* 10, 117. doi: 10.1186/1476-4598-10-117

Yoo, A., Sun, A., Li, L., Shcheglovitov, A., Portmann, T., Li, Y., et al. (2011). MicroRNA-mediated conversion of human fibroblasts to neurons. *Nature* 476, 228–231. doi: 10.1038/nature10323

Zampetaki, A., Willeit, P., Tilling, L., Drozdov, I., Prokopi, M., Renard, J., et al. (2012). Prospective study on circulating MicroRNAs and risk of myocardial infarction. *J. Am. Coll. Cardiol.* 60, 290–294. doi: 10.1016/j.jacc.2012.03.056

Zhang, Y., Liu, D., Chen, X., Li, J., Li, L., Bian, Z., et al. (2010). Secreted monocytic miR-150 enhances targeted endothelial cell migration. *Mol. Cell* 39, 133–144. doi: 10.1016/j.molcel.2010.06.010

Genetic variants in microRNA genes: impact on microRNA expression, function, and disease

Sophia Cammaerts[1], Mojca Strazisar[1], Peter De Rijk[1] and Jurgen Del Favero[1,2]*

[1] Applied Molecular Genomics Unit, Department of Molecular Genetics, VIB, University of Antwerp, Antwerp, Belgium,
[2] Multiplicom N.V., Niel, Belgium

MicroRNAs (miRNAs) are important regulators of gene expression and like any other gene, their coding sequences are subject to genetic variation. Variants in miRNA genes can have profound effects on miRNA functionality at all levels, including miRNA transcription, maturation, and target specificity, and as such they can also contribute to disease. The impact of variants in miRNA genes is the focus of the present review. To put these effects into context, we first discuss the requirements of miRNA transcripts for maturation. In the last part an overview of available databases and tools and experimental approaches to investigate miRNA variants related to human disease is presented.

Keywords: microRNA, genetic variants, expression, function, disease

Introduction

miRNAs are short, non-protein-coding RNA molecules that mediate post-transcriptional regulation by affecting mRNA stability and translational repression or activation (Vasudevan et al., 2007; Filipowicz et al., 2008).

miRNAs were first discovered in 1993 in *Caenorhabditis elegans* (Lee et al., 1993). In 2000–2001 many other miRNA genes were identified in *C. elegans* and miRNAs were shown to be widely present in other species (Pasquinelli et al., 2000; Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001). Since then, given their impact on expression, miRNAs have been extensively studied and the miRBase database, the central miRNA sequence repository, continues to expand, collecting novel miRNA genes in a wide variety of species (Kozomara and Griffiths-Jones, 2014). Even in the human genome, with 1881 precursors and 2588 mature miRNA sequences deposited in miRBase version 21, many novel miRNA genes are being identified and published continuously (Friedländer et al., 2014; Cheng et al., 2015; Londin et al., 2015). Despite the extensive research, we are still uncovering aspects of the miRNA maturation process and we have to date only a limited understanding of the functions of specific miRNAs.

miRNAs add a layer of complexity to gene regulation. They target transcripts mainly by complementarity to the seed region, nucleotides (nt) 2–7 of the mature miRNA molecule (Bartel, 2009). This minimal requirement for complementarity results in targeting of many mRNAs by a single miRNA and targeting of one mRNA by several miRNAs (Lim et al., 2005; Peter, 2010; Wu et al., 2010). In addition, miRNAs can have sequence and length variability (isomiRs), potentially resulting in altered targeting capacity and/or specificity. One of the aspects that further influences the complexity of the miRNA repertoire are genetic variants.

This review provides an overview of the current knowledge of the influence of genetic variants on miRNA biogenesis and function. First, miRNA biogenesis and requirements of the miRNA processing enzymes are described, followed by a discussion of isomiRs and their functional implications. In the second part, we highlight the effect of genetic variants on the expression and functioning of miRNAs. In the last part, approaches that are used to identify miRNAs involved in diseases are discussed, with a focus on genetic strategies. Furthermore, we present available tools, databases, and experimental approaches to aid functional characterization studies of disease-associated miRNA variants.

MicroRNA Biogenesis and Generation of IsomiRs

miRNA genes can be classified in several categories according to their location in the genome. A majority of the currently known human miRNA genes deposited in miRBase are intergenic (68%). Of the intragenic miRNAs, most are intronic (12% of all genes). The remaining genes are located in repeats, lncRNAs, UTRs, or coding regions of host genes (Londin et al., 2015). miRNA genes are often located close to other miRNA genes in so-called clusters. miRNAs located in host genes or clustered with other miRNAs can be cotranscribed, which is supported by good correlation of their expression patterns for several genes (Baskerville and Bartel, 2005). Alternatively, intronic miRNAs can be transcribed independently from their host gene and polycistronic miRNA transcripts can undergo alternative splicing to yield specific miRNA expression (Ramalingam et al., 2014).

In the canonical biogenesis pathway, primary miRNA transcripts (pri-miRNAs) are cleaved in the nucleus by the Microprocessor complex, a complex consisting of ribonuclease Drosha and its cofactor DGCR8, to release the shorter precursor miRNA (pre-miRNA). This precursor molecule is transported to the cytosol by the nucleocytoplasmic transport protein Exportin-5, where it is cleaved by the ribonuclease Dicer to result in a mature miRNA duplex consisting of the mature 5p and 3p strands. The duplex is loaded onto an Argonaute protein, the core component of the RNA induced silencing complex (RISC). One of the strands is discarded, while the other mediates posttranscriptional regulation by base pairing to target mRNAs.

There are different types of target sites. Canonical sites for miRNA targeting are seed matches (nt 2–7 of the mature miRNA) with an adenine opposite from the first nt of the mature miRNA and/or complementarity to the eighth nt of the miRNA. This seed match can be supplemented with complementarity to nt 13–16 (3′ supplementary sites). Extended sites of complementarity at the 3′ end of the miRNA can also compensate for seed mismatches (3′ compensatory sites) (Bartel, 2009). Though a majority of the miRNA-target interactions include seed matches, many non-seed interactions are also observed (Helwak et al., 2013). Similarly, though mRNAs are the predominant miRNA targets, 30% of all miRNA targets are other classes of RNA, such as rRNA, tRNA, snRNA, miRNA, lincRNA, and pseudogenes (Helwak et al., 2013).

Mirtrons, miRNAs encoded within short introns of host genes, bypass processing by the Microprocessor complex. By splicing the mirtron is released and can be, if needed after 5′ or 3′ tail trimming, transported to the cytosol to be further processed by Dicer. In the study of Ladewig et al. (2012), 240 human mirtrons were identified, indicating that bypassing Drosha is not a rare phenomenon. Other non-canonical miRNA biogenesis pathways have also been discovered and were recently reviewed by Abdelfattah et al. (2014). In the next sections, we will focus on the general miRNA characteristics and substrate requirements of the key enzymes of the canonical miRNA maturation pathway.

Maturation Enzyme Requirements
Microprocessor Complex
Drosha, a nuclear ribonuclease, is the enzyme that cleaves the miRNA hairpin from the primary transcript, whereby it generates a precursor with 3′ end overhangs of 2 nt (Lee et al., 2003). An essential cofactor for its pri-miRNA processing activity is the protein DGRC8. The complex containing Drosha and DGCR8, also called the Microprocessor complex, is sufficient for pri-miRNA processing (Denli et al., 2004; Gregory et al., 2004; Han et al., 2004; Landthaler et al., 2004). The capacity of DGCR8 to bind pri-miRNAs is necessary for Drosha cleavage (Yeom et al., 2006). DCGR8 trimerizes upon binding of pri-miRNAs and formation of this protein-pri-miRNA complex may allow specific recognition of pri-miRNAs as opposed to other types of RNA and trigger subsequent cleavage by Drosha (Faller et al., 2010).

Several common features of pri-miRNAs were determined by thermodynamic profiling of secondary structure predictions on a set of human and fly pri-miRNAs: the transcript contains a local hairpin structure with an imperfectly paired stem of ~33 base pairs (bp), consisting of an upper stem (~22 bp, region of the mature miRNA duplex) and a lower stem (~11 bp), a terminal loop region and flanking segments that are usually single-stranded or have large bulges or internal loops (Han et al., 2006). Essential requirements of pri-miRNAs for Drosha processing were experimentally determined. The stem-loop structure within the pri-miRNA needs to be larger than the pre-miRNA: it needs an extended stem (i.e., the lower stem, ~11 bp) and an unstructured region flanking the hairpin (Lee et al., 2003; Chen et al., 2004; Zeng and Cullen, 2005; Zeng et al., 2005; Han et al., 2006; Auyeung et al., 2013). *In vitro* cleavage assays showed that Drosha was able to process pri-miR-30a and pri-miR-16-1 with as little as ~20–25 nt of flanking regions (Lee et al., 2003; Han et al., 2004). However, for correct *in vivo* processing longer flanking regions are required (Chen et al., 2004). The presence of a large unstructured terminal loop region is beneficial for Drosha processing, as reduction of the size of the predicted terminal loop results in reduced processing (Zeng et al., 2005; Han et al., 2006; Zhang and Zeng, 2010). These properties correspond with the *in silico* determined characteristics of known miRNA hairpins.

Two different models have been proposed to explain how Drosha determines where to cleave the pri-miRNA. One study proposed that the Microprocessor mainly determines the distance from the terminal loop region of the hairpin and

cleaves ∼22 bp upstream (Zeng et al., 2005). Another hypothesis proposed that Drosha predominantly cleaves ∼11 bp away from the junction between the lower hairpin stem and the flanking regions (Han et al., 2006). A recent study showed that the complex determines the distance to both ssRNA–dsRNA junctions and that both distances need to be optimal in order to result in precise cleavage (Ma et al., 2013).

Next to the structural prerequisites of the hairpin, position-specific sequence motifs present in the hairpin flanking segments of the primary transcript (UG motif, CNNC motif) and in the loop region (UGU or GUG motif) can enhance processing efficiency in human cells. Nearly 80% of the human pri-miRNAs that are conserved between human and mice contain at least one of these motifs (Auyeung et al., 2013).

Exportin-5

The transport of pre-miRNAs from the nucleus to the cytosol is mediated by Exportin-5 in the presence of cofactor RanGTP (Yi et al., 2003; Bohnsack et al., 2004; Lund et al., 2004). The RNA structure is the main determinant for precursor binding: a stem of at least 18 bp and a blunt end or a 3′ overhang, such as the overhang created by Drosha (Lee et al., 2003), is preferential (Zeng and Cullen, 2004). Upon binding, the 3′ 2 nt overhang of the precursor and a large part of the stem is bound by Exportin-5-RanGTP in a sequence-independent manner, which protects the precursor from degradation by nucleases (Bohnsack et al., 2004; Zeng and Cullen, 2004; Okada et al., 2009).

Dicer

Dicer, a ribonuclease that processes dsRNA into duplexes with a length of ∼22 nt, processes pre-miRNAs into mature miRNA duplexes (Bernstein et al., 2001; Grishok et al., 2001; Hutvágner et al., 2001; Ketting et al., 2001; Knight and Bass, 2001). Dicer cleaves its substrates gradually with a preference to start at the termini of the RNA duplex and generates 2 nt 3′ overhangs (Zhang et al., 2002, 2004). The structure of the termini contributes to the size of the end product, with 2 nt 3′ overhangs resulting in products less than 24 nt, while blunt ends result in longer fragments (Vermeulen et al., 2005). Dicer efficiency is affected by several substrate parameters, such as the size of the loop (in case of pre-miRNAs) and the sequence and the size of the overhangs, with 3′ overhangs of 2 nt being the most efficient substrate for human Dicer (Vermeulen et al., 2005; Lund and Dahlberg, 2006; Zhang and Zeng, 2010; Park et al., 2011). Dicer binds both the 5′ and 3′ end of the precursor molecule and mainly cleaves at a distance of 22 nt from the 5′ end (Park et al., 2011). The accuracy of the cleavage seems to depend on the distance between this canonical cleavage site and the nearest bulge or the terminal loop (Gu et al., 2012). The Drosha-mediated generation of pre-miRNAs with a 3′ overhang of 2 nt and a stem of ∼22 bp thus provides suitable substrates for Dicer. This results in the removal of the loop by Dicer and acquirement of mature miRNA duplexes of ∼22 nt bearing 2 nt overhangs at the 3′ end of both strands, on one side generated by Drosha and at the other side generated by Dicer.

Argonaute

The mature miRNA duplex is loaded onto an Argonaute protein by the RISC loading complex. This complex matures to RISC when one of the strands, termed the passenger strand, is removed (Czech and Hannon, 2011). The other strand, termed the guide strand, can then, in complex with Argonaute, target RNAs via sequence complementarity for posttranscriptional regulation. The process of strand selection is asymmetric: for many miRNAs one of the strands is preferentially retained. This bias is in part explained by the relative stability of the 5′ end of both strands, where the strand with the lowest stability is preferentially retained within RISC (Khvorova et al., 2003; Schwarz et al., 2003). Another contributing factor is the sequence composition of the strands: for human miRNA duplexes with a large difference in expression between both strands, the highest expressed strand has a bias toward a uracil at position 1 and a high purine content, while the lower expressed strand has a bias toward a cytosine at position 1 and a high pyrimidine content (Hu et al., 2009). However, as the dominant strand can vary between tissues (Cloonan et al., 2011), other, yet unknown, factors than sequence and structure of the miRNA duplex itself must be contributing to this variable selection bias.

Proteins Involved in Regulation of Biogenesis

In addition to the biogenesis key proteins described above, a number of other proteins are known to be involved in the regulation of miRNA biogenesis. The following examples highlight the relevance of the terminal loop region in this regulation. The first example is the protein KSRP. This protein binds directly to the terminal loop of a set of miRNAs and enhances their processing by Drosha and Dicer (Trabucchi et al., 2009). Another example is the regulation by Lin28 and terminal uridyl transferase TUTase4. Lin28 binds a GGAG sequence motif in the terminal loop of specific pre-miRNAs, including let-7 family members, upon which TUTase4 binds to the Lin28-pre-miRNA complex and uridylates the 3′ end of the pre-miRNA. This oligo-uridylation prevents processing by Dicer (Heo et al., 2008, 2009; Hagan et al., 2009). In addition to the Lin28-mediated inhibitory role of TUTase4, this enzyme (and other terminal uridyl transferases) also mono-uridylates a subset of pre-let-7 family members and other miRNAs with a 3′ 1 nt overhang, promoting Dicer processing (Heo et al., 2012). An overview of all proteins involved in the regulation of miRNA biogenesis can be found in Siomi and Siomi (2010).

IsomiR Repertoire

Mature miRNAs originating from one arm of a pre-miRNA can have sequence and length heterogeneity *in vivo*, with sequences and lengths closely related to, but different from the canonical miRNA sequence reported in miRBase. Such alternative sequences are termed isomiRs (Morin et al., 2008). Different types of isomiRs have been observed: templated additions or deletions at the 5′ and/or 3′ end of the miRNA, non-templated additions at the 3′ end and substitutions within the sequence (Landgraf et al., 2007; Morin et al., 2008; Martí et al., 2010). Alternative processing by Drosha or

Dicer, Dicer-independent Argonaute2-mediated cleavage and exonuclease degradation of the 3′ terminus of miRNAs can be a source of templated isomiR genesis. Non-templated changes can be established by nucleotidyl transferases, posttranscriptional editing or due to the presence of genetic variants within the transcript (Morin et al., 2008; Cloonan et al., 2011; Neilsen et al., 2012; Lee et al., 2013; Ma et al., 2013).

Since miRNAs target transcripts via imperfect sequence complementarity, isomiRs with a changed sequence for the nucleotides contributing to the target specificity are expected to affect the target spectrum of the miRNA. Given that the majority of miRNA-target interactions include matches to the miRNA seed (Helwak et al., 2013), isomiRs with an altered seed sequence, such as 5′ isomiRs, may have a large influence on target specificity. Sets of isomiRs with an altered seed sequence were predicted to gain the capacity to regulate many additional genes and/or lose the capacity to regulate a subset of genes compared to the canonical miRNA (Gong et al., 2012; Tan et al., 2014). Gain or (partial) loss of function of isomiRs with altered seed sequence and even of isomiRs with a preserved seed has been confirmed experimentally (Kawahara et al., 2007; Gong et al., 2012; Chan et al., 2013; Tan et al., 2014). The size of the effect will, however, be largely determined by the absolute and relative expression and stability of the isomiR compared to the canonical miRNA *in vivo* and its binding efficiency to RISC, which can also be affected by the sequence variability (Chan et al., 2013; Llorens et al., 2013). Alternatively, it has been suggested that isomiRs, which often show a high expression correlation with their canonical miRNAs, may function to regulate the same pathways as the canonical miRNA and that the sequence heterogeneity within isomiR populations may act to reduce off-target effects (Cloonan et al., 2011). The broader functional implications of isomiRs, their contribution to gene expression regulation and how their expression is controlled remains to be further elucidated.

Impact of Genetic Variants on miRNA Expression and Function

As presented above, miRNA transcripts need to fulfill structural and sequence prerequisites in order to result in expression of the correct mature miRNA sequence(s). Sequence variability in miRNA genes can therefore influence both the expression level as the functionality of the miRNA and consequently will result in differential regulation of their target genes.

Large-scale *in silico* analyses of single nucleotide polymorphisms (SNPs) in human miRNA genes have demonstrated that miRNA genes have lower SNP densities than their flanking regions or the human genome (Saunders et al., 2007; Gong et al., 2012; Han and Zheng, 2013). Within the miRNA gene the mature sequence has a lower SNP density than the precursor, with the seed having the lowest SNP density, reflecting their functional importance. In addition, it was shown that there is a negative correlation between the number of SNPs a miRNA gene harbors and the number of diseases the miRNA is associated with (Han and Zheng, 2013). This again highlights the

importance of genetic variants in miRNA genes related to their function and their involvement in human diseases.

Genetic variants can affect miRNAs on several levels. Variants in miRNA promoter regions and other regulatory regions may result in an altered transcription rate. Variants in splice sites of the host gene (for intronic miRNAs) or of the polycistron (clustered miRNAs) could result in aberrant expression patterns. Next, variants within the miRNA transcript can have an effect on miRNA maturation in multiple aspects. They can change the binding affinity of the miRNA hairpin to biogenesis enzymes or accessory proteins (sequence motifs or structural motifs that are changed due to the modified underlying sequence). Variants can lead to altered processing accuracy or to changed frequency of alternative cleavage sites of biogenesis enzymes. They can also lead to altered strand loading bias into RISC. These changes in maturation can all result in altered expression of the canonical miRNA and its existing isomiRs, resulting in deregulation of target genes. It may also result in production of novel isomiRs, which can lead to altered functionality of the miRNA. Lastly, genetic variants within the mature sequence can affect target specificity by generating isomiRs.

Given the potentially huge impact of genetic variants on the tightly regulated miRNA repertoire, and the importance of miRNA-mediated gene regulation, it is not surprising that genetic variants have been found to be causal for or associated with human diseases (Mencía et al., 2009; Hughes et al., 2011). Variants in miRNA biogenesis enzymes and in miRNA binding sites can also lead to impaired miRNA regulation. However, the latter category of variants would likely have a smaller impact than the former two categories, because only the targeting of one gene by one miRNA would be disrupted in case the target does not contain multiple binding sites for that miRNA. Genetic variants in miRNA genes, biogenesis genes and target binding sites associated with human diseases were recently reviewed by Kawahara (2014). Here, we will focus on illustrating the different effects genetic variants can have on miRNA transcription, maturation, and targeting with examples from human miRNA variants studied in disease where possible.

Variants Influencing miRNA Transcription or Splicing

Though bio-informatics approaches have been used to predict miRNA gene promoters (Saini et al., 2007; Marsico et al., 2013), many miRNA gene promoters have not yet been experimentally validated. Below we describe an example of a variant in a miRNA promoter, which was experimentally delineated. The second example is a case of a variant in a host gene promoter for an intronic miRNA cluster. However, experimental investigation of this type of variants can be complicated by the presence of an independent miRNA promoter in addition to the host gene promoter.

SNP rs57095329 is located in the promoter region of MIR146A, 17 kb upstream of the pre-miR-146a sequence. This variant was found to be associated with systemic lupus erythematosus (SLE). The risk allele reduces the binding to transcription factor Ets-1, another SLE susceptibility gene, and

results in decreased promoter activity. Consistent with this mode of action, risk allele carriers have lower expression of miR-146a-5p (Luo et al., 2011).

SNP rs999885 is located in the promoter region of the protein-coding gene MCM7. This variant was associated with a decreased risk of chronic hepatitis B infection, but also with an increased risk of hepatocellular carcinoma (HCC) in individuals with chronic hepatitis B virus infection. The miR-106b-25 cluster, coding for miR-106b, miR-93, and miR-25, is located in the 13th intron of MCM7. HCC patients that carry the risk allele have higher expression of the miRNA transcript in non-tumor liver tissue than non-carriers (Liu et al., 2012). Recently it was shown that this cluster can be transcribed independently of MCM7 via different promoters and that the miRNA polycistronic transcript can undergo alternative splicing (Ramalingam et al., 2014). Therefore it would be very interesting to uncover whether and how this variant influences the relative and absolute expression of the different mature miRNAs in this cluster.

To the best of our knowledge, variants in splice sites that affect miRNA expression in human disease have not been published yet. Nevertheless, the effect of host gene splice variants on mature miRNA expression has been demonstrated by mutagenesis experiments (Janas et al., 2011).

Variants Influencing miRNA Maturation

Variants within the pri-miRNA can affect miRNA expression levels by inducing changes in the maturation process, as described above. This effect can take place at several steps in the maturation pathway: Drosha processing, Dicer processing, and/or altering strand preference.

An example of a variant with and effect on the Drosha processing step is rs2910164. This variant is located in the seed sequence of miR-146a-3p and was predicted to lead to a mismatch in the hairpin stem. The heterozygous variant genotype was associated with increased risk of papillary thyroid carcinoma (PTC). After overexpression of pri-miR-146a with the G allele or the C allele in cells, the C allele resulted in nearly twofold reduction of pre-miR-146a compared to the G allele. The reduced production of pre-miR-146a from pri-miR-146a with a C allele was also confirmed by an *in vitro* processing assay (Jazdzewski et al., 2008). An example of a pri-miRNA variant affecting the first processing step is a variant located four bases upstream of the miR-510-5p sequence. This variant was identified in a study screening X-linked miRNA genes in male schizophrenia patients and control individuals. The variant was predicted to alter the secondary structure around the Drosha processing site. Functional validation showed that the variant results in increased expression of pre-miR-510, miR-510-5p, and miR-510-3p (Feng et al., 2009; Sun et al., 2009).

A variant affecting Dicer processing is rs546098287, located within the seed sequence of miR-96-3p. This variant was identified as the causative mutation in an Italian family with non-syndromic hearing loss (Soldà et al., 2012). Two segregating point mutations in miR-96-5p were previously

identified in two Spanish families with non-syndromic hearing loss, demonstrating the importance of this gene in the disease pathogenesis (Mencía et al., 2009). The mutation in miR-96-3p reduces the processing from pre-miR-96 to both miR-96-5p and miR-96-3p. This is likely induced by a structural change, because the variant was predicted to enlarge an internal loop in the hairpin stem and when a second variant was introduced to restore the original stem structure, the expression was largely brought back to normal levels (Soldà et al., 2012).

An example of a variant influencing the relative strand abundance is a variant in MIR133A2, identified in a patient with atrial fibrillation. The variant is located at the 3′ end of the miR-133a-3p sequence and was predicted to alter the secondary structure of the hairpin at the base of the miR-133a duplex. miR-133a was found to be highly expressed in atrial tissue, with the 3p strand being the dominant one, while miR-133a-5p represented less than 1% of all the miR-133a reads. Functional validation experiments demonstrated that the variant resulted in increased miR-133a-5p expression, but had no significant effect on miR-133a-3p expression. The variant thus increases the relative abundance of the 5p strand compared to the 3p strand (Ohanian et al., 2013).

The cases described above are all variants located either within the precursor sequence or close to it. A study by Diederichs and Haber (2006) reported the functional validation of 15 pri-miRNA variants ranging from 5 to 133 nt outside the pre-miRNA sequence, identified in cancer cell lines. Despite some of these variants had predicted structural changes at or near the base of the miRNA hairpin, none of the variants displayed detectable differences in mature miRNA expression. On the other hand, for the variant rs11134527, which is located 96 nt outside the MIR218-2 hairpin sequence, it was shown that it does change the expression of miR-218-5p significantly (Gao et al., 2013). Therefore variants located further outside the hairpin can still affect miRNA biogenesis, although variants closer to the hairpin are more likely to have a higher impact.

miRNA Variants Influencing Targeting

Beside resulting in altered expression, miRNA variants can also result in altered target specificity by creating isomiRs.

An interesting example of a variant creating a polymorphic isomiR is rs2910164 in miR-146a-3p. This variant also affects Drosha processing as described above (Jazdzewski et al., 2008). The isomiR and canonical miR-146a-3p are predicted to share only a small set of target genes. Transcriptome analysis in healthy and in tumor tissue of PTC patients with GG or GC genotype showed that hundreds of genes (358 in unaffected tissue, 575 genes in tumor tissue) were differentially expressed (Jazdzewski et al., 2009). Even though this does not provide direct evidence that all those genes are direct targets of the isomiR, it does give insight in the large-scale downstream changes isomiRs may induce.

Direct evidence of altered targeting was shown in a study investigating the effect of eight seed SNPs and three mature sequence SNPs (among which one deletion) identified in dbSNP in nine miRNA genes. Four of the tested miRNAs (miR-627-5p, miR-379-5p, miR-499-3p, and miR-124-3p) partially or

completely lost their potential to suppress their original target due to a SNP in their seed sequence. The seed sequence variant rs2620381 in miR-627-5p not only resulted in loss of targeting of SEMA3F, but also in gained capacity to suppress ATP6V0E1 (Gong et al., 2012). In addition, miR-379-5p, miR-940, and miR-34a-5p containing variants in the mature sequence also displayed partial or total loss of function of their original mRNA targets tested.

Studying miRNAs Involved in Disease

miRNA Profiling and Genetic Studies

To study if and which miRNAs play a role in the pathogenesis of a disease, two main strategies can be used: a miRNA profiling approach and a genetic approach.

In the miRNA profiling approach, the affected tissue of a group of patients is subjected to miRNA expression profiling and compared to the expression profile of either the same tissue in healthy individuals, tissue of patients with a different disease or subphenotype, or to adjacent non-affected tissue of the same patients (Calin et al., 2005; Kan et al., 2012; Feliciano et al., 2013; Huang et al., 2014). The tested miRNAs can range from selected candidate miRNAs to a genome-wide approach and different technologies are available, such as RT-qPCR, microarray based methods or RNA sequencing (Pritchard et al., 2012). The miRNA expression profiles are compared between the disease phenotype and the controls in order to determine whether and which miRNAs are differentially expressed. Resulting signatures may be used as diagnostic or prognostic biomarkers (Calin and Croce, 2006) and can provide clues about the role of miRNAs in the disease mechanism. Additionally, interesting miRNAs can be further investigated to dissect their mode of action.

In the genetic approach, the starting point is a linkage or an association analysis (genome-wide or candidate gene approach) including but not necessarily limited to miRNA genes between patients and control individuals (Mencía et al., 2009; Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium, 2011; Zhang et al., 2015). When a genetic variant in or near a miRNA gene is associated with or is causal for the disease, functional validation can be initiated to assess its effect and its role in the pathogenesis of the disease. Variants in miRNA genes linked to the disease could also be used for disease diagnosis, as exemplified by the patent on rare miRNA gene variants identified in schizophrenia patients (Sommer and Rossi, 2012). The genetic screening approach focused on miRNA genes is not yet as frequently applied on a large scale as genome-wide miRNA profiling.

Both strategies have limitations and benefits. General considerations regarding miRNA profiling were reviewed by Pritchard et al. (2012). In the study of disease, genome-wide miRNA profiling can provide a global insight in which miRNAs may play a role in the disease by generating disease-specific miRNA expression signatures. An advantage of this approach is that the deregulated miRNA provides a first indication of its possible functional role in the disease. The main limitation,

however, is that the differential expression of a miRNA in itself does not provide an indication of causality: the altered expression can be a causal factor, a consequence, or could even be unrelated to the disease pathogenesis (Kan et al., 2012). Another constraint is that the affected tissue has to be accessible. This is especially problematic in the study of neurological diseases. Alternatively, brain tissue of deceased patients and healthy individuals, if available, can be used. However, though many miRNAs seem to be relatively stable (Bail et al., 2010; Winter and Diederichs, 2011), the half-lives of some miRNAs are less than 1 hour (Sethi and Lukiw, 2009). This implicates that when brain samples with different postmortem delay from death to tissue processing are compared, variability may arise due to differential instability of the miRNAs in the samples. The restricted access to postmortem brain samples can also result in suboptimal sample sizes. A much more accessible proxy for unavailable affected tissues is blood. While this may provide a good avenue for determining diagnostic or prognostic biomarkers, it is not clear whether the miRNAs deregulated in blood would also be deregulated in the affected tissue. In addition, several factors can influence miRNA expression in blood, such as age, drug use, and other diseases (Margis et al., 2011; de Boer et al., 2013; Meder et al., 2014).

The main advantage of a genetic approach is that when a variant is associated with disease risk, it gives an indication that the miRNA involved has a primary effect in the disease pathomechanism (large or small contributing effect), which cannot be determined by expression profiling. This advantage also has a downside: while the association or linkage analysis may indicate a miRNA variant is likely to be involved in the disease, it is not at all clear whether the identified variant has a functional effect and whether the expression of that miRNA is altered in the disease state. Another asset of the genetic approach is that diseases for which the main affected tissue is not accessible can still be studied because genomic DNA can be isolated from more accessible patient material, such as blood or saliva. There is, however, still a chance of missing a mutation, if the patient is mosaic for the mutation and the load of the mutation is lower in the examined tissue than in the affected tissue.

Both approaches are complementary and should both be used to determine the involvement of a miRNA in relation to a disease: genetic approaches can pinpoint the causal or risk factors involved in the disease, while expression studies reveal whether and how the expression of the disease-associated miRNAs is altered in the affected tissue. Neither approach is sufficient to determine both. An example of a study combining both approaches is the study of Calin et al. (2005) where miRNA profiling in affected tissue of patients was performed, followed by variant screening in a set of relevant miRNA genes. An expression signature of thirteen miRNA genes was found to be able to distinguish between patient subgroups. In two of these genes, patient-specific variants were found. One of the variants was a functional germline mutation in pri-miR-15a/16-1. On the other hand, when a genetic variant is associated with a disease, usually only the expression of that particular miRNA is assessed in patient material. The profiling and genetic approach has also been

combined on a large scale by miRNA profiling in blood samples of patients with different diseases and investigating whether the deregulated miRNA genes were located within 250 kb of published GWAS significant SNPs (Keller et al., 2011). While on average 103 deregulated miRNAs were identified in each disease, of all the deregulated miRNAs they found only six miRNA genes in proximity to a GWAS hit associated with the same disease. Combining the genetic and the profiling approach on the same set of patients could thus provide an invaluable amount of information on both DNA and RNA level from which the disease-contributing miRNAs can be separated from the miRNAs that are probably deregulated due to secondary disease processes.

Databases and Tools for Genetic Approaches

To provide direct evidence that a genetic variant in or near a miRNA gene and the deregulated miRNA expression in disease tissue are linked, detailed functional validation is required. Functional validation is labor-intensive, costly, and time-consuming, therefore researchers usually prioritize the potential candidate variants using *in silico* approaches. In this section, we will focus on available databases and tools for the impact prediction of genetic variants on miRNA function (**Tables 1** and **2**). Different types of resources for miRNA research have been reviewed by Vlachos and Hatzigeorgiou (2013).

A first step in the assessment of the relevance of a genetic variant or a set of genetic variants is determining whether the variant is located near or in a specific region of a miRNA gene. For a single variant of interest, the variant location may be assessed in a genome browser to determine the position of the variant compared to neighboring miRNA genes. To evaluate the exact location of the variant relative to the different miRNA gene regions (e.g., seed, mature, terminal loop), one also has

to compare this location with the information provided about the gene in miRBase (Kozomara and Griffiths-Jones, 2014). This procedure is tedious and error-prone, especially when investigating multiple miRNA variants. For variant lists such as those derived from massively parallel sequencing projects, the first annotation step can be done with any of the annotation tools available to determine which (miRNA) genes, if any, are located near the variants.

Online repositories compiling information about known SNPs in miRNA genes, such as miRNASNP (Gong et al., 2012), miRNA SNiPer (Zorc et al., 2015), and miRvar (Bhartiya et al., 2011), can be used to extract exact location information for these variants relative to the miRNA hairpin. A limitation for the last two databases is that the search needs to be initiated by using the miRNA gene name instead of the SNP ID. In addition, databases cannot be used when investigating novel variants and are less suitable when investigating multiple variants in several miRNA genes. To eliminate the time-consuming processes related to assessing the position of variants in the context of miRNA genes and what their predicted structural impact is, we recently developed the freely available software miRVaS[1]. Required input is the genomic location of a (known or novel) variant or of a list of variants.

The next possible step in determining the potential relevance of a genetic variant is predicting its effect on the secondary structure of the miRNA. For known SNPs within miRNA genes, structure prediction results of the miRNA hairpin with and without the variant can be found in the miRNASNP database (Gong et al., 2012). The difference in free energy is calculated and a predicted effect on the miRNA expression is provided, based on the assumption that variants destabilizing the hairpin

---

[1]http://mirvas.bioinf.be/

TABLE 1 | Overview of described databases.

Name	Use	URL
miRBase	Database for miRNA sequences and gene annotations	http://www.mirbase.org/
miRNASNP	Database for miRNA gene variants, location annotation, RNA secondary structure impact prediction, target impact prediction for seed variants	http://bioinfo.life.hust.edu.cn/miRNASNP2/index.php
miRNA SNiPer	Database for miRNA gene variants, location annotation	http://www.integratomics-time.com/miRNA-SNiPer/
miRvar	Database for miRNA gene variants, location annotation, predicted impact on Dicer and RISC loading	http://genome.igib.res.in/mirlovd/home.php
PolymiRTS Database	Database with miRNA seed variants and predicted effect on targeting	http://compbio.uthsc.edu/miRSNP/
dPORE-miRNA	Database with miRNA promoter variants and their predicted effect on transcription binding	http://cbrc.kaust.edu.sa/dpore/

The Use column refers to the usage described in the text. Most of the databases also incorporate other information not mentioned here.

TABLE 2 | Overview of described software.

Name	Use	URL
miRVaS	Location annotation, RNA secondary structure impact prediction for miRNA gene variants	http://mirvas.bioinf.be/
Mfold	RNA secondary structure prediction	http://mfold.rna.albany.edu/
RNAfold	RNA secondary structure prediction	http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi
PHDcleav	Prediction of Dicer processing sites	http://www.imtech.res.in/raghava/phdcleav/
RISC binder	Prediction of preferential strand loaded into RISC	http://crdd.osdd.net:8081/RISCbinder/

will reduce the expression and stabilizing variants will increase the expression of the mature miRNA (Gong et al., 2012). For novel variants the RNA secondary structures can be predicted for wild type and variant sequences using web-based RNA structure prediction tools, such as the Mfold or RNAfold web servers (Zuker, 2003; Gruber et al., 2008), and compared. Again, this process is time-consuming and not suitable for the analysis of large variant sets. miRVaS[1] automates this type of analysis, predicting structural impact for lists of variants, needing only the genomic location and alternative allele of the variant.

If a variant is located in a miRNA seed sequence, the effect on target binding can also be assessed *in silico*. Both PolymiRTS database (Bhattacharya et al., 2014) and miRNASNP (Gong et al., 2012) contain lists of predicted gain and loss of target binding sites for seed sequence variants that can be browsed. To assess the impact of variants in miRNA promoter regions on transcription factor binding, the dPORE-miRNA database can be queried (Schmeier et al., 2011). Algorithms to predict Dicer cleavage patterns and preferential strand loading into RISC, such as PHDcleav and RISC binder (Ahmed et al., 2009, 2013) could also be useful to assess the impact of variants. The predicted effect on Dicer cleavage and RISC binding induced by variants is incorporated in the miRvar database (Bhartiya et al., 2011).

Functional Characterization of a Disease-Associated miRNA Variant

Using tools and databases, a set of genetic variants can be prioritized based on their potential functional relevance. The most promising candidates can then be chosen for direct functional validation in cell-based assays. The assays used depend on the location of the variant and the hypothesis.

For variants in miRNA promoter regions, the wild type and variant promoter can be cloned into a promoter-less reporter gene vector and transfected (transiently or stably) into cell lines to assess whether there is a difference in promoter activity due to the variant and whether the variant promoter has different binding affinity to transcription factors (Luo et al., 2011).

To assess the impact of a genetic variant on miRNA biogenesis and/or targeting, cell lines overexpressing wild type and mutant miRNAs are usually established. The constructs should include the full precursor gene and upstream and downstream flanking sequences. The inclusion of flanking regions is crucial: the biogenesis machinery in the cell may not efficiently recognize or process the transcript when the flanking regions are absent or too small, as evidenced by the study of Chen et al. (2004) where only constructs for miR-223 with at least 40 nt of flanking region resulted in detectable miR-223 expression. Using up- and downstream flanks of 125 nt resulted in mature miRNA expression for all 13 tested miRNAs in this study, showing that the incorporation of longer flanks in miRNA constructs is generally a good strategy. To assess the effect on biogenesis, miRNA expression analysis is performed. Comparison of wild type and variant miRNA levels can be done at the level of the mature miRNA (Shen et al., 2009), assessing the total effect of the variant on the whole biogenesis process, or it can be done on several maturation levels of the miRNA, to deduce the exact step of the biogenesis that is affected by the variant (Duan et al., 2007).

If a variant is located in the mature miRNA sequence, targeting may be affected. Investigating which genes and pathways are affected due to the variant may provide insight in the disease mechanism by identifying new pathways and/or by confirming pathways previously hypothesized to be involved in the disease. Investigation and validation of direct interactions between miRNA and target RNAs can be done using reporter assays. Hereto, the potential binding site is cloned downstream of a reporter gene and cells are cotransfected with the miRNA and the reporter vector to assess whether the miRNA (or the variant miRNA) can target the site and thus reduce the reporter gene activity (Duan et al., 2007; Gong et al., 2012). While this allows confirmation of direct interactions, overexpression of the binding site and the miRNA may also lead to validation of interactions that may not take place endogenously in the cell type or tissue of interest, due to, for instance, absence of colocalization, or co-expression.

Investigation of variant induced targeting alterations can also be performed on a large scale by subjecting the wild type and variant miRNA expressing cells to transcriptome and/or proteome profiling. Single genes or proteins can be analyzed for differential expression. Pathway analysis can also identify which set of genes are influenced by the variant miRNA and may lead to biological meaningful findings about the variant miRNA and the pathways in the cell it affects in the disease state (Jazdzewski et al., 2009). If a variant is located outside the mature miRNA sequence, this approach can still be used if the variant affects mature miRNA expression, as the altered expression of the miRNA will also deregulate its target genes (Strazisar et al., 2015). However, in contrast to the reporter gene assays, these large-scale approaches cannot distinguish between direct and indirect targets of the miRNA, so target prediction algorithms need to be used to determine which of the deregulated genes can potentially be targeted directly by the miRNA.

Different crosslinking and immunoprecipitation (CLIP) approaches were developed and applied for global miRNA–RNA target interaction identification by Argonaute precipitation, such as HITS-CLIP and PAR-CLIP (Ule et al., 2005; Chi et al., 2009; Hafner et al., 2010). In these methods, RNA and proteins are crosslinked in cells by UV irradiation (in case of PAR-CLIP after incorporation of photoactivatable nucleosides). After cell lysis and partial RNA digestion, Argonaute-RNA complexes are pulled down and purified stringently. Subsequently, coprecipitated RNA is sequenced to identify which miRNAs and which target RNA regions were bound to Argonaute. Because miRNA and target RNAs are not physically linked to each other during the procedure, *in silico* target prediction is required to assess the interactions between the identified RNA sequences. These approaches thus provide enrichment for actual target RNA sites bound to Argonaute and limits the target prediction to these sites and to the detected miRNAs.

A method that can identify direct miRNA-target interactions on a large scale is the crosslinking, ligation, and sequencing of hybrids (CLASH) technology (Kudla et al., 2011; Helwak et al., 2013). Similar to the CLIP techniques, cells are UV irradiated

to mediate crosslinking, after which Argonaute-RNA complexes are pulled down and RNA is partially digested. Then a ligation step is performed to ligate miRNAs to their target RNAs. The RNA is sequenced to identify chimeric RNAs. The non-chimeric reads provide the same information as the CLIP experiments, while the chimeric RNAs contain linked miRNA and target RNA fragments and hence provide direct proof of miRNA-target interactions. This technology may thus be used to directly identify whether mature miRNA variants alter targeting capacity of a disease-associated miRNA.

Together, these approaches can be used to assess what the effect of the genetic variant is on the expression and the functionality of the miRNA.

Concluding Remarks

In this review we summarized the current knowledge of the miRNA maturation process, focusing on the prerequisites of the miRNA transcript to be processed. From these requirements it is clear that the RNA structure, and thus also the underlying sequence determining the structure, are of paramount importance for maturation into mature miRNAs. Thus genetic variants in miRNA genes can have large effects, not only on the miRNA gene itself, but due to its role as a fine-tuner of gene expression and translation, also on its downstream targets. In line with this, different miRNA variants have been found to be involved or hypothesized to be involved in human disease. We also highlighted the biological relevance of genetic variants located within or near miRNA genes and provided an overview of *in silico* and experimental approaches to investigate the effect of these variants on miRNA expression and function. This in turn could then be correlated to changes in miRNA and target

gene abundance in the affected tissue of patients, as a first step in the understanding of the role of the miRNA in the disease pathomechanism.

As we are now coming to realize that the miRNA repertoire is much more complex than previously appreciated with the recognition of the wide-spread presence of isomiRs and their functional relevance, the possibilities of intronic miRNAs being transcribed independently from their host genes and even alternative splicing, many more questions arise. Though many aspects of miRNA biogenesis and functioning have already been unraveled, more investigations will need to be done to be able to fully grasp the regulation and the genesis of the correct expression pattern of the full miRNA repertoire in the cell. We also anticipate that, especially for the investigation of genetic variants in disease, we will need to undertake comprehensive and integrative approaches to be able to fully appreciate how certain genetic variants can undermine this regulatory system and lead to disease, while others remain completely harmless.

Author Contributions

All authors contributed to the conception of the work. SC wrote the manuscript. MS, PD, and JD provided input and revised the manuscript. All authors approved the final manuscript.

Acknowledgments

References

Abdelfattah, A. M., Park, C., and Choi, M. Y. (2014). Update on non-canonical microRNAs. *Biomol. Concepts* 5, 275–287. doi: 10.1515/bmc-2014-0012

Ahmed, F., Ansari, H. R., and Raghava, G. P. S. (2009). Prediction of guide strand of microRNAs from its sequence and secondary structure. *BMC Bioinformatics* 10:105. doi: 10.1186/1471-2105-10-105

Ahmed, F., Kaundal, R., and Raghava, G. P. S. (2013). PHDcleav: a SVM based method for predicting human Dicer cleavage sites using sequence and secondary structure of miRNA precursors. *BMC Bioinformatics* 14(Suppl. 14):S9. doi: 10.1186/1471-2105-14-S14-S9

Auyeung, V. C., Ulitsky, I., McGeary, S. E., and Bartel, D. P. (2013). Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. *Cell* 152, 844–858. doi: 10.1016/j.cell.2013.01.031

Bail, S., Swerdel, M., Liu, H., Jiao, X., Goff, L. A., Hart, R. P., et al. (2010). Differential regulation of microRNA stability. *RNA* 16, 1032–1039. doi: 10.1261/rna.1851510

Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233. doi: 10.1016/j.cell.2009.01.002

Baskerville, S., and Bartel, D. P. (2005). Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA* 11, 241–247. doi: 10.1261/rna.7240905

Bernstein, E., Caudy, A. A., Hammond, S. M., and Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409, 363–366. doi: 10.1038/35053110

Bhartiya, D., Laddha, S. V., Mukhopadhyay, A., and Scaria, V. (2011). miRvar: a comprehensive database for genomic variations in microRNAs. *Hum. Mutat.* 32, E2226–E2245. doi: 10.1002/humu.21482

Bhattacharya, A., Ziebarth, J. D., and Cui, Y. (2014). PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic Acids Res.* 42, D86–D91. doi: 10.1093/nar/gkt1028

Bohnsack, M. T., Czaplinski, K., and Gorlich, D. (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* 10, 185–191. doi: 10.1261/rna.5167604

Calin, G. A., and Croce, C. M. (2006). MicroRNA signatures in human cancers. *Nat. Rev. Cancer* 6, 857–866. doi: 10.1038/nrc1997

Calin, G. A., Ferracin, M., Cimmino, A., Di Leva, G., Shimizu, M., Wojcik, S. E., et al. (2005). A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N. Engl. J. Med.* 353, 1793–1801. doi: 10.1056/NEJMoa050995

Chan, Y.-T., Lin, Y.-C., Lin, R.-J., Kuo, H.-H., Thang, W.-C., Chiu, K.-P., et al. (2013). Concordant and discordant regulation of target genes by miR-31 and its isoforms. *PLoS ONE* 8:e58169. doi: 10.1371/journal.pone.0058169

Chen, C.-Z., Li, L., Lodish, H. F., and Bartel, D. P. (2004). MicroRNAs modulate hematopoietic lineage differentiation. *Science* 303, 83–86. doi: 10.1126/science.1091903

Cheng, W.-C., Chung, I.-F., Tsai, C.-F., Huang, T.-S., Chen, C.-Y., Wang, S.-C., et al. (2015). YM500v2: a small RNA sequencing (smRNA-seq) database for human cancer miRNome research. *Nucleic Acids Res.* 43, D862–D867. doi: 10.1093/nar/gku1156

Chi, S. W., Zang, J. B., Mele, A., and Darnell, R. B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* 460, 479–486. doi: 10.1038/nature08170

Cloonan, N., Wani, S., Xu, Q., Gu, J., Lea, K., Heater, S., et al. (2011). MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome Biol.* 12:R126. doi: 10.1186/gb-2011-12-12-r126

Czech, B., and Hannon, G. J. (2011). Small RNA sorting: matchmaking for Argonautes. *Nat. Rev. Genet.* 12, 19–31. doi: 10.1038/nrg2916

de Boer, H. C., van Solingen, C., Prins, J., Duijs, J. M. G. J., Huisman, M. V., Rabelink, T. J., et al. (2013). Aspirin treatment hampers the use of plasma microRNA-126 as a biomarker for the progression of vascular disease. *Eur. Heart J.* 34, 3451–3457. doi: 10.1093/eurheartj/eht007

Denli, A. M., Tops, B. B. J., Plasterk, R. H. A., Ketting, R. F., and Hannon, G. J. (2004). Processing of primary microRNAs by the Microprocessor complex. *Nature* 432, 231–235. doi: 10.1038/nature03049

Diederichs, S., and Haber, D. A. (2006). Sequence variations of microRNAs in human cancer: alterations in predicted secondary structure do not affect processing. *Cancer Res.* 66, 6097–6104. doi: 10.1158/0008-5472.CAN-06-0537

Duan, R., Pak, C., and Jin, P. (2007). Single nucleotide polymorphism associated with mature miR-125a alters the processing of pri-miRNA. *Hum. Mol. Genet.* 16, 1124–1131. doi: 10.1093/hmg/ddm062

Faller, M., Toso, D., Matsunaga, M., Atanasov, I., Senturia, R., Chen, Y., et al. (2010). DGCR8 recognizes primary transcripts of microRNAs through highly cooperative binding and formation of higher-order structures. *RNA* 16, 1570–1583. doi: 10.1261/rna.2111310

Feliciano, A., Castellvi, J., Artero-Castro, A., Leal, J. A., Romagosa, C., Hernández-Losa, J., et al. (2013). miR-125b acts as a tumor suppressor in breast tumorigenesis via its novel direct targets ENPEP, CK2-α, CCNJ, and MEGF9. *PLoS ONE* 8:e76247. doi: 10.1371/journal.pone.0076247

Feng, J., Sun, G., Yan, J., Noltner, K., Li, W., Buzin, C. H., et al. (2009). Evidence for X-chromosomal schizophrenia associated with microRNA alterations. *PLoS ONE* 4:e6121. doi: 10.1371/journal.pone.0006121

Filipowicz, W., Bhattacharyya, S. N., and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat. Rev. Genet.* 9, 102–114. doi: 10.1038/nrg2290

Friedländer, M. R., Lizano, E., Houben, A. J. S., Bezdan, D., Báñez-Coronel, M., Kudla, G., et al. (2014). Evidence for the biogenesis of more than 1,000 novel human microRNAs. *Genome Biol.* 15:R57. doi: 10.1186/gb-2014-15-4-r57

Gao, X., Yang, L., Ma, Y., Yang, J., Zhang, G., Huang, G., et al. (2013). No association of functional variant in pri-miR-218 and risk of congenital heart disease in a Chinese population. *Gene* 523, 173–177. doi: 10.1016/j.gene.2013.03.119

Gong, J., Tong, Y., Zhang, H.-M., Wang, K., Hu, T., Shan, G., et al. (2012). Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. *Hum. Mutat.* 33, 254–263. doi: 10.1002/humu.21641

Gregory, R. I., Yan, K.-P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., et al. (2004). The Microprocessor complex mediates the genesis of microRNAs. *Nature* 432, 235–240. doi: 10.1038/nature03120

Grishok, A., Pasquinelli, A. E., Conte, D., Li, N., Parrish, S., Ha, I., et al. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* 106, 23–34. doi: 10.1016/S0092-8674(01)00431-7

Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R., and Hofacker, I. L. (2008). The Vienna RNA websuite. *Nucleic Acids Res.* 36, W70–W74. doi: 10.1093/nar/gkn188

Gu, S., Jin, L., Zhang, Y., Huang, Y., Zhang, F., Valdmanis, P. N., et al. (2012). The loop position of shRNAs and pre-miRNAs is critical for the accuracy of dicer processing in vivo. *Cell* 151, 900–911. doi: 10.1016/j.cell.2012.09.042

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141, 129–141. doi: 10.1016/j.cell.2010.03.009

Hagan, J. P., Piskounova, E., and Gregory, R. I. (2009). Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* 16, 1021–1025. doi: 10.1038/nsmb.1676

Han, J., Lee, Y., Yeom, K.-H., Kim, Y.-K., Jin, H., and Kim, V. N. (2004). The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.* 18, 3016–3027. doi: 10.1101/gad.1262504

Han, J., Lee, Y., Yeom, K.-H., Nam, J.-W., Heo, I., Rhee, J.-K., et al. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* 125, 887–901. doi: 10.1016/j.cell.2006.03.043

Han, M., and Zheng, Y. (2013). Comprehensive analysis of single nucleotide polymorphisms in human microRNAs. *PLoS ONE* 8:e78028. doi: 10.1371/journal.pone.0078028

Helwak, A., Kudla, G., Dudnakova, T., and Tollervey, D. (2013). Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* 153, 654–665. doi: 10.1016/j.cell.2013.03.043

Heo, I., Ha, M., Lim, J., Yoon, M.-J., Park, J.-E., Kwon, S. C., et al. (2012). Mono-uridylation of pre-microRNA as a key step in the biogenesis of group II let-7 microRNAs. *Cell* 151, 521–532. doi: 10.1016/j.cell.2012.09.022

Heo, I., Joo, C., Cho, J., Ha, M., Han, J., and Kim, V. N. (2008). Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol. Cell* 32, 276–284. doi: 10.1016/j.molcel.2008.09.014

Heo, I., Joo, C., Kim, Y.-K., Ha, M., Yoon, M.-J., Cho, J., et al. (2009). TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell* 138, 696–708. doi: 10.1016/j.cell.2009.08.002

Hu, H. Y., Yan, Z., Xu, Y., Hu, H., Menzel, C., Zhou, Y. H., et al. (2009). Sequence features associated with microRNA strand selection in humans and flies. *BMC Genomics* 10:413. doi: 10.1186/1471-2164-10-413

Huang, Y., Liu, Y., Li, L., Su, B., Yang, L., Fan, W., et al. (2014). Involvement of inflammation-related miR-155 and miR-146a in diabetic nephropathy: implications for glomerular endothelial injury. *BMC Nephrol.* 15:142. doi: 10.1186/1471-2369-15-142

Hughes, A. E., Bradley, D. T., Campbell, M., Lechner, J., Dash, D. P., Simpson, D. A., et al. (2011). Mutation altering the miR-184 seed region causes familial keratoconus with cataract. *Am. J. Hum. Genet.* 89, 628–633. doi: 10.1016/j.ajhg.2011.09.014

Hutvágner, G., McLachlan, J., Pasquinelli, A. E., Bálint, E., Tuschl, T., and Zamore, P. D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* 293, 834–838. doi: 10.1126/science.1062961

Janas, M. M., Khaled, M., Schubert, S., Bernstein, J. G., Golan, D., Veguilla, R. A., et al. (2011). Feed-forward microprocessing and splicing activities at a microRNA-containing intron. *PLoS Genet.* 7:e1002330. doi: 10.1371/journal.pgen.1002330

Jazdzewski, K., Liyanarachchi, S., Swierniak, M., Pachucki, J., Ringel, M. D., Jarzab, B., et al. (2009). Polymorphic mature microRNAs from passenger strand of pre-miR-146a contribute to thyroid cancer. *Proc. Natl. Acad. Sci. U.S.A.* 106, 1502–1505. doi: 10.1073/pnas.0812591106

Jazdzewski, K., Murray, E. L., Franssila, K., Jarzab, B., Schoenberg, D. R., and de la Chapelle, A. (2008). Common SNP in pre-miR-146a decreases mature miR expression and predisposes to papillary thyroid carcinoma. *Proc. Natl. Acad. Sci. U.S.A.* 105, 7269–7274. doi: 10.1073/pnas.0802682105

Kan, A. A., van Erp, S., Derijck, A. A. H. A., de Wit, M., Hessel, E. V. S., O'Duibhir, E., et al. (2012). Genome-wide microRNA profiling of human temporal lobe epilepsy identifies modulators of the immune response. *Cell. Mol. Life Sci.* 69, 3127–3145. doi: 10.1007/s00018-012-0992-7

Kawahara, Y. (2014). Human diseases caused by germline and somatic abnormalities in microRNA and microRNA-related genes. *Congenit. Anom. (Kyoto)* 54, 12–21. doi: 10.1111/cga.12043

Kawahara, Y., Zinshteyn, B., Sethupathy, P., Iizasa, H., Hatzigeorgiou, A. G., and Nishikura, K. (2007). Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science* 315, 1137–1140. doi: 10.1126/science.1138050

Keller, A., Leidinger, P., Bauer, A., Elsharawy, A., Haas, J., Backes, C., et al. (2011). Toward the blood-borne miRNome of human diseases. *Nat. Methods* 8, 841–843. doi: 10.1038/nmeth.1682

Ketting, R. F., Fischer, S. E., Bernstein, E., Sijen, T., Hannon, G. J., and Plasterk, R. H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* 15, 2654–2659. doi: 10.1101/gad.927801

Khvorova, A., Reynolds, A., and Jayasena, S. D. (2003). Functional siRNAs and miRNAs exhibit strand bias. *Cell* 115, 209–216. doi: 10.1016/S0092-8674(03)00801-8

Knight, S. W., and Bass, B. L. (2001). A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science* 293, 2269–2271. doi: 10.1126/science.1062039

Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42, D68–D73. doi: 10.1093/nar/gkt1181

Kudla, G., Granneman, S., Hahn, D., Beggs, J. D., and Tollervey, D. (2011). Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10010–10015. doi: 10.1073/pnas.1017386108

Ladewig, E., Okamura, K., Flynt, A. S., Westholm, J. O., and Lai, E. C. (2012). Discovery of hundreds of mirtrons in mouse and human small RNA data. *Genome Res.* 22, 1634–1645. doi: 10.1101/gr.133553.111

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. *Science* 294, 853–858. doi: 10.1126/science.1064921

Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., et al. (2007). A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* 129, 1401–1414. doi: 10.1016/j.cell.2007.04.040

Landthaler, M., Yalcin, A., and Tuschl, T. (2004). The human DiGeorge syndrome critical region gene 8 and Its *D. melanogaster* homolog are required for miRNA biogenesis. *Curr. Biol.* 14, 2162–2167. doi: 10.1016/j.cub.2004.11.001

Lau, N. C., Lim, L. P., Weinstein, E. G., and Bartel, D. P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 294, 858–862. doi: 10.1126/science.1065062

Lee, H. Y., Zhou, K., Smith, A. M., Noland, C. L., and Doudna, J. A. (2013). Differential roles of human Dicer-binding proteins TRBP and PACT in small RNA processing. *Nucleic Acids Res.* 41, 6568–6576. doi: 10.1093/nar/gkt361

Lee, R. C., and Ambros, V. (2001). An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* 294, 862–864. doi: 10.1126/science.1065329

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75, 843–854. doi: 10.1016/0092-8674(93)90529-Y

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., et al. (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature* 425, 415–419. doi: 10.1038/nature01957

Lim, L. P., Lau, N. C., Garrett-Engele, P., Grimson, A., Schelter, J. M., Castle, J., et al. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 433, 769–773. doi: 10.1038/nature03315

Liu, Y., Zhang, Y., Wen, J., Liu, L., Zhai, X., Liu, J., et al. (2012). A genetic variant in the promoter region of miR-106b-25 cluster and risk of HBV infection and hepatocellular carcinoma. *PLoS ONE* 7:e32230. doi: 10.1371/journal.pone.0032230

Llorens, F., Bañez-Coronel, M., Pantano, L., del Río, J. A., Ferrer, I., Estivill, X., et al. (2013). A highly expressed miR-101 isomiR is a functional silencing small RNA. *BMC Genomics* 14:104. doi: 10.1186/1471-2164-14-104

Londin, E., Loher, P., Telonis, A. G., Quann, K., Clark, P., Jing, Y., et al. (2015). Analysis of 13 cell types reveals evidence for the expression of numerous novel primate- and tissue-specific microRNAs. *Proc. Natl. Acad. Sci. U.S.A.* 112, E1106–E1115. doi: 10.1073/pnas.1420955112

Lund, E., and Dahlberg, J. E. (2006). Substrate selectivity of exportin 5 and Dicer in the biogenesis of microRNAs. *Cold Spring Harb. Symp. Quant. Biol.* 71, 59–66. doi: 10.1101/sqb.2006.71.050

Lund, E., Güttinger, S., Calado, A., Dahlberg, J. E., and Kutay, U. (2004). Nuclear export of microRNA precursors. *Science* 303, 95–98. doi: 10.1126/science.1090599

Luo, X., Yang, W., Ye, D.-Q., Cui, H., Zhang, Y., Hirankarn, N., et al. (2011). A functional variant in microRNA-146a promoter modulates its expression and confers disease risk for systemic lupus erythematosus. *PLoS Genet.* 7:e1002128. doi: 10.1371/journal.pgen.1002128

Ma, H., Wu, Y., Choi, J.-G., and Wu, H. (2013). Lower and upper stem-single-stranded RNA junctions together determine the Drosha cleavage site. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20687–20692. doi: 10.1073/pnas.1311639110

Margis, R., Margis, R., and Rieder, C. R. M. (2011). Identification of blood microRNAs associated to Parkinsonis disease. *J. Biotechnol.* 152, 96–101. doi: 10.1016/j.jbiotec.2011.01.023

Marsico, A., Huska, M. R., Lasserre, J., Hu, H., Vucicevic, D., Musahl, A., et al. (2013). PROmiRNA: a new miRNA promoter recognition method uncovers the complex regulation of intronic miRNAs. *Genome Biol.* 14:R84. doi: 10.1186/gb-2013-14-8-r84

Martí, E., Pantano, L., Bañez-Coronel, M., Llorens, F., Miñones-Moyano, E., Porta, S., et al. (2010). A myriad of miRNA variants in control and Huntington's disease brain regions detected by massively parallel sequencing. *Nucleic Acids Res.* 38, 7219–7235. doi: 10.1093/nar/gkq575

Meder, B., Backes, C., Haas, J., Leidinger, P., Stähler, C., Großmann, T., et al. (2014). Influence of the confounding factors age and sex on microRNA profiles from peripheral blood. *Clin. Chem.* 60, 1200–1208. doi: 10.1373/clinchem.2014.224238

Mencía, A., Modamio-Høybjør, S., Redshaw, N., Morín, M., Mayo-Merino, F., Olavarrieta, L., et al. (2009). Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. *Nat. Genet.* 41, 609–613. doi: 10.1038/ng.355

Morin, R. D., O'Connor, M. D., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A.-L., et al. (2008). Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res.* 18, 610–621. doi: 10.1101/gr.7179508

Neilsen, C. T., Goodall, G. J., and Bracken, C. P. (2012). IsomiRs–the overlooked repertoire in the dynamic microRNAome. *Trends Genet.* 28, 544–549. doi: 10.1016/j.tig.2012.07.005

Ohanian, M., Humphreys, D. T., Anderson, E., Preiss, T., and Fatkin, D. (2013). A heterozygous variant in the human cardiac miR-133 gene, MIR133A2, alters miRNA duplex processing and strand abundance. *BMC Genet.* 14:18. doi: 10.1186/1471-2156-14-18

Okada, C., Yamashita, E., Lee, S. J., Shibata, S., Katahira, J., Nakagawa, A., et al. (2009). A high-resolution structure of the pre-microRNA nuclear export machinery. *Science* 326, 1275–1279. doi: 10.1126/science.1178705

Park, J.-E., Heo, I, Tian, Y., Simanshu, D. K., Chang, H., Jee, D., et al. (2011). Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature* 475, 201–205. doi: 10.1038/nature10198

Pasquinelli, A. E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B., et al. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* 408, 86–89. doi: 10.1038/35040556

Peter, M. E. (2010). Targeting of mRNAs by multiple miRNAs: the next step. *Oncogene* 29, 2161–2164. doi: 10.1038/onc.2010.59

Pritchard, C. C., Cheng, H. H., and Tewari, M. (2012). MicroRNA profiling: approaches and considerations. *Nat. Rev. Genet.* 13, 358–369. doi: 10.1038/nrg3198

Ramalingam, P., Palanichamy, J. K., Singh, A., Das, P., Bhagat, M., Kassab, M. A., et al. (2014). Biogenesis of intronic miRNAs located in clusters by independent transcription and alternative splicing. *RNA* 20, 76–87. doi: 10.1261/rna.041814.113

Saini, H. K., Griffiths-Jones, S., and Enright, A. J. (2007). Genomic analysis of human microRNA transcripts. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17719–17724. doi: 10.1073/pnas.0703890104

Saunders, M. A., Liang, H., and Li, W.-H. (2007). Human polymorphism at microRNAs and microRNA target sites. *Proc. Natl. Acad. Sci. U.S.A.* 104, 3300–3305. doi: 10.1073/pnas.0611347104

Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. (2011). Genome-wide association study identifies five new schizophrenia loci. *Nat. Genet.* 43, 969–976. doi: 10.1038/ng.940

Schmeier, S., Schaefer, U., MacPherson, C. R., and Bajic, V. B. (2011). dPORE-miRNA: polymorphic regulation of microRNA genes. *PLoS ONE* 6:e16657. doi: 10.1371/journal.pone.0016657

Schwarz, D. S., Hutvágner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P. D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. *Cell* 115, 199–208. doi: 10.1016/S0092-8674(03)00759-1

Sethi, P., and Lukiw, W. J. (2009). Micro-RNA abundance and stability in human brain: specific alterations in Alzheimer's disease temporal lobe neocortex. *Neurosci. Lett.* 459, 100–104. doi: 10.1016/j.neulet.2009.04.052

Shen, J., Ambrosone, C. B., and Zhao, H. (2009). Novel genetic variants in microRNA genes and familial breast cancer. *Int. J. Cancer* 124, 1178–1182. doi: 10.1002/ijc.24008

Siomi, H., and Siomi, M. C. (2010). Posttranscriptional regulation of microRNA biogenesis in animals. *Mol. Cell* 38, 323–332. doi: 10.1016/j.molcel.2010.03.013

Soldà, G., Robusto, M., Primignani, P., Castorina, P., Benzoni, E., Cesarani, A., et al. (2012). A novel mutation within the MIR96 gene causes non-syndromic

inherited hearing loss in an Italian family by altering pre-miRNA processing. *Hum. Mol. Genet.* 21, 577–585. doi: 10.1093/hmg/ddr493

Sommer, S. S., and Rossi, J. J. (2012). Micro RNAs and their methods of use for the treatment and diagnosis of schizophrenia and schizophrenia spectrum disorders. *U.S. Patent No* 8252534(B2). Washington, DC: U.S. Patent and Trademark Office.

Strazisar, M., Cammaerts, S., van der Ven, K., Forero, D. A., Lenaerts, A.-S., Nordin, A., et al. (2015). MIR137 variants identified in psychiatric patients affect synaptogenesis and neuronal transmission gene sets. *Mol. Psychiatry.* 20, 472–481. doi: 10.1038/mp.2014.53

Sun, G., Yan, J., Noltner, K., Feng, J., Li, H., Sarkis, D. A., et al. (2009). SNPs in human miRNA genes affect biogenesis and function. *RNA* 15, 1640–1651. doi: 10.1261/rna.1560209

Tan, G. C., Chan, E., Molnar, A., Sarkar, R., Alexieva, D., Isa, I. M., et al. (2014). 5' isomiR variation is of functional and evolutionary importance. *Nucleic Acids Res.* 42, 9424–9435. doi: 10.1093/nar/gku656

Trabucchi, M., Briata, P., Garcia-Mayoral, M., Haase, A. D., Filipowicz, W., Ramos, A., et al. (2009). The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature* 459, 1010–1014. doi: 10.1038/nature08025

Ule, J., Jensen, K., Mele, A., and Darnell, R. B. (2005). CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods* 37, 376–386. doi: 10.1016/j.ymeth.2005.07.018

Vasudevan, S., Tong, Y., and Steitz, J. A. (2007). Switching from repression to activation: microRNAs can up-regulate translation. *Science* 318, 1931–1934. doi: 10.1126/science.1149460

Vermeulen, A., Behlen, L., Reynolds, A., Wolfson, A., Marshall, W. S., Karpilow, J., et al. (2005). The contributions of dsRNA structure to Dicer specificity and efficiency. *RNA* 11, 674–682. doi: 10.1261/rna.7272305

Vlachos, I. S., and Hatzigeorgiou, A. G. (2013). Online resources for miRNA analysis. *Clin. Biochem.* 46, 879–900. doi: 10.1016/j.clinbiochem.2013.03.006

Winter, J., and Diederichs, S. (2011). Argonaute proteins regulate microRNA stability: increased microRNA abundance by Argonaute proteins is due to microRNA stabilization. *RNA Biol* 8, 1149–1157. doi: 10.4161/rna.8.6.17665

Wu, S., Huang, S., Ding, J., Zhao, Y., Liang, L., Liu, T., et al. (2010). Multiple microRNAs modulate p21Cip1/Waf1 expression by directly targeting its 3' untranslated region. *Oncogene* 29, 2302–2308. doi: 10.1038/onc.2010.34

Yeom, K.-H., Lee, Y., Han, J., Suh, M. R., and Kim, V. N. (2006). Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing. *Nucleic Acids Res.* 34, 4622–4629. doi: 10.1093/nar/gkl458

Yi, R., Qin, Y., Macara, I. G., and Cullen, B. R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* 17, 3011–3016. doi: 10.1101/gad.1158803

Zeng, Y., and Cullen, B. R. (2004). Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. *Nucleic Acids Res.* 32, 4776–4785. doi: 10.1093/nar/gkh824

Zeng, Y., and Cullen, B. R. (2005). Efficient processing of primary microRNA hairpins by Drosha requires flanking nonstructured RNA sequences. *J. Biol. Chem.* 280, 27595–27603. doi: 10.1074/jbc.M504714200

Zeng, Y., Yi, R., and Cullen, B. R. (2005). Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *EMBO J.* 24, 138–148. doi: 10.1038/sj.emboj.7600491

Zhang, A.-Q., Gu, W., Zeng, L., Zhang, L.-Y., Du, D.-Y., Zhang, M., et al. (2015). Genetic variants of microRNA sequences and susceptibility to sepsis in patients with major blunt trauma. *Ann. Surg.* 261, 189–196. doi: 10.1097/SLA.0000000000000687

Zhang, H., Kolb, F. A., Brondani, V., Billy, E., and Filipowicz, W. (2002). Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J.* 21, 5875–5885. doi: 10.1093/emboj/cdf582

Zhang, H., Kolb, F. A., Jaskiewicz, L., Westhof, E., and Filipowicz, W. (2004). Single processing center models for human Dicer and bacterial RNase III. *Cell* 118, 57–68. doi: 10.1016/j.cell.2004.06.017

Zhang, X., and Zeng, Y. (2010). The terminal loop region controls microRNA processing by Drosha and Dicer. *Nucleic Acids Res.* 38, 7689–7697. doi: 10.1093/nar/gkq645

Zorc, M., Obsteter, J., Dovc, P., and Kunej, T. (2015). Genetic variability of microRNA genes in 15 animal species. *J. Genomics* 3, 51–56. doi: 10.7150/jgen.11246

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415. doi: 10.1093/nar/gkg595

Noncoding RNAs in human saliva as potential disease biomarkers

Xianzhi Lin [1†], Hsien-Chun Lo [1†], David T. W. Wong [2, 3, 4] and Xinshu Xiao [1, 2, 4*]

[1] Department of Integrative Biology and Physiology, University of California, Los Angeles, Los Angeles, CA, USA, [2] Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA, USA, [3] School of Dentistry, University of California, Los Angeles, Los Angeles, CA, USA, [4] Jonnson Comprehensive Cancer Center, University of California, Los Angeles, Los Angeles, CA, USA

Human saliva emerged as a research material in as early as the 17th century when investigators sought to understand the basis for salivary secretion (Garrett, 1975). Over the centuries, the focus of salivary research has evolved greatly and a wide range of topics has been examined (Garrett, 1975; Schipper et al., 2007). It is now known that the functions of saliva include at least lubrication, digestion of food, remineralization, prevention of demineralization, protection against microbial and viral infection, speech facilitation, and maintenance of oral and general health (Schipper et al., 2007; Malathi et al., 2014).

One active research area related to human saliva is the discovery of biomarker molecules for a variety of diseases. Compared to other body fluids, saliva is easily accessible in a non-invasive manner. However, it is also immediately exposed to the outside environment, thus may be confounded by a wide variety of environmental factors. Nevertheless, previous studies have established that at least some human molecules in saliva are highly stable and potential biomarkers have been examined for a number of oral and systemic diseases (Bonne and Wong, 2012; Schafer et al., 2014).

Key Components of Human Saliva

Saliva consists of both cellular and fluid contents. Epithelial cells, leukocytes, and erythrocytes are the three major human cell types, which co-exist with bacterial cells in human whole saliva (Aps et al., 2002). The fluid content of saliva is primarily generated by the salivary glands, but with additional contributions from blood, oral tissue, bacteria, viruses, and food remnants (Schipper et al., 2007). It mainly consists of water, macromolecules (such as glycoproteins, enzymes), small organic molecules, inorganic components (e.g., electrolytes), and metabolites from oral bacteria (Almstahl and Wikstrom, 2003; Aps and Martens, 2005; Schipper et al., 2007).

Many biomarker studies focused on profiling and quantification of proteins or RNA molecules in saliva. Since the diseases investigated for salivary biomarkers are often systemic, it is of great interest to identify circulating protein or RNA molecules that may have originated from disease-relevant cells (such as tumor cells). Such molecules reside outside of the cells in saliva and are often captured in cell-free saliva (CFS), the fraction of saliva with cellular contents removed (often by centrifugation). Most of the salivary RNAs appear to be highly degraded compared to full-length mRNAs in cellular compartments, possibly due to presence of RNA degradation enzymes in saliva and other body fluids (for circulating mRNAs) (Park et al., 2007). Notably, certain miRNA and mRNA molecules were shown to be highly stable, possible owing to protection by exosomes or protein complexes (Park et al., 2006, 2009; Palanisamy and Wong, 2010; Palanisamy et al., 2010).

Technologies for Salivary RNA Profiling

About a decade ago, microarrays were applied to characterize the global profile of mRNAs in saliva (Li et al., 2004; Park et al., 2007). These studies revealed that there were over one thousand distinct mRNA molecules in human CFS (Li et al., 2004). In addition to mRNAs derived from coding genes, many noncoding RNAs (ncRNAs) were also detected. Data from these studies demonstrate that there are hundreds of microRNAs (miRNAs) in human saliva, and most of them likely exist in exosomes (Michael et al., 2010; Gallo et al., 2012).

However, microarray techniques have inherent limitations, such as the dependence on gene annotation and cross-hybridization noise. In recent years, more powerful techniques based on next generation sequencing (NGS) revealed additional coding and ncRNA species in human saliva (Spielmann et al., 2012; Bahn et al., 2015). In contrast to the hybridization-based microarrays, RNA sequencing (RNA-Seq) offers single nucleotide information, high sensitivity and accuracy in transcript detection, and the capability to detect novel RNA species and transcript isoforms (Lee et al., 2011, 2013; Li et al., 2012). An increasing number of bioinformatic tools are emerging for analysis of RNA-Seq data, ranging from rapid short read aligners to detailed examination of RNA expression patterns (Oshlack et al., 2010). Owing to these improvements, the catalog of human genes, especially ncRNA genes, has been greatly expanded (refer to Sai Lakshmi and Agrawal, 2008; Kozomara and Griffiths-Jones, 2014; Xie et al., 2014 for ncRNA databases).

ncRNA Molecules in Saliva

In 2012, the Wong group reported the first global characterization of the human salivary transcriptome using high-throughput RNA-Seq (Spielmann et al., 2012). This study demonstrated that saliva harbors a wide variety of RNA species. More than 4000 distinct RNA molecules derived from coding or noncoding human RNAs were identified, including a small number of miRNAs. This study established that the RNA content in saliva is very diverse, which should be fully explored in future biomarker studies.

Recently, another in-depth analysis of human salivary extracellular ncRNA revealed novel insights regarding its RNA content and provided a comparative view of salivary ncRNAs relative to those of other body fluids (Bahn et al., 2015). Using human CFS, this study confirmed previous findings that miRNAs are stably and abundantly present in saliva (Patel et al., 2011), often harbored within exosomes (Gallo et al., 2012). miRNA expression profiles of healthy individuals were quantified and compared. Highly concordant miRNA expression was observed across individuals. Furthermore, considerable similarity was observed between miRNA expression levels of saliva and other body fluids (blood, cerebral spinal fluid (CSF)). Thus, these data suggest that salivary miRNAs could serve as candidate biomarkers, at least with equivalent promise as those derived from more invasive fluids.

A surprising observation from this study was the relative abundance of human piwi-interacting RNAs (piRNAs) in saliva.

piRNAs are small ncRNAs typically ~26–32 nt in length observed in germ cells of both vertebrates and invertebrates (Aravin et al., 2006; Girard et al., 2006; Grivna et al., 2006; Lau et al., 2006; Watanabe et al., 2006; Das et al., 2008). piRNAs are known to target transposons and repress their mobility (Das et al., 2008; Malone and Hannon, 2009). The number of abundant piRNAs is less than that of miRNAs in saliva, despite the large number of annotated piRNAs in various databases (Sai Lakshmi and Agrawal, 2008; Bahn et al., 2015). Nevertheless, piRNA expression levels were highly concordant between healthy individuals, similarly as miRNA levels. However, in contrast to the consistent expression profile of miRNAs across body fluids, piRNAs were highly exclusive to saliva with very low abundance in blood or CSF. These observations indicate that salivary piRNAs may have originated from cells in the oral mucosa or salivary glands, rather than circulating from systemic organs via blood. Nevertheless, salivary piRNAs may impose systemic functional impact, which needs to be further investigated.

Another novel finding in this study was the discovery of circular RNAs (circRNAs) in CFS, which is the first report of the presence of circRNAs in an extracellular fluid (Bahn et al., 2015). CircRNAs were originally identified in RNA viruses (Sanger et al., 1976; Kos et al., 1986). Later, intracellular circRNAs generated from specific exons of coding genes were reported (Nigro et al., 1991; Cocquerelle et al., 1992; Capel et al., 1993). Recent studies demonstrated that circRNAs exist in many different cell types and species (Salzman et al., 2012, 2013; Jeck et al., 2013; Memczak et al., 2013). Some circRNAs are likely noncoding (Capel et al., 1993; Memczak et al., 2013; Guo et al., 2014), but others may code for proteins (Wang and Wang, 2015). The function of most circRNAs remains unknown. Two circRNAs were shown to function as miRNA sponges (Hansen et al., 2013; Memczak et al., 2013). However, this function may not apply to the majority of other circRNAs as they lack bioinformatic evidence of significant miRNA complementarity (Guo et al., 2014). The discovery of circRNAs in CFS indicates that this type of molecule may have extracellular function and should be considered as a type of candidate biomarker (Li et al., 2015).

Although mRNAs are highly degraded in saliva and other body fluids, small ncRNAs are often stable with reproducible expression across individuals. Indeed, miRNAs have been extensively studied in blood and other body fluids as potential disease biomarkers (Chen et al., 2008; Gilad et al., 2008; Mitchell et al., 2008; Wang et al., 2009, 2010; Fichtlscherer et al., 2010; Li et al., 2010; Liu et al., 2011). The similarity between miRNA profiles of saliva and other body fluids (Weber et al., 2010; Bahn et al., 2015) strongly supports the potential of using miRNAs (and possibly other ncRNAs) from human CFS as biomarkers for various human diseases.

Salivary ncRNAs as Potential Biomarkers for Diseases

Although at an early stage, salivary ncRNA studies have revealed potential disease biomarkers. Thus, far, most studies focused on miRNA expression in saliva. **Table 1** summarizes a number of

TABLE 1 | Potential salivary miRNA biomarkers in cancer.

miRNA[a]	AUC[b]	p-value	Up/Down-regulated[c]	Statistical test
HEALTHY CONTROL VS. ORAL SQUAMOUS CELL CARCINOMA (Park et al., 2009)				
miR-200a	0.65	0.01	Down	Mann-Whitney U
miR-125a	0.62	0.03	Down	Mann-Whitney U
BENIGN VS. MALIGNANT PAROTID GLAND TUMORS[d] (Matse et al., 2013)				
miR-132	0.9	0.003	Up	Wilcoxon 2-sided
miR-15b	0.9	0.0019	Up	Wilcoxon 2-sided
miR-140-5p	0.9	0.0003	Up	Wilcoxon 2-sided
miR-223	0.9	0.0542	Up	Wilcoxon 2-sided
HEALTHY CONTROL VS. ESOPHAGEAL CANCER (Xie et al., 2013)				
miR-10b-3p	0.702	0.013	Up (4.5)	Mann-Whitney U or Kruskall-Wallis H
miR-144	0.671	0.036	Up (6.7)	Mann-Whitney U or Kruskall-Wallis H
miR-21	0.698	0.015	Up (9.7)	Mann-Whitney U or Kruskall-Wallis H
miR-451	0.725	0.006	Up (5.4)	Mann-Whitney U or Kruskall-Wallis H
HEALTHY CONTROL VS. RESECTABLE PANCREATIC CANCER (Xie et al., 2015)				
miR-3679-5p	0.673	0.008	Down (2.162)	Mann-Whitney U
miR-940	0.68	0.006	Up (2.586)	Mann-Whitney U
BENIGN PANCREATIC TUMORS VS. RESECTABLE PANCREATIC CANCER (Xie et al., 2015)				
miR-3679-5p	0.716	0.007	Down (4.708)	Mann-Whitney U
miR-940	0.729	0.004	Up (3.02)	Mann-Whitney U
HEALTHY CONTROL AND BENIGN PANCREATIC TUMORS VS. RESECTABLE PANCREATIC CANCER (Xie et al., 2015)				
miR-3679-5p	0.688	0.002	Down (3.012)	Mann-Whitney U
miR-940	0.696	0.0001	Up (2.716)	Mann-Whitney U

[a]Validated by qRT-PCR.
[b]Area under Receiver Operating Characteristic (ROC) curve.
[c]Change in disease relative to control. The number in the parentheses represents the fold change.
[d]A combination of four miRNA exhibits discriminating power.

studies where miRNAs were assessed as putative biomarkers for oral squamous cell carcinoma (OSCC) (Park et al., 2009), parotid gland tumors (Matse et al., 2013), and esophageal cancer (Xie et al., 2013). In addition to oral and esophageal diseases, salivary ncRNAs were also examined as potential biomarkers for systemic diseases. In a clinical study focusing on Sjögren's Syndrome, a chronic autoimmune disease, the authors observed different miRNA expression patterns in minor salivary glands of Sjögren's Syndrome patient compared to healthy individuals (Alevizos et al., 2011). The disease group can be clearly distinguished from the normal group using the miRNA expression profile by principal components and hierarchical clustering analyses. A very recent study focused on pancreatic cancer, using samples of patients with pancreatic cancer, benign pancreatic tumor or healthy controls (Xie et al., 2015). The authors observed significant down-regulation of miR-3679-5p and up-regulation of miR-940 in the cancer group compared to the other groups (**Table 1**), suggesting salivary miRNA may potentially be used for early detection of pancreatic cancer.

In addition to human ncRNAs, exogenous ncRNAs in saliva may also serve as potential disease biomarkers. The human-associated microbial communities have profound impact on the individual's physiological outcome (Human Microbiome Project Consortium, 2012). In human saliva, over 1500 bacteria have been identified and completely sequenced [Human Oral Microbiome Database; http://www.homd.org/]. Some studies have shown that saliva can be used to detect microbial infection (Schafer et al., 2014). In addition, both DNA and RNA viruses were detected in human saliva from viral infected hosts (Liou et al., 1992; Chen et al., 1997; Vieira et al., 1997; Shugars et al., 2001; Hermida et al., 2002; Mackiewicz et al., 2004; Goncalves et al., 2005; Balamane et al., 2010; Pride et al., 2012), thus could serve as biomarkers of viral infection. Thus, far, little is known regarding the landscape and function of exogenous ncRNAs in saliva.

Future Challenges and Perspectives

A comprehensive ncRNA expression profile is emerging for human saliva including the presence of miRNAs, piRNAs, and circular RNAs (Ogawa et al., 2013; Bahn et al., 2015). More RNA species may be discovered in the future given the rapid evolution of new technologies and powerful bioinformatic methods. The value of saliva as a body fluid for biomarker discovery is just becoming widely recognized. However, there are a number of challenges in this field, most of which are general to usage of any body fluid in biomarker discoveries. One challenge lies in the unbiased isolation of short and long RNA molecules from saliva samples. Although this topic is under intensive investigation, improved methods that can retain most RNA species unbiasedly in an operator-independent manner are highly desired.

Another challenge is accurate quantification of ncRNA abundance, which is key to biomarker assessment. RNA yield from different samples may vary greatly, which calls for effective experimental and bioinformatic methods for normalization of RNA expression. Most RNA-Seq studies discussed above calculated RNA expression levels by normalizing the number of reads of a particular RNA molecule against the total number of mapped reads (i.e., the RPKM measure Mortazavi et al., 2008). However, to estimate the absolute concentration of an RNA molecule in a sample, synthetic spike-in RNAs with known concentration should be added to the RNA sample before library generation. This approach necessitates accurate measurement of RNA concentration of the sample and synthesis of a large number of spike-in RNAs with varying sequence contents and concentrations (see Williams et al., 2013 for a demonstration of this approach). This challenging approach, though highly desirable and necessary for clinical usage of a biomarker, has not been widely adopted.

A third major challenge is a better understanding of the biogenesis pathways of human ncRNAs in saliva, which constitutes the basis to assess whether and to what extent ncRNA expression can reflect a person's health condition. Salivary RNAs could be derived from systemic organs or local cells of the oral cavity. Cellular origins of candidate biomarkers for various diseases should be further examined to substantiate our understanding of the validity of the biomarkers. Indeed, the presence, origin, and functional roles of disease biomarkers are all essential questions general to studies of different types of biomarkers and diseases. A valid disease biomarker should be directly involved in disease mechanisms or indirectly associated/correlated with key pathways driving the pathogenesis of disease. The ultimate question is how knowledge gained in biomarker studies could be utilized to develop effective strategies for disease prevention and treatment, which closely relies on a clear understanding of disease mechanisms.

Acknowledgments

References

Alevizos, I., Alexander, S., Turner, R. J., and Illei, G. G. (2011). MicroRNA expression profiles as biomarkers of minor salivary gland inflammation and dysfunction in Sjogren's syndrome. *Arthritis Rheum.* 63, 535–544. doi: 10.1002/art.30131

Almstahl, A., and Wikstrom, M. (2003). Electrolytes in stimulated whole saliva in individuals with hyposalivation of different origins. *Arch. Oral Biol.* 48, 337–344. doi: 10.1016/S0003-9969(02)00200-5

Aps, J. K., and Martens, L. C. (2005). Review: the physiology of saliva and transfer of drugs into saliva. *Forensic Sci. Int.* 150, 119–131. doi: 10.1016/j.forsciint.2004.10.026

Aps, J. K., Van Den Maagdenberg, K., Delanghe, J. R., and Martens, L. C. (2002). Flow cytometry as a new method to quantify the cellular content of human saliva and its relation to gingivitis. *Clin. Chim. Acta* 321, 35–41. doi: 10.1016/S0009-9981(02)00062-1

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., et al. (2006). A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* 442, 203–207. doi: 10.1038/nature04916

Bahn, J. H., Zhang, Q., Li, F., Chan, T. M., Lin, X., Kim, Y., et al. (2015). The landscape of microRNA, Piwi-interacting RNA, and circular RNA in human saliva. *Clin. Chem.* 61, 221–230. doi: 10.1373/clinchem.2014.230433

Balamane, M., Winters, M. A., Dalai, S. C., Freeman, A. H., Traves, M. W., Israelski, D. M., et al. (2010). Detection of HIV-1 in Saliva: implications for case-identification, clinical monitoring and surveillance for drug resistance. *Open Virol. J.* 4, 88–93. doi: 10.2174/1874357901004010088

Bonne, N. J., and Wong, D. T. (2012). Salivary biomarker development using genomic, proteomic and metabolomic approaches. *Genome Med.* 4, 82. doi: 10.1186/gm383

Capel, B., Swain, A., Nicolis, S., Hacker, A., Walter, M., Koopman, P., et al. (1993). Circular transcripts of the testis-determining gene Sry in adult mouse testis. *Cell* 73, 1019–1030. doi: 10.1016/0092-8674(93)90279-Y

Chen, M., Sonnerborg, A., Johansson, B., and Sallberg, M. (1997). Detection of hepatitis G virus (GB virus C) RNA in human saliva. *J. Clin. Microbiol.* 35, 973–975.

Chen, X., Ba, Y., Ma, L., Cai, X., Yin, Y., Wang, K., et al. (2008). Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res.* 18, 997–1006. doi: 10.1038/cr.2008.282

Cocquerelle, C., Daubersies, P., Majerus, M. A., Kerckaert, J. P., and Bailleul, B. (1992). Splicing with inverted order of exons occurs proximal to large introns. *EMBO J.* 11, 1095–1098.

Das, P. P., Bagijn, M. P., Goldstein, L. D., Woolford, J. R., Lehrbach, N. J., Sapetschnig, A., et al. (2008). Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the Caenorhabditis elegans germline. *Mol. Cell* 31, 79–90. doi: 10.1016/j.molcel.2008.06.003

Fichtlscherer, S., De Rosa, S., Fox, H., Schwietz, T., Fischer, A., Liebetrau, C., et al. (2010). Circulating microRNAs in patients with coronary artery disease. *Circ. Res.* 107, 677–684. doi: 10.1161/CIRCRESAHA.109.215566

Gallo, A., Tandon, M., Alevizos, I., and Illei, G. G. (2012). The majority of microRNAs detectable in serum and saliva is concentrated in exosomes. *PLoS ONE* 7:e30679. doi: 10.1371/journal.pone.0030679

Garrett, J. R. (1975). Changing attitudes on salivary secretion—a short history on spit. *Proc. R. Soc. Med.* 68, 553–560.

Gilad, S., Meiri, E., Yogev, Y., Benjamin, S., Lebanony, D., Yerushalmi, N., et al. (2008). Serum microRNAs are promising novel biomarkers. *PLoS ONE* 3:e3148. doi: 10.1371/journal.pone.0003148

Girard, A., Sachidanandam, R., Hannon, G. J., and Carmell, M. A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 442, 199–202. doi: 10.1038/nature04917

Goncalves, P. L., Cunha, C. B., Busek, S. C., Oliveira, G. C., Ribeiro-Rodrigues, R., and Pereira, F. E. (2005). Detection of hepatitis C virus RNA in saliva samples from patients with seric anti-HCV antibodies. *Braz. J. Infect. Dis.* 9, 28–34. doi: 10.1590/S1413-86702005000100006

Grivna, S. T., Beyret, E., Wang, Z., and Lin, H. (2006). A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev.* 20, 1709–1714. doi: 10.1101/gad.1434406

Guo, J. U., Agarwal, V., Guo, H., and Bartel, D. P. (2014). Expanded identification and characterization of mammalian circular RNAs. *Genome Biol.* 15:409. doi: 10.1186/s13059-014-0409-z

Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., et al. (2013). Natural RNA circles function as efficient microRNA sponges. *Nature* 495, 384–388. doi: 10.1038/nature11993

Hermida, M., Ferreiro, M. C., Barral, S., Laredo, R., Castro, A., and Diz Dios, P. (2002). Detection of HCV RNA in saliva of patients with hepatitis C virus infection by using a highly sensitive test. *J. Virol. Methods* 101, 29–35. doi: 10.1016/S0166-0934(01)00417-7

Human Microbiome Project Consortium. (2012). Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214. doi: 10.1038/nature11234

Jeck, W. R., Sorrentino, J. A., Wang, K., Slevin, M. K., Burd, C. E., Liu, J., et al. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* 19, 141–157. doi: 10.1261/rna.035667.112

Kos, A., Dijkema, R., Arnberg, A. C., Van der Meide, P. H., and Schellekens, H. (1986). The hepatitis delta (delta) virus possesses a circular RNA. *Nature* 323, 558–560. doi: 10.1038/323558a0

Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42, D68–D73. doi: 10.1093/nar/gkt1181

Lau, N. C., Seto, A. G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D. P., et al. (2006). Characterization of the piRNA complex from rat testes. *Science* 313, 363–367. doi: 10.1126/science.1130164

Lee, J. H., Ang, J. K., and Xiao, X. (2013). Analysis and design of RNA sequencing experiments for identifying RNA editing and other single-nucleotide variants. *RNA* 19, 725–732. doi: 10.1261/rna.037903.112

Lee, J. H., Gao, C., Peng, G., Greer, C., Ren, S., Wang, Y., et al. (2011). Analysis of transcriptome complexity through RNA sequencing in normal and failing murine hearts. *Circ. Res.* 109, 1332–1341. doi: 10.1161/CIRCRESAHA.111.249433

Li, G., Bahn, J. H., Lee, J. H., Peng, G., Chen, Z., Nelson, S. F., et al. (2012). Identification of allele-specific alternative mRNA processing via transcriptome sequencing. *Nucleic Acids Res.* 40, e104. doi: 10.1093/nar/gks280

Li, L. M., Hu, Z. B., Zhou, Z. X., Chen, X., Liu, F. Y., Zhang, J. F., et al. (2010). Serum microRNA profiles serve as novel biomarkers for HBV infection and diagnosis of HBV-positive hepatocarcinoma. *Cancer Res.* 70, 9798–9807. doi: 10.1158/0008-5472.CAN-10-1001

Li, P., Chen, S., Chen, H., Mo, X., Li, T., Shao, Y., et al. (2015). Using circular RNA as a novel type of biomarker in the screening of gastric cancer. *Clin. Chim. Acta* 444C, 132–136. doi: 10.1016/j.cca.2015.02.018

Li, Y., Zhou, X., St John, M. A., and Wong, D. T. (2004). RNA profiling of cell-free saliva using microarray technology. *J. Dent. Res.* 83, 199–203. doi: 10.1177/154405910408300303

Liou, T. C., Chang, T. T., Young, K. C., Lin, X. Z., Lin, C. Y., and Wu, H. L. (1992). Detection of HCV RNA in saliva, urine, seminal fluid, and ascites. *J. Med. Virol.* 37, 197–202. doi: 10.1002/jmv.1890370309

Liu, R., Zhang, C., Hu, Z., Li, G., Wang, C., Yang, C., et al. (2011). A five-microRNA signature identified from genome-wide serum microRNA expression profiling serves as a fingerprint for gastric cancer diagnosis. *Eur. J. Cancer* 47, 784–791. doi: 10.1016/j.ejca.2010.10.025

Mackiewicz, V., Dussaix, E., Le Petitcorps, M. F., and Roque-Afonso, A. M. (2004). Detection of hepatitis A virus RNA in saliva. *J. Clin. Microbiol.* 42, 4329–4331. doi: 10.1128/JCM.42.9.4329-4331.2004

Malathi, N., Mythili, S., and Vasanthi, H. R. (2014). Salivary diagnostics: a brief review. *ISRN Dent.* 2014:158786. doi: 10.1155/2014/158786

Malone, C. D., and Hannon, G. J. (2009). Molecular evolution of piRNA and transposon control pathways in Drosophila. *Cold Spring Harb. Symp. Quant. Biol.* 74, 225–234. doi: 10.1101/sqb.2009.74.052

Matse, J. H., Yoshizawa, J., Wang, X., Elashoff, D., Bolscher, J. G., Veerman, E. C., et al. (2013). Discovery and prevalidation of salivary extracellular microRNA biomarkers panel for the noninvasive detection of benign and malignant parotid gland tumors. *Clin. Cancer Res.* 19, 3032–3038. doi: 10.1158/1078-0432.CCR-12-3505

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495, 333–338. doi: 10.1038/nature11928

Michael, A., Bajracharya, S. D., Yuen, P. S., Zhou, H., Star, R. A., Illei, G. G., et al. (2010). Exosomes from human saliva as a source of microRNA biomarkers. *Oral Dis.* 16, 34–38. doi: 10.1111/j.1601-0825.2009.01604.x

Mitchell, P. S., Parkin, R. K., Kroh, E. M., Fritz, B. R., Wyman, S. K., Pogosova-Agadjanyan, E. L., et al. (2008). Circulating microRNAs as stable blood-based markers for cancer detection. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10513–10518. doi: 10.1073/pnas.0804549105

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226

Nigro, J. M., Cho, K. R., Fearon, E. R., Kern, S. E., Ruppert, J. M., Oliner, J. D., et al. (1991). Scrambled exons. *Cell* 64, 607–613. doi: 10.1016/0092-8674(91)90244-S

Ogawa, Y., Taketomi, Y., Murakami, M., Tsujimoto, M., and Yanoshita, R. (2013). Small RNA transcriptomes of two types of exosomes in human whole saliva determined by next generation sequencing. *Biol. Pharm. Bull.* 36, 66–75. doi: 10.1248/bpb.b12-00607

Oshlack, A., Robinson, M. D., and Young, M. D. (2010). From RNA-seq reads to differential expression results. *Genome Biol.* 11:220. doi: 10.1186/gb-2010-11-12-220

Palanisamy, V., Sharma, S., Deshpande, A., Zhou, H., Gimzewski, J., and Wong, D. T. (2010). Nanostructural and transcriptomic analyses of human saliva derived exosomes. *PLoS ONE* 5:e8577. doi: 10.1371/journal.pone.0008577

Palanisamy, V., and Wong, D. T. (2010). Transcriptomic analyses of saliva. *Methods Mol. Biol.* 666, 43–51. doi: 10.1007/978-1-60761-820-1_4

Park, N. J., Li, Y., Yu, T., Brinkman, B. M., and Wong, D. T. (2006). Characterization of RNA in saliva. *Clin. Chem.* 52, 988–994. doi: 10.1373/clinchem.2005.063206

Park, N. J., Zhou, H., Elashoff, D., Henson, B. S., Kastratovic, D. A., Abemayor, E., et al. (2009). Salivary microRNA: discovery, characterization, and clinical utility for oral cancer detection. *Clin. Cancer Res.* 15, 5473–5477. doi: 10.1158/1078-0432.CCR-09-0736

Park, N. J., Zhou, X., Yu, T., Brinkman, B. M., Zimmermann, B. G., Palanisamy, V., et al. (2007). Characterization of salivary RNA by cDNA library analysis. *Arch. Oral Biol.* 52, 30–35. doi: 10.1016/j.archoralbio.2006.08.014

Patel, R. S., Jakymiw, A., Yao, B., Pauley, B. A., Carcamo, W. C., Katz, J., et al. (2011). High resolution of microRNA signatures in human whole saliva. *Arch. Oral Biol.* 56, 1506–1513. doi: 10.1016/j.archoralbio.2011.05.015

Pride, D. T., Salzman, J., Haynes, M., Rohwer, F., Davis-Long, C., White, R. A. III, et al. (2012). Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome. *ISME J.* 6, 915–926. doi: 10.1038/ismej.2011.169

Sai Lakshmi, S., and Agrawal, S. (2008). piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. *Nucleic Acids Res.* 36, D173–D177. doi: 10.1093/nar/gkm696

Salzman, J., Chen, R. E., Olsen, M. N., Wang, P. L., and Brown, P. O. (2013). Cell-type specific features of circular RNA expression. *PLoS Genet.* 9:e1003777. doi: 10.1371/journal.pgen.1003777

Salzman, J., Gawad, C., Wang, P. L., Lacayo, N., and Brown, P. O. (2012). Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS ONE* 7:e30733. doi: 10.1371/journal.pone.0030733

Sanger, H. L., Klotz, G., Riesner, D., Gross, H. J., and Kleinschmidt, A. K. (1976). Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc. Natl. Acad. Sci. U.S.A.* 73, 3852–3856. doi: 10.1073/pnas.73.11.3852

Schafer, C. A., Schafer, J. J., Yakob, M., Lima, P., Camargo, P., and Wong, D. T. (2014). Saliva diagnostics: utilizing oral fluids to determine health status. *Monogr. Oral Sci.* 24, 88–98. doi: 10.1159/000358791

Schipper, R. G., Silletti, E., and Vingerhoeds, M. H. (2007). Saliva as research material: biochemical, physicochemical and practical aspects. *Arch. Oral Biol.* 52, 1114–1135. doi: 10.1016/j.archoralbio.2007.06.009

Shugars, D. C., Patton, L. L., Freel, S. A., Gray, L. R., Vollmer, R. T., Eron, J. J. Jr., et al. (2001). Hyper-excretion of human immunodeficiency virus type 1 RNA in saliva. *J. Dent. Res.* 80, 414–420. doi: 10.1177/00220345010800020301

Spielmann, N., Ilsley, D., Gu, J., Lea, K., Brockman, J., Heater, S., et al. (2012). The human salivary RNA transcriptome revealed by massively parallel sequencing. *Clin. Chem.* 58, 1314–1321. doi: 10.1373/clinchem.2011.176941

Vieira, J., Huang, M. L., Koelle, D. M., and Corey, L. (1997). Transmissible Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in saliva of men with a history of Kaposi's sarcoma. *J. Virol.* 71, 7083–7087.

Wang, G. K., Zhu, J. Q., Zhang, J. T., Li, Q., Li, Y., He, J., et al. (2010). Circulating microRNA: a novel potential biomarker for early diagnosis of acute myocardial infarction in humans. *Eur. Heart J.* 31, 659–666. doi: 10.1093/eurheartj/ehq013

Wang, K., Zhang, S., Marzolf, B., Troisch, P., Brightman, A., Hu, Z., et al. (2009). Circulating microRNAs, potential biomarkers for drug-induced liver injury. *Proc. Natl. Acad. Sci. U.S.A.* 106, 4402–4407. doi: 10.1073/pnas.0813371106

Wang, Y., and Wang, Z. (2015). Efficient backsplicing produces translatable circular mRNAs. *RNA* 21, 172–179. doi: 10.1261/rna.048272.114

Watanabe, T., Takeda, A., Tsukiyama, T., Mise, K., Okuno, T., Sasaki, H., et al. (2006). Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev.* 20, 1732–1743. doi: 10.1101/gad.1425706

Weber, J. A., Baxter, D. H., Zhang, S., Huang, D. Y., Huang, K. H., Lee, M. J., et al. (2010). The microRNA spectrum in 12 body fluids. *Clin. Chem.* 56, 1733–1741. doi: 10.1373/clinchem.2010.147405

Williams, Z., Ben-Dov, I. Z., Elias, R., Mihailovic, A., Brown, M., Rosenwaks, Z., et al. (2013). Comprehensive profiling of circulating microRNA via small RNA sequencing of cDNA libraries reveals biomarker potential and limitations. *Proc. Natl. Acad. Sci. U.S.A.* 110, 4255–4260. doi: 10.1073/pnas.1214046110

Xie, C., Yuan, J., Li, H., Li, M., Zhao, G., Bu, D., et al. (2014). NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic Acids Res.* 42, D98–D103. doi: 10.1093/nar/gkt1222

Xie, Z., Chen, G., Zhang, X., Li, D., Huang, J., Yang, C., et al. (2013). Salivary microRNAs as promising biomarkers for detection of esophageal cancer. *PLoS ONE* 8:e57502. doi: 10.1371/journal.pone.0057502

Xie, Z., Yin, X., Gong, B., Nie, W., Wu, B., Zhang, X., et al. (2015). Salivary microRNAs show potential as a noninvasive biomarker for detecting resectable pancreatic cancer. *Cancer Prev. Res. (Phila.)* 8, 165–173. doi: 10.1158/1940-6207.CAPR-14-0192

Conflict of Interest Statement: David T. W. Wong is co-founder of RNAmeTRIX Inc., a molecular diagnostic company. He holds equity in RNAmeTRIX, and serves as a company Director and Scientific Advisor. The University of California also holds equity in RNAmeTRIX. Intellectual property that David Wong invented and which was patented by the University of California has been licensed to RNAmeTRIX. Additionally, he is a consultant to PeriRx. The other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Phage display and targeting peptides: surface functionalization of nanocarriers for delivery of small non-coding RNAs

Babak Bakhshinejad *

Department of Genetics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran

Introduction

Small non-coding RNAs are known as a clinically relevant category of non-coding RNAs (ncRNAs) that have gained growing attention for their therapeutic values. miRNA and siRNA are of the most important and well-documented types of small ncRNAs. These regulatory RNA molecules are short–approximately 22 nucleotides long–with a similar mechanism of generation in which they are excised from longer double-stranded RNA precursor molecules. Both miRNA and siRNA target protein-coding genes via an antisense-based strategy and follow the same processing manner to reach gene-silencing effect (Chen et al., 2015). Functionally, small ncRNAs play a determining part in the modulation of cellular gene expression. The impressive role of these molecules as the controlling machinery of cells has proposed promising capacities for their potential translation into the clinic. In this context, tuning the regulatory components (such as ncRNAs) rather than the regulated components (protein-coding gens) has been suggested to be a more convenient and more effective approach to correct many cellular dysfunctions (Taft et al., 2010). This emerging concept symbolizes the dramatic potential of small ncRNAs, as a significant part of the regulatory apparatus of cells, to treat human diseases.

Small ncRNA Delivery: Active Targeting and the Dream of Magic Bullet

Within the recent years, nanotechnology-based approaches have emerged as potentially powerful and efficient systems for cellular delivery of ncRNAs. The exploitation of nanocarriers shows promise to bypass a variety of biological barriers for systematically administered RNA therapeutics (Miele et al., 2012; Segovia et al., 2015). Nanoparticles provide protection from serum nucleases for RNA-based therapeutic molecules, decrease their sequestration by phagocytes of the reticulo-endothelial system (RES) and are able to carry both hydrophilic and hydrophobic substances (Aagaard and Rossi, 2007; Peer et al., 2007). The excellent characteristics of nanocarrier platforms aid in enhancing overall bioavailability, extending *in vivo* stability, and improving the cellular delivery of RNA therapeutics such as small ncRNAs.

For the successful delivery of small ncRNA, one of the most important problems that needs to be solved is targeting of these drugs into the desired sites. The delivery of small ncRNAs to the diseased cells, tissues or organs not only raises their silencing potency, but also leads to a considerable reduction of side effects by avoiding normal cells (Daka and Peer, 2012). Active targeting plays a highly critical part in achieving efficient *in vivo* biodistribution and optimized systemic delivery

of ncRNA-based therapeutics. Active targeting becomes possible when the surface of nanocarriers that are embedded with small ncRNAs is modified through the covalent attachment of a cell/tissue-specific ligand. These surface-conjugated ligands will specifically interact with receptors that are expressed on the surface of target cells (Bertrand et al., 2014). In this manner, the delivery of therapeutic small ncRNA molecules will be poised to materialize the dream of constructing magic bullet; the concept that was theoretically fathered by Paul Ehrlich to describe targeted transportation of therapeutic agents to the desired sites of the body. Peptides are known as one of the most interesting categories of targeting ligands that have recently received huge attention. Due to some unique properties, peptides have become favorable targeting ligands for formulating powerful and sophisticated drug delivery platforms.

Phage Display Libraries: A Treasure of Cell-targeting Peptides

Phage display that is defined as the ability to genetically modify bacteriophages for introducing specific amino acids onto their surface is a worthwhile tool in biomedicine. The birth of phage display, as a kind of virus-based molecular chimera, goes back to three decades ago when Smith for the first time reported the expression of a foreign polypeptide on the surface of phage particles (Smith, 1985). The power of phage display for gaining momentum as a revolutionary strategy for gene/drug delivery studies comes from two distinctive features: (i) establishing a physical connection between the phenotype (the displayed ligand) and the genotype (the DNA sequence encoding the displayed ligand) within the same viral particle and (ii) producing very large libraries of ligands displayed on the surface of the phage particles. The most widely-used phage display libraries are random peptide libraries (RPL). In simple terms, a random peptide library is a collection of phages carrying on their surface extremely diverse types of peptides.

Due to the phenotype-genotype link, phage peptide libraries can be screened against any target of interest for selecting target-specific binders. This screening strategy displays similarity to the approach of natural selection but with distinctive difference of being performed in the test tube instead of nature (Bakhshinejad and Sadeghizadeh, 2014). Affinity selection is a hallmark of phage display screening through which binding peptides specific to any target including organic and inorganic structures can be obtained. The procedure of affinity-based library screening is called "panning." Whole cells in the culture are one of the most important targets used in panning studies. Cell panning involves incubating the phage library with target cells. Some phages interact with and bind to the target with higher affinity. The unbound or weakly-bound phages are then washed away and the bound phages are recovered. Subsequently, the recovered phages are amplified through infecting host bacterial cells. This process is repeated for several rounds to ultimately isolate the most strongly binding phages whose surface displayed peptides are identified by DNA sequencing (Bakhshinejad et al., 2014).

Phage Library Selected Peptides and Targeted Delivery of miRNA/siRNA-based Nanocarriers

Currently, RPLs are known to be one of the major sources of clinically relevant peptides. The marriage of combinatorial peptide chemistry and phage display has turned phage peptide libraries to a treasure for the discovery of useful targeting peptides. As we are beginning to more deeply understand the biological functions of ncRNAs in cellular processes and thereby their vital roles in the pathogenesis of numerous disorders, phage peptide libraries are attracting increasing attention for the delivery of small ncRNAs. Cell-specific peptides derived from phage libraries have been demonstrated to be capable of acting as important targeting components in nanocarriers and delivering both miRNA and siRNA into target cells. A variety of nanocarriers have been used for the delivery of small ncRNAs (Conde et al., 2015). Liposomes, polymers, micelles, protamine, and virus-like particles (VLPs) are delivery platforms that in different studies have represented potential to achieve delivery of miRNA and siRNA to target cells/tissues. **Table 1** indicates a list of studies in which peptides selected by phage display screening have been used for targeted delivery of miRNA/siRNA-containing nanocarriers to desired cells.

Cancer is a disease that has been the focus of much investigation for the delivery of small ncRNA-based therapeutics. Peptides obtained from screening of phage libraries against various tumor cells have served to target nanocarriers with encapsulated ncRNAs to the cells of interest. In line with this, several attempts have been made to identify phage-borne peptides specific to different tumors. A number of these peptides have indicated favorable outcomes for the delivery of peptide-targeted miRNA/siRNA-containing nanocarriers to malignant cells. DS 4-3, breast cancer-specific peptides, conjugated to branched polyethylenimine (bPEI) have been indicated to deliver the tumor suppressor miR-145 to metastatic breast cancer cells, thereby inhibiting tumor cell growth and suppressing cell invasion (Lee et al., 2013). In a novel strategy for intracellular delivery of siRNA into cancer cells, a research group led by Petrenko targeted liposomes by using preselected intact phage proteins (Bedi et al., 2011). In this work, not only cancer-avid peptide but the whole fusion phage coat protein in combination with its displayed peptide was used as a targeting ligand. This tumor-specific complex (phage fusion protein and peptide) was isolated from affinity screening of a phage library against MCF-7 breast cancer cells followed by being inserted into the lipid bilayer of liposome. The result was construction of a siRNA-encapsulated phage protein-targeted liposome that specifically down-regulated the expression of PRDM14 gene–a gene with important roles in the carcinogenesis of breast cancer–and inhibited the synthesis of its protein. This strategy remarkably overcomes the problem of purification of targeting peptides and their subsequent conjugation to nanocarriers.

Phage display selected peptides have also been exploited for targeted delivery of small ncRNAs involved in disorders

TABLE 1 | Studies in which peptides obtained through screening of phage display libraries have been used to functionalize the surface of nanocarriers for targeted delivery of miRNA and siRNA.

Non-coding RNA	Nanocarrier	Targeting peptide sequence	Target	Reference
miR-145	Branched polyethylenimine (bPEI) polymer	CRIMRILRILKLAR RIMRILRILKLARC	Breast cancer	Lee et al., 2013
PRDM14 siRNA	Liposome	DMPGTVLP (attached to M13 phage fusion protein pVIII)	Breast cancer	Bedi et al., 2011
siRNA cocktail against cyclin family members	VLPs of MS2 bacteriophage	SFSIIHTPILPL	Hepatocellular carcinoma	Ashley et al., 2011
GAPDH siRNA	Phage-like nanoparticles/Nanophage	VEEGGYIAA DWRGDSMDS	Breast cancer	Bedi et al., 2013
fluorescein-labeled siRNA	Protamine	KYLAYPDSVHIW	LNCaP – prostate cancer	Qin et al., 2011
Fas siRNA	Poly (cystamine bisacrylamide-diaminohexane, CBA-DAH) polymer	WLSEAGPVVTVRALRGTGSW	Primary cardiomyocyte	Nam et al., 2010
VEGF siRNA	Poly (cystamine bisacrylamide-diaminohexane, CBA-DAH) polymer	WLSEAGPVVTVRALRGTGSW	Cardiomyocyte	Nam et al., 2012
SHP1 and Fas siRNA	Poly (cystamine bisacrylamide-diaminohexane, CBA-DAH) polymer	WLSEAGPVVTVRALRGTGSW	Cardiomyocyte	Nam et al., 2011
miR-92a	Polyelectrolyte complex micelles	VHPKQHR	Endothelial cells	Kuo et al., 2014

other than cancer. Cardiovascular diseases, as one of the leading causes of death worldwide, have attracted attention to obtain benefits from achievements of phage display technology in the development of targeted therapy approaches. Since cardiomyocyte apoptosis is an important player in several heart diseases, the inhibition of this type of cell death has been viewed as a useful strategy for potential treatment of pathological conditions linked to cardiovascular system. This can be achieved through abolishing the activity of those factors whose function stimulates apoptosis. The conjugation of a primary cardiomyocyte (PCM) specific peptide selected by phage display to a polymer-based nanocarrier was used for the delivery of siRNA specific to Fas (an apoptosis inducer) to cardiomyocytes under hypoxic conditions (Nam et al., 2010). This PCM-targeted siRNA delivery system led to the prevention of apoptosis through down-regulating the Fas gene. Sometimes the inhibitors of ncRNAs, not ncRNAs themselves, have to be delivered to target cells. The exploitation of miRNA inhibitors is a prime example of miRNA-based scenarios that can be used for therapeutic purposes. As dysregulated miRNAs are associated with some disorders, the ability to inhibit the cellular activity of these miRNAs can provide new opportunities to treat human diseases. In this context, a micelle-based nanocarrier targeted with a phage

display selected peptide specific to vascular endothelial cells was shown to successfully deliver miR-92a inhibitor to endothelial cells; a strategy that is helpful for cell-specific delivery and targeted inhibition of miRNAs that contribute to the pathology of atherosclerosis (Kuo et al., 2014).

Conclusion

The capacity of small ncRNAs in specifically silencing the expression of disease-associated genes underlines their importance as novel forms of therapeutic intervention in numerous human disorders. Nanocarriers, due to some tunable features, represent huge potential to overcome delivery issues concerned with miRNA/siRNA-based pharmaceutical platforms. Recent findings highlight the significance of phage display methodology in providing a wide variety of peptides to be used for targeting of nanocarriers to desired cells. Although still in its infancy, phage peptide libraries are receiving growing attention for targeted cellular delivery of miRNA/siRNA therapeutic cargoes. Active targeting of nanocarriers through surface modification with phage selected peptides will provide new opportunities for taking advantage of the tremendous potential of small ncRNAs in

the clinical setting. The quickly advancing role played by phage peptide libraries in the territory of targeted delivery of RNA therapeutics has roots in the surprising diversity of these libraries and the convenience of their screening over a wide variety of cellular targets that cover numerous human diseases. Cell-targeting peptides derived from phage display libraries have raised hopes for researchers to take great strides toward the cherished goal of targeted delivery of miRNAs and siRNAs into the same cells that should take these therapeutic agents.

References

Aagaard, L., and Rossi, J. J. (2007). RNAi therapeutics: principles, prospects and challenges. *Adv. Drug Deliv. Rev.* 59, 75–86. doi: 10.1016/j.addr.2007.03.005

Ashley, C. E., Carnes, E. C., Phillips, G. K., Durfee, P. N., Buley, M. D., Lino, C. A., et al. (2011). Cell-specific delivery of diverse cargos by bacteriophage MS2 virus-like particles. *ACS Nano* 5, 5729–5745. doi: 10.1021/nn201397z

Bakhshinejad, B., Karimi, M., and Sadeghizadeh, M. (2014). Bacteriophages and medical oncology: targeted gene therapy of cancer. *Med. Oncol.* 31, 110. doi: 10.1007/s12032-014-0110-9

Bakhshinejad, B., and Sadeghizadeh, M. (2014). Bacteriophages and development of nanomaterials for neural regeneration. *Neural Regen. Res.* 9, 1955–1958. doi: 10.4103/1673-5374.145371

Bedi, D., Gillespie, J. W., Petrenko, V. A. Jr., Ebner, A., Leitner, M., Hinterdorfer, P., et al. (2013). Targeted delivery of siRNA into breast cancer cells via phage fusion proteins. *Mol. Pharm.* 10, 551–559. doi: 10.1021/mp3006006

Bedi, D., Musacchio, T., Fagbohun, O. A., Gillespie, J. W., Deinnocentes, P., Bird, R. C., et al. (2011). Delivery of siRNA into breast cancer cells via phage fusion protein-targeted liposomes. *Nanomedicine* 7, 315–323. doi: 10.1016/j.nano.2010.10.004

Bertrand, N., Wu, J., Xu, X., Kamaly, N., and Farokhzad, O. C. (2014). Cancer nanotechnology: the impact of passive and active targeting in the era of modern cancer biology. *Adv. Drug Deliv. Rev.* 66, 2–25. doi: 10.1016/j.addr.2013.11.009

Chen, Y., Gao, D. Y., and Huang, L. (2015). *In vivo* delivery of miRNAs for cancer therapy: challenges and strategies. *Adv. Drug Deliv. Rev.* 81, 128–141. doi: 10.1016/j.addr.2014.05.009

Conde, J., Edelman, E. R., and Artzi, N. (2015). Target-responsive DNA/RNA nanomaterials for microRNA sensing and inhibition: the jack-of-all-trades in cancer nanotheranostics? *Adv. Drug Deliv. Rev.* 81, 169–183. doi: 10.1016/j.addr.2014.09.003

Daka, A., and Peer, D. (2012). RNAi-based nanomedicines for targeted personalized therapy. *Adv. Drug Deliv. Rev.* 64, 1508–1521. doi: 10.1016/j.addr.2012.08.014

Kuo, C. H., Leon, L., Chung, E. J., Huang, R. T., Sontag, T. J., Reardon, C. A., et al. (2014). Inhibition of atherosclerosis-promoting microRNAs via targeted polyelectrolyte complex micelles. *J. Mater. Chem. B Mater. Biol. Med.* 2, 8142–8153. doi: 10.1039/C4TB00977K

Lee, H. J., Namgung, R., Kim, W. J., Kim, J. I., and Park, I.-K. (2013). Targeted delivery of MicroRNA-145 to Metastatic breast cancer by peptide conjugated branched PEI gene carrier. *Macromol. Res.* 21, 1201–1209. doi: 10.1007/s13233-013-1161-z

Miele, E., Spinelli, G. P., Di Fabrizio, E., Ferretti, E., Tomao, S., and Gulino, A. (2012). Nanoparticle-based delivery of small interfering RNA: challenges for cancer therapy. *Int. J. Nanomedicine* 7, 3637–3657. doi: 10.2147/IJN.S23696

Nam, H. Y., Kim, J., Kim, S. W., and Bull, D. A. (2012). Cell targeting peptide conjugation to siRNA polyplexes for effective gene silencing in cardiomyocytes. *Mol. Pharm.* 9, 1302–1309. doi: 10.1021/mp200589z

Nam, H. Y., Kim, J., Kim, S., Yockman, J. W., Kim, S. W., and Bull, D. A. (2011). Cell penetrating peptide conjugated bioreducible polymer for siRNA delivery. *Biomaterials* 32, 5213–5222. doi: 10.1016/j.biomaterials.2011.03.058

Nam, H. Y., Mcginn, A., Kim, P. H., Kim, S. W., and Bull, D. A. (2010). Primary cardiomyocyte-targeted bioreducible polymer for efficient gene delivery to the myocardium. *Biomaterials* 31, 8081–8087. doi: 10.1016/j.biomaterials.2010.07.025

Peer, D., Karp, J. M., Hong, S., Farokhzad, O. C., Margalit, R., and Langer, R. (2007). Nanocarriers as an emerging platform for cancer therapy. *Nat. Nanotechnol.* 2, 751–760. doi: 10.1038/nnano.2007.387

Qin, B., Tai, W., Shukla, R. S., and Cheng, K. (2011). Identification of a LNCaP-specific binding peptide using phage display. *Pharm. Res.* 28, 2422–2434. doi: 10.1007/s11095-011-0469-7

Segovia, N., Pont, M., Oliva, N., Ramos, V., Borros, S., and Artzi, N. (2015). Hydrogel doped with nanoparticles for local sustained release of siRNA in breast cancer. *Adv. Healthc. Mater.* 4, 271–280. doi: 10.1002/adhm.201400235

Smith, G. P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 228, 1315–1317. doi: 10.1126/science.4001944

Taft, R. J., Pang, K. C., Mercer, T. R., Dinger, M., and Mattick, J. S. (2010). Non-coding RNAs: regulators of disease. *J. Pathol.* 220, 126–139. doi: 10.1002/path.2638

The role of long non-coding RNAs in genome formatting and expression

Pierre-Olivier Angrand, Constance Vennin, Xuefen Le Bourhis and Eric Adriaenssens*

Cell Plasticity and Cancer – Inserm U908, University of Lille, Lille, France

Long non-coding RNAs (lncRNAs) are transcripts without protein-coding potential but having a pivotal role in numerous biological functions. Long non-coding RNAs act as regulators at different levels of gene expression including chromatin organization, transcriptional regulation, and post-transcriptional control. Misregulation of lncRNAs expression has been found to be associated to cancer and other human disorders. Here, we review the different types of lncRNAs, their mechanisms of action on genome formatting and expression and emphasized on the multifaceted action of the H19 lncRNA.

Keywords: lncRNAs, H19, chromatin organization, transcriptional regulation, post-transcriptional control

The advent of DNA tilling arrays and deep sequencing technologies has revealed that a much larger part of the genome is transcribed into RNAs than previously assumed. It is estimated that up to 70% of the genome is transcribed but only 2% of the human genome codes for proteins (Bertone et al., 2004; Birney et al., 2007; Kapranov et al., 2007; ENCODE Project Consortium, 2012) and RNAs without coding potential are collectively referred as non-coding RNAs (ncRNAs).

Non-coding RNAs include the well-known ribosomal (r) RNAs, ribozymes, transfer (t) RNAs, small nuclear (sn) RNAs, telomere-associated RNAs (TERRA, TERC), as well as a plethora of far less characterized RNAs. Based on their size, these ncRNAs are subdivided into two groups: small ncRNAs (<200 nt) and long ncRNAs [lncRNA (>200 nt)]. Small ncRNAs, such as microRNAs (miRs), small interfering RNAs (siRNAs), or PIWI-interacting RNAs (piRNAs) received much attention and were shown to mainly act as negative regulators of gene expression. In contrast, lncRNAs represent a more functionally diverse class of transcripts. LncRNAs are found in a large diversity of animals species (Guttman et al., 2009; Jia et al., 2010; Pauli et al., 2012), but also in plants (Swiezewski et al., 2009), yeast (Houseley et al., 2008), and even in prokaryotes (Bernstein et al., 1993) and viruses (Reeves et al., 2007). LncRNAs remains poorly conserved among species (Pang et al., 2006; Derrien et al., 2012). However, accumulating evidences indicate that this RNA class plays an important role in a variety of biological processes and may be involved in cancer and other human diseases (Wapinski and Chang, 2011; Tano and Akimitsu, 2012).

Majority of lncRNAs are 5′ capped, 3′ polyadenylated, multi-exonic and are subjected to transcriptional regulation as coding mRNAs (Carninci et al., 2005; Guttman et al., 2010; Cabili et al., 2011; Derrien et al., 2012). Some of the lncRNAs such as XIST, MALAT1, or NEAT1 are almost exclusively localized in the nucleus (Brown et al., 1992; Hutchinson et al., 2007), whereas others are mostly found in the cytoplasm (Coccia et al., 1992; Yoon et al., 2012). In term of genomic organization, lncRNAs can be classified according to their proximity to protein coding genes into five categories: sense, when overlapping one or more exons of another transcript; antisense, when overlapping one or more exons of another transcript on the opposite strand; bidirectional, when its expression and the expression of the neighboring coding transcript on the opposite strand are initiated in close proximity; intronic, when raising from an intron of another transcript; or intergenic, when produced from an independent transcription unit in the interval between two protein coding

genes. This crude classification illustrates that lncRNA expression may be controlled by different molecular mechanisms, but it does account neither for their modes of action nor for their cellular functions.

While only a limited number of lncRNAs has been studied, numerous evidences indicate that lncRNAs interact with a plethora of proteins. Furthermore, homologous Watson–Crick base pairing provides an efficient way by which lncRNAs may selectively interact with other nucleic acid species. It is believed that lncRNAs are involved in a diversity of cellular functions through gene expression regulation at different levels including chromatin organization, transcriptional regulation, and post-transcriptional mRNA processing (Mercer et al., 2009; Wilusz et al., 2009).

To complicate matters further, Anderson et al. (2015) recently described that a conserved micropeptide is encoded by a skeletal muscle-specific RNA previously annotated as a putative long non-coding RNA. This finding leads to the proposal that several lncRNAs could also have a biological function through the production of micropeptides.

LncRNAs in the Control of mRNA Processing

The ability of lncRNAs to recognize complementary sequences allows the regulation of mRNA processing at various steps, including degradation, splicing, translation, or transport (**Figure 1**).

FIGURE 1 | Functional mechanisms of long non-coding (lncRNA) action at the post-transcriptional levels. (A) mRNA stabilization. Base pairing between specific regions of a long non-coding antisense RNA and its sense transcript induces stabilization of the target mRNA and increases protein abundance. **(B)** mRNA degradation. Staufen double-stranded RNA-binding protein 1 (STAU1)-mediated mRNA decay is induced when base pairing is formed between the mRNA and a lncRNA. **(C)** Ribosome targeting. Through homologous base pairing with mRNAs and interactions with ribosomal proteins lncRNAs target transcripts to ribosomes or prevent translation. **(D)** Regulation of splicing. Base pairing between mRNAs and lncRNAs may prevent splicing by masking the splicing sites. In addition, lncRNAs are also implicated in the formation and maintenance of nuclear structures involved in alternative splicing of nascent transcripts. **(E)** miR sponge. By sequestering miRs through base pairing formations, lncRNAs affect the expression of the miR target genes. **(F)** Precursor of miRs. LncRNAs can serve as a source of miRs after processing. LncRNAs are shown in red, whereas mRNAs are in blue. See text for examples.

Base pairing between defined regions of the human β-site APP-cleaving enzyme 1 (BACE1) transcript and its antisense lncRNA BACE1-AS induces the mRNA stabilization and consequently the increase in BACE1 protein abundance (Faghihi et al., 2008). Similarly, the lncRNA TINCR (terminal differentiation-induced ncRNA) interacts with a range of differentiation mRNAs including FLG, LOR, ALOXE3, ALOX12B, ABCA12, CASP14, or ELOVL3, to increase their stability (Kretz et al., 2013). In contrast, the recognition of mRNAs by other lncRNAs, such as half-STAU1-binding site RNAs (1/2sbsR-NAs) decrease target mRNA stability by inducing STAU1 recruitment and the STAU1-mediated mRNA decay pathway (Gong and Maquat, 2011).

The translational process may also be modulated positively or negatively by lncRNA–mRNA pairing. For example, the antisense lncRNA ULCH-AS1 (ubiquitin carboxy-terminal hydrolase L1 antisense RNA 1) enhances ULCH mRNA translation (Carrieri et al., 2012), whereas lincRNA-p21 or pseudo-NOS suppress target mRNA translation (Korneev et al., 1999; Yoon et al., 2012).

The lncRNA MALAT1 (metastasis associated lung adenocarcinoma transcript 1) regulates pre-mRNA alternative splicing by modulating active serine/arginine splicing factors levels (Tripathi et al., 2010). In this case, the modulation of the mRNA processing is not achieved by a lncRNA–mRNA pairing mechanism but rather by the MALAT1-mediated modulation of the distribution of various splicing factors in nuclear speckle domains. However, antisense transcripts may also affect alternative splicing of their sense transcripts by virtue of masking splice sites by base complementarity (Krystal et al., 1990; Khochbin et al., 1992; Beltran et al., 2008). For example, a specific isoform of the lncRNA NPPA-AS is capable of down-regulating the intron-retained NPPA (atriuretic peptide precursor A) mRNA variant through RNA duplex formation between the sense and antisense transcripts (Annilo et al., 2009).

LncRNAs and the Connection with the MicroRNA World

Some lncRNAs act on post-transcriptional regulation through the modulation of the microRNA (miR) pathways. MiRs, a large class of small ncRNA, function by annealing to complementary sites in the coding sequences or 3′-untranslated regions (UTRs) of target mRNAs where they favor the recruitment of protein factors that impair translation and/or promote transcript degradation leading to a decrease in protein abundance (Baek et al., 2008; Bartel, 2009). Specifically, one mechanism by which the BACE1-AS lncRNA enhances BACE1 sense mRNA stability could be by masking the binding site for miR-485-5p (Faghihi et al., 2010). Rather than competing for miR-binding sites, a number of lncRNAs contain miR-binding sites in their sequence and therefore act as "sponges" to sequester miRs away from their mRNA targets. The pseudogene PTENP1 previously considered as biologically inactive was found to sequester miRs, consequently affecting their action on target gene regulation (Poliseno et al., 2010). In particular, the 3′-UTR of the PTENP1 lncRNA binds the same set of miRs targeting the tumor suppressor gene PTEN, then reducing the downregulation of this transcript and thus enhancing PTEN protein abundance. A number of other lncRNAs, including KRASP1, linc-MD1, HULC, or linc-ROR were shown to control mRNA activity through a miR sponge mechanism (Poliseno et al., 2010; Wang et al., 2010, 2013; Cesana et al., 2011). These examples illustrate that lncRNAs could counteract miR actions, but lncRNAs can themselves give rise to miRs and thus favor post-translational control by miR pathways as it is the case for the mouse Dlk1–Dio3 cluster or the BIC lncRNA (Eis et al., 2005; Hagan et al., 2009). Within the Dlk1–Dio3 cluster, *Meg3/Gtl2* contains in its last intron the evolutionarily conserved microRNA miR-770 whereas *Meg8* transcripts have the intron-encoded miR-341, miR-1188, and miR-370. Similarly, miR-155 is processed from sequences present in BIC lncRNA that accumulates in lymphoma cells.

LncRNAs in the Transcriptional Control

A number of evidences indicate that lncRNAs can act at the level of transcription either negatively or positively through a variety of molecular mechanisms (**Figure 2**). The dihydrofolate reductase (DHFR) gene contains a major and a minor promoter. The minor promoter gives rise to a lncRNA that forms a stable triplex lncRNA-DNA association at the major DHFR promoter and interacts with the general transcription factor II B (TFIIB) leading to the dissociation of the transcriptional preinitiation complex at this major promoter and then reducing DHFR expression (Martianov et al., 2007).

Other lncRNAs act as decoys to negatively control transcription by titrating transcription factors away from their cognate promoters. The lncRNA PANDAR (promoter of CDKN1A antisense DNA damage activated RNA) is induced in a TP53-dependent manner and inhibits apoptotic gene expression to favor cell-cycle arrest through direct interaction with, and sequestration of NFYA, a transcription factor controlling the apoptotic program upon DNA damage (Hung et al., 2011). Similarly, the lncRNA GAS5 (growth arrest-specific 5) contains an RNA motif derived from a stem-loop structure mimicking a DNA motif corresponding to the glucocorticoid response element. GAS5 binds to the DNA-binding domain of the glucocorticoid receptor, acts as a decoy glucocorticoid response element and is thus competing with DNA sites for binding to the glucocorticoid receptor (Kino et al., 2010).

Rather than acting as molecular decoys, lncRNA could modulate transcription by recruiting factors at target gene promoters or acting as transcription factor co-activators. For example, a lncRNA produced at the 5′ regulatory region of the cyclin D1 (CCND1) gene in response to genotoxic stress tethers and modulates the activity of the RNA-binding protein TLS (translocated in liposarcoma) which in turn inhibits the activity of the histone acetyltransferases CBP (CREB binding protein) and EP300, leading to CCND1 transcriptional repression (Wang et al., 2008). The lncRNA Evf-2 (DLX6-AS1) forms a stable complex with the homeodomain-containing protein DLX2 to induce expression of
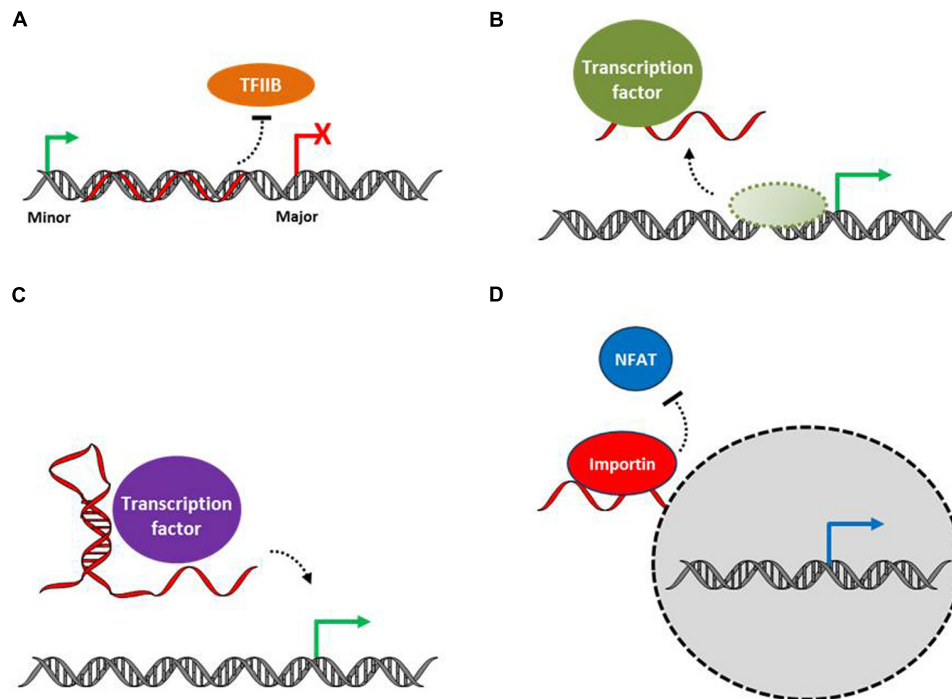
FIGURE 2 | Functional mechanism of action at the levels of transcriptional regulation. (A) LncRNA may regulate transcription by virtue of RNA–DNA triplex formation preventing the formation of the transcription initiation complex at promoters. **(B)** LncRNAs can act as decoys by titrating transcription factors away from their cognate promoters. **(C)** LncRNAs can regulate transcription through the targeting of transcription factors to promoters or acting as co-factors involved in transcription factor activity. **(D)** LncRNA can also control transcription factor trafficking. LncRNAs are shown in red.

the adjacent genes at the DLX5/6 locus (Feng et al., 2006). In this later case, the Evf-2 lncRNA functions as a co-factor regulating transcription factor activity.

Other lncRNAs regulate transcription by controlling transcription factor trafficking. As such, the lncRNA NRON (non-protein coding RNA, repressor of NFAT) interacts with importin-beta family members to inhibit nuclear translocation of the inactive dephosphorylated nuclear factor of activated T cells (NFAT) trans-activator (Willingham et al., 2005).

LncRNAs and Epigenetics

LncRNAs have been implicated in the control of gene expression through the recruitment of epigenetic modifiers at specific genomic loci. In eukaryotic chromatin, epigenetic regulation is conveyed by covalent modifications of DNA (methylation, hydroxymethylation), modifications of histone tails (acetylation, methylation, phosphorylation, ubiquitinylation), and the incorporation of various histone variants. These modifications locally change chromatin organization and regulate gene expression without changes in the DNA sequence. A number of evidences indicate that lncRNAs, acting as guides targeting enzymes involved in chromatin modifications, are part of this picture (**Figure 3**).

The lncRNA HOTAIR (HOX transcript antisense RNA) is transcribed from the HOXC locus and targets Polycomb

Repressive Complex 2 (PRC2) to silence distantly located genes, including genes at the HOXD locus and 100s of other genes on various chromosomes (Rinn et al., 2007; Zhang et al., 2015). Components of PRC2 trimethylate lysine 27 of histone H3 (H3K27me3) establishing the silent chromatin state (Völkel and Angrand, 2007; Völkel et al., 2015). Interestingly, HOTAIR also binds the LSD1–CoREST complex which possesses a lysine 4 of histone H3 demethylase activity, thus removing an active H3K4me2 chromatin mark (Tsai et al., 2010). Furthermore, deletion analysis of HOTAIR revealed that distinct parts of the lncRNA interact with PRC2 and LSD1 indicating that HOTAIR is able to bridge two independent chromatin modifying activities at a target locus. Indeed, the knockdown of HOTAIR is responsible for the concomitant loss of occupancy of PRC2 and LSD1, and concurrent loss of H3K27me3 and gain of H3K4me2 at target loci. Then, HOTAIR acts as an RNA scaffold targeting two different histone modification activities involved in heterochromatin formation.

The interplay between one lncRNA and different chromatin modifying complexes is also found at the INK4A tumor-suppressor locus. The antisense lncRNA ANRIL (antisense non-coding RNA in the INK4 locus, CDKN2B-AS) which is produced by the INK4B/ARF/INK4A locus binds specifically two Polycomb proteins, CBX7 (PRC1) and SUZ12 (PRC2). Disruption of interaction with both PRC1 and PRC2 proteins impacts the transcriptional repression at the INK4B locus in *cis* (Yap et al., 2010; Kotake et al., 2011). As another example, the lncRNA
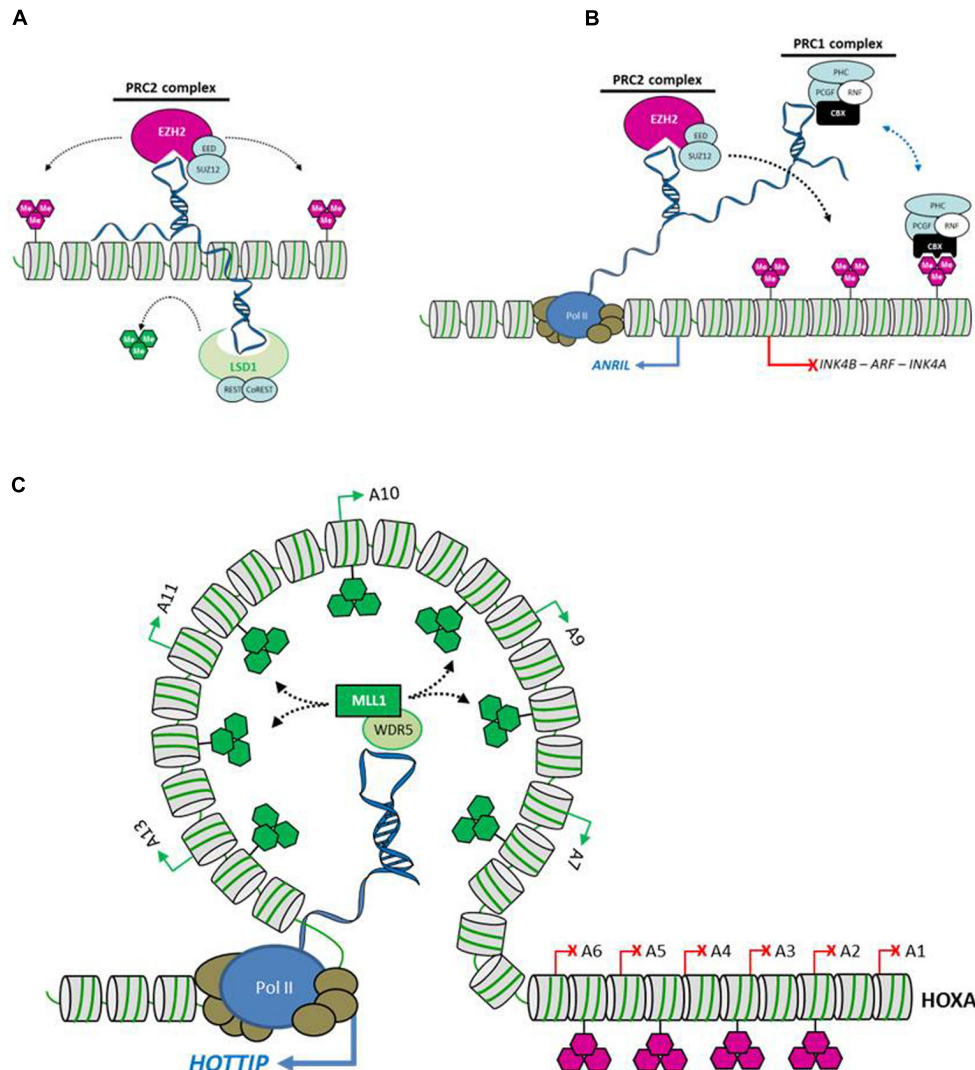
FIGURE 3 | Examples of lncRNAs controlling chromatin organization.
(A) HOTAIR (HOX transcript antisense RNA) represses transcription in
trans by recruiting two different chromatin modifying activities. The
Polycomb Repressive Complex 2 (PRC2) produces the repressive
H3K27me3 marks, whereas the LSD1-CoREST complex is responsible
for the removal of the active H3K4me2/3 marks. **(B)** The ANRIL lncRNA
represses transcription in *cis* at the INK4B/ARF/INK4A locus by
recruiting the Polycomb repressive complexes PRC1 and PRC2. **(C)** The
HOTTIP (HOXA transcript at the distal tip) lncRNA activates genes by
recruiting the histone modifier complex WDR5-MLL which is responsible
for H3K4me3 methylation, and by mediating long-range chromatin
looping at one extremity of the HOXA locus. Purple hexagons represent
H3K27me3 repressive marks, whereas green hexagons illustrate
H3K4me3 activating marks.

KCNQ1OT1 (KCNQ1 opposite strand/antisense transcript 1)
mediates bidirectional silencing by interacting with chromatin
and recruiting the PRC2 complex, as well as the histone methyl-
transferase G9a (EHMT2), resulting in an increase in the repres-
sive histone modifications H3K27me3 and H3K9me3 at the
KCNQ1 domain (Pandey et al., 2008). Thus, similar to HOTAIR
and ANRIL, KCNQ1OT1 represents a prototype of a scaffold
RNA recruiting multiple sets of chromatin modifying activities
involved in target gene silencing. Approximately 20% of lncR-
NAs, including HOTAIR, ANRIL, KCNQ1OT1, but also XIST,
RepA, HEIH, PCAT-1, H19, or linc-UBC1 (Zhao et al., 2008;
Maenner et al., 2010; Prensner et al., 2011; Yang et al., 2011; Luo

et al., 2013; He et al., 2013), are believed to guide PRC2 activity to
target genes, indicating that lncRNA-mediated targeting of PRC2
at chromatin is a widely used strategy to repress gene expression
through a chromatin reorganization mechanism (Khalil et al.,
2009).

In contrast, the lncRNA HOTTIP (HOXA transcript at the
distal tip) mediates transcriptional activation by controlling
chromatin modification and organization (Wang et al., 2011).
HOTTIP is produced from the 5′-end of the HOXA locus,
downstream of HOXA13. The knockdown of HOTTIP decreases
expression of HOXA genes in *cis*, with an efficacy that correlates
with the proximity of the HOXA genes relative to the HOTTIP

FIGURE 4 | Schematic representation of the transcriptional complexity at the H19/IGF2 locus. Non-coding transcripts at the H19/IGF2 are shown as blue squares when they are expressed. Coding genes are in green, when expressed. The differences in gene expression between the paternal and maternal alleles are shown. The DNA methylation status of the regulatory elements ICR (imprinting control region) and DMRs (differentially methylated regions) is indicated for the paternal and maternal alleles.

transcriptional unit. At the target genes, knockdown of HOTTIP results in the loss of activating H3K4me3 and H3K4me2 epigenetic marks, together with the decreases in occupancy of the MLL1 protein complex responsible for the establishment of these histone modifications. Furthermore, chromosome conformation capture carbon copy (C5) assays revealed abundant long-range looping interactions, bridging the transcribed target HOXA genes into proximity of the HOTTIP transcriptional unit. Thus, the mechanism by which the lncRNA HOTTIP controls HOXA expression relies on its potential to guide the histone methyltransferase MLL1 at target HOXA gene promoters, and on the formation of chromatin loops that connect distantly expressed HOXA genes to HOTTIP transcripts.
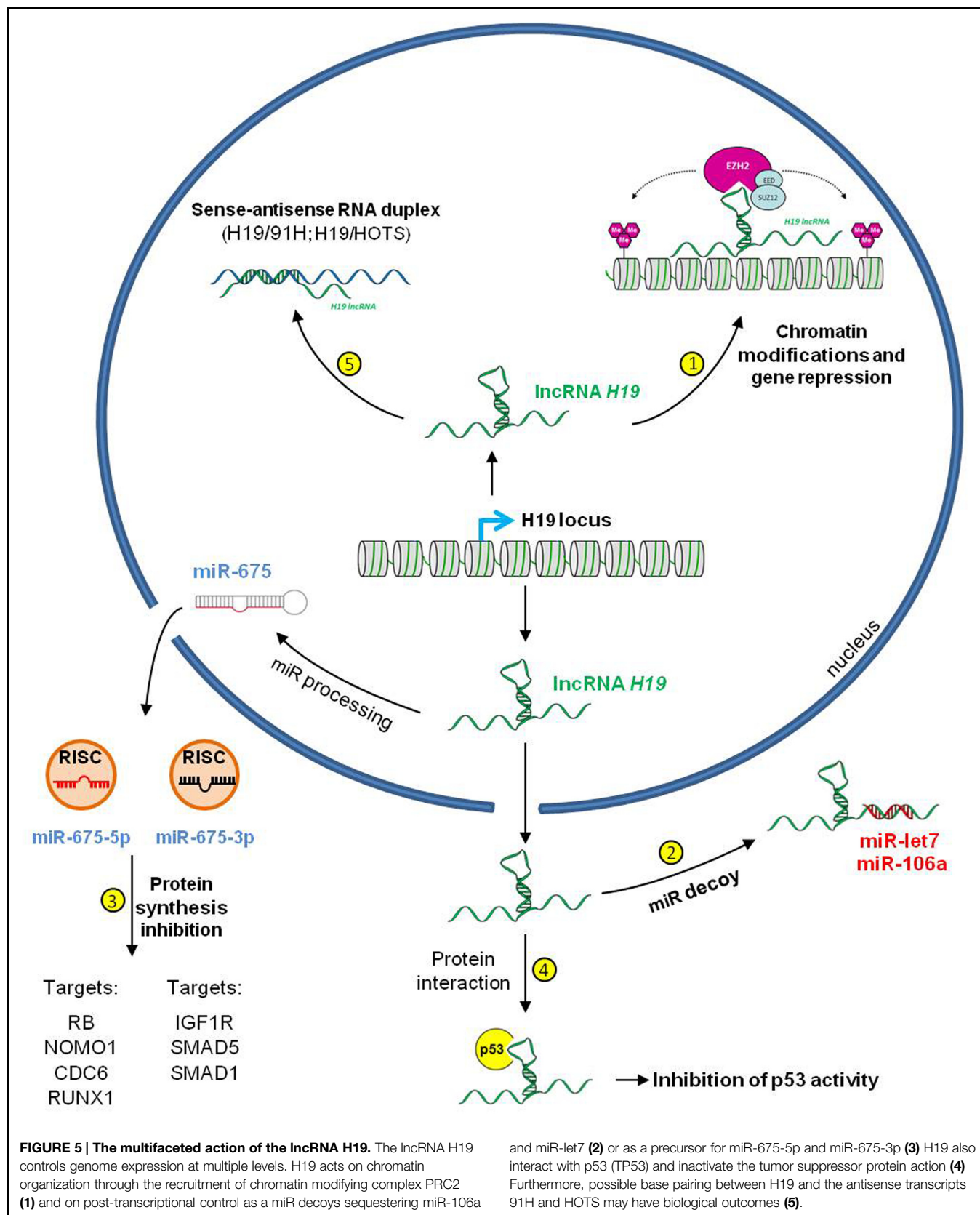
A role of lncRNAs in chromatin loop formation has also been described for the lncRNA CCAT1-L (Xiang et al., 2014). Indeed, CCAT1-L, is transcribed from a locus upstream of MYC and plays a role in MYC transcriptional regulation by promoting long-range chromatin looping.

Thus, lncRNAs, through the recruitment of chromatin modifiers and/or the induction of chromatin loops will modulate the chromatin conformation and will format the genome in a particular configuration. This lncRNA-mediated genome formatting emerges as a crucial and fundamental mechanism by which lncRNA may act on gene expression programs.

H19, a Prototype of a Multitask lncRNA

As discussed above, lncRNAs can regulate genome expression through different molecular mechanisms. However, several lncRNAs use multiple strategies that, in combination, may be required for their biological function. The action of the lncRNA H19 on gene expression illustrates the complexity of the combinatorial mechanisms of regulation achieved by a single lncRNA. H19 was the first lncRNA discovered (Brannan et al., 1990). Furthermore, H19 and its neighboring IGF2 gene located at position 11p15.5 are subjected to genomic imprinting and the study of the gene regulation at this locus serves as a model for understanding the molecular mechanisms involved in this genomic regulation. In addition, alterations of gene expression at the H19/IGF2 locus are associated to malignancies and developmental disorders. Loss of heterozygosity including loss of imprinting could be responsible for a loss of expression or a biallelic expression of these genes. Patients suffering from Beckwith–Wiedemann syndrome (BWS, OMIM 130650; Choufani et al., 2010) exhibit a loss of H19 expression and a biallelic expression of IGF2. BWS is associated with fetal and postnatal overgrowth and increased risk of embryonic or childhood cancers such as Wilm's tumors. Loss of IGF2 expression with a biallelic H19 expression is responsible for 20 to 60% of cases of Silver–Russel syndrome (SRS, OMIM 180860; Penaherrera et al., 2010). SRS is an intrauterine growth delay associated to an altered postnatal growth with facial dysmorphia and corporal asymmetry. Numerous studies including ours indicate that H19 may play a key role in tumorigenesis and could contribute to tumor progression and aggressiveness. H19 overexpression has also been reported in various cancer tissues including breast (Adriaenssens et al., 1998; Lottin et al., 2002), bladder (Cooper et al., 1996), lung (Kondo et al., 1995), and esophageal cancers (Hibi et al., 1996). Several lines of evidence indicate that H19 could play a role in tumor invasion and angiogenesis. In breast cancer, the oncogenic role of H19 has been well established (Berteaux et al., 2005), even if the precise molecular mechanisms involved in tumorigenesis are not yet fully understood.

FIGURE 5 | The multifaceted action of the lncRNA H19. The lncRNA H19 controls genome expression at multiple levels. H19 acts on chromatin organization through the recruitment of chromatin modifying complex PRC2 **(1)** and on post-transcriptional control as a miR decoys sequestering miR-106a and miR-let7 **(2)** or as a precursor for miR-675-5p and miR-675-3p **(3)** H19 also interact with p53 (TP53) and inactivate the tumor suppressor protein action **(4)** Furthermore, possible base pairing between H19 and the antisense transcripts 91H and HOTS may have biological outcomes **(5)**.

At the H19/IGF2 locus, both genes share a common set of enhancers located downstream of the H19 gene (**Figure 4**). The ICR (imprinting control region), located 2 kbp upstream of the H19 promoter, controls the monoallelic expression of H19 and IGF2 by insulating communication between the 3′ enhancers and the IGF2 promoter. The chromatin insulator property of the H19/IGF2 ICR is regulated by the insulator CTCF (CCTC-binding factor), which binds specifically to the unmethylated maternal allele. On the paternal allele, the ICR methylation does not allow CTCF binding and leads to IGF2 expression (reviewed in Lewis and Murrell, 2004). The H19/IGF2 locus contains other differentially methylated regions (DMRs), with DMR1 being a methylation-sensitive silencer and DMR2 being a methylation-sensitive activator (Constancia et al., 2000; Murrell et al., 2004). CTCF binding to the maternal ICR regulates its interaction with matrix attachment region 3 (MAR3) and DMR1 at IGF2, thus forming a tight loop around the maternal IGF2 locus which may contribute to its silencing. These interactions restrict the physical access of distal enhancers to the IGF2 promoter (Weber et al., 2003; Murrell et al., 2004; Kurukuti et al., 2006). Furthermore, several lncRNAs are produced at the H19/IGF2 locus adding further complexity to the locus regulation. The first antisense transcript at the H19/IGF2 locus is the lncRNA IGF2-AS (3–4 kb) discovered in 1991 in chicken (Rivkin et al., 1993; Moore et al., 1997). IGF2-AS and IGF2 are coregulated at the transcriptional levels but the function of this IGF2-AS lncRNA remains unclear. The lncRNA 91H (about 120 kb) is transcribed from the maternal allele (Berteaux et al., 2008). Recently, at the same position, a new protein coding gene HOTS (6 kbp) has been described (Onyango and Feinberg, 2011) but the relationship between the HOTS and 91H is still not clear. However, these two transcripts are transcribed in an antisense orientation compared to H19. An additional lncRNA produced by the H19/IGF2 locus has been identified (Court

et al., 2011). This PIHit (paternally expressed IGF2/H19 intergenic transcript) lncRNA is a 5–6 kb transcript expressed from the paternal allele after birth. Thus, the genomic organization of coding and non-coding transcripts illustrates the complexity of the interleaved networks of lncRNAs expressed from the H19/IGF2 locus.

To complicate matters further, H19 lncRNA mechanisms of action appear to be extremely diverse, acting at various levels (**Figure 5**). H19 has been shown to guide chromatin modifying enzymes to specific loci. In particular, Luo et al. (2013) have shown that H19 binds to and recruit the histone methyltransferase EZH2 at the E-cadherin promoter, leading to an increase in H3K27me3 repressive marks and to the silencing of the E-cadherin gene in bladder cancer. PRC2 protein members are not the only chromatin modifying factors interacting with H19 since it has been shown that this lncRNA physically binds to the methyl-CpG-binding domain protein 1 (MBD1). The H19-MBD1 complex is then recruited at several imprinted genes including IGF2, SLC38A4, and PEG1 (Monnier et al., 2013). This recruitment also induces methylation at lysine 9 of histone H3 (H3K9me3), probably via the additional interaction with an H3K9 histone methylransferase.

The multifaceted action of H19 is also illustrated by its dual interaction with miR pathways. On one hand, the lncRNA H19 acts as miR sponge to sequester miR-106a as well as the mir-let7 family members (Kallen et al., 2013; Imig et al., 2015). On the other end, H19 serves as a precursor of miR-675 that will in turn, post-translationally regulate a number of targets involved in cell tumorigenicity, including RB, IGFR1, SMAD1, SMAD5, CDC6, NOMO1, or RUNX1 (Cai and Cullen, 2007; Tsang et al., 2010; Gao et al., 2012; Keniry et al., 2012; Dey et al., 2014; Zhuang et al., 2014). The role of H19 in tumor progression could also be mediated through its interaction with the tumor-suppressor

TABLE 1 | Examples of long non-coding RNAs (lncRNAs) associated with human disorders.

lncRNA	Cancer/disease	Mechanisms of action	Reference
ANRIL	Neurofibromatosis type 1, prostate cancer, melanoma, acute lymphoblastic leukemia	Chromatin modification via the recruitment of the Polycomb Repressive Complex 2 (PRC2) at the INKB/ARF/INK4A tumor suppressor locus	Pasmant et al. (2007), Pasmant et al. (2011), Yap et al. (2010), Iacobucci et al. (2011)
HOTAIR	Hepatocellular carcinoma, colorectal cancer, breast cancer, glioblastomas	Chromatin modification via the recruitment of PRC2 and LSD1 in *trans*.	Gupta et al. (2010), Kogo et al. (2011), Yang et al. (2011), Zhang et al. (2015)
H19	Colorectal, gastric, breast, lung, esophageal, bladder, pancreas, ovary cancers	Chromatin modification via the recruitment of PRC2; Decoy for miR-Let-7; source of miR-675; TP53 inactivation	Kondo et al. (1995), Cooper et al. (1996), Hibi et al. (1996), Lottin et al. (2002), Berteaux et al. (2005), Tsang et al. (2010), Yang et al. (2012), Luo et al. (2013), Ma et al. (2014), Zhuang et al. (2014)
HEIH	Hepatocellular carcinoma	Chromatin modification via the recruitment of PRC2	Yang et al. (2011)
PCAT-1	Prostate cancer	Chromatin modification via the recruitment of PRC2	Prensner et al. (2011)
linc-UBC1	Bladder cancer	Chromatin modification via the recruitment of PRC2	He et al. (2013)
BACE1-AS	Alzheimer's disease	Increase in mRNA stability	Faghihi et al. (2008)
GAS5	Breast, bladder cancers	Decoy for the glucocorticoid receptor; regulation of CDK6 expression	Mourtada-Maarabouni et al. (2009), Kino et al. (2010), Liu et al. (2013)
PTENP1	Prostate cancer	miR decoy	Poliseno et al. (2010)
KCNA2-AS	Neuropathic pain	Decrease of KCNA2 expression	Zhao et al. (2013)
MIAT	Schizophrenia	Component of the nuclear matrix involved in mRNA splicing	Barry et al. (2014), Ishizuka et al. (2014)
MALAT1	Lung cancer	Alternative splicing regulation	Schmidt et al. (2011)

TP53 protein. This association results in partial TP53 inactivation (Yang et al., 2012).

Several evidences also indicate that the H19 lncRNA controls IGF2 expression at the translational and/or post-translational levels (Li et al., 1998), suggesting that other mechanisms by which H19 exerts its action remain to be deciphered. Similarly, the possible role of RNA duplex formation between H19 and the antisense transcripts 91H and HOTS requires investigations.

LncRNAs in Human Diseases

Given the wide range of molecular actions achieved by the lncRNAs and their roles in various physiological processes, it is not surprising that they have been shown to be involved in many human diseases. A number of data indicate that alterations of lncRNA expression lead to tumorigenesis through changes at the chromatin, transcriptional or post-transcriptional levels that impact target genes expression (**Table 1**). Since lncRNAs are regulating a different cellular pathways, growing evidences suggest that they could play a role in a large number of other human disorders including metabolic diseases, neurodegenerative and psychiatric disorders, cardiovascular and immune dysfunctions (Taft et al., 2010; Esteller, 2011; Harries, 2012; Shi et al., 2013; Clark and Blackshaw, 2014).

Perspectives and Concluding Remarks

LncRNAs represent a large part of the transcriptome and a very heterogeneous class of transcripts in terms of genomic organization and modes of action. Many of them are considered as key regulators of gene expression and thus, lncRNAs constitute an additional layer controlling the cellular programs. LncRNAs regulate diverse expression steps at the levels

of chromatin rearrangement, transcriptional control, and/or post-transcriptional processing. By these actions, lncRNAs are involved in numerous physiological functions and in many cases lncRNA alterations are associated with human disorders.

The fact that lncRNAs can be deregulated in tumors and other human pathologies, make them attractive candidates as biomarkers and as targets for therapy. LncRNAs may be down-regulated at the RNA levels by targeting their sequence. As so, short interfering RNAs (siRNAs) designed to perfectly match exact stretches of nucleotides, guarantee a high degree of specificity leading to lncRNA degradation. The power of the siRNA approach is illustrated by the success of a number of preclinical studies where siRNAs targeted mRNAs (Kaur et al., 2014). Similar approaches can thus be envisioned to target non-coding RNAs. Indeed, siRNAs have also been used to target miRs, leading to heart regeneration in an *in vivo* mouse model (Aguirre et al., 2014) and the use of siRNAs has been proposed in a therapeutic strategy targeting the lncRNA HOTAIR in endometrial carcinoma (Huang et al., 2014). Similarly, antisense oligonucleotides, single-strand DNA, or RNA molecules of 8 to 50 nucleotides can be used to target lncRNA. Specifically, *in vivo* and *in vitro* experiments revealed that antisense oligonucleotides directed against the lncRNA MALAT1 inhibit its expression and drastically reduce lung cancer metastasis (Gutschner et al., 2013; Tripathi et al., 2013).

In this context, further exploration in the complexity of the lncRNA world promises the emergence of novel therapeutic opportunities.

References

Adriaenssens, E., Dumont, L., Lottin, S., Bolle, D., Leprêtre, A., Delobelle, A., et al. (1998). H19 overexpression in breast adenocarcinoma stromal cells is associated with tumor values and steroid receptor status but independent of p53 and Ki-67 expression. *Am. J. Pathol.* 153, 1597–1607. doi: 10.1016/S0002-9440(10)65748-3

Aguirre, A., Montserrat, N., Zacchigna, S., Nivet, E., Hishida, T., Krause, M. N., et al. (2014). In vivo activation of a conserved microRNA program induces mammalian heart regeneration. *Cell Stem Cell* 15, 589–604. doi: 10.1016/j.stem.2014.10.003

Anderson, D. M., Anderson, K. M., Chang, C. L., Makarewich, C. A., Nelson, B. R., McAnally, J. R., et al. (2015). A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* 160, 595–606. doi: 10.1016/j.cell.2015.01.009

Annilo, T., Kepp, K., and Laan, M. (2009). Natural antisense transcript of natriuretic peptide precursor A (NPPA): structural organization and modulation of NPPA expression. *BMC Mol. Biol.* 10:81. doi: 10.1186/1471-2199-10-81

Baek, D., Villén, J., Shin, C., Camargo, F. D., Gygi, S. P., and Bartel, D. P. (2008). The impact of microRNAs on protein output. *Nature* 455, 64–71. doi: 10.1038/nature07242

Barry, G., Briggs, J. A., Vanichkina, D. P., Poth, E. M., Beveridge, N. J., Ratnu, V. S., et al. (2014). The long non-coding RNA Gomafu is acutely regulated in response to neuronal activation and involved in schizophrenia-associated alternative splicing. *Mol. Psychiatry* 19, 486–494. doi: 10.1038/mp.2013.45

Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233. doi: 10.1016/j.cell.2009.01.002

Beltran, M., Puig, I., Peña, C., García, J. M., Alvarez, A. B., Peña, R., et al. (2008). A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes Dev.* 22, 756–769. doi: 10.1101/gad.455708

Bernstein, H. D., Zopf, D., Freymann, D. M., and Walter, P. (1993). Functional substitution of the signal recognition particle 54-kDa subunit by its *Escherichia coli* homolog. *Proc. Natl. Acad. Sci. U.S.A.* 90, 5229–5233. doi: 10.1073/pnas.90.11.5229

Berteaux, N., Aptel, N., Cathala, G., Genton, C., Coll, J., Daccache, A., et al. (2008). A novel H19 antisense RNA overexpressed in breast cancer contributes to paternal IGF2 expression. *Mol. Cell. Biol.* 28, 6731–6745. doi: 10.1128/MCB.02103-07

Berteaux, N., Lottin, S., Monté, D., Pinte, S., Quatannens, B., Coll, J., et al. (2005). H19 mRNA-like noncoding RNA promotes breast cancer cell proliferation through positive control by E2F1. *J. Biol. Chem.* 280, 29625–22936. doi: 10.1074/jbc.M504033200

Bertone, P., Stolc, V., Royce, T. E., Rozowsky, J. S., Urban, A. E., Zhu, X., et al. (2004). Global identification of human transcribed sequences with genome tiling arrays. *Science* 306, 2242–2246. doi: 10.1126/science.1103388

Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigó, R., Gingeras, T. R., Margulies, E. H., et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799–816. doi: 10.1038/nature05874

Brannan, C. I., Dees, E. C., Ingram, R. S., and Tilghman, S. M. (1990). The product of the H19 gene may function as an RNA. *Mol. Cell. Biol.* 10, 28–36.

Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafrenière, R. G., Xing, Y., Lawrence, J., et al. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* 71, 527–542. doi: 10.1016/0092-8674(92)90520-M

Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., and Regev, A, et al. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25, 1915–1927. doi: 10.1101/gad.17446611

Cai, X., and Cullen, B. R. (2007). The imprinted H19 noncoding RNA is a primary microRNA precursor. *RNA* 13, 313–316. doi: 10.1261/rna.351707

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., et al. (2005). FANTOM consortium; RIKEN genome exploration research group and genome science group (genome network project core group). The transcriptional landscape of the mammalian genome. *Science* 309, 1559–1563.

Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele S., et al. (2012). Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* 491, 454–457. doi: 10.1038/nature11508

Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., et al. (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* 147, 358–369. doi: 10.1016/j.cell.2011.09.028

Choufani, S., Shuman, C., and Weksberg, R. (2010). Beckwith-Wiedemann syndrome. *Am. J. Med. Genet. C Semin. Med. Genet.* 154C, 343–354. doi: 10.1002/ajmg.c.30267

Clark, B. S., and Blackshaw, S. (2014). Long non-coding RNA-dependent transcriptional regulation in neuronal development and disease. *Front. Genet.* 5:164. doi: 10.3389/fgene.2014.00164

Coccia, E. M., Cicala, C., Charlesworth, A., Ciccarelli, C., Rossi, G. B., Philipson, L., et al. (1992). Regulation and expression of a growth arrest-specific gene (gas5) during growth, differentiation, and development. *Mol. Cell. Biol.* 12, 3514–3521.

Constancia, M., Dean, W., Lopes, S., Moore, T., Kelsey, G., and Reik, W. (2000). Deletion of a silencer element in Igf2 results in loss of imprinting independent of H19. *Nat. Genet.* 26, 203–206. doi: 10.1038/79930

Cooper, J., Fischer, M., Komitowski, D., Shevelev, A., Schulze, E., Ariel, I., et al. (1996). Developmentally imprinted genes as markers for bladder tumor progression. *J. Urol.* 155, 2120–2127. doi: 10.1016/S0022-5347(01)66120-2

Court, F., Baniol, M., Hagege, H., Petit, J. S., Lelay-Taha, M. N., Carbonell, F., et al. (2011). Long-range chromatin interactions at the mouse Igf2/H19 locus reveal a novel paternally expressed long non-coding RNA. *Nucleic Acids Res.* 39, 5893–5906. doi: 10.1093/nar/gkr209

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 22, 1775–1189. doi: 10.1101/gr.132159.111

Dey, B. K., Pfeifer, K., and Dutta, A. (2014). 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.* 28, 491–501. doi: 10.1101/gad.234419.113

Eis, P. S., Tam, W., Sun, L., Chadburn, A., Li, Z., Gomez, M. F., et al. (2005). Accumulation of miR-155 and BIC RNA in human B cell lymphomas. *Proc. Natl. Acad. Sci. U.S.A.* 102, 3627–3632. doi: 10.1073/pnas.0500613102

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. doi: 10.1038/nature11247

Esteller, M. (2011). Non-coding RNAs in human disease. *Nat. Rev. Genet.* 12, 861–874. doi: 10.1038/nrg3074

Faghihi, M. A., Modarresi, F., Khalil, A. M., Wood, D. E., Sahagan, B. G., Morgan, T. E., et al. (2008). Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feedforward regulation of β-secretase. *Nat. Med.* 14, 723–730. doi: 10.1038/nm1784

Faghihi, M. A., Zhang, M., Huang, J., Modarresi, F., Van der Brug, M. P., Nalls, M. A., et al. (2010). Evidence for natural antisense transcript-mediated inhibition of microRNA function. *Genome Biol.* 11, R56. doi: 10.1186/gb-2010-11-5-r56

Feng, J., Bi, C., Clark, B. S., Mady, R., Shah, P., and Kohtz, J. D. (2006). The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes Dev.* 20, 1470–1484. doi: 10.1101/gad.1416106

Gao, W. L., Liu, M., Yang, Y., Yang, H., Liao, Q., Bai, Y., et al. (2012). The imprinted H19 gene regulates human placental trophoblast cell proliferation via encoding miR-675 that targets Nodal Modulator 1 (NOMO1). *RNA Biol.* 9, 1002–1010. doi: 10.4161/rna.20807

Gong, C., and Maquat, L. E. (2011). lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* 470, 284–288. doi: 10.1038/nature09701

Gupta, R. A., Shah, N., Wang, K. C., Kim, J., Horlings, H. M., Wong, D. J., et al. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464, 1071–1076. doi: 10.1038/nature0897

Gutschner, T., Hämmerle, M., Eissmann, M., Hsu, J., Kim, Y., Hung, G., et al. (2013). The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res.* 73, 1180–1189. doi: 10.1158/0008-5472.CAN-12-2850

Guttman, M., Amit, I., Garber, M., French, C., Lin, M. F., Feldser, D., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458, 223–227. doi: 10.1038/nature07672

Guttman, M., Garber, M., Levin, J. Z., Donaghey, J., Robinson, J., Adiconis, X., et al. (2010). Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat. Biotechnol.* 28, 503–510. doi: 10.1038/nbt.1633

Hagan, J. P., O'Neill, B. L., Stewart, C. L., Kozlov, S. V., and Croce, C. M. (2009). At least ten genes define the imprinted Dlk1-Dio3 cluster on mouse chromosome 12qF1. *PLoS ONE* 4:e4352. doi: 10.1371/journal.pone.0004352

Harries, L. W. (2012). Long non-coding RNAs and human disease. *Biochem. Soc. Trans.* 40, 902–906. doi: 10.1042/BST20120020

He, W., Cai, Q., Sun, F., Zhong, G., Wang, P., Liu, H., et al. (2013). linc-UBC1 physically associates with polycomb repressive complex 2 (PRC2) and acts as a negative prognostic factor for lymph node metastasis and survival in bladder cancer. *Biochim. Biophys. Acta* 1832, 1528–1537. doi: 10.1016/j.bbadis.2013.05.010

Hibi, K., Nakamura, H., Hirai, A., Fujikake, Y., Kasai, Y., Akiyama, S., et al. (1996). Loss of H19 imprinting in esophageal cancer. *Cancer Res.* 56, 480–482.

Houseley, J., Rubbi, L., Grunstein, M., Tollervey, D., and Vogelauer, M. (2008). A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster. *Mol. Cell* 32, 685–695. doi: 10.1016/j.molcel.2008.09.027

Huang, J., Ke, P., Guo, L., Wang, W., Tan, H., Liang, Y., et al. (2014). Lentivirus-mediated RNA interference targeting the long noncoding RNA HOTAIR inhibits proliferation and invasion of endometrial carcinoma cells in vitro and in vivo. *Int. J. Gynecol. Cancer* 24, 635–642. doi: 10.1097/IGC.0000000000000121

Hung, T., Wang, Y., Lin, M. F., Koegel, A. K., Kotake, Y., Grant, G. D., et al. (2011). Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat. Genet.* 43, 621–629. doi: 10.1038/ng.848

Hutchinson, J. N., Ensminger, A. W., Clemson, C. M., Lynch, C. R., Lawrence, J. B., and Chess, A. (2007). A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* 8:39. doi: 10.1186/1471-2164-8-39

Iacobucci, I., Sazzini, M., Garagnani, P., Ferrari, A., Boattini, A., Lonetti, A., et al. (2011). A polymorphism in the chromosome 9p21 ANRIL locus is associated to Philadelphia positive acute lymphoblastic leukemia. *Leuk. Res.* 35, 1052–1059. doi: 10.1016/j.leukres.2011.02.020

Imig, J., Brunschweiger, A., Brümmer, A., Guennewig, B., Mittal, N., Kishore, S., et al. (2015). miR-CLIP capture of a miRNA targetome uncovers a lincRNA

H19-miR-106a interaction. *Nat. Chem. Biol.* 11, 107–114. doi: 10.1038/nchembio.1713

Ishizuka, A., Hasegawa, Y., Ishida, K., Yanaka, K., and Nakagawa, S. (2014). Formation of nuclear bodies by the lncRNA Gomafu-associating proteins Celf3 and SF1. *Genes Cells* 19, 704–721. doi: 10.1111/gtc.12169

Jia, H., Osak, M., Bogu, G. K., Stanton, L. W., Johnson, R., and Lipovich, L. (2010). Genome-wide computational identification and manual annotation of human long noncoding RNA genes. *RNA* 16, 1478–1487. doi: 10.1261/rna.1951310

Kallen, A. N., Zhou, X. B., Xu, J., Qiao, C., Ma, J., Yan, L., et al. (2013). The imprinted H19 lncRNA antagonizes let-7 microRNAs. *Mol. Cell* 52, 101–112. doi: 10.1016/j.molcel.2013.08.027

Kapranov, P., Willingham, A. T., and Gingeras, T. R. (2007). Genome-wide transcription and the implications for genomic organization. *Nat. Rev. Genet.* 8, 413–423. doi: 10.1038/nrg2083

Kaur, I. P., Chopra, K., Rishi, P., Puri, S., and Sharma, G. (2014). Small RNAs: the qualified candidates for gene manipulation in diverse clinical pathologies. *Crit. Rev. Ther. Drug. Carrier Syst.* 31, 305–329. doi: 10.1615/CritRevTherDrugCarrierSyst.2014007943

Keniry, A., Oxley, D., Monnier, P., Kyba, M., Dandolo, L., Smits, G., et al. (2012). The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. *Nat. Cell Biol.* 14, 659–665. doi: 10.1038/ncb2521

Khalil, A. M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11667–11672. doi: 10.1073/pnas.0904715106

Khochbin, S., Brocard, M. P., Grunwald, D., and Lawrence, J. J. (1992). Antisense RNA and p53 regulation in induced murine cell differentiation. *Ann. N. Y. Acad. Sci.* 660, 77–87. doi: 10.1111/j.1749-6632.1992.tb21060.x

Kino, T., Hurt, D. E., Ichijo, T., Nader, N., and Chrousos, G. P. (2010). Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci. Signal.* 3:ra8. doi: 10.1126/scisignal.2000568

Kogo, R., Shimamura, T., Mimori, K., Kawahara, K., Imoto, S., Sudo, T., et al. (2011). Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res.* 71, 6320–6326. doi: 10.1158/0008-5472.CAN-11-1021

Kondo, M., Suzuki, H., Ueda, R., Osada, H., Takagi, K., Takahashi, T., et al. (1995). Frequent loss of imprinting of the H19 gene is often associated with its overexpression in human lung cancers. *Oncogene* 10, 1193–1198. doi: 10.1378/chest.13-2736û7.13

Korneev, S. A., Park, J. H., and O'Shea, M. (1999). Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an NOS pseudogene. *J. Neurosci.* 19, 7711–7720.

Kotake, Y., Nakagawa, T., Kitagawa, K., Suzuki, S., Liu, N., Kitagawa, M., et al. (2011). Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene* 30, 1956–1962. doi: 10.1038/onc.2010.568

Kretz, M., Siprashvili, Z., Chu, C., Webster, D. E., Zehnder, A., Qu, K., et al. (2013). Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature* 493, 231–235. doi: 10.1038/nature11661

Krystal, G. W., Armstrong, B. C., and Battey, J. F. (1990). N-myc mRNA forms an RNA-RNA duplex with endogenous antisense transcripts. *Mol. Cell. Biol.* 10, 4180–4191.

Kurukuti, S., Tiwari, V. K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., et al. (2006). CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc. Natl. Acad. Sci. U.S.A.* 103, 10684–10649. doi: 10.1073/pnas.0600326103

Lewis, A., and Murrell, A. (2004). Genomic imprinting: CTCF protects the boundaries. *Curr. Biol.* 14, R284–R286. doi: 10.1016/j.cub.2004.03.026

Li, Y. M., Franklin, G., Cui, H. M., Svensson, K., He, X. B., Adam, G., et al. (1998). The H19 transcript is associated with polysomes and may regulate IGF2 expression in trans. *J. Biol. Chem.* 273, 28247–28252. doi: 10.1074/jbc.273.43.28247

Liu, Z., Wang, W., Jiang, J., Bao, E., Xu, D., Zeng, Y., et al. (2013). Downregulation of GAS5 promotes bladder cancer cell proliferation, partly by regulating CDK6. *PLoS ONE* 8:e73991. doi: 10.1371/journal.pone.0073991

Lottin, S., Adriaenssens, E., Dupressoir, T., Berteaux, N., Montpellier, C., Coll, J., et al. (2002). Overexpression of an ectopic H19 gene enhances the tumorigenic properties of breast cancer cells. *Carcinogenesis* 23, 1885–1895. doi: 10.1093/carcin/23.11.1885

Luo, M., Li, Z., Wang, W., Zeng, Y., Liu, Z., and Qiu, J. (2013). Long noncoding RNA H19 increases bladder cancer metastasis by associating with EZH2 and inhibiting E-cadherin expression. *Cancer Lett.* 333, 213–221. doi: 10.1016/j.canlet.2013.01.033

Ma, C., Nong, K., Zhu, H., Wang, W., Huang, X., Yuan, Z., et al. (2014). H19 promotes pancreatic cancer metastasis by derepressing let-7's suppression on its target HMGA2-mediated EMT. *Tumour Biol.* 35, 9163–9169. doi: 10.1007/s13277-014-2185-5

Maenner, S., Blaud, M., Fouillen, L., Savoye, A., Marchand, V., Dubois, A., et al. (2010). 2-D structure of the A region of Xist RNA and its implication for PRC2 association. *PLoS Biol.* 8:e1000276. doi: 10.1371/journal.pbio.1000276

Martianov, I., Ramadass, A., Serra Barros, A., Chow, N., and Akoulitchev, A. (2007). Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* 445, 666–670. doi: 10.1038/nature05519

Mercer, T. R., Dinger, M. E., and Mattick, J. S. (2009). Long noncoding RNAs: insights into functions. *Nat. Rev. Genet.* 10, 155–159. doi: 10.1038/nrg2521

Monnier, P., Martinet, C., Pontis, J., Stancheva, I., Ait-Si-Ali, S., and Dandolo, L. (2013). H19 lncRNA controls gene expression of the Imprinted Gene Network by recruiting MBD1. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20693–20698. doi: 10.1073/pnas.1310201110

Moore, T., Constancia, M., Zubair, M., Bailleul, B., Feil, R., Sasaki, H., et al. (1997). Multiple imprinted sense and antisense transcripts, differential methylation and tandem repeats in a putative imprinting control region upstream of mouse Igf2. *Proc. Natl. Acad. Sci. U.S.A.* 94, 12509–12514. doi: 10.1073/pnas.94.23.12509

Mourtada-Maarabouni, M., Pickard, M. R., Hedge, V. L., Farzaneh, F., and Williams, G. T. (2009). GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. *Oncogene* 28, 195–208. doi: 10.1038/onc.2008.373

Murrell, A., Heeson, S., and Reik, W. (2004). Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. *Nat. Genet.* 36, 889–893. doi: 10.1038/ng1402

Onyango, P., and Feinberg, A. P. (2011). A nucleolar protein, H19 opposite tumor suppressor (HOTS), is a tumor growth inhibitor encoded by a human imprinted H19 antisense transcript. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16759–16764. doi: 10.1073/pnas.1110904108

Pandey, R. R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., and Komorowski, J., et al. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell* 32, 232–246. doi: 10.1016/j.molcel.2008.08.022

Pang, K. C., Frith, M. C., and Mattick, J. S. (2006). Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet.* 22, 1–5. doi: 10.1016/j.tig.2005.10.003

Pasmant, E., Laurendeau, I., Héron, D., Vidaud, M., Vidaud, D., and Bièche, I. (2007). Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. *Cancer Res.* 67, 3963–3969. doi: 10.1158/0008-5472.CAN-06-2004

Pasmant, E., Sabbagh, A., Masliah-Planchon, J., Ortonne, N., Laurendeau, I., Melin, L., et al. (2011). Role of noncoding RNA ANRIL in genesis of plexiform neurofibromas in neurofibromatosis type 1. *J. Natl. Cancer Inst.* 103, 1713–1722. doi: 10.1093/jnci/djr416

Pauli, A., Valen, E., Lin, M. F., Garber, M., Vastenhouw, N. L., Levin, J. Z., et al. (2012). Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res.* 22, 577–591. doi: 10.1101/gr.133009.111

Penaherrera, M. S., Weindler, S., Van Allen, M. I., Yong, S. L., Metzger, D. L., McGillivray, B., et al. (2010). Methylation profiling in individuals with Russell-Silver syndrome. *Am. J. Med. Genet.* 152A, 347–355. doi: 10.1002/ajmg.a.33204

Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W. J., and Pandolfi, P. P. (2010). A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 465, 1033–1038. doi: 10.1038/nature09144

Prensner, J. R., Iyer, M. K., Balbin, O. A., Dhanasekaran, S. M., Cao, Q., Brenner, J. C., et al. (2011). Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat. Biotechnol.* 29, 742–749. doi: 10.1038/nbt.1914

Reeves, M. B., Davies, A. A., McSharry, B. P., Wilkinson, G. W., and Sinclair, J. H. (2007). Complex I binding by a virally encoded RNA regulates mitochondria-induced cell death. *Science* 316, 1345–1348. doi: 10.1126/science.1142984

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., et al. (2007). Functional demarcation of active and silent chroatin domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311–1323. doi: 10.1016/j.cell.2007.05.022

Rivkin, M., Rosen, K. M., and Villa-Komaroff, L. (1993). Identification of an anti-sense transcript from the IGF-II locus in mouse. *Mol. Reprod. Dev.* 35, 394–397. doi: 10.1002/mrd.1080350413

Schmidt, L. H., Spieker, T., Koschmieder, S., Schäffers, S., Humberg, J., Jungen, D., et al. (2011). The long noncoding MALAT-1 RNA indicates a poor prognosis in non-small cell lung cancer and induces migration and tumor growth. *J. Thorac. Oncol.* 6, 1984–1992. doi: 10.1097/JTO.0b013e3182307eac

Shi, X., Sun, M., Liu, H., Yao, Y., and Song, Y. (2013). Long non-coding RNAs: a new frontier in the study of human diseases. *Cancer Lett.* 339, 159–166. doi: 10.1016/j.canlet.2013.06.013

Swiezewski, S., Liu, F., Magusin, A., and Dean, C. (2009). Coldinduced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target. *Nature* 462, 799–802. doi: 10.1038/nature08618

Taft, R. J., Pang, K. C., Mercer, T. R., Dinger, M., and Mattick, J. S. (2010). Non-coding RNAs: regulators of disease. *J. Pathol.* 220, 126–139. doi: 10.1002/path.2638

Tano, K., and Akimitsu, N. (2012). Long non-coding RNAs in cancer progression. *Front. Genet.* 3:219. doi: 10.3389/fgene.2012.00219

Tripathi, V., Ellis, J. D., Shen, Z., Song, D. Y., Pan, Q., Watt, A. T., et al. (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* 39, 925–938. doi: 10.1016/j.molcel.2010.08.011

Tripathi, V., Shen, Z., Chakraborty, A., Giri, S., Freier, S. M., Wu, X., et al. (2013). Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. *PLoS Genet.* 9:e1003368. doi: 10.1371/journal.pgen.1003368

Tsai, M. C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lan, F., et al. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329, 689–693. doi: 10.1126/science.1192002

Tsang, W. P., Ng, E. K., Ng, S. S., Jin, H., Yu, J., Sung, J. J., et al. (2010). Oncofetal H19-derived miR-675 regulates tumor suppressor RB in human colorectal cancer. *Carcinogenesis* 31, 350–358. doi: 10.1093/carcin/bgp181

Völkel, P., and Angrand, P. O. (2007). The control of histone lysine methylation in epigenetic regulation. *Biochimie* 89, 1–20. doi: 10.1016/j.biochi.2006.07.009

Völkel, P., Dupret, B., Le Bourhis, X., and Angrand, P. O. (2015). Diverse involvement of EZH2 in cancer epigenetics. *Am. J. Transl. Res.* 7, 175–193.

Wang, J., Liu, X., Wu, H., Ni, P., Gu, Z., Qiao, Y., et al. (2010). CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. *Nucleic Acids Res.* 38, 5366–5383. doi: 10.1093/nar/gkq285

Wang, K. C., Yang, Y. W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 472, 120–124. doi: 10.1038/nature09819

Wang, X., Arai, S., Song, X., Reichart, D., Du, K., Pascual, G., et al. (2008). Induced ncRNAs allosterically modify RNA-binding proteins *in cis* to inhibit transcription. *Nature* 454, 126–130. doi: 10.1038/nature06992

Wang, Y., Xu, Z., Jiang, J., Xu, C., Kang, J., Xiao, L., et al. (2013). Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. *Dev. Cell.* 25, 69–80. doi: 10.1016/j.devcel.2013.03.002

Wapinski, O., and Chang, H. Y. (2011). Long noncoding RNAs and human disease. *Trends Cell. Biol.* 21, 354–361. doi: 10.1016/j.tcb.2011.04.001

Weber, M., Hagège, H., Murrell, A., Brunel, C., Reik, W., Cathala, G., et al. (2003). Genomic imprinting controls matrix attachment regions in the Igf2 gene. *Mol. Cell. Biol.* 23, 8953–8959. doi: 10.1128/MCB.23.24.8953-8959.2003

Willingham, A. T., Orth, A. P., Batalov, S., Peters, E. C., Wen, B. G., Aza-Blanc, P., et al. (2005). A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science* 309, 1570–1573. doi: 10.1126/science.1115901

Wilusz, J. E., Sunwoo, H., and Spector, D. L. (2009). Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev.* 23, 1494–1504. doi: 10.1101/gad.1800909

Xiang, J. F., Yin, Q. F., Chen, T., Zhang, Y., Zhang, X. O., Wu, Z., et al. (2014). Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res* 24, 513–531. doi: 10.1038/cr.2014.35

Yang, F., Bi, J., Xue, X., Zheng, L., Zhi, K., Hua, J., et al. (2012). Up-regulated long non-coding RNA H19 contributes to proliferation of gastric cancer cells. *FEBS J.* 279, 3159–3165. doi: 10.1111/j.1742-4658.2012.08694.x

Yang, F., Zhang, L., Huo, X. S., Yuan, J. H., Xu, D., Yuan, S. X., et al. (2011). Long noncoding RNA high expression in hepatocellular carcinoma facilitates tumor growth through enhancer of zeste homolog 2 in humans. *Hepatology* 54, 1679–1689. doi: 10.1002/hep.24563

Yap, K. L., Li, S., Muñoz-Cabello, A. M., Raguz, S., Zeng, L., Mujtaba, S., et al. (2010). Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol. Cell* 38, 662–674. doi: 10.1016/j.molcel.2010.03.021

Yoon, J. H., Abdelmohsen, K., Srikantan, S., Yang, X., Martindale, J. L., De, S., et al. (2012). LincRNA-p21 suppresses target mRNA translation. *Mol. Cell* 47, 648–655. doi: 10.1016/j.molcel.2012.06.027

Zhang, K., Sun, X., Zhou, X., Han, L., Chen, L., Shi, Z., et al. (2015). Long non-coding RNA HOTAIR promotes glioblastoma cell cycle progression in an EZH2 dependent manner. *Oncotarget* 6, 537–546.

Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J., and Lee, J. T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 322, 750–756. doi: 10.1126/science.1163045

Zhao, X., Tang, Z., Zhang, H., Atianjoh, F. E., Zhao, J. Y., Liang, L., et al. (2013). A long noncoding RNA contributes to neuropathic pain by silencing Kcna2 in primary afferent neurons. *Nat. Neurosci.* 16, 1024–1031. doi: 10.1038/nn.3438

Zhuang, M., Gao, W., Xu, J., Wang, P., and Shu, Y. (2014). The long non-coding RNA H19-derived miR-675 modulates human gastric cancer cell proliferation by targeting tumor suppressor RUNX1. *Biochem. Biophys. Res. Commun.* 448, 315–322. doi: 10.1016/j.bbrc.2013.12.126

Changes in expression of the long non-coding RNA *FMR4* associate with altered gene expression during differentiation of human neural precursor cells

Veronica J. Peschansky[1], Chiara Pastori[1], Zane Zeier[1], Dario Motti[2], Katya Wentzel[1], Dmitry Velmeshev[1], Marco Magistri[1], John L. Bixby[2,3,4,5], Vance P. Lemmon[2,3,4], José P. Silva[1] and Claes Wahlestedt[1]*

[1] Center for Therapeutic Innovation and Department of Psychiatry and Behavioral Sciences, Miller School of Medicine, University of Miami, Miami, FL, USA, [2] Miami Project to Cure Paralysis, University of Miami, Miami, FL, USA, [3] Department of Neurological Surgery, Miller School of Medicine, University of Miami, Miami, FL, USA, [4] Center for Computational Science, University of Miami, Coral Gables, FL, USA, [5] Department of Molecular and Cellular Pharmacology, Miller School of Medicine, University of Miami, Miami, FL, USA

CGG repeat expansions in the Fragile X mental retardation 1 (*FMR1*) gene are responsible for a family of associated disorders characterized by either intellectual disability and autism Fragile X Syndrome (FXS), or adult-onset neurodegeneration Fragile X-associated Tremor/Ataxia Syndrome. However, the *FMR1* locus is complex and encodes several long non-coding RNAs, whose expression is altered by repeat expansion mutations. The role of these lncRNAs is thus far unknown; therefore we investigated the functionality of *FMR4*, which we previously identified. "Full"-length expansions of the *FMR1* triplet repeat cause silencing of both *FMR1* and *FMR4*, thus we are interested in potential loss-of-function that may add to phenotypic manifestation of FXS. Since the two transcripts do not exhibit *cis*-regulation of one another, we examined the potential for *FMR4* to regulate target genes at distal genomic loci using gene expression microarrays. We identified *FMR4*-responsive genes, including the methyl-CpG-binding domain protein 4 (*MBD4*). Furthermore, we found that in differentiating human neural precursor cells, *FMR4* expression is developmentally regulated in opposition to expression of both *FMR1* (which is expected to share a bidirectional promoter with *FMR4*) and *MBD4*. We therefore propose that *FMR4*'s function is as a gene-regulatory lncRNA and that this transcript may function in normal development. Closer examination of *FMR4* increases our understanding of the role of regulatory lncRNA and the consequences of *FMR1* repeat expansions.

Keywords: lncRNA, intellectual disability, epigenetics, differentiation, chromatin remodeling, *MBD4*, *FMR4*, Fragile X

Introduction

Fragile X Syndrome, FXTAS, and FXPOI are X-linked disorders that arise from expansions in a CGG-repeat region in the 5′-UTR of the *FMR1* gene. Normal *FMR1* alleles contain 6–54 repeats, expansions from 55 to 200 repeats are considered "premutations" and all larger repeat sizes are known as the "full mutations." Individuals with a premutation may develop the adult-onset neurodegenerative disorder known as FXTAS, while women carrying the premutation are at risk for FXPOI. Only the full mutation leads to FXS, which is a common cause of inherited intellectual disability and autism (Oostra and Willemsen, 2003). *FMR1* premutations result in overproduction of toxic, expanded mRNAs that contribute to the development of FXPOI and FXTAS pathology (Tassone et al., 2000; Kenneson et al., 2001). Full mutations lead to DNA and repressive histone methylation of the *FMR1* locus (Sutcliffe et al., 1992; Hornstra et al., 1993; Coffee et al., 1999, 2002; Kumari and Usdin, 2010). Thus FXS derives from the loss of *FMR1* mRNA and protein FMRP. We and others have identified four non-coding transcripts with abnormal expression in response to Fragile X repeat expansions at the *FMR1* locus (Ladd et al., 2007; Khalil et al., 2008; Pastori et al., 2014), but their role in FXS/FXTAS/FXPOI phenotypes remains to be determined.

The vast majority of the human transcriptome is comprised of either long [>200 nucleotides (nt)] or short ncRNAs (Cheng et al., 2005; Banfai et al., 2012; Djebali et al., 2012). While short ncRNAs typically regulate gene expression through posttranscriptional mechanisms or by interfering with translation (Rother and Meister, 2011; Fabian and Sonenberg, 2012), lncRNAs (which can be many kilobases long) often act in *cis* or *trans* to regulate gene expression at their locus of origin or elsewhere in the genome, respectively. Evidence suggests that lncRNAs perform scaffolding functions by recruiting epigenetic complexes or ribonucleoproteins that cause chromatin remodeling (Wang and Chang, 2011). Other lncRNAs act post-transcriptionally by targeting mRNAs or translational machinery. Regardless of the mechanism, a growing body of evidence implicates lncRNAs in a myriad of normal cellular functions such as DNA damage response and mitosis (Tsai et al., 2010; Yap et al., 2010; Hung et al., 2011; Kotake et al., 2011; Wang and Chang, 2011) and in diseases, such as cancer (Hajjari et al., 2014).

Recent attention has focused more specifically on the role of lncRNAs in neurodevelopmental programs and diseases of the nervous system. For example, lncRNAs are involved in the differentiation of neural cell types, and synaptic signaling and maturation (Mercer et al., 2010; Qureshi et al., 2010). In addition, both short and long ncRNAs are known to be involved in Prader–Willi syndrome, which is a developmental disorder caused by paternal deletion of a maternally imprinted region and can present with metabolic dysregulation including circadian rhythm defects (Sahoo et al., 2008; De Smith et al., 2009; Powell et al., 2013). Both syndromic and non-syndromic ASD susceptibility loci also contain aberrantly expressed lncRNAs that may contribute to disease (Velmeshev et al., 2013; Ziats

and Rennert, 2013). Dysfunction of lncRNAs has also been linked to pathogenesis of neurodegenerative disorders including Alzheimer's disease (Faghihi et al., 2008) and spinocerebellar ataxia type 7, another repeat expansion disorder (Sopher et al., 2011). In sum, there is a growing body of evidence supporting the involvement lncRNAs in both the normal and diseased nervous system, spurring further mechanistic inquiries.

FMR4, an untranslated, antisense lncRNA originating at the *FMR1* gene locus was shown to have anti-apoptotic functions in HEK293T and HeLa cells but to have no effect on expression of *FMR1* (Khalil et al., 2008). Here, we describe *FMR4*'s function as a regulator of gene expression in *trans* by identifying mRNA expression changes induced by *FMR4*. In particular, these effects in HEK293T cells are mirrored by discordant developmental regulation between *FMR4* and one of its targets, *MBD4*, in hNPCs.

Materials and Methods

HEK293T Cell Culture, Transfection, and RNA Extraction

HEK293T cells were cultured in DMEM with 10% FBS. In overexpression experiments, cells were transfected with pcDNA3.1-FMR4 or the empty pcDNA3.1 control vector using Lipfectamine 2000CD. For knockdown experiments, the siRNA FMR4(C) (Khalil et al., 2008), versus Silencer Negative Control siRNA #1 (Ambion) were used with the Lipofectamine RNAiMAX transfection reagent, according to manufacturer's instructions (Invitrogen). For microarray experiments, 1×10^6 cells were plated and transfected with 0.5 μg plasmid or 40 nM siRNA on the same day, and incubated for 6 or 24 h after transfection. For validation, *FMR4* was knocked down using two sequential siRNA transfections over 72 h. RNA was extracted using Trizol (Invitrogen) and the RNeasy Mini Kit, and treated with DNAse on-column using the RNAse-free DNAse Set (Qiagen) according to manufacturer's instructions.

Microarray Hybridization and Analysis

At 6 or 24 h post transfection, RNA was extracted and samples were submitted to the Hussman Institute for Human Genomics Center for Genome Technology for microarray analysis using Affymetrix Human Gene 2.0 ST Arrays. Total RNA samples were first prepared using the Ambion WT Expression Kit (cat# 4411974). Briefly, the kit generates sense-strand cDNA from total RNA using a reverse transcription priming method that specifically primes non-ribosomal RNA, including both poly(A) and non-poly(A) mRNA. Next, samples are fragmented and labeled using the Affymetrix GeneChip WT Terminal Labeling Kit (cat# 902280). Final yield was hybridized onto the array, washed and stained using the Affymetrix GeneChip Hybridization, Wash, and Stain Kit (cat# 900720). Arrays were scanned using GeneChip Scanner 3000 7G system. Background subtraction, GC-RMA normalization and quality control were performed using the Affymetrix GeneChip Command Console Software and the bioconductor package from R. Data have been archived in

the Gene Expression Omnibus at the National Center for Biotechnology Information, and assigned the accession number GSE70817.

cDNA Synthesis and Quantitative Real-Time PCR (qPCR)

cDNA was synthesized using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) with 500 ng of total RNA per reaction. Gene-specific *FMR4* cDNA was primed separately ("FMR4 RT": ATTGCTGGCAGTCGTTTCTT), in order to specifically detect the antisense transcript and prevent capture of overlapping sense transcripts arising from that genomic region. Random hexamer-primed cDNA libraries were used for detection of all other genes. *FMR4* RNA expression was quantified using SYBR Green quantitative real-time PCR (qPCR) with the following primers, validated by melting curve: "FMR4 FW" – ACCAAACCAAACCAAACCAA and "FMR4 REV" – GTGGGAAATCAAATGCATCC. Commercially available TaqMan probes (Invitrogen) were used for all other transcripts (MBD4, cat# Hs00187498_m1; FMR1, cat# Hs00924547_m1; MALAT1, cat# Hs01910177_s1). The endogenous control was glyceraldehyde 3-phosphate dehydrogenase (GAPDH, cat# 4326317E) where necessary, and data were analyzed by the $\Delta\Delta$Ct method. For polyA detection, cDNA was synthesized from 1 μg total RNA using the High Capacity cDNA kit and oligodT primers at 50 nM final concentration. In the noRT control, we omitted the reverse-transcriptase enzyme from the cDNA synthesis reaction. Reverse transcription products were amplified using qPCR as described above and visualized on a 1.5% agarose gel.

Human Neural Precursor Cell (hNPC) Culture, Differentiation, and Viral Transduction

hNPCs used in this study were derived from human fetal brains collected from third trimester aborted fetuses received from the Birth Defects Research Lab at the University of Washington in Seattle. This work was classified as "Non-Human Subject Research" by the Human Subject Research Office at the University of Miami, and therefore was not subject to Institutional Review Board approval. Briefly, brain tissues were dissected and dissociated using the trypsin-based Neural Tissue Dissociation Kit (Miltenyi Biotec, cat #130-093-231). Five million single cells were seeded into each 75-mm tissue culture flask in proliferation media supplemented with B27 (1X of proprietary formula), human Epidermal Growth Factor (hEGF) at 20 ng/mL, human Fibroblast Growth Factor (hFGF) at 10 ng/mL, heparin at 20 ug/mL, GlutaMax (1X, Gibco), and Primocin at 0.1 mg/mL (InvivoGen). Neural stem cells formed neurosphere colonies after approximately 7 DIV, while other cells remained in suspension as single cells or formed a monolayer on the flask surface. Neurospheres can be maintained as such in suspension for several months using proliferation media, or hNPCs can be differentiated, forming a mixture of neurons and astrocytes. Neurospheres were transduced with a lentiviral vector expressing *FMR4* (pLentiCMV/TO-mCherry-FMR4) or the control vector (pLentiCMV/TO-mCherry), and

collected for qPCR analysis after 2 days. To differentiate, neurospheres are dissociated into single cells with Accutase and cultured in Advanced DMEM/F12 media without hFGF or hEGF, but with Bottenstein's N2 at 1X (Invitrogen, proprietary formula), 2.5% fetal bovine serum, heparin at 20 μg/mL, GlutaMax (1X, Gibco) and Primocin at 0.1 mg/mL (InvivoGen).

Subcellular Fractionation

Neurospheres were collected by centrifugation at 250 × *g* in order to form 50 μL pellets. Pellets were washed with PBS and fractionated with the NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Scientific, cat# 78835) according to manufacturer's instructions. Briefly, cells were sequentially lysed and centrifuged to first separate pelleted nuclei from cytoplasm, then to separate chromatin from nucleoplasmic components. RNA was extracted from solid chromatin with 1 mL of Trizol, and with 0.75 mL Trizol LS for every 0.25 mL of liquid fraction. cDNA prepared using random primers was used for TaqMan qPCR (*MALAT1, GAPDH,* and *FMR1*), while gene-specific priming and SYBR Green was used for *FMR4* as noted above. In each case, starting material for cDNA reactions was 2 μL total RNA (not equal mass of RNA), to enable comparison of absolute quantities of each transcript between compartments. Relative quantification (RQ) for each transcript in each compartment was calculated from Cq values by qPCR. Individual fraction RQ values were normalized to the total detected amount for each transcript.

Results

Fmr4 Induces Genome-Wide Changes in Gene Expression

Previous studies of the 2.4 kb antisense lncRNA *FMR4* described no *cis*-regulation of *FMR1* (Khalil et al., 2008); therefore we hypothesized that *FMR4* would regulate gene expression in *trans*, which is a well-documented function of other lncRNAs (Rinn et al., 2007; Kino et al., 2010; Miyagawa et al., 2012). In order to comprehensively measure gene regulation in response to *FMR4* at the mRNA level, we treated HEK293T cells with either an siRNA against *FMR4*, a scrambled control siRNA (knockdown), pcDNA3.1-*FMR4*, or the empty pcDNA3.1 vector (overexpression) and processed for microarray hybridization after 6 or 24 h. Using LIMMA, a linear modeling approach (Smyth, 2004), we identified differential expression of over 3,700 genes between *FMR4* knockdown, overexpression and their respective control conditions, and characterized the pattern of target gene expression relative to *FMR4*. To this end, we used the Cluster Affinity tool of TIGR MultiExperiment Viewer to identify genes with opposite behavior in the knockdown condition compared to the overexpression condition. This strategy narrowed our focus to the 238 transcripts represented in **Figure 1A** and Supplementary Figure S1, which are further classified by whether they are concordant or discordant with respect to *FMR4* changes. This analysis yielded 155 and 83 target genes with concordant and discordant changes in mRNA expression
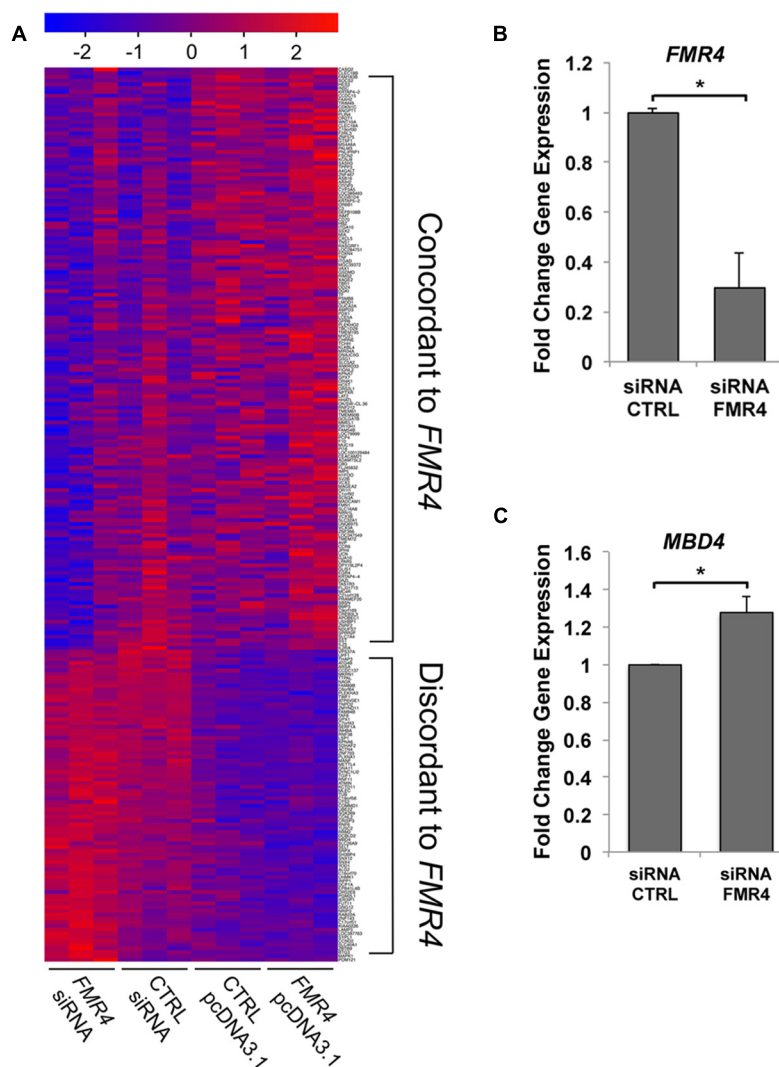
FIGURE 1 | Microarray analysis of mRNA expression changes in response to *FMR4* knockdown and overexpression reveal *trans*-regulatory targets. (A) Microarray analysis of HEK293T cells with *FMR4* overexpression or knockdown (*n* = 3). LIMMA analysis and Cluster Affinity identified genes concordantly or discordantly regulated compared to *FMR4*. **(B)** *FMR4* was knocked down with siRNA in HEK293T cells in order to validate changes in a putative *FMR4*-sensitive gene identified by microarray **(C)**. (*n* = 6, *p < 0.05 by Student's *t*-test).

relative to *FMR4*, respectively. These data support the idea that *FMR4* is a regulator of gene expression through *trans*-activity.

Pathway Analysis and *FMR4* Target Validation

We then analyzed the cohort of concordant and discordant *FMR4*-responsive genes with GeneGo Metacore (**Tables 1** and **2**). Cell cycle regulation and apoptosis were highly ranked biological processes affected by *FMR4*, which is consistent with our earlier findings (Khalil et al., 2008). Additionally, *FMR4*-sensitive genes are enriched in developmental processes in general and neurodevelopmental processes in particular (e.g., adrenergic and opioid signaling, Wnt pathway, cytoskeletal elements, synaptogenesis). Informed by the pathway analysis and previous insights into *FMR4*'s

function, we used HEK293T cells with *FMR4* knocked down by sequential siRNA transfections (**Figure 1B**) to validate a 28% increase in methyl CpG-binding domain (*MBD4*; **Figure 1C**).

Developmental Regulation of Gene Expression by *FMR4* in hNPCs

To investigate the putative role of *FMR4* in neurodevelopment and Fragile X-associated neurological disorders, we used an *in vitro* model system consisting of human fetal-derived neurospheres (hNSs). These cells can be maintained as precursor in hNSs, or induced to differentiate into a mixed culture of early neurons and glia (see Materials and Methods). This system has the advantage of being a primary culture of human brain cells (critical for studying a primate-specific transcript *in vitro*)

TABLE 1 | Top GeneGo pathways represented by *FMR4*-responsive genes identified by microarray analysis.

Category	GeneGo pathway	*p*-value
Cell adhesion	Chemokines and adhesion	0.10161
	Integrin-mediated cell adhesion and migration	0.11458
Cell cycle	Regulation of G1/S transition (part 2)	0.00000
	Nucleocytoplasmic transport of CDK/Cyclins	0.03480
	Spindle assembly and chromosome separation	0.08019
Development	Mu-type opioid receptor signaling via beta-arrestin	0.00000
	S1P1 receptor signaling via beta-arrestin	0.00000
	Beta-adrenergic receptors transactivation of EGFR	0.00000
	Ligand-independent activation of ESR1 and ESR2	0.00001
	Gastrin in cell growth and proliferation	0.00002
	Alpha-2 adrenergic receptor activation of ERK	0.00002
	Regulation of epithelial-to-mesenchymal transition (EMT)	0.04670
	Lipoxin inhibitory action on PDGF, EGF, and LTD4 signaling	0.08717
	A3 receptor signaling	0.11683
Immune response	Oncostatin M signaling via MAPK in mouse cells	0.00000
	Oncostatin M signaling via MAPK in human cells	0.00000
	IL-15 signaling	0.00002
	Histamine H1 receptor signaling in immune response	0.02758
Transport	RAN regulation pathway	0.04453
N/A	Inhibitory action of Lipoxin A4 on PDGF, EGF, and LTD4 signaling	0.08485

TABLE 2 | Top GeneGo process networks represented by *FMR4*-responsive genes identified by microarray analysis.

Category	GeneGo process networks	*p*-value
Apoptosis	Death domain receptors and caspases in apoptosis	0.02682
Cell adhesion	Platelet aggregation	0.02999
	Leucocyte chemotaxis	0.04076
Cell cycle	G1/S Interleukin regulation	0.00000
	G1-S Growth factor regulation	0.00002
Cytoskeleton	Actin filaments	0.00022
	Actin filaments	0.03088
Development	Hemopoiesis, erythropoietin pathway	0.00000
	Neurogenesis and synaptogenesis	0.08499
	Neuromuscular junction	0.11290
	Wnt/beta-catenin, notch, VEGF, IP3, and integrin signaling	0.11674
DNA Damage	Checkpoint	0.00006
Immune response	BCR pathway	0.00008
Inflammation	MIF signaling	0.00000
	IL-2 signaling	0.00003
	Protein C signaling	0.06668
	Kallikrein–kinin system	0.09179
Reproduction	FSH-beta signaling pathway	0.00001
Signal	CREM pathway	0.00002
Transduction	Neuropeptide signaling pathways	0.01407

without the need for reprogramming, as would be the case with embryonic or induced pluripotent stem cells. We focused on the relationship between *FMR4* and its putative target gene, *MBD4*, which we identified by microarray analysis. *MBD4* is a transcriptional repressor involved in DNA repair (Ballestar and Wolffe, 2001; Kondo et al., 2005). Expression of *MBD4* developmentally regulates several tissues (Ruddock-D'cruz et al., 2008; Zhang et al., 2014a), and its aberrant expression in hippocampal GABAergic neurons in psychiatric disease may be linked to abnormal differentiation in these cells (Benes et al., 2009).

The *FMR1* locus is crucial to normal brain development; thus, we wanted to determine whether *FMR4* expression is dependent on developmental stage. We extracted RNA from undissociated hNS ["0 days *in vitro*" (DIV)] and cultured cells from dissociated hNS up to five DIV in differentiation media. At both time points, we measured *FMR4* expression as well as that of *FMR1*, to determine whether these transcripts are independently regulated. We observed that *FMR4* expression
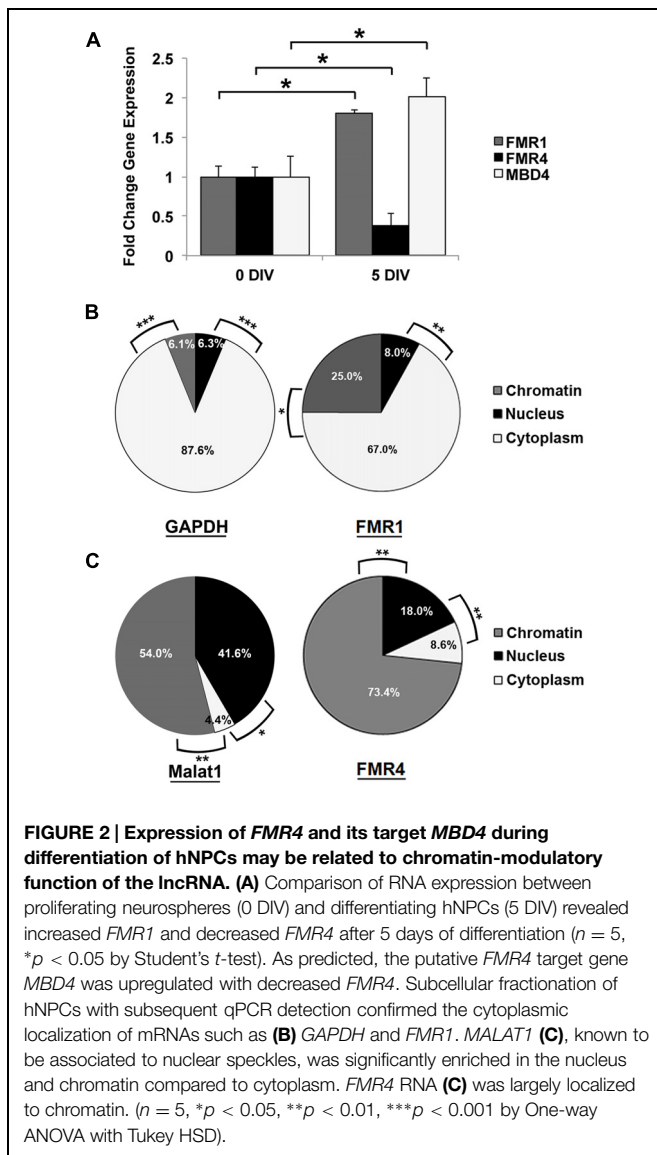
is significantly decreased in differentiating cells at 5 DIV while *FMR1* is increased at the same time point (**Figure 2A**). We then measured *MBD4* and found that it is also developmentally regulated. As indicated by the microarray analysis, *MBD4* was increased at a time point when *FMR4* expression is low (**Figure 2A**). We also observed changes in expression level of other *FMR4* target genes in undifferentiated, proliferating hNPCs transduced with an mCherry-tagged lentiviral vector expressing *FMR4* (Supplementary Figure S2A). We found that *FMR4* overexpression significantly upregulated two putative targets, the deubiquitinase *YOD1* and the G-protein subunit *GNG12*, and downregulated the ribonucleotide reductase *RRM2* (Supplementary Figure S2B). These data corroborate our finding that *FMR4* regulates gene expression in *trans*.

Molecular Mechanisms of *FMR4*

To better understand the molecular role of *FMR4*, we performed subcellular fractionation of hNS. As expected, *GAPDH* and *FMR1* mRNA were highest in the cytoplasmic fraction (87.6 and 67.0%, respectively) where they are translated (Duszczyk et al., 2011) (**Figure 2B**). Some lncRNAs are expressed predominantly in the nucleus. A well-known example of this is the lncRNA *NEAT2/MALAT1* (Gutschner et al., 2013). We measured this transcript as a positive control for nuclear transcripts, and found it enriched in the nucleoplasm (41.6%) and chromatin (54.0%) relative to the cytoplasm (**Figure 2C**). *FMR4* RNA was primarily localized (73.4%) to chromatin (**Figure 2C**), which is consistent with a transcriptional regulatory function of *FMR4*.

FIGURE 2 | Expression of *FMR4* and its target *MBD4* during differentiation of hNPCs may be related to chromatin-modulatory function of the lncRNA. (A) Comparison of RNA expression between proliferating neurospheres (0 DIV) and differentiating hNPCs (5 DIV) revealed increased *FMR1* and decreased *FMR4* after 5 days of differentiation ($n = 5$, *$p < 0.05$ by Student's *t*-test). As predicted, the putative *FMR4* target gene *MBD4* was upregulated with decreased *FMR4*. Subcellular fractionation of hNPCs with subsequent qPCR detection confirmed the cytoplasmic localization of mRNAs such as **(B)** *GAPDH* and *FMR1*. *MALAT1* **(C)**, known to be associated to nuclear speckles, was significantly enriched in the nucleus and chromatin compared to cytoplasm. *FMR4* RNA **(C)** was largely localized to chromatin. ($n = 5$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ by One-way ANOVA with Tukey HSD).

Since RNA polyadenylation frequently targets mature RNA for export into the cytoplasm (Zhang et al., 2014b); many lncRNAs are not polyadenylated. Therefore, to establish whether *FMR4* can be polyadenylated, we performed first-strand cDNA synthesis with oligodT primers, thereby capturing polyadenylated transcripts from total hNS RNA. We observed *via* qPCR that in addition to *FMR1* and *GAPDH* mRNAs (which are expected to be polyadenylated), a detectable portion of *FMR4* is also polyadenylated (Supplementary Figure S3). These data suggest that a fraction of *FMR4* may be stabilized by polyadenylation; this would increase the molecule's half-life and permit its diffusion to distant genomic loci.

Discussion

Loss of *FMR1* mRNA and FMRP function has been widely studied, but there is relatively little known about ncRNA

encoded by the locus. Here we show that one such ncRNA, *FMR4*, may regulate mRNA expression genome-wide *via* a developmentally regulated transcriptional mechanism, thereby impacting important biological processes.

Similar to *FMR4, trans*-acting lncRNAs [such as *HOTAIR, MALAT1,* and *GAS5* (Rinn et al., 2007; Kino et al., 2010; Miyagawa et al., 2012)] affect loci far from their genomic locus of origin. In this study, we confirmed that knockdown and overexpression of *FMR4* causes changes in genes involved in proliferation and differentiation. As a chromatin-associated transcript, *FMR4* may act at the transcriptional level by forming complexes with histone modifying enzymes or by directly targeting mRNA stability, splicing, or editing. It remains to be seen whether these observations are dependent on RNA-protein interactions, and whether they result from direct epigenetic changes or *via* downstream effects.

Our data show that *FMR4* is developmentally regulated in an hNPC model. After 5 days of *in vitro* differentiation, *FMR4* expression is significantly reduced, while that of both *FMR1* and the *FMR4* target gene *MBD4* is increased. Since *FMR4* and *FMR1* do not interact in *cis* but are discordantly expressed with differentiation of hNPCs, the bidirectional promoter responsible for expression of both of these transcripts might be activated in only one direction at a time. An alternative possibility is that *FMR4* RNA is degraded at a higher rate during this period while *FMR1* is not. Nevertheless, it would be interesting to determine whether transcription factors normally regulate this locus as a whole or target each transcript individually. It is also unclear whether the decrease in *FMR4* lncRNA expression contributes to disease in addition to loss of *FMR1*, or is an artifact of the locus-wide, full mutation-induced epigenetic changes causing transcriptional repression. Future studies will be necessary to distinguish between these possible mechanisms and consequences of *FMR4* regulation.

Discordant regulation of *MBD4* and *FMR4* leads us to conclude that *FMR4* regulates expression of genes at distal loci, as the same relationship was identified by our overexpression and knockdown studies. Our independent validation studies confirm this was not due to false positive identification, which is a common problem in genome-wide analyses such as microarray studies. *FMR4*'s localization to the chromatin fraction points to a transcriptional mechanism for this effect, however, studies evaluating the direct binding of *FMR4* to nucleic acids or proteins are warranted. Such studies could suggest a direct role for *FMR4* in specifically regulating *MBD4* and gene expression more broadly, although this cannot be concluded definitively based solely on changes in mRNA levels. We acknowledge that differences in RNA processing or any number of effects downstream of *FMR4* could be responsible for the observed differential gene expression, therefore it would be useful to examine precise interactions between *FMR4* and its targets. Chromatin immunoprecipitation experiments could, for example, establish an interaction between *FMR4* and promoter regions of target genes. Based on the function of other lncRNAs, one could speculate that

FMR4 helps form three-dimensional interactions between distant sequence elements such as promoters and enhancers (Ma et al., 2014), and that *FMR4* polyadenylation targets this transcript for further RNA processing (Norbury, 2013). Answering these questions would help our understanding of *FMR4* function in particular and continue the rapid expansion of evidence on the elements that govern lncRNA actions in general.

In this study we have described novel facets of *FMR4* functionality as it relates to neurodevelopment, and suggest that perturbation in the expression of this lncRNA may contribute to pathogenesis of the Fragile X repeat expansion-associated disorders. We report that *trans*-regulatory activity of *FMR4* is corroborated by changes in target gene expression during differentiation of hNPCs. With this new information, we have further developed the evidence supporting the role of primate-specific lncRNAs in complex developmental programs and opened new avenues of research into the causes and therapies for Fragile X-associated neurological disorders.

Funding

Author Contributions

VP, CP, DM, KW, DV, and MM contributed to the acquisition, analysis, and interpretation of data. VP, CP, ZZ, JB, VL, JS, and CW made substantial contributions to the conception and design of the work. All authors participated in drafting and/or revising the manuscript for intellectual content, approval for publication and agree to be accountable for all aspects of the work.

Acknowledgments

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fgene.2015.00263

References

Ballestar, E., and Wolffe, A. P. (2001). Methyl-CpG-binding proteins. Targeting specific gene repression. *Eur. J. Biochem.* 268, 1–6. doi: 10.1046/j.1432-1327.2001.01869.x

Banfai, B., Jia, H., Khatun, J., Wood, E., Risk, B., Gundling, W. E., et al. (2012). Long noncoding RNAs are rarely translated in two human cell lines. *Genome Res.* 22, 1646–1657. doi: 10.1101/gr.134767.111

Benes, F. M., Lim, B., and Subburaju, S. (2009). Site-specific regulation of cell cycle and DNA repair in post-mitotic GABA cells in schizophrenic versus bipolars. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11731–11736. doi: 10.1073/pnas.0903066106

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., et al. (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* 308, 1149–1154. doi: 10.1126/science.1108625

Coffee, B., Zhang, F., Ceman, S., Warren, S. T., and Reines, D. (2002). Histone modifications depict an aberrantly heterochromatinized FMR1 gene in fragile x syndrome. *Am. J. Hum. Genet.* 71, 923–932. doi: 10.1086/342931

Coffee, B., Zhang, F., Warren, S. T., and Reines, D. (1999). Acetylated histones are associated with FMR1 in normal but not fragile X-syndrome cells. *Nat. Genet.* 22, 98–101. doi: 10.1038/8807

De Smith, A. J., Purmann, C., Walters, R. G., Ellis, R. J., Holder, S. E., Van Haelst, M. M., et al. (2009). A deletion of the HBII-85 class of small nucleolar RNAs (snoRNAs) is associated with hyperphagia, obesity and hypogonadism. *Hum. Mol. Genet.* 18, 3257–3265. doi: 10.1093/hmg/ddp263

Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., et al. (2012). Landscape of transcription in human cells. *Nature* 489, 101–108. doi: 10.1038/nature11233

Duszczyk, M. M., Wutz, A., Rybin, V., and Sattler, M. (2011). The Xist RNA A-repeat comprises a novel AUCG tetraloop fold and a platform for multimerization. *RNA* 17, 1973–1982. doi: 10.1261/rna.2747411

Fabian, M. R., and Sonenberg, N. (2012). The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nat. Struct. Mol. Biol.* 19, 586–593. doi: 10.1038/nsmb.2296

Faghihi, M. A., Modarresi, F., Khalil, A. M., Wood, D. E., Sahagan, B. G., Morgan, T. E., et al. (2008). Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nat. Med.* 14, 723–730. doi: 10.1038/nm1784

Gutschner, T., Hammerle, M., and Diederichs, S. (2013). MALAT1 – a paradigm for long noncoding RNA function in cancer. *J. Mol. Med. (Berl.)* 91, 791–801. doi: 10.1007/s00109-013-1028-y

Hajjari, M., Khoshnevisan, A., and Shin, Y. K. (2014). Molecular function and regulation of long non-coding RNAs: paradigms with potential roles in cancer. *Tumour Biol.* 35, 10645–10663. doi: 10.1007/s13277-014-2636-z

Hornstra, I. K., Nelson, D. L., Warren, S. T., and Yang, T. P. (1993). High resolution methylation analysis of the FMR1 gene trinucleotide repeat region in fragile X syndrome. *Hum. Mol. Genet.* 2, 1659–1665. doi: 10.1093/hmg/2.10.1659

Hung, T., Wang, Y., Lin, M. F., Koegel, A. K., Kotake, Y., Grant, G. D., et al. (2011). Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat. Genet.* 43, 621–629. doi: 10.1038/ng.848

Kenneson, A., Zhang, F., Hagedorn, C. H., and Warren, S. T. (2001). Reduced FMRP and increased FMR1 transcription is proportionally associated with CGG repeat number in intermediate-length and premutation carriers. *Hum. Mol. Genet.* 10, 1449–1454. doi: 10.1093/hmg/10.14.1449

Khalil, A. M., Faghihi, M. A., Modarresi, F., Brothers, S. P., and Wahlestedt, C. (2008). A novel RNA transcript with antiapoptotic function is silenced in fragile X syndrome. *PLoS ONE* 3:e1486. doi: 10.1371/journal.pone.0001486

Kino, T., Hurt, D. E., Ichijo, T., Nader, N., and Chrousos, G. P. (2010). Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci. Signal.* 3:ra8. doi: 10.1126/scisignal.2000568

Kondo, E., Gu, Z., Horii, A., and Fukushige, S. (2005). The thymine DNA glycosylase MBD4 represses transcription and is associated with methylated p16(INK4a) and hMLH1 genes. *Mol. Cell. Biol.* 25, 4388–4396. doi: 10.1128/MCB.25.11.4388-4396.2005

Kotake, Y., Nakagawa, T., Kitagawa, K., Suzuki, S., Liu, N., Kitagawa, M., et al. (2011). Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene* 30, 1956–1962. doi: 10.1038/onc.2010.568

Kumari, D., and Usdin, K. (2010). The distribution of repressive histone modifications on silenced FMR1 alleles provides clues to the mechanism of gene silencing in fragile X syndrome. *Hum. Mol. Genet.* 19, 4634–4642. doi: 10.1093/hmg/ddq394

Ladd, P. D., Smith, L. E., Rabaia, N. A., Moore, J. M., Georges, S. A., Hansen, R. S., et al. (2007). An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals. *Hum. Mol. Genet.* 16, 3174–3187. doi: 10.1093/hmg/ddm293

Ma, W., Ay, F., Lee, C., Gulsoy, G., Deng, X., Cook, S., et al. (2014). Fine-scale chromatin interaction maps reveal the cis-regulatory landscape of human lincRNA genes. *Nat. Methods* 12, 71–78. doi: 10.1038/nmeth.3205

Mercer, T. R., Qureshi, I. A., Gokhan, S., Dinger, M. E., Li, G., Mattick, J. S., et al. (2010). Long noncoding RNAs in neuronal-glial fate specification and oligodendrocyte lineage maturation. *BMC Neurosci.* 11:14. doi: 10.1186/1471-2202-11-14

Miyagawa, R., Tano, K., Mizuno, R., Nakamura, Y., Ijiri, K., Rakwal, R., et al. (2012). Identification of cis- and trans-acting factors involved in the localization of MALAT-1 noncoding RNA to nuclear speckles. *RNA* 18, 738–751. doi: 10.1261/rna.028639.111

Norbury, C. J. (2013). Cytoplasmic RNA: a case of the tail wagging the dog. *Nat. Rev. Mol. Cell Biol.* 14, 643–653. doi: 10.1038/nrm3645

Oostra, B. A., and Willemsen, R. (2003). A fragile balance: FMR1 expression levels. *Hum. Mol. Genet.* 12, R249–R257. doi: 10.1093/hmg/ddg298

Pastori, C., Peschansky, V. J., Barbouth, D., Mehta, A., Silva, J. P., and Wahlestedt, C. (2014). Comprehensive analysis of the transcriptional landscape of the human FMR1 gene reveals two new long noncoding RNAs differentially expressed in Fragile X syndrome and Fragile X-associated tremor/ataxia syndrome. *Hum. Genet.* 133, 59–67. doi: 10.1007/s00439-013-1356-6

Powell, W. T., Coulson, R. L., Crary, F. K., Wong, S. S., Ach, R. A., Tsang, P., et al. (2013). A Prader-Willi locus lncRNA cloud modulates diurnal genes and energy expenditure. *Hum. Mol. Genet.* 22, 4318–4328. doi: 10.1093/hmg/ddt281

Qureshi, I. A., Mattick, J. S., and Mehler, M. F. (2010). Long non-coding RNAs in nervous system function and disease. *Brain Res.* 1338, 20–35. doi: 10.1016/j.brainres.2010.03.110

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311–1323. doi: 10.1016/j.cell.2007.05.022

Rother, S., and Meister, G. (2011). Small RNAs derived from longer non-coding RNAs. *Biochimie* 93, 1905–1915. doi: 10.1016/j.biochi.2011.07.032

Ruddock-D'cruz, N. T., Xue, J., Wilson, K. J., Heffernan, C., Prashadkumar, S., Cooney, M. A., et al. (2008). Dynamic changes in the localization of five members of the methyl binding domain (MBD) gene family during murine and bovine preimplantation embryo development. *Mol. Reprod. Dev.* 75, 48–59. doi: 10.1002/mrd.20712

Sahoo, T., Del Gaudio, D., German, J. R., Shinawi, M., Peters, S. U., Person, R. E., et al. (2008). Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nat. Genet.* 40, 719–721. doi: 10.1038/ng.158

Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3, 3. doi: 10.2202/1544-6115.1027

Sopher, B. L., Ladd, P. D., Pineda, V. V., Libby, R. T., Sunkin, S. M., Hurley, J. B., et al. (2011). CTCF regulates ataxin-7 expression through promotion of a convergently transcribed, antisense noncoding RNA. *Neuron* 70, 1071–1084. doi: 10.1016/j.neuron.2011.05.027

Sutcliffe, J. S., Nelson, D. L., Zhang, F., Pieretti, M., Caskey, C. T., Saxe, D., et al. (1992). DNA methylation represses FMR-1 transcription in fragile X syndrome. *Hum. Mol. Genet.* 1, 397–400. doi: 10.1093/hmg/1.6.397

Tassone, F., Hagerman, R. J., Taylor, A. K., Gane, L. W., Godfrey, T. E., and Hagerman, P. J. (2000). Elevated levels of FMR1 mRNA in carrier males: a new mechanism of involvement in the fragile-X syndrome. *Am. J. Hum. Genet.* 66, 6–15. doi: 10.1086/302720

Tsai, M. C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lan, F., et al. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329, 689–693. doi: 10.1126/science.1192002

Velmeshev, D., Magistri, M., and Faghihi, M. A. (2013). Expression of non-protein-coding antisense RNAs in genomic regions related to autism spectrum disorders. *Mol. Autism* 4, 32. doi: 10.1186/2040-2392-4-32

Wang, K. C., and Chang, H. Y. (2011). Molecular mechanisms of long noncoding RNAs. *Mol. Cell.* 43, 904–914. doi: 10.1016/j.molcel.2011.08.018

Yap, K. L., Li, S., Munoz-Cabello, A. M., Raguz, S., Zeng, L., Mujtaba, S., et al. (2010). Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol. Cell.* 38, 662–674. doi: 10.1016/j.molcel.2010.03.021

Zhang, L. J., Zhu, Y. N., Gao, Y., Liu, S. Y., Zhai, B., Li, C. H., et al. (2014a). The MBD4 gene plays an important role in porcine adipocyte differentiation. *Cell Physiol. Biochem.* 34, 1216–1226. doi: 10.1159/000366333

Zhang, Y., Yang, L., and Chen, L. L. (2014b). Life without A tail: new formats of long noncoding RNAs. *Int. J. Biochem. Cell Biol.* 54, 338–349. doi: 10.1016/j.biocel.2013.10.009

Ziats, M. N., and Rennert, O. M. (2013). Aberrant expression of long noncoding RNAs in autistic brain. *J. Mol. Neurosci.* 49, 589–593. doi: 10.1007/s12031-012-9880-8

Long non-coding RNA SOX2OT: expression signature, splicing patterns, and emerging roles in pluripotency and tumorigenesis

Alireza Shahryari [1], Marie Saghaeian Jazi [2], Nader M. Samaei [3] and Seyed J. Mowla [4*]

[1] Stem Cell Research Center, Golestan University of Medical Sciences, Gorgan, Iran, [2] Department of Molecular Medicine, Faculty of Advanced Medical Technologies, Golestan University of Medical Sciences, Gorgan, Iran, [3] Department of Medical Genetics, Faculty of Advanced Medical Technologies, Golestan University of Medical Sciences, Gorgan, Iran, [4] Department of Molecular Genetics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran

SOX2 overlapping transcript (SOX2OT) is a long non-coding RNA which harbors one of the major regulators of pluripotency, SOX2 gene, in its intronic region. *SOX2OT* gene is mapped to human chromosome 3q26.3 (Chr3q26.3) locus and is extended in a high conserved region of over 700 kb. Little is known about the exact role of SOX2OT; however, recent studies have demonstrated a positive role for it in transcription regulation of *SOX2* gene. Similar to SOX2, SOX2OT is highly expressed in embryonic stem cells and down-regulated upon the induction of differentiation. SOX2OT is dynamically regulated during the embryogenesis of vertebrates, and delimited to the brain in adult mice and human. Recently, the disregulation of SOX2OT expression and its concomitant expression with SOX2 have become highlighted in some somatic cancers including esophageal squamous cell carcinoma, lung squamous cell carcinoma, and breast cancer. Interestingly, SOX2OT is differentially spliced into multiple mRNA-like transcripts in stem and cancer cells. In this review, we are describing the structural and functional features of *SOX2OT*, with an emphasis on its expression signature, its splicing patterns and its critical function in the regulation of *SOX2* expression during development and tumorigenesis.

Keywords: lncRNA, SOX2OT, splicing pattern, expression signature, pluripotency, tumorigenesis, stem cell

Introduction

According to the recent genome-wide studies, most of the human genome is transcribed, yielding a complex network of large and small RNA molecules in human cells. However, only 1–2% of the transcripts have the capacity for protein translation (Kapranov et al., 2007; Guttman et al., 2009). The new class of long (or large) non-coding RNAs (lncRNAs) comprises the most proportion of the human transcriptome. Little is known about the exact functional roles of lncRNAs in human. Nevertheless, some recent studies have reported dysregulations of lncRNAs in several human disorders. LncRNAs key roles in the regulation of pluripotency, stem cells differentiation, and tumorigenesis are emerging (Perez et al., 2007; Gupta et al., 2010; Loewer et al., 2010; Esteller, 2011; Ng et al., 2011; Prensner and Chinnaiyan, 2011). Furthermore, a number of studies have achieved toward a therapeutic effect for some genetic disorders by targeting an lncRNA *in vitro* and *in vivo* (Gupta et al., 2010; Gutschner et al., 2013; Meng et al., 2015).

SOX2 is a HMG-box transcription factor which is essential for the maintenance of self-renewal and the pluripotency of undifferentiated embryonic stem cells (Avilion et al., 2003; Fong et al., 2008). More interestingly, SOX2 along with OCT4, c-Myc and Klf4 plays a critical role in the generation of induced pluripotent stem cells (iPSC) from both adult human and mouse somatic cells (Takahashi and Yamanaka, 2006; Takahashi et al., 2007). Recently, it has been suggested that SOX2 promotes tumor initiation and controls cancer stem cell properties in squamous cell carcinoma (SCC) of the skin tumors (Boumahdi et al., 2014). The single-exon *SOX2* gene was mapped to the human chromosome 3q26.3 (Chr3q26.3) locus, where it is embedded within the intronic region of a multi-exon lncRNA, known as *SOX2 overlapping transcript (SOX2OT)* which transcribed in the same orientation as *SOX2* (Fantes et al., 2003). While little is known about the exact role of *SOX2OT*, recent studies have demonstrated a positive role for it in the regulation of *SOX2* gene in human stem cells (Amaral et al., 2009; Shahryari et al., 2014).

Human *SOX2OT* gene has a high nucleotide identity with its ortholog in mouse and other vertebrates, demonstrating its high degree of evolutionarily conservation. The multi-exon SOX2OT has no open reading frame (ORF), but is spliced into several mRNA-like transcripts with the longest one of approximately 3.5 kb in human (Amaral et al., 2009; Shahryari et al., 2014). Close concomitant expression patterns of SOX2OT and SOX2 in stem cells and some human cancers, have all suggested that they may be co-regulated and involved in similar molecular pathways. Accordingly, some recent reports have demonstrated the transcriptional regulation of *SOX2* by SOX2OT (Amaral et al., 2009; Askarian-Amiri et al., 2014; Hou et al., 2014; Shahryari et al., 2014).

In this review, we have delineated the complex structure and functional features of *SOX2OT* locus, with more emphasis on its expression and splicing patterns, and its potential role in the regulation of SOX2 expression during the development and cancer progression.

Genomic Architecture of *SOX2OT* Gene Region in Vertebrates

SOX2OT gene with the official symbol of *SOX2-OT* (also known as NCRNA00043) was originally mapped to the human chromosome 3q26.3-q27 locus [current location of NC_000003.12 (181056680..181742228)], and harbors one of the main regulators of pluripotency, *SOX2* gene [also known as *ANOP3* and *MCOPS3*, with current location of NC_000003.12 (181711924..181714436)], in its intronic region overlapping in the same transcriptional orientation (Fantes et al., 2003). *SOX2OT* gene is located and extended in a highly conserved region of over 700 kb in human and other vertebrates (Fantes et al., 2003; Amaral et al., 2009; **Figure 1A**).

Amplification of several genomic regions at 3q26-qter chromosome is associated with multiple human cancers (Massion et al., 2002; Jiang et al., 2004). The gene amplification events in those regions, particularly in q26–q29 region of chromosome 3, are present in the multiple types of SCCs of different tissues including lung, head and neck, esophagus, and cervix (Gebhart

and Liehr, 2000; Balsara and Testa, 2002; Bass et al., 2009). Interestingly, a 2 Mb gained/amplified genomic region in 3q26.3 which encompasses *SOX2* and *SOX2OT* has been reported in lung SCC (Hussenet et al., 2010).

Chromatin modification maps of chromosome 3q26.3-q27 acquired by chromatin immune precipitation sequencing (ChIP-Seq) data represented several transcription start sites (TSSs) for *SOX2OT* gene (Mikkelsen et al., 2007; Amaral et al., 2009). Those promoter regions embedded within 1–7 non-coding highly conserved sequence blocks in vertebrates known as highly conserved elements (HCEs), probably are associated with the regulatory region of *SOX2OT*. These blocks of transposon-free regions with over 5 kb long have remained resistant to transposon invasion throughout vertebrate evolution and encompassed regulatory sequences controlling the expression of genes that are involved in early development (Simons et al., 2006; Amaral et al., 2009).

Interestingly, analysis of the alternative TSSs of *Sox2ot* orthologous in various vertebrates demonstrated the existence of a distal promoter, located over 500 kb upstream of the *SOX2OT* sequence in mouse and human (*SOX2OT* refers to human and *Sox2ot* to non-human). This distal promoter region which is associated with the transposon-free region, highly positional conserved elements, and histone modification marks of promoters, created a novel isoform of Sox2ot termed Sox2dot (Sox2 distal overlapping transcript) which has HCE 1 with an enhancer-like function in the mouse's developing forebrain (Amaral et al., 2009). Here, we bioinformatically analyzed the potential binding sites for transcription factors in a highly conserved genomic regions upstream of *SOX2OT* and *SOX2DOT*. As illustrated in **Figure 1**, the data represent the existence of binding sites of several transcription factors involved in cancer progression as well as stem cells pluripotency and differentiation in those regulatory regions. Noticeably, the number and distribution of binding sites of some transaction factors belonging to POU domain and HMG-box families is surprising (**Figures 1B,C**).

Primary sequence analysis of *sox2ot* in vertebrates including fish, reptiles, amphibians and mammals highlighted some highly conserved regions, including a 400-nt segment in exons near to *SOX2* gene, as well as an upstream region with more than 90% identity between mouse and human genomes. However, there is only a low degree of conservation when full length sequence of *SOX2OT* gene (∼750 kb) is compared among different species (Amaral et al., 2009; **Figure 2**).

Splicing Patterns of *SOX2OT* Gene in Human and Other Vertebrates

Protein-coding capacity parameters including ORF length, synonymous versus non-synonymous base substitution rates, and similarity to known proteins demonstrated that human and mouse SOX2OT/Sox2ot full-length sequences have no significant protein-coding potential. Nevertheless, there is a possibility for generation of some small peptides, encoded by some transcripts (Dinger et al., 2008; Amaral et al., 2009). Mark signs of mRNAs including a lot of cap and poly Adenine signals suggest that
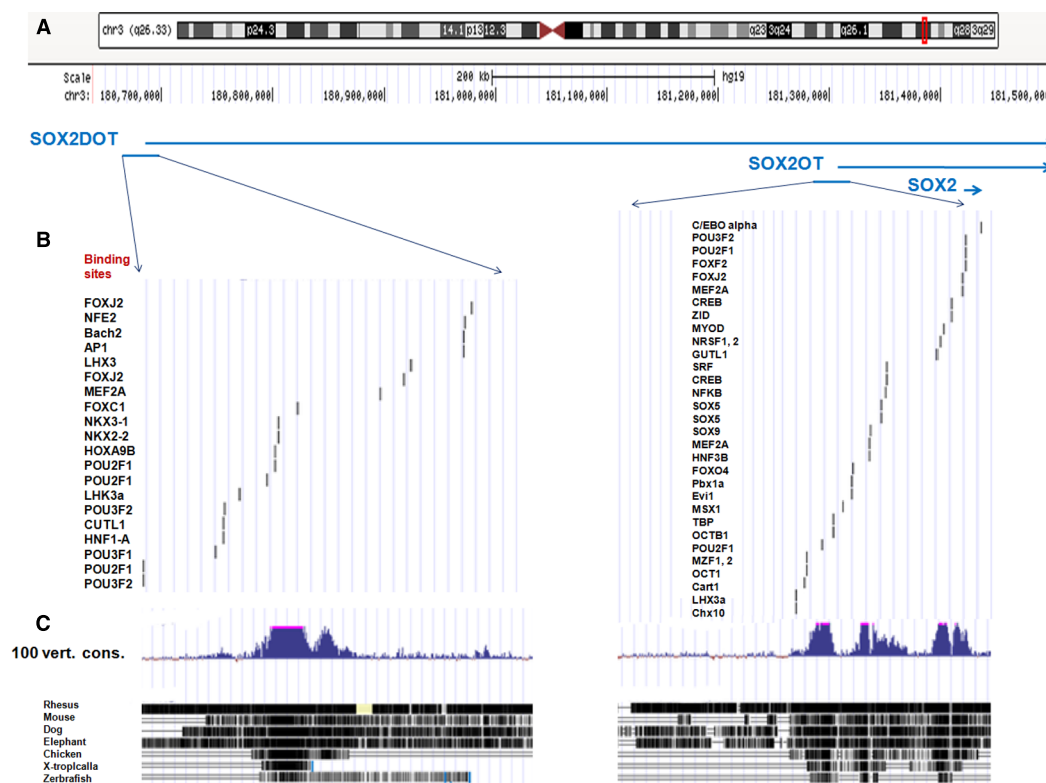
FIGURE 1 | Genomic architecture of Chr3q26.33 region in human and vertebrates. (A) The banding pattern of chromosome 3 and location of *SOX2OT* locus of 3q26.33 is presented according to the UCSC genome browser (h19 assembly). **(B)** The conserved transcription factor binding sites is presented at upstream of human genomic regions of SOX2OT and the isoform of SOX2DOT. The binding sites distribution for multiple transcription factors of POU domain and HMG-box families is noticeable. **(C)** A high degree of conservation at upstream of genomic regions of SOX2OT and SOX2DOT in 100 vertebrates is presented, using Multiz alignment program (adopted from http://genome.ucsc.edu).
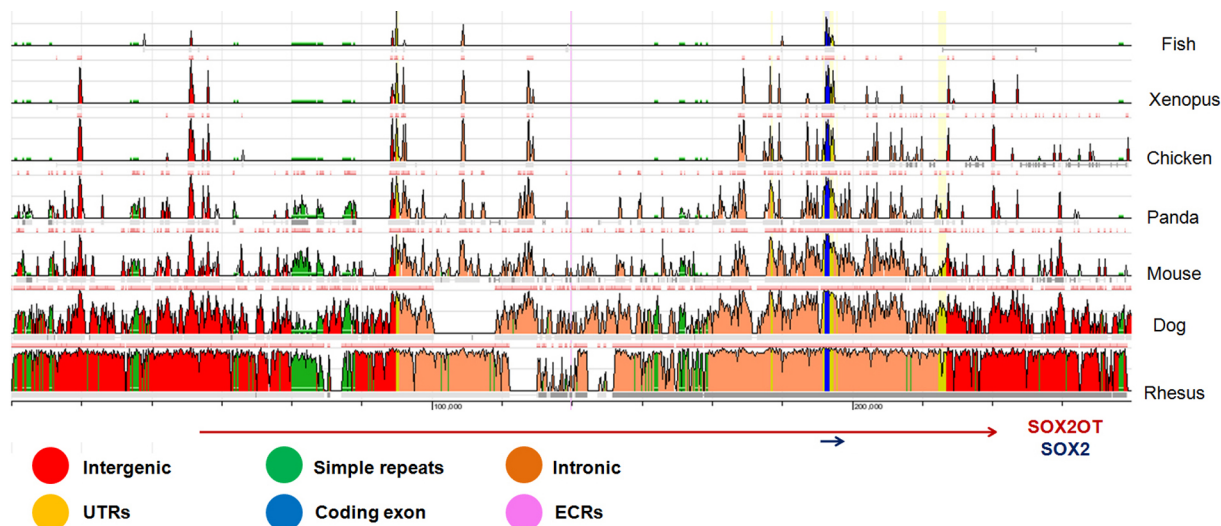
FIGURE 2 | A schematic representation of comparative genomics locus of *SOX2OT* in different species. The evolutionary conserved regions (ECR) of chr3:181236624-18150288 adapted from ECR browser (http://ecrbrowser.dcode.org) on human (hg19) is presented. The data is compared with some well-known vertebrates including rhesus, dog, mouse, chicken, *Xenopus*, and fish. The ECR length and similarity considered as browser default. The UCSC known genes of *SOX2OT* and *SOX2* locating in the region is shown at the bottom. Notice that the most conserved regions are concentrated around SOX2 overlapping region.

FIGURE 3 | Schematic representative of human *SOX2OT* gene and its splicing patterns. (A) Multi-exon gene of SOX2OT is located on human Chr3q26.3 and is extending in a 750 kb genomic region, where it holds the single-exon of SOX2 gene in its intronic region in the same strand and orientation. **(B)** The splicing patterns of SOX2OT and **(C)** SOX2DOT isoforms are presented, respectively.

SOX2OT gene is transcribed by RNA polymerase II enzyme, and produces a mRNA-like lncRNA transcript (Numata et al., 2003; Amaral et al., 2009).
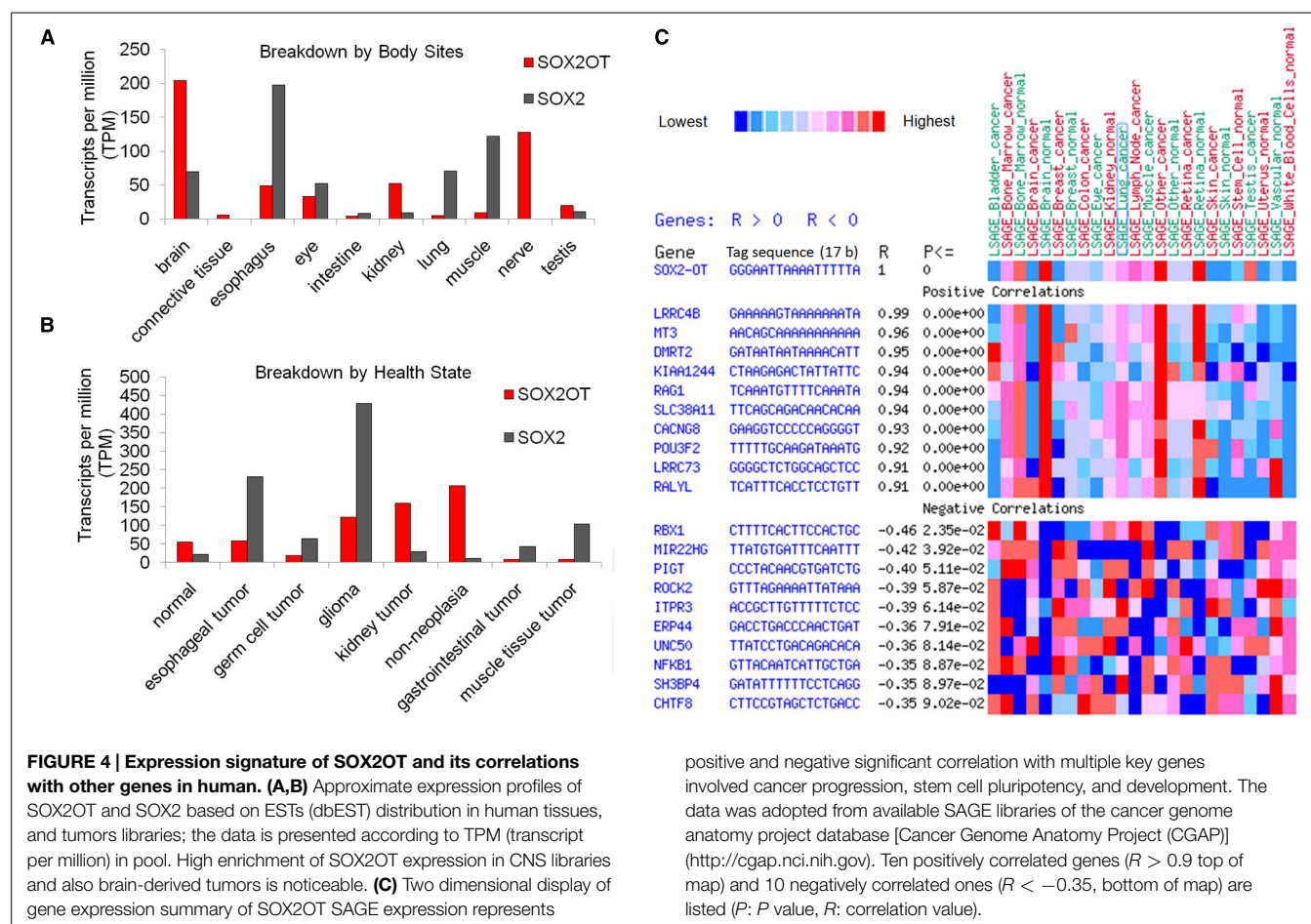
Human and mouse SOX2OT have multiple TSSs, and several alternatively spliced variants and polyadenylation sites have already been reported for them (Amaral et al., 2009). Several full-length clones of mouse sox2ot have been registered with a wide range of sizes, from 638 nucleotides (GenBank accession no. BY721402) to an approximately 3.5 kb form (accession no. AK031919). The various sizes of the registered cDNA clones are in accordance with the Northern blot data obtained from several mouse tissues. While the most abundant isoform of sox2ot posses a size of ∼3 kb, several other rare ones with approximate sizes of 1, 4, 6, and >10 kb have also been reported in some mouse tissues. In zebrafish embryo, Northern blot analysis revealed an abundant 2.5 kb transcript variant and two other less abundant transcripts of 1.5 and 6 kb (Amaral et al., 2009).

As we have previously reported, SOX2OT is spliced into several transcript variants, including SOX2OT, SOX2OT-S1, and SOX2OT-S2 which co-upregulated with master regulators of pluripotency, SOX2 and OCT4, in esophageal squamous cell carcinoma (ESCC). SOX2OT-S1 (Accession no: JN711430, GI: 379031002) lacks exon 4 of the main transcript, whereas SOX2OT-S2 (SOX2OT-S2; Accession no: JN882275, GI: 379031003) lacks exons 3 and 4. In addition to the experimentally approved novel transcripts, human EST database (dbEST) also provided some ESTs with GenBank accession numbers BX423294.2, BX442540.2, BX459910.2, DA268964.1, and DA282731.1 which are related to the novel sequence of exon 3-exon five junction in SOX2OT-S1, and DA308672.1 which is related to the novel sequence of exon 2-exon five junction in SOX2OT-S2 variant (Shahryari et al., 2014).

More than 15 different *Major Class* of introns (GT-AG), at least 13 spliced variants, and six TSSs were presented for SOX2OT using bioinformatics analysis and AceView annotation (Amaral et al., 2009). Our group has been also identified several novel variants of SOX2DOT, which demonstrates a complex pattern of TSSs and alternative splicing of SOX2OT (**Figure 3A**). According to the validated NCBI Reference Sequence (RefSeqs), splicing patterns of SOX2OT, as illustrated in **Figure 3**, generates at least six transcript variants. Among those, three variants are generated from alternative splicing of SOX2OT, while the other three ones are originated from SOX2DOT (**Figures 3B,C**).

Expression Signature of SOX2OT in Somatic, Stem, and Cancer Cells

Sox2ot isoforms are widely expressed in whole embryo and newborn mouse, but in adult tissues their expression is primarily restricted to brain. It is also expressed at lower levels in tissues where Sox2 is also expressed, such as lung, as well as in tissues were sox2 is not expressed, such as testis. Nevertheless, sox2dot isoform is exclusively expressed in adult mouse brain tissues. Concomitant with Sox2, Sox2ot is mainly expressed in mouse embryonic stem cells and down-regulated during the course of differentiation. Nevertheless, only Sox2ot is upregulated during the late mouse embryoid body differentiation events. Moreover, expression of Sox2 and Sox2ot are coregulated during mouse neurosphere differentiation *in vitro*. Accordingly, Sox2dot isoform is also upregulated upon the induction of differentiation in neurospheres. Similar to mouse, Sox2ot and sox2 are also dynamically regulated during embryogenesis of other vertebrate, including

FIGURE 4 | Expression signature of SOX2OT and its correlations with other genes in human. (A,B) Approximate expression profiles of SOX2OT and SOX2 based on ESTs (dbEST) distribution in human tissues, and tumors libraries; the data is presented according to TPM (transcript per million) in pool. High enrichment of SOX2OT expression in CNS libraries and also brain-derived tumors is noticeable. **(C)** Two dimensional display of gene expression summary of SOX2OT SAGE expression represents positive and negative significant correlation with multiple key genes involved cancer progression, stem cell pluripotency, and development. The data was adopted from available SAGE libraries of the cancer genome anatomy project database [Cancer Genome Anatomy Project (CGAP)] (http://cgap.nci.nih.gov). Ten positively correlated genes ($R > 0.9$ top of map) and 10 negatively correlated ones ($R < -0.35$, bottom of map) are listed (P: P value, R: correlation value).

chicken and zebrafish (Mercer et al., 2008; Amaral et al., 2009).

The lncRNA SOX2OT is co-upregulated with master regulators of pluripotency, SOX2 and OCT4, in ESCC. The qRT-PCR analysis revealed a high level of SOX2OT expression in tumor samples of ESCC, compared to the apparently non-tumor marginal tissues from the same patients, which suggested a potential part for it in tumorigenesis of esophagus (Shahryari et al., 2014).

A concomitant expression pattern of *SOX2OT* with that of *SOX2* and *OCT4* genes is reported in a pluripotent cell line, NT2. SOX2OT and its variants also proved to have a distinct expression pattern during neural differentiation of NT2 cells. The expression pattern of SOX2OT variants was similar to those of SOX2 and OCT4, and downregulated upon the induction of neural differentiation. However, in contrast to a complete shut-down of SOX2 and OCT4 expression, a low expression of SOX2OT and its variants is persisted in later time points of differentiation (Shahryari et al., 2014).

Distinct differences in the expression patterns of SOX2OT and SOX2 were observed in breast cancer tissue samples. Analysis of the genome-wide RNA transcript profiles from the Cancer Genome Atlas (breast invasive carcinoma gene expression) by RNA Seq data set in 1106 samples of breast cancer tissues revealed the concordant expression of SOX2OT and SOX2 in this somatic

cancer. SOX2OT and SOX2 are highly expressed in estrogen receptor positive (ER+) breast cancer cell lines, in comparison with the ER− ones. In ER+ breast cancer cell lines, expression of SOX2OT is positively correlated with SOX2 expression level, albeit at lower levels. Moreover, SOX2OT and SOX2 are co-upregulated in suspension culture conditions of breast cancer cell lines which advocates the growth of cellular subpopulation with cancer stem cell-like properties (Askarian-Amiri et al., 2014).

Overexpression of both SOX2OT and SOX2 has been reported in human primary lung cancer tissues, in comparison with the corresponding non-tumor samples. Furthermore, SOX2OT demonstrated a significant high expression level in SCC of the lung, compared with adenocarcinoma ones. There was a positive correlation between SOX2OT and SOX2 expression levels in the same lung cancer tissue samples (Hou et al., 2014).

In order to expand our knowledge of expression regulations, we reviewed some resources on gene expression profile of SOX2OT and SOX2. Exploring the expressed sequence tags (ESTs) profiles which are available from NCBI, demonstrated the expression patterns of SOX2 and its overlapping transcript in multiple pools of different human tissues and tumors. The data represent the possibility of SOX2OT and SOX2 expression in a wide list of human tissues including brain, connective tissue, esophagus, eye, intestine, kidney, lung, muscle, nerve, and testis. More

interestingly, the data hint the possibility of upregulation of SOX2OT expression in glioma and kidney tumors. In agreement with the data reported by Amaral et al. (2009) our results also revealed a high enrichment of SOX2OT expression in CNS libraries (**Figures 4A,B**). The high expression of SOX2OT and some other lncRNAs in CNS tissues suggests a potential role for them in animal brain development and function (Amaral and Mattick, 2008).

We also evaluated cancer genome anatomy project resources [Cancer Genome Anatomy Project (CGAP)] to find out a correlation between the expression signatures of *SOX2OT* with that of other genes. Based on SAGE (Serial Analysis of Gene Expression) data, SOX2OT represented a significant positive and negative correlation with multiple key genes involved in neuronal development (e.g., *LRRC4B*) addressing its function in CNS development. Furthermore, cancer associated genes (e.g., *ROCK2*, *NFKB*) are also significantly correlated with SOX2OT in SAGE libraries; which highlighted the potential function of SOX2OT in cancer progression. Noticeably, a significant positive correlation of *POU3F2* transcription factor which had multiple binding sites in genomic regulatory region of SOX2OT was observed (**Figure 4C**).

The Potential Roles of SOX2OT in Pluripotency and/or Tumorigenesis Through Regulation of SOX2 Expression

Transcription factor SOX2 regulates the expression of more than one thousand genes in stem cells where small changes of its expression strikingly alter the self-renewal and pluripotency properties; hence SOX2 acts role as a molecular *rheostat* in those cells (Boyer et al., 2005; Boer et al., 2007; Kopp et al., 2008; Amaral et al., 2009; Mandalos et al., 2014). Recent evidences have demonstrated that gene amplification and/or aberrant expression level of SOX2 play a role in the development and tumorigenesis of many types of cancer including pancreatic carcinoma, prostate, breast, lung, gastric, and esophagus cancers (Gure et al., 2000; Sattler et al., 2000; Sanada et al., 2006; Rodriguez-Pinilla et al., 2007; Chen et al., 2008; Jia et al., 2011; Hütz et al., 2013). SOX2 is also involved in the proliferation and anchorage-independent growth of esophageal and lung cell lines. SOX2-driven tumors expressed both squamous differentiation and pluripotency markers which introduced SOX2 as a lineage-survival oncogene in SCC of both lung and esophagus (Bass et al., 2009). Nevertheless, the exact regulation of SOX2 in pathway-dependent pluripotency and tumorigenesis has not been fully addressed yet.

LncRNAs have been suggested to regulate the expression of neighboring overlapped or antisense genes via different mechanisms (Mercer et al., 2009; Hung and Chang, 2010). The location of *SOX2* gene within the intronic region of *SOX2OT* gene proposed a possibility for SOX2 expression regulation by SOX2OT. This hypothesis is more approved by several experimental approaches obtained from gene expression alteration during stem cell differentiation or carcinogenesis, and also by manipulation of SOX2OT expression *in vitro* (Amaral

et al., 2009; Askarian-Amiri et al., 2014; Hou et al., 2014; Shahryari et al., 2014). Similar dynamic regulation of sox2ot transcripts and sox2 proposed a conserved role for sox2ot in vertebrate embryogenesis and neuronal system development (Amaral et al., 2009).

Using the RNA interference strategy, our group performed a functional assay on SOX2OT, where the data supported our hypothesis on the existence of a positive regulation of SOX2 and OCT4 by SOX2OT (Shahryari et al., 2014). In line with the data, Askarian-Amiri et al. (2014) demonstrated that ectopic expression of SOX2OT caused increased SOX2 expression level. They also demonstrated that the enriched suspension culture of breast cancer cells, which favors stem cell growth, exhibited upregulation of both SOX2 and SOX2OT expression, in comparison to the original adherent cells (Askarian-Amiri et al., 2014).

Furthermore, SOX2OT exerts regulatory function in cell cycle progression; hence its association with carcinogenesis of human tumors of breast (Askarian-Amiri et al., 2014), esophagus (Shahryari et al., 2014), and lung (Hussenet et al., 2010; Hou et al., 2014) cancers is not surprising. SOX2OT controls lung cancer cell proliferation, and represents a novel prognostic indicator for this cancer (Hou et al., 2014). The knocking down of SOX2OT caused induction of G2/M arrest, prohibition of S phase entry and inhibited cell proliferation which correlated with reduced protein levels of Cyclin B1 and Cdc2 in human lung cancer cell lines. SOX2OT moderated lung cancer cell cycle progression through regulating EZH2 expression level; albeit any evidence of physical interaction between them has not been observed (Hou et al., 2014). EZH2 (a histone-lysine *N*-methyltransferase enzyme) is a major component of the polycomb repressive complex 2 (PRC2) which is involved in maintaining the transcriptional repressive state of its target genes (Cardoso et al., 2000; Cao et al., 2002).

High expression levels of SOX2OT and SOX2 are associated with estrogen receptor status and tamoxifen sensitivity of breast cancer cells (Askarian-Amiri et al., 2014). SOX2OT and SOX2 co-upregulation has been reported in lung tumor tissues, particularly in squamous cell lung carcinoma (Hussenet et al., 2010; Hou et al., 2014), which is related to 3q26.33 genomic amplification (Hou et al., 2014). A statistically significant correlation coefficient between SOX2 and SOX2OT in cancer tissues (Askarian-Amiri et al., 2014; Hou et al., 2014; Shahryari et al., 2014), suggested the possibility of SOX2OT role in the regulation of SOX2 expression.

Altogether, current evidences indicate a functional association between SOX2OT and SOX2 in tumorigenesis, cellular differentiation, and pluripotency (**Table 1**). Yet, more remains to be investigated on the mechanisms underlying this regulation.

Concluding Remarks

According to recent achievements, a large number of lncRNAs primarily exert their biological functions through induction of epigenetic events including DNA methylation or histon modifications in their target genes. This is mediated by the well-known chromatin modifying complexes of PRC1 and PRC2, as well as other related complexes in a *cis*- or *trans*-

Reference	SOX2OT association with pluripotency or carcinogenesis	SOX2OT status
Amaral et al. (2009)	Vertebrate embryos	Upregulated
	Embryonic body differentiation	Downregulated
Hussenet et al. (2010)	Squamous cell carcinoma of lung with 3q.26 amplification	Upregulated
Askarian-Amiri et al. (2014)	Estrogen positive to estrogen negative breast cancer tissue	Upregulated
	Tamoxifen resistant breast cancer cell lines	Downregulated
	Suspension culture enriched breast cancer cell	Upregulated
Shahryari et al. (2014)	Esophagus squamous cell tumor to normal tissue	Upregulated
	NT2 neuronal like differentiation	Downregulated
Hou et al. (2014)	Lung cancer of squamous and adenocarcinoma to normal tissue	Upregulated

acting manner (Prensner and Chinnaiyan, 2011; Wang and Chang, 2011; Brockdorff, 2013). Multiple lncRNAs including HOTAIR, ANCR, and ANRIL are able to recruit PRC1 or PRC2 complexes to genomic regulatory regions of their target genes to reshape/regulate the chromatin state/their expression (Gupta et al., 2010; Aguilo et al., 2011; Kretz et al., 2012).

LncRNA ANRIL is involved in various mechanisms of epigenetic regulation including triggering a repression of INK4 locus by SUZ12 in PRC2 (Kotake et al., 2011), an induction of chromatin silencing of the CDKN2A/B genes through interaction with CBX7 in PRC1 (Yap et al., 2010), and an alteration of DNA methylation of the locus in differentiated cells (Yu et al., 2008). Genomic association of *SOX2* and *SOX2OT* remarkably resembles that of *ANRIL* and *CDKN2B*. Similarly, the lncRNA *ANRIL* holds the protein-coding gene *CDKN2B* in its intronic region, albeit in the antisense/opposite strand.

A brain specific lncRNA known as RMST which is involved in modulating neurogenesis physically interacts with SOX2. By acting as a transcriptional coregulator, RMST helps SOX2 to bind to regulatory regions of that of target genes which have a role in the regulation of neural stem cell fate (Ng et al., 2013). Although recent studies on SOX2OT and SOX2 have not claimed the existence of any physical interaction between them, the functional assays obtained from both knockdown and overexpression events have demonstrated that SOX2OT has a positive effect on SOX2 expression (Askarian-Amiri et al., 2014; Shahryari et al., 2014). As it was mentioned above, SOX2OT regulated the expression of EZH2 (in PRC2); however, the exact mechanism of regulation of SOX2 expression by SOX2OT mediated either by regulating PRC2 or other molecular mechanism remained largely questionable.

Several isoforms of Sox2ot which originated from alternative TSSs are associated with chromatin modifications characteristic of well-known promoters in HCEs. These isoforms have tissue or cell type specific signature, and are differentially regulated (Kimura et al., 2006; Denoeud et al., 2007; Amaral et al., 2009). This event is more prominent in SOX2DOT isoform which has a specific tissue expression pattern restricted to the adult mouse brain. SOX2DOT also demonstrates different expression patterns during differentiation of ESCs and neurospheres. The existence

of alternative splicing and alternative TSSs suggests that the different transcripts of Sox2ot might have differential regulation and function (Amaral et al., 2009).

Moreover, according to the sequences registered for SOX2OT in EST database of NCBI, it is deduced that SOX2OT could have more than three splicing variants with a unique tissue or cell type specific expression signature. Moreover, the isoform of SOX2DOT indicates a more complex splicing pattern for SOX2OT. Altogether, the overlapped expression of SOX2OT with SOX2, and the conserved association between them in different developmental systems of vertebrates, and also in human cancer and stem cells all support the existence of a complex functional regulatory relationship. The latter could be a consequence of having similar regulatory elements that regulate the expression of both Sox2ot and Sox2 (Amaral et al., 2009; Askarian-Amiri et al., 2014; Shahryari et al., 2014).

Several conserved genomic regions upstream of *SOX2OT* and *SOX2DOT* serve as the binding sites for key transcription factors responsible for controlling the pluripotency as well as tumorigenesis processes. This observation along with the observed correlations between the expression of SOX2OT variants with that of key genes promoting those events, all suggested a key role for SOX2OT in pluripotency and tumorigenesis.

In this review we have provided insights into structural characteristics, epigenetic modifications, and splicing patterns of *SOX2OT* gene. Furthermore, the expression patterns of its variants and their emerging roles in stem cell biology and tumorigenesis is discussed. It is clear that SOX2OT has a positive regulatory effect on SOX2 expression; however, the exact molecular mechanism remains to be elucidated. Specifying SOX2OT-dependent molecular pathways in organ tissue culture or engineered animal models may identify more common pathways between development, pluripotency and tumorigenesis.

In conclusion, current evidences support the idea that the lncRNA SOX2OT is a key regulatory molecule in mediating pluripotency and tumorigenesis events, probably through regulation of SOX2 expression. The positive effect of

SOX2OT upon SOX2 expression also supports a role for it in promoting generation of iPSCs. SOX2OT has a potential to be employed as a novel prognostic indicator/therapeutic target of several human cancers including breast, lung and esophagus cancers.

References

Aguilo, F., Zhou, M. M., and Walsh, M. J. (2011). Long noncoding RNA, polycomb, and the ghosts haunting INK4b-ARF-INK4a expression. *Cancer Res.* 71, 5365–5369. doi: 10.1158/0008-5472.CAN-10-4379

Amaral, P. P. and Mattick, J. S. (2008). Noncoding RNA in development. *Mamm. Genome* 19, 454–492. doi: 10.1007/s00335-008-9136-7

Amaral, P. P., Neyt, C., Wilkins, S. J., Askarian-Amiri, M. E., Sunkin, S. M., Perkins, A. C., et al. (2009). Complex architecture and regulated expression of the *Sox2ot* locus during vertebrate development. *RNA* 15, 2013–2027. doi: 10.1261/rna.1705309

Askarian-Amiri, M. E., Seyfoddin, V., Smart, C. E., Wang, J., Kim, J. E., Hansji, H., et al. (2014). Emerging role of long non-coding RNA *SOX2OT* in *SOX2* regulation in breast cancer. *PLoS ONE* 9:e102140. doi: 10.1371/journal.pone.0102140

Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* 17, 126–140. doi: 10.1101/gad.224503

Balsara, B. R., and Testa, J. R. (2002). Chromosomal imbalances in human lung cancer. *Oncogene* 21, 6877–6883. doi: 10.1038/sj.onc.1205836

Bass, A. J., Watanabe, H., Mermel, C. H., Yu, S., Perner, S., and Verhaak, R. G. (2009). SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nat. Genet.* 41, 1238–1242. doi: 10.1038/ng.465

Boer, B., Kopp, J., Mallanna, S., Desler, M., Chakravarthy, H., Wilder, P. J., et al. (2007). Elevating the levels of Sox2 in embryonal carcinoma cells and embryonic stem cells inhibits the expression of Sox2:Oct-3/4 target genes. *Nucleic Acids Res.* 35, 1773–1786. doi: 10.1093/nar/gkm059

Boumahdi, S., Driessens, G., Lapouge, G., Rorive, S., Nassar, D., Le Mercier, M., et al. (2014). SOX2 controls tumour initiation and cancer stem-cell functions in squamous-cell carcinoma. *Nature* 511, 246–250. doi: 10.1038/nature13305

Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122, 947–956. doi: 10.1016/j.cell.2005.08.020

Brockdorff, N. (2013). Noncoding RNA and Polycomb recruitment. *RNA* 19, 429–442. doi: 10.1261/rna.037598.112

Cao, R, Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., et al. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* 298, 1039–1043. doi: 10.1126/science.1076997

Cardoso, C., Mignon, C., Hetet, G., Grandchamps, B., Fontes, M., and Colleaux L. (2000). The human EZH2 gene: genomic organisation and revised mapping in 7q35 within the critical region for malignant myeloid disorders. *Eur. J. Hum. Genet.* 8, 174–180. doi: 10.1038/sj.ejhg.5200439

Chen, Y., Shi, L., Zhang, L., Li, R., Liang, J., Yu, W., et al. (2008). The molecular mechanism governing the oncogenic potential of SOX2 in breast cancer. *J. Biol. Chem.* 283, 17969–17978. doi: 10.1074/jbc.M802917200

Denoeud, F., Kapranov, P., Ucla, C., Frankish, A., Castelo, R., Drenkow, J., et al. (2007). Prominent use of distal 5′ transcription start sites and discovery of a large number of additional exons in ENCODE regions. *Genome Res.* 17, 746–759. doi: 10.1101/gr.5660607

Dinger, M. E., Pang, K. C., Mercer, T. R., and Mattick, J. S. (2008). Differentiating protein-coding and noncoding RNA: challenges and ambiguities. *PLoS Comput. Biol.* 4:e1000176. doi: 10.1371/journal.pcbi.1000176

Esteller, M. (2011). Non-coding RNAs in human disease. *Nat. Rev.* 12, 861–874. doi: 10.1038/nrg3074

Fantes, J., Ragge, N. K., Lynch, S. A., McGill, N. I., Collin, J. R., Howard-Peebles, P. N., et al. (2003). Mutations in SOX2 cause anophthalmia. *Nat. Genet.* 33, 461–463. doi: 10.1038/ng1120

Fong, H., Hohenstein, K. A., and Donovan, P. J. (2008). Regulation of self-renewal and pluripotency by Sox2 in human embryonic stem cells. *Stem Cells* 26, 1931–1938. doi: 10.1634/stemcells.2007-1002

Gebhart, E., and Liehr, T. (2000). Patterns of genomic imbalances in human solid tumors. *Int. J. Oncol.* 162, 383–399. doi: 10.3892/ijo.16.2.383

Gupta, R. A., Shah, N., Wang, K. C., Kim, J., Horlings, H. M., and Wong, D. J. (2010). Long non-coding RNA *HOTAIR* reprograms chromatin state to promote cancer metastasis. *Nature* 464, 1071–1076. doi: 10.1038/nature08975

Gure, A., Stockert, E., Scanlan, M., Keresztes, R., Jager, D., Altorki, N. K., et al. (2000). Serological identification of embryonic neural proteins as highly immunogenic tumor antigens in small cell lung cancer. *Proc. Natl. Acad. Sci. U.S.A.* 97, 4198–4203. doi: 10.1073/pnas.97.8.4198

Gutschner, T., Hämmerle, M., Eissmann, M., Hsu, J., Kim, Y., Hung, G., et al. (2013). The noncoding RNA *MALAT1* is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res.* 73, 1180–1189. doi: 10.1158/0008-5472.CAN-12-2850

Guttman, M., Amit, I., Garber, M., French, C., Lin, M. F., Feldser, D., et al. (2009). Chromatin signature reveals over a thousand highly conserved large noncoding RNAs in mammals. *Nature* 458, 223–227. doi: 10.1038/nature07672

Hou, Z., Zhao, W., Zhou, J., Shen, L., Zhan, P., Xu, C., et al. (2014). A long noncoding RNA Sox2ot regulates lung cancer cell proliferation and is a prognostic indicator of poor survival. *Int. J. Biochem. Cell Biol.* 53, 380–388. doi: 10.1016/j.biocel.2014.06.004

Hung, T., and Chang, H. Y. (2010). Long noncoding RNA in genome regulation: prospects and mechanisms. *RNA Biol.* 7, 582–585. doi: 10.4161/rna.7.5.13216

Hussenet, T., Dali, S., Exinger, J., Monga, B., Jost, B., Dembele, D., et al. (2010). SOX2 is an oncogene activated by recurrent 3q26.3 amplifications in human lung squamous cell carcinomas. *PLoS ONE* 5:e8960. doi: 10.1371/journal.pone.0008960

Hütz, K., Mejías-Luque, R., Farsakova, K., Ogris, M., Krebs, S., Anton, M., et al. (2013). The stem cell factor SOX2 regulates the tumorigenic potential in human gastric cancer cells. *Carcinogenesis* 35, 942–950. doi: 10.1093/carcin/bgt410

Jia, X., Li, X., Xu, Y., Zhang, S., Mou, W., Liu, Y., et al. (2011). SOX2 promotes tumorigenesis and increases the anti-apoptotic property of human prostate cancer cell. *J. Mol. Cell Biol.* 3, 230–238. doi: 10.1093/jmcb/mjr002

Jiang, F., Yin, Z., Caraway, N. P., Li, R., and Katz, R. L. (2004). Genomic profiles in stage I primary non small cell lung cancer using comparative genomic hybridization analysis of cDNA microarrays. *Neoplasia* 6, 623–635. doi: 10.1593/neo.04142

Kapranov, P., Cheng, J., Dike, S., Nix, D. A., Duttagupta, R., Willingham, A. T., et al. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* 316, 1484–1488. doi: 10.1126/science.1138341

Kimura, K., Wakamatsu, A., Suzuki, Y., Ota, T., Nishikawa, T., Yamashita, R., et al. (2006). Diversification of transcriptional modulation: largescale identification and characterization of putative alternative promoters of human genes. *Genome Res.* 16, 55–65. doi: 10.1101/gr.4039406

Kopp, J. L., Ormsbee, B. D., Desler, M., and Rizzino, A. (2008). Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. *Stem Cells* 26, 903–911. doi: 10.1634/stemcells.2007-0951

Kotake, Y., Nakagawa, T., Kitagawa, K., Suzuki, S., Liu, N., Kitagawa, M., et al. (2011). Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15^{INK4B} tumor suppressor gene. *Oncogene* 30, 1956–1962. doi: 10.1038/onc.2010.568

Kretz, M., Webster, D. E., Flockhart, R. J., Lee, C. S., Zehnder, A., Lopez-Pajares, V., et al. (2012). Suppression of progenitor differentiation requires the long noncoding RNA ANCR. *Gene Dev.* 26, 338–343. doi: 10.1101/gad.182121.111

Loewer, S., Cabili, M. N., Guttman, M., Loh, Y. H., Thomas, K., Park, I. H., et al. (2010). Large intergenic non-coding RNA-RoR modulates reprogramming of human. *Nat. Genet.* 42, 1113–1117. doi: 10.1038/ng.710

Mandalos, N., Rhinn, M., Granchi, Z., Karampelas, I., Mitsiadis, T., Economides, A. N., et al. (2014). Sox2 acts as a rheostat of epithelial to mesenchymal transition during neural crest development. *Front. Physiol.* 5:345. doi: 10.3389/fphys.2014.00345

Massion, P. P., Kuo, W. L., Stokoe, D., Olshen, A. B., Treseler, P. A., Chin, K., et al. (2002). Genomic copy number analysis of non-small cell lung cancer using array comparative genomic hybridization: implications of the phosphatidylinositol 3-kinase pathway. *Cancer Res.* 62, 3636–3640.

Meng, L., Ward, A. J., Chun, S., Bennett, C. F., Beaudet, A. L., and Rigo, F. (2015). Towards a therapy for Angelman syndrome by targeting a long non-coding RNA. *Nature* 518, 409–412. doi: 10.1038/nature13975

Mercer, T. R., Dinger, M. E., and Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* 10, 155–159. doi: 10.1038/nrg2521

Mercer, T. R., Dinger, M. E., Sunkin, S. M., Mehler, M. F., and Mattick, J. S. (2008). Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl. Acad. Sci. U.S.A.* 105, 716–721. doi: 10.1073/pnas.0706729105

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., et al. (2007). Genomewide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448, 553–560. doi: 10.1038/nature06008

Ng, S. Y., Bogu, G. K., Soh, B. S., and Stanton, L. W. (2013). The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis. *Mol. Cell* 51, 349–359. doi: 10.1016/j.molcel.2013.07.017

Ng, S. Y., Johnson, R., and Stanton, L. W. (2011). Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO J.* 31, 522–533. doi: 10.1038/emboj.2011.459

Numata, K., Kanai, A., Saito, R., Kondo, S., Adachi, J., Wilming, L. G., et al. (2003). Identification of putative noncoding RNAs among the RIKEN mouse full-length cDNA collection. *Genome Res.* 13, 1301–1306. doi: 10.1101/gr.1011603

Perez, D. S., Hoage, T. R., Pritchett, J. R., Ducharme-Smith, A. L., Halling, M. L., Ganapathiraju, S. C., et al. (2007). Long, abundantly expressed non-coding transcripts are altered in cancer. *Hum. Mol. Genet.* 17, 642–655. doi: 10.1093/hmg/ddm336

Prensner, J. R., and Chinnaiyan, A. M. (2011). The emergence of lncRNAs in cancer biology. *Cancer Discov.* 1, 391–407. doi: 10.1158/2159-8290.CD-11-0209

Rodriguez-Pinilla, S., Sarrio, D., Moreno-Bueno, G., Rodriguez-Gil, Y., Martinez, M. A., Hernandez, L., et al. (2007). Sox2: a possible driver of the basal-like phenotype in sporadic breast cancer. *Mod. Pathol.* 20, 474–481. doi: 10.1038/modpathol.3800760

Sanada, Y., Yoshida, K., Ohara, M., Oeda, M., Konishi, K., and Tsutani Y. (2006). Histopathologic evaluation of stepwise progression of pancreatic carcinoma with immunohistochemical analysis of gastric epithelial transcription factor SOX2: comparison of expression patterns between invasive components and cancerous or nonneoplastic intraductal components. *Pancreas* 32, 164–170. doi: 10.1097/01.mpa.0000202947.80117.a0

Sattler, H., Lensch, R., Rohde, V., Zimmer, E., Meese, E., Bonkhoff, H., et al. (2000). Novel amplification unit at chromosome 3q25–q27 in human prostate cancer. *Prostate* 45, 207–215. doi: 10.1002/1097-0045(20001101)45:3<207::AID-PROS2>3.0.CO;2-H

Shahryari, A., Rafiee, M. R., Fouani, Y., Oliae, N. A., Samaei, N. M., Shafiee, M., et al. (2014). Two novel splice variants of SOX2OT, SOX2OT-S1, and SOX2OT-S2 are coupregulated with SOX2 and OCT4 in esophageal squamous cell carcinoma. *Stem Cells* 32, 126–134. doi: 10.1002/stem.1542

Simons, C., Pheasant, M., Makunin, I. V., and Mattick, J. S. (2006). Transposon free regions in mammalian genomes. *Genome Res.* 16, 164–172. doi: 10.1101/gr.4624306

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., et al. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131, 861–872. doi: 10.1016/j.cell.2007.11.019

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663–676. doi: 10.1016/j.cell.2006.07.024

Wang, K. C., and Chang, H. Y. (2011). Molecular mechanisms of long noncoding RNAs. *Mol. Cell* 43, 904–914. doi: 10.1016/j.molcel.2011.08.018

Yap, K. L., Li, S., Munoz-Cabello, A. M., Raguz, S., Zeng, L., Mujtaba, S., et al. (2010). Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol. Cell.* 38, 662–674. doi: 10.1016/j.molcel.2010.03.021

Yu, W., Gius, D., Onyango, P., Muldoon-Jacobs, K., Karp, J., Feinberg, A. P., et al. (2008). Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* 451, 202–206. doi: 10.1038/nature06468

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Genetics

Exosomal lncRNA-p21 levels may help to distinguish prostate cancer from benign disease

Mustafa Işın[1], Ege Uysaler[1], Emre Özgür[1], Hikmet Köseoğlu[2], Öner Şanlı[3], Ömer B. Yücel[2], Uğur Gezer[1] and Nejat Dalay[1]*

[1] Department of Basic Oncology, Oncology Institute, Istanbul University, Istanbul, Turkey, [2] Department of Urology, School of Medicine, Istanbul Hospital, Başkent University, Istanbul, Turkey, [3] Department of Urology, Istanbul Medical Faculty, Istanbul University, Istanbul, Turkey

Exosomes are membranous vesicles containing various biomolecules including lncRNAs which are involved in cellular communication and are secreted from many cells including cancer cells. In our study, investigated the exosomal GAS5 and lincRNA-p21 lncRNA levels in urine samples from 30 patients with prostate cancer (PCa) and 49 patients with benign prostatic hyperplasia. Quantification of lncRNA molecules was performed by real-time PCR. We observed a significant difference in the exosomal lincRNA-p21 levels between PCa and BPH patients whereas the GAS5 levels did not reveal a difference. Our data suggest that the discriminative potential of exosomal lincRNA-p21 levels may help to improve the diagnostic prediction of the malignant state for patients with PCa.

Keywords: prostate cancer, exosome, lncRNA, benign prostatic hyperplasia, non-invasive diagnosis

Introduction

Prostate cancer (PCa) is the second most common malignancy and accounts for 15% of the cancers in men. It is the fifth leading cause of death and almost 70% of the cases occur in the developed countries (Ferlay et al., 2015). The diagnosis of PCa is performed via histopathological evaluation of biopsy samples which has several known disadvantages like bleeding and infection (Loeb et al., 2013). Although use of the prostate specific antigen (PSA) as a diagnostic marker has improved the detection and management of PCa (Lippi et al., 2009) its low specificity and lack of other predictive parameters for the progression of the disease makes the stratification of the patients with high risk or indolent PCa difficult (Roddam et al., 2005; Nogueira et al., 2010).

Recently, a prostate specific lncRNA, the PCa Gene 3 (PCA3), has been approved as an additional test to determine the need for biopsies in PCa. Unfortunately, negative PCA3 results in indolent cancer carriers and high grade prostatic intraepithelial neoplasia (HGPIN; Morote et al., 2010; Auprich et al., 2011a) renders the biomarker insufficient. The level of TMPRSS2-ERG fusion transcripts, in cells collected from urine after digital examination, have also been evaluated together with the PCA3 levels for a more accurate diagnosis (Leyten et al., 2014). However, none of these approaches can satisfactorily distinguish high-risk from indolent cancer. Therefore identification of new biomarkers to correctly identify patients needing more aggressive treatment would help to prevent individuals with localized tumors from getting unnecessary biopsies and from the side effects of overtreatment.

Exosomes are small membranous vesicles originating from the endosomal compartment which function as messengers in intercellular communication (Simpson et al., 2008; Bang and Thum, 2012; Gezer et al., 2014). They are secreted and released by cells and bind to the receptors on recipient cells, thereby transferring the signal. Exosome secretion is an evolutionary conserved cellular

mechanism dating back to Archaea but their role as cellular messengers has been described only recently (Taylor et al., 2006; Deatherage and Cookson, 2012). Exosomes contain proteins and various types of RNA molecules including lncRNAs. Non-exonic transcripts as a whole constitute the majority of non-ribosomal RNA molecules in the cell and lincRNAs constitute a significant portion of this fraction (Kapranov et al., 2010). Exosomes secreted from prostate can be detected in semen (Ronquist and Brody, 1985) and urine (Pisitkun et al., 2004). In recent years it has been shown that PCA3 and several other microRNAs are present in exosomes isolated from PCa patients (Dijkstra et al., 2014; Huang et al., 2015).

LincRNA-p21 and GAS5 lncRNA act as tumor suppressor molecules in the cellular machinery (Schneider et al., 1988; Huarte et al., 2010). Expression of lincRNA-p21 is stimulated by the p53 tumor suppressor protein and upon transcription it suppresses expression of the genes transcriptionally regulated by p53 by binding to the hnRNP-K complex (Huarte et al., 2010). GAS5 plays a role in the induction of apoptosis. It suppresses several antiapoptotic genes by binding to the glucocorticoid receptor (GR) and hence prevents GR from binding to the glucocorticoid response elements on the target DNA molecule (Kino et al., 2010). Recent data indicate that lncRNA molecules may exhibit tissue- and disease-specific expression which can provide important potential biomarkers specific to the particular cancer types (Yaman Agaoglu et al., 2011; Gezer et al., 2014). However, it should be noted that lncRNA molecules are associated both with cancer and pluripotency which can be a confounding factor (St. Laurent et al., 2013). In this study, we aimed to evaluate the diagnostic utility of exosomal lincRNA-p21 and GAS5 levels in individuals with benign prostatic hyperplasia (BPH) and PCa.

Materials and Methods

Exosomal RNA Isolation

Thirty patients with PCa (median age 64 ± 6) and 49 patients with BPH (median age 68 ± 9) were enrolled in the study. The study was approved by the Ethics Committee of the Istanbul Faculty of Medicine and informed consent was obtained from the participants. All patients were Caucasian and the disease scores of the patients are given in **Table 1**. Urine samples from patients were collected after digital rectal examination and were centrifuged for 10 min at 1000 rpm in a conical tube to remove cell and debris. Supernatant was transferred to another conical tube and centrifuged at 2000 rpm for 10 min to remove remaining debris and bacteria. 10 ml of cell-free urine was then transferred to a new 15 ml tube and stored at $-80°C$ until use. Exosomal RNA was extracted according to the manufacturer's protocol using the "Urine Exosome RNA Isolation Kit" (Norgen Biotek, Thorold, ON, Canada).

Quantification of lncRNAs

Exosomal RNA isolated from urine samples were used for cDNA synthesis using the First Strand cDNA Synthesis kit (Thermo Scientific, West Palm Beach, FL, USA) according to the manufacturers' instructions. The real-time amplification of lncRNA molecules was performed using the LightCycler 480 system

TABLE 1 | Disease scores of the patients.

Benign prostatic hyperplasia patients	*n* = 49
IPSS Score (Barry et al., 1992)	
0–7	0
8–19	24
20–35	17
Unknown	8

Prostate cancer patients	*n* = 30
Gleason score (Epstein, 2010)	
6 (3+3)	19
7 (3+4)	3
7 (4+3)	4
Unknown	4

TABLE 2 | The primer sequences used in the study.

Gene	Primer sequence	Reference
LincRNA-p21	F 5′- GGGTGGCTCACTCTTCTGGC -3′ R 5′- TGGCCTTGCCCGGGCTTGTC -3′	Huarte et al. (2010)
GAS5	F 5′-CTTCTGGGCTCAAGTGATCCT -3′ R 5′- TTGTGCCATGAGACTCCATCAG -3′	Khalil et al. (2009)
GAPDH	F 5′-GCTCTCTGCTCCTCCTGTTC -3′ R 5′- ACGACCAAATCCGTTGACTC-3′	Khalil et al. (2009)

(Roche, Germany). SYBR Green (Roche) was used as the fluorescent molecule. The primer sequences are shown in **Table 2**.

The PCR reaction included an initial "hot start" for 10 min., followed by 45 cycles of amplification. Each cycle consisted of a denaturation step at 95°C for 10 s, annealing starting at 60°C for 20 s and decreasing by 2°C every two cycles down to 55°C, and amplification at 72°C for 30 s. For quantification of lncRNAs, the $\Delta\Delta Ct$ method was used. The GAPDH gene was used as the reference. All experiments were performed twice and the mean values were calculated.

Statistical Analyses

SPSS® ver.21.0 statistical program was used for statistical analysis. Kruskal–Wallis or Mann–Whitney U tests were used when appropriate to compare the parameters (lncRNA and PSA levels). A p-value ≤ 0.05 was considered statistically significant.

Results

Exosomal levels of GAS5 and lncRNA-p21 lncRNAs were evaluated in the urine samples of 49 patients diagnosed with BPH and 30 patients with PCa. The distribution of exosomal lncRNA levels are shown in **Figure 1**. The lincRNA-p21 levels were significantly higher in PCa than in BPH (median; 0.163 vs 0.071; $p = 0.016$, AUC: 0.663). Exosomal GAS5 levels were found to be similar in the two disease groups (median; 1.197 vs 1.235 and $p = 0.127$). The data are shown in **Table 3**.

The PSA levels were higher in the PCa group than in the patients with BPH ($p < 0.001$, median values; 7.7 and 2.16 respectively). In the BPH group no correlation was observed between the
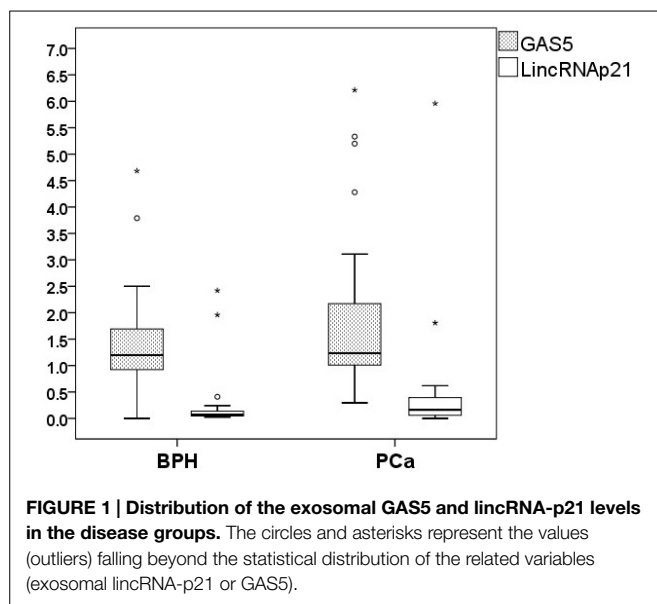
FIGURE 1 | Distribution of the exosomal GAS5 and lincRNA-p21 levels in the disease groups. The circles and asterisks represent the values (outliers) falling beyond the statistical distribution of the related variables (exosomal lincRNA-p21 or GAS5).

TABLE 3 | Median values of exosomal lncRNAs and their statistical significances.

Median	BPH	PCa	Significance (p-value)
GAS5	1.197	1.235	0.127
LincRNA-p21	0.071	0.163	0.016

AUC value of LincRNA-p21: 0.663, CI:95%

FIGURE 2 | The ROC curve for the LincRNA-p21 and PSA combination.

TABLE 4 | Sensitivity and specificity of exosomal lncRNA-p21 and of the lincRNA-p21/PSA combination.

	LincRNA-p21	LincRNA-p21 and PSA
Sensitivity	67%	52%
Specificity	63%	94%

IPSS (International Prostate Symptom Score) score and GAS5 or lincRNA-p21 levels. Likewise, no correlation between the clinical stage (Gleason score) and exosomal lncRNA levels was observed in the PCa group. There was no correlation between the PSA levels and the IPSS score in patients with BPH but a correlation was observed between the PSA values and the Gleason score in the PCa group ($p = 0.123$ and 0.049, respectively).

The sensitivity and specificity of lincRNA-p21 and lincRNA-p21 in combination with PSA were calculated using a cut off value of 2.5 ng/ml for PSA and 0.181 for exosomal lincRNA-p21 expression (Figure 2). The specificity for predicting PCa increased from 63 to 94% when the two parameters were combined while the specificity did not change (Table 4).

Discussion

Our study is the first report, revealing presence of lincRNA-p21 and GAS5 lncRNA molecules in the exosomes derived from urine samples. Circulating GAS5 and lincRNA-p21 have been previously detected in B-cell malignancies (Isin et al., 2014) but exosomal GAS5 and lincRNA-p21 molecules have previously been only reported in exosomes secreted from HeLa and MCF-7 cell lines but not in human tumors (Gezer et al., 2014). Cellular lincRNA-p21 expression has been suggested to affect global gene expression in different cancers by modulating mRNA translation and supressing the p53 and Wnt/β-catenin signaling pathways (Dimitrova et al., 2014; Wang et al., 2014). Cellular GAS5

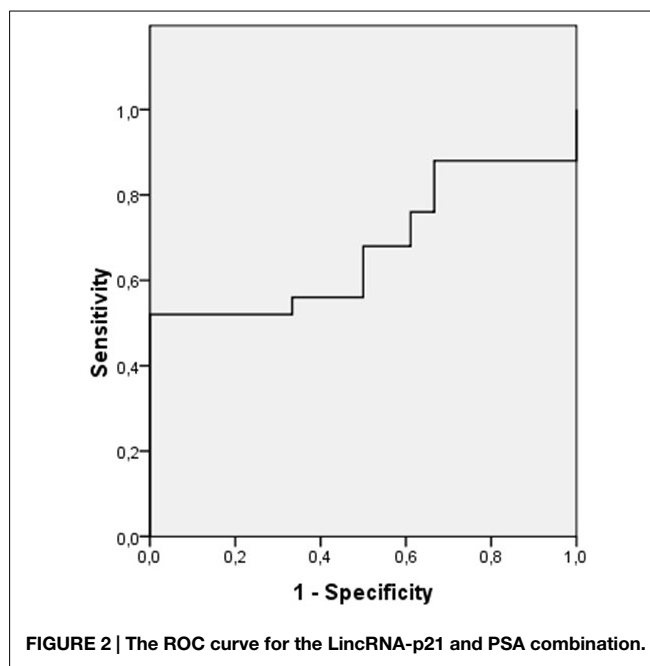expression suggest a role of GAS5 in the regulation of apoptosis in breast cancer cell lines and tumors (Pickard and Williams, 2014) and an inverse association with the mTOR expression in PCa cell lines (Yacqub-Usman et al., 2015). In the present study, we observed significantly higher levels of exosomal lincRNA-p21 levels in the patients with PCa.

It has been shown that one in six of the prostatectomy specimens may contain indolent cancers which usually do not progress to clinical detection during the lifetime of the patient (Epstein et al., 1994). On the other hand, several studies suggest that more than 50% of the cancers which are initially diagnosed as localized tumors are actually advanced at the time of treatment (Chun et al., 2008; Jeldres et al., 2008).

The PSA test is not sensitive enough to predict the presence, extent and risk of recurrence of PCa (Albertsen, 2010; Zeliadt et al., 2010; Friedrich, 2011). Therefore, there is a definite need for non-invasive diagnostic biomarkers which can distinguish the low- and high-risk patients in the clinical decision-making. Even though non-invasive detection of PCA3 levels in urine after digital rectal examination may provide some useful information on the need of a repeated biopsy, evaluation of PCA3 also fails to be specific (Bradley et al., 2013). Dijkstra et al. (2014) have shown that the diagnostic performance of PCA3 in exosomes was found to be more successful than in urine where the PCA3 score was normalized with PSA mRNA in order to achieve a higher performance. Although their study group is quite small the authors suggested an advantage of exosomal PCA3 evaluation

while indicating that their data should be validated in larger clinical cohorts. The discriminative capacity achieved by the exosomal PCA3 in this study (AUC: 0.524) is lower than lincRNA-p21 (AUC: 0.663, CI: 95%) in the present study. These data indicate that analysis of exosomal lincRNA-p21 in urine performs better than PCA3 in detecting PCa.

Recently, there is an active controversy over decreasing the cutoff level of PCA3 from 35 to 25 (Nakanishi et al., 2008; Auprich et al., 2011b; de la Taille et al., 2011; Ploussard et al., 2011) which is expected to increase its diagnostic sensitivity. A meta-analysis by Luo et al. (2014) reported that the sensitivity of PCA3 for detecting PCa ranges from 46.9 to 82.3% and specificity from 55 to 92%. In our study the specificity of lincRNA-p21 for PCa was 94% when combined with PSA.

Expression of TMPRSS2-ERG has also been analyzed in urine samples (Leyten et al., 2014). It has been reported that combination of PCA3 and TMPRSS2-ERG expression increased the sensitivity of detecting PCa. However, the biomarker pair still failed to detect indolent tumors with a Gleason score of ≥ 7.

A new study investigating exosomal microRNA molecules derived from plasma samples in 29 castration resistant PCa (CRPCa) patients reported two significant miRNA molecules (miR-1290 and miR-375) which were later evaluated in 100 patients with CRPCa, as a prognostic marker significantly associated with poor prognosis, needing prospective validation (Huang et al., 2015).

In absence of reliable markers for detection and classification of PCa urinary exosomal lncRNAs may can provide an alternative, and non-invasive source of biomarkers. The stability and longevity of the RNA molecules is ideal for non-invasive diagnosis and characterization of the tumors. Our study for the first time demonstrates that detection of exosomal lncRNAs in urine may act as suitable biomarkers with potential utility of therapeutic implications. LincRNA-p21 provides a promising marker with therapeutic potential for the detection and stratification of PCa. Further studies with larger patient groups are needed to validate the therapeutic utility of exosomal lincRNA-p21 levels in urine.

References

Albertsen, P. C. (2010). Efficacy vs effectiveness in prostate-specific antigen screening. *J. Natl. Cancer Inst.* 102, 288–289. doi: 10.1093/jnci/djq002

Auprich, M., Bjartell, A., Chun, F. K., de la Taille, A., Freedland, S. J., Haese, A., et al. (2011a). Contemporary role of prostate cancer antigen 3 in the management of prostate cancer. *Eur. Urol.* 60, 1045–1054. doi: 10.1016/j.eururo.2011.08.003

Auprich, M., Chun, F. K., Ward, J. F., Pummer, K., Babaian, R., Agustin, H., et al. (2011b). Critical assessment of preoperative urinary prostate cancer antigen 3 on the accuracy of prostate cancer staging. *Eur. Urol.* 59, 96–105. doi: 10.1016/j.eururo.2010.10.024

Bang, C., and Thum, T. (2012). Exosomes: new players in cell-cell communication. *Int. J. Biochem. Cell Biol.* 44, 2060–2064. doi: 10.1016/j.biocel.2012.08.007

Barry, M. J., Fowler, F. J Jr, O'Leary, M. P., Bruskewitz, R. C., Holtgrewe, H. L., Mebust, W. K., et al. (1992). The American Urological Association symptom index for benign prostatic hyperplasia. The Measurement Committee of the American Urological Association. *J. Urol.* 148, 1549–1557; discussion 1564.

Bradley, L. A., Palomaki, G. E., Gutman, S., Samson, D., and Aronson, N. (2013). Comparative effectiveness review: prostate cancer antigen 3 testing for the diagnosis and management of prostate cancer. *J. Urol.* 190, 389–398. doi: 10.1016/j.juro.2013.02.005

Chun, F. K., Haese, A., Ahyai, S. A., Walz, J., Suardi, N., Capitanio, U., et al. (2008). Critical assessment of tools to predict clinically insignificant prostate cancer at radical prostatectomy in contemporary men. *Cancer* 113, 701–709. doi: 10.1002/cncr.23610

de la Taille, A., Irani, J., Graefen, M., Chun, F., de Reijke, T., Kil, P., et al. (2011). Clinical evaluation of the PCA3 assay in guiding initial biopsy decisions. *J. Urol.* 185, 2119–2125. doi: 10.1016/j.juro.2011.01.075

Deatherage, B. L., and Cookson, B. T. (2012). Membrane vesicle release in bacteria, eukaryotes, and archaea: a conserved yet underappreciated aspect of microbial life. *Infect. Immun.* 80, 1948–1957. doi: 10.1128/IAI.06014-11

Dijkstra, S., Birker, I. L., Smit, F. P., Leyten, G. H., de Reijke, T. M., van Oort, I. M., et al. (2014). Prostate cancer biomarker profiles in urinary sediments and exosomes. *J. Urol.* 191, 1132–1138. doi: 10.1016/j.juro.2013.11.001

Dimitrova, N., Zamudio, J. R., Jong, R. M., Soukup, D., Resnick, R., Sarma, K., et al. (2014). LincRNA-p21 activates p21 in cis to promote Polycomb target gene expression and to enforce the G1/S checkpoint. *Mol. Cell* 54, 777–790. doi: 10.1016/j.molcel.2014.04.025

Epstein, J. I. (2010). An update of the Gleason grading system. *J. Urol.* 183, 433–440. doi: 10.1016/j.juro.2009.10.046

Epstein, J. I., Walsh, P. C., Carmichael, M., and Brendler, C. B. (1994). Pathologic and clinical findings to predict tumor extent of nonpalpable (stage T1c) prostate cancer. *JAMA* 271, 368–374. doi: 10.1001/jama.1994.03510290050036

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., et al. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* 136, E359–E386. doi: 10.1002/ijc.29210

Friedrich, M. J. (2011). Debate continues on use of PSA testing for early detection of prostate cancer. *JAMA* 305, 2273–2275. doi: 10.1001/jama.2011.777

Gezer, U., Özgür, E., Cetinkaya, M., Isin, M., and Dalay, N. (2014). Long non-coding RNAs with low expression levels in cells are enriched in secreted exosomes. *Cell Biol. Int.* 38, 1076–1079. doi: 10.1002/cbin.10301

Huang, X., Yuan, T., Liang, M., Du, M., Xia, S., Dittmar, R., et al. (2015). Exosomal miR-1290 and miR-375 as prognostic markers in castration-resistant prostate cancer. *Eur. Urol.* 67, 33–41. doi: 10.1016/j.eururo.2014.07.035

Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M. J., Kenzelmann-Broz, D., et al. (2010). A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* 142, 409–419. doi: 10.1016/j.cell.2010.06.040

Isin, M., Ozgur, E., Cetin, G., Erten, N., Aktan, M., Gezer, U., et al. (2014). Investigation of circulating lncRNAs in B-cell neoplasms. *Clin. Chim. Acta.* 431, 255–259. doi: 10.1016/j.cca.2014.02.010

Jeldres, C., Suardi, N., Walz, J., Hutterer, G. C., Ahyai, S., Lattouf, J. B., et al. (2008). Validation of the contemporary epstein criteria for insignificant prostate cancer in European men. *Eur. Urol.* 54, 1306–1313. doi: 10.1016/j.eururo.2007.11.057

Kapranov, P., St Laurent, G., Raz, T., Ozsolak, F., Reynolds, C. P., and Sorensen, P. H. B. (2010). The majority of total nuclear-encoded non-ribosomal RNA in a human cell is 'dark matter' un-annotated RNA. *BMC Biol.* 8:149. doi: 10.1186/1741-7007-8-149

Khalil, A. M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11667–11672. doi: 10.1073/pnas.0904715106

Kino, T., Hurt, D. E., Ichijo, T., Nader, N., and Chrousos, G. P. (2010). Noncoding RNA GAS5 is a growth arrest and starvation-associated repressor of the glucocorticoid receptor. *Sci. Signal.* 3, 8. doi: 10.1126/scisignal.2000568

Leyten, G. H., Hessels, D., Jannink, S. A., Smit, F. P., de Jong, H., Cornel, E. B., et al. (2014). Prospective multicentre evaluation of PCA3 and TMPRSS2-ERG gene fusions as diagnostic and prognostic urinary biomarkers for prostate cancer. *Eur. Urol.* 65, 534–542. doi: 10.1016/j.eururo.2012.11.014

Lippi., G, Montagnana, M., Guidi, G. C., and Plebani, M. (2009). Prostate specific antigen-based screening for prostate cancer in the third millennium: useful or hype? *Ann. Med.* 41, 480–489. doi: 10.1080/07853890903156468

Loeb, S., Vellekoop, A., Ahmed, H. U., Catto, J., Emberton, M., Nam, R., et al. (2013). Systematic review of complications of prostate biopsy. *Eur. Urol.* 64, 876–892. doi: 10.1016/j.eururo.2013.05.049

Luo, Y., Gou, X., Huang, P., and Mou, C. (2014). Prostate cancer antigen 3 test for prostate biopsy decision: a systematic review and meta analysis. *Chin. Med. J.* 127, 1768–1774.

Morote, J., Rigau, M., Garcia, M., Mir, C., Ballesteros, C., Planas, J., et al. (2010). Behavior of the PCA3 gene in the urine of men with high grade prostatic intraepithelial neoplasia. *World J. Urol.* 28, 677–680. doi: 10.1007/s00345-010-0580-0

Nakanishi, H., Groskopf, J., Fritsche, H. A., Bhadkamkar, V., Blase, A., Kumar, S. V., et al. (2008). PCA3 molecular urine assay correlates with prostate cancer tumor volume: implication in selecting candidates for active surveillance. *J. Urol.* 179, 1804–1809; discussion 1809–1810. doi: 10.1016/j.juro.2008.01.013

Nogueira, L., Corradi, R., and Eastham, J. A. (2010). Other biomarkers for detecting prostate cancer. *BJU Int.* 105, 166–169. doi: 10.1111/j.1464-410X.2009.09088.x

Pickard, M. R., and Williams, G. T. (2014). Regulation of apoptosis by the long non-coding RNA GAS5 in breast cancer cells: implications for chemotherapy. *Breast Cancer Res. Treat.* 145, 359–370. doi: 10.1007/s10549-014-2974-y

Pisitkun, T., Shen, R. F., and Knepper, M. A. (2004). Identification and proteomic profiling of exosomes in human urine. *Proc. Natl. Acad. Sci. U.S.A.* 101, 13368–13373. doi: 10.1073/pnas.0403453101

Ploussard, G., Durand, X., Xylinas, E., Moutereau, S., Radulescu, C., Forque, A., et al. (2011). Prostate cancer antigen 3 score accurately predicts tumour volume and might help in selecting prostate cancer patients for active surveillance. *Eur. Urol.* 59, 422–429. doi: 10.1016/j.eururo.2010.11.044

Roddam, A. W., Duffy, M. J., Hamdy, F. C., Ward, A. M., Patnick, J., Price, C. P., et al. (2005). NHS Prostate Cancer Risk Management Programme. Use of prostate-specific antigen (PSA) isoforms for the detection of prostate cancer in men with a PSA level of 2-10 ng/ml: systematic review and meta-analysis. *Eur. Urol.* 48, 386–399. doi: 10.1016/j.eururo.2005.04.015

Ronquist, G., and Brody, I. (1985). The prostasome: its secretion, and function in man. *Biochim. Biophys. Acta Rev. Biomembr.* 822, 203–218. doi: 10.1016/0304-4157(85)90008-5

Schneider, C., King, R. M., and Philipson, L. (1988). Genes specifically expressed at growth arrest of mammalian cells. *Cell* 54, 787–793. doi: 10.1016/S0092-8674(88)91065-3

Simpson, R. J., Jensen, S. S., and Lim, J. W. (2008). Proteomic profiling of exosomes: current perspectives. *Proteomics* 8, 4083–4099. doi: 10.1002/pmic.200800109

St. Laurent, G., Shtokalo, D., Dong, B., Tackett, M. R., Fan, X., Lazorthes, S., et al. (2013). VlincRNAs controlled by retroviral elements are a hallmark of pluripotency and cancer. *Genome Biol.* 14, R73. doi: 10.1186/gb-2013-14-7-r73

Taylor, D. D., Akyol, S., and Gercel-Taylor, C. (2006). Pregnancy-associated exosomes and their modulation of T cell signaling. *J. Biol. Chem.* 176, 1534–1542. doi: 10.4049/jimmunol.176.3.1534

Wang, G., Li, Z., Zhao, Q., Zhu, Y., Zhao, C., Li, X., et al. (2014). LincRNA-p21 enhances the sensitivity of radiotherapy for human colorectal cancer by targeting the Wnt/β-catenin signaling pathway. *Oncol. Rep.* 31, 1839–1845. doi: 10.3892/or.2014.3047

Yacqub-Usman, K., Pickard, M. R., and Williams, G. T. (2015). Reciprocal regulation of GAS5 lncRNA levels and mTOR inhibitor action in prostate cancer cells. *Prostate* 75, 693–705. doi: 10.1002/pros.22952

Yaman Agaoglu, F., Kovancilar, M., Dizdar, Y., Darendeliler, E., Holdenrieder, S., Dalay, N., et al. (2011). Investigation of miR-21, miR-141, and miR-221 in blood circulation of patients with prostate cancer. *Tumour Biol.* 32, 583–588. doi: 10.1007/s13277-011-0154-9

Zeliadt, S. B., Hoffman, R. M., Etzioni, R., Ginger, V. A., and Lin, D. W. (2010). What happens after an elevated PSA test: the experience of 13, 591 veterans. *J. Gen. Intern. Med.* 25, 1205–1210. doi: 10.1007/s11606-010-1468-9

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read, for greatest visibility

COLLABORATIVE PEER-REVIEW
Designed to be rigorous – yet also collaborative, fair and constructive

FAST PUBLICATION
Average 85 days from submission to publication (across all journals)

COPYRIGHT TO AUTHORS
No limit to article distribution and re-use

TRANSPARENT
Editors and reviewers acknowledged by name on published articles

SUPPORT
By our Swiss-based editorial team

IMPACT METRICS
Advanced metrics track your article's impact

GLOBAL SPREAD
5'100'000+ monthly article views and downloads

LOOP RESEARCH NETWORK
Our network increases readership for your article