NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

SCHOOL OF SCIENCE
DEPARTMENT OF INFORMATICS & TELECOMMUNICATIONS

POSTGRADUATE PROGRAM
"INFORMATION TECHNOLOGIES IN MEDICINE AND BIOLOGY"

MASTER THESIS

A computational approach to identify long non-coding RNAs acting as microRNA sponges

Anna A. Karavangeli

Supervisor: **Prof. Artemis Hatzigeorgiou,** Professor of Bioinformatics,
Department of Electrical & Computer Engineering,
Telecommunications and Networks, University of Thessaly

ATHENS

JUNE 2019

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
"ΤΕΧΝΟΛΟΓΙΕΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΣΤΗΝ ΙΑΤΡΙΚΗ ΚΑΙ ΤΗ ΒΙΟΛΟΓΙΑ"

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Υπολογιστική μέθοδος για την αναγνώριση μακρών μη κωδικών RNA που δρουν σαν «σφουγγάρια» των microRNA

Άννα Α. Καραβαγγέλη

Επιβλέπουσα : **Άρτεμις Χατζηγεωργίου,** Καθηγήτρια Βιοπληροφορικής,
Τμήμα Μηχανικών Η/Υ, Τηλεπικοινωνιών και Δικτύων του
Πανεπιστημίου Θεσσαλίας

ΑΘΗΝΑ

ΙΟΥΝΙΟΣ 2019

MASTER THESIS

A computational approach to identify long non-coding RNAs acting as microRNA sponges

Anna A. Karavangeli

SRN.: ΠΙΒ0174

Supervisor : **Prof. Artemis Hatzigeorgiou,** Professor of Bioinformatics, Department of Electrical & Computer Engineering, Telecommunications and Networks, University of Thessaly

EXAMINATION COMMITTEE:

Prof. Artemis Hatzigeorgiou, Professor of Bioinformatics, Department of Electrical & Computer Engineering, Telecommunications and Networks, University of Thessaly

Martin Reczko, Head of the bioinformatics group of the genomics facility, Alexander Fleming

Dr. Ema Anastasiadou, Researcher-Lecturer Level, Bioimedical Research Foundation of the Academy of Athens (BRFAA)

June 2019

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Υπολογιστική μέθοδος για την αναγνώριση μακρών μη κωδικών RNA που δρουν σαν «σφουγγάρια» των microRNA

Άννα Α. Καραβαγγέλη
Α.Μ.: ΠΙΒ0174

Επιβλέπουσα : **Άρτεμις Χατζηγεωργίου,** Καθηγήτρια Βιοπληροφορικής,
Τμήμα Μηχανικών Η/Υ, Τηλεπικοινωνιών και Δικτύων του
Πανεπιστημίου Θεσσαλίας

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ: **Καθ. Άρτεμις Χατζηγεωργίου,** Καθηγήτρια
Βιοπληροφορικής,
Τμήμα Μηχανικών Η/Υ, Τηλεπικοινωνιών και Δικτύων του
Πανεπιστημίου Θεσσαλίας
Martin Reczko, Επικεφαλής Τμήματος Βιοπληροφορικής
Ερευνητικού Κέντρου Βιοϊατρικών Επιστημών, Αλέξανδρος
Φλέμινγκ
Δρ. Έμα Αναστασιάδου, Ερευνήτρια Δ',
Ιδρυμα Ιατροβιολογικών Ερευνών Ακαδημίας Αθηνών
(ΙΙΒΕΑΑ)

Ιούνιος 2019

ABSTRACT

Long non-coding RNAs (lncRNAs) are transcribed non-coding RNAs (ncRNAs) that are more than 200 nucleotides long. Among their reported functions, their ability to act as molecular “sponges” for microRNAs (miRNAs) in several physiological processes and pathological conditions has gained attention over the past few years. According to the ceRNA hypothesis, by sequestering miRNAs, lncRNAs can reduce the number of the available miRNAs that target mRNAs and indirectly prevent target gene repression.

In order to investigate in what extent lncRNAs reduce the amount of miRNAs available to other targets and whether individual lncRNAs can be characterized as “sponges”, a mathematical quantitative model of binding site competition was employed. Argonaute Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (AGO-PAR-CLIP), RNA-Seq and small RNA-Seq (sRNA-Seq) experiments were used to identify targets and quantify target and miRNA abundances for to 2 different tissues. Site occupancies (fraction of sites bound by the miRNA) for protein coding targets were predicted in the presence and absence of lncRNAs. Increased occupancies of miRNA targeted mRNAs, observed in the lack of lncRNAs, lead to potential lncRNA “sponges”.

This analysis resulted in the identification of 8 lncRNAs acting as potential sponges for 38 miRNAs. The abundance of most individual targets was insufficient to alter miRNA levels and the reported sponges were highly expressed. Two well-studied nuclear lncRNAs, XIST and MALAT1, were among them suggesting additional functionalities of lncRNAs in certain settings, but also highlighting the need for separate analysis of nuclear and cytoplasmic RNAs.

SUBJECT AREA: Computational Biology, Bioinformatics

KEYWORDS: microRNA, lncRNA, molecular sponges, competition, mathematical modeling

ΠΕΡΙΛΗΨΗ

Τα μακρά μη κωδικά RNAs (lncRNAs) είναι μεταγραφόμενα μη κωδικά RNAs (ncRNAs) με μήκος πάνω από 200 νουκλεοτίδια. Μεταξύ των αναγνωρισμένων λειτουργιών τους, η ικανότητά τους να δρουν σαν μοριακά “σφουγγάρια” για τα microRNAs (miRNAs) σε διάφορες φυσιολογικές διεργασίες και παθολογικές καταστάσεις έχει κερδίσει προσοχή τα τελευταία χρόνια. Σύμφωνα με την υπόθεση του ceRNA, με την πρόσδεση των miRNAs, τα lncRNAs μπορούν να μειώσουν τον αριθμό των διαθέσιμων miRNAs που στοχεύουν mRNAs και να αποτρέψουν έμμεσα τη καταστολή του γονιδίου στόχου.

Για να διερευνηθεί σε ποιο βαθμό τα lncRNAs μειώνουν την ποσότητα των miRNAs που είναι διαθέσιμα σε άλλους στόχους και αν μεμονωμένα lncRNAs μπορούν να χαρακτηριστούν ως "σφουγγάρια", χρησιμοποιήθηκε ένα μαθηματικό ποσοτικό μοντέλο για τον ανταγωνισμό των θέσεων πρόσδεσης. Argonaute Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (AGO-PAR-CLIP), RNA-Seq και small RNA-Seq (sRNA-Seq) πειράματα χρησιμοποιήθηκαν για τον προσδιορισμό των στόχων και την ποσοτικοποίηση της αφθονίας τόσο των ίδιων όσο και των miRNA για 2 διαφορετικούς ιστούς. Το ποσοστό των θέσεων που δεσμεύονται από τα miRNA για πρωτεϊνικούς στόχους προβλέφθηκε παρουσία και απουσία των lncRNA. Αυξημένα ποσοστά πρόσδεσης των mRNA στόχων, που παρατηρούνται απουσία των lncRNA, οδηγούν σε πιθανά lncRNA «σφουγγάρια».

Η ανάλυση αυτή είχε σαν αποτέλεσμα την ταυτοποίηση 8 lncRNAs με πιθανή λειτουργία σφουγγαριού για 38 miRNAs. Η αφθονία των περισσότερων επιμέρους στόχων ήταν ανεπαρκής για να μεταβάλει τα επίπεδα του miRNA και τα αναφερθέντα σφουγγάρια εκφράζονταν σε μεγάλο βαθμό. Δύο καλά μελετημένα πυρηνικά lncRNAs, το XIST και το MALAT1, αναγνωρίστηκαν μεταξύ των άλλων, υποδεικνύοντας επιπλέον λειτουργίες των lncRNA σε συγκεκριμένα περιβάλλοντα, αλλά και υπογραμμίζοντας την ανάγκη ξεχωριστής ανάλυσης πυρηνικών και κυτταροπλασματικών RNAs.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Υπολογιστική Βιολογία, Βιοπληροφορική

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: microRNA, lncRNA, μοριακά σφουγγάρια, ανταγωνισμός, μαθηματική μοντελοποίηση

ACKNOWLEDGEMENTS

I would like to sincerely thank my supervisor, Prof. Artemis Hatzigeorgiou for giving me the opportunity to work on this exciting topic and in such a positive environment.

I would also like to express my gratitude to Dimitra Karagkouni for her guidance, help and support all this time.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

1. INTRODUCTION

1.1 miRNAs

microRNAs (miRNAs) are a class of small non-coding RNAs (~22 nucleotides long) that regulate their protein coding targets by causing degradation or/and translational repression [1, 2] (Figure 1). While translational repression seems to be the predominant way of gene regulation in animals, this is not the case in plants where direct cleavage of the targets is more common [3]. Short “seed” sequences at the 5′ ends of miRNAs (nucleotides 2–8) are most critical for miRNA binding. Mature miRNAs are loaded into a protein of the Argonaut family (AGO), a member of the RNA-induced silencing complex (RISC), and guide the complex to microRNA response elements (MREs) on targeted transcripts. MREs are mainly located on the 3’ untranslated region (3’UTR) of mRNAs but functional binding sites have been identified within the 5’ untranslated region (5’UTR) and the coding region (CDS) as well [4].

miRNAs are implicated in various developmental processes including differentiation, proliferation and apoptosis [1, 3] as well as in stress response [5, 6]. As a consequence of their regulatory role in important processes, deregulation of miRNAs often results in pathological conditions such as carcinogenesis, metabolic disorders etc. [7, 8]. Their implication in a wide range of physiological processes and disease is consistent with the fact that more than half of all mRNAs are targets of miRNAs [9]. Each miRNA is predicted to regulate up to hundreds of targets and each target can be regulated by multiple miRNAs. Thus, miRNAs and their respective targets shape a large regulatory network.

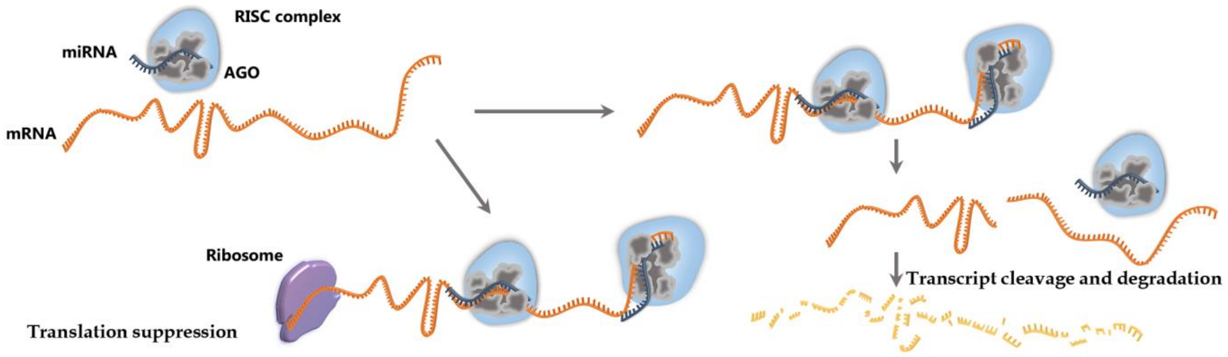

Figure 1: miRNA functionality. miRNAs are loaded into AGO and guide the RISC complex to complementary sequences on targeted transcripts, causing their degradation or/and translational repression [10].

1.2 lncRNAs

Less than 2% of the human genome encodes proteins. The vast majority of the transcribed genome produces non-coding RNAs. Long noncoding RNAs (lncRNAs) are among the various classes of non-coding RNAs and are defined as transcripts more than 200 nucleotides long with no coding potential, although it has been been demonstrated that some transcripts that are annotated as lncRNAs actually encode for small proteins [11, 12].

Based on their location with respect to protein coding genes they are mainly classified as [13]:

- antisense RNAs (transcripts that intersect any exon of a protein-coding locus on the opposite strand)
- lincRNAs (intergenic transcripts)
- sense overlapping (transcripts that contain a coding gene within an intron on the same strand)
- sense intronic (transcripts that reside within introns of a coding gene, but do not intersect any exons)
- processed transcripts (transcripts that do not contain an open reading frame)

Subcellular localization of lncRNAs varies, with transcripts observed in the nucleus, cytoplasm or both [14-16]. For example RNA FISH demonstrated that the lncRNA XIST, a key regulator of X inactivation [15], accumulates on the inactive X-chromosome [15, 17, 18], while the lncRNA GAS5 can be found both in the nucleus and at the cytoplasm.[15, 19]. Localization patterns may be indicative of their molecular role and thus are investigated by numerous teams [15, 16, 20-24]. The conventional view of lncRNAs as nuclear-enriched, epigenetic regulators [20-22] has been challenged by subsequent research showing that the number and importance of cytoplasmic lncRNAs has been underestimated [15, 16, 24, 25]. Benoit Bouvrette and colleagues found 75% of lncRNAs present in cytoplasmic fractions of human and Drosophila cells [16] and Cabili and colleagues reported complex localization of lncRNAs [15], which could be divided into five categories: large nuclear foci, large nuclear foci with single molecules scattered through the nucleus, predominantly nuclear without foci, cytoplasmic and nuclear, and predominantly cytoplasmic [15].

1.2.1 Functionality of lncRNAs

Although the majority of lncRNAs have yet uncharacterized functions, there are some well-studied examples that offer insights in the various ways lncRNAs operate in the cell. As reviewed in [26] based on their function lncRNAs can be broadly classified into two categories: Those that influence the expression and/or chromatin state of nearby genes and those that execute an array of functions throughout the cell away from their transcription site.

Class 1: Those that influence the expression and/or chromatin state of nearby genes.
Establishment of repressive or activating chromatin.
The most well studied example is that of the lncRNA Xist. In female mammals, one of the two X chromosomes is transcriptionally silenced for dosage compensation upon transcription of Xist from only one X chromosome, which will later become the inactive X (Xi). Following its induction, Xist spreads across the entire Xi resulting in the transcriptional silencing of almost the entire chromosome [16].

Transcriptional interference.
In some cases, the mere transcription of an lncRNA locus is enough to infer expression regulation of a neighbor gene independently of the specific lncRNA that is transcribed. The transcribed lncRNA may use the same promoter as another protein coding gene [27] or it may compete for RNA polymerase II and transcription factors reducing their availability for other neighbor genes [28, 29].

Presence of regulatory DNA elements in the lncRNA loci.
Genetic analyses of the lincRNA-p21 locus, uncovered that this lncRNA locus actually

functions to regulate the gene Cdkn1a in cis [30]. Modulation of gene expression due to the regulatory DNA elements in the gene body of the lncRNA was observed even in tissues in which lincRNA-p21is not expressed [31].

Class 2: Those that execute an array of functions throughout the cell away from their transcription site.

Transcription regulation through direct binding or recruitment of protein complexes.

lincRNA-EPS has been reported to interact directly with the promoters of distant genes and recuit the hnRNPL(Heterogeneous nuclear ribonucleoprotein L), resulting in the transcriptional repression of its targets [18]. Another example is that of the lncRNA HOTAIR, that has been proposed to act as a scaffold that recruits chromatin modifying complexes to the HOXD locus to establish a repressed chromatin state [32].

Organization of nuclear architecture

Some lncRNAs seem to orchestrate transcription, RNA processing or other gene expression related functions, through the organization of nuclear architecture. Such an example is the metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) that has been proposed to act as a scaffold for the positioning of nuclear speckles at active gene loci [33]. Nuclear speckles are dynamic nuclear compartments enriched in pre-mRNA splicing factors such as spliceosomal subunits, small nuclear ribonucleoproteins (snRNPs), and serine/arginine-rich proteins that are connected in places by a thin fibril [34]. While the involvement of MALAT1 to splicing has been suggested due to its association with the nuclear speckles, Malat1-deficient mice do not exhibit measureable splicing abnormalities [35], leaving room for alternative hypothesis for its functions.

Interaction with proteins and RNAs

Another way that lncRNAs may infer regulation is through their binding to other proteins or RNAs and the consequent modulation of their activity or abundance. The lncRNA NORAD for example, is targeted by the PUMILIO1 (PUM1) and PUMILIO2 (PUM2) proteins that normally bind to specific mRNAs and accelerate their decay or inhibit their translation [36]. Through this binding, NORAD reduces the amount of the PUMILIO proteins that is available to interact with their other mRNA targets. Inactivation of NORAD resulted in PUMILIO hyperactivity and downregulation of PUMILIO target transcripts [37, 38]. A similar mechanism is proposed for the lncRNA regulation of miRNAs [39].

1.2.2 lncRNA targeting by microRNAs

There have been a number of publications identifying lncRNA - miRNA interactions through computational and/or experimental approaches. In situ hybridization and confocal microscopy revealed MALAT1 targeting in the nucleus by MiR-9 [40]. Negative correlation between the expression of two miRNAs (miR-192 and miR-204) and their target lncRNA HOTTIP was observed in hepatocellular carcinoma [41]. Extensive evidence of lncRNA - miRNA interactions emerge from the analysis of crosslinking experiments, followed by sequencing (CLIP-Seq) and several algorithms and databases have been developed to predict and catalogue them [42-44]. Databases with lncRNA – miRNA interactions are reported below:

DIANA – LncBase v2 [42]

An extensive database including more than 70.000 miRNA:lncRNA experimentally supported interactions, derived from manually curated publications and the analysis of

153 AGO CLIP-Seq libraries. It also provides a module for *in silico* taget prediction with the DIANA-microT algorithm.

starBase v2.0 [43]
A database with comprehensive RNA–RNA and protein–RNA interaction networks in normal tissues and cancer cells derived from the analysis of 108 CLIP-Seq experiments. By taking into account common miRNA interactions with ncRNAs (lncRNAs, pseudogenes, circRNAs) and mRNAs they form ceRNA pairs.

NPInter v3.0 [44]
A database with experimentally verified interactions between ncRNAs (especially lncRNAs) and other biomolecules (proteins, mRNAs, miRNAs and genomic DNAs). Interactions derive from manually curated publications, high-throughput technologies and in silico predictions supported by AGO CLIP-Seq.

1.2.3 Artificial miRNA sponges

Reverse genetic approaches that act to inhibit microRNAs were initially developed in order to better understand the miRNA functions. These approaches involved introduction of antisense oligonucleotides [45, 46] or overexpression of transgenic reporters that contain miRNA binding sites [47] to achieve miRNA "sponging" and derepression of its targets. The use of artificial antisense RNAs as miRNA sponges is broad in molecular research, and progress has been made towards their application as a new class of drug [48]. Oligonucleotide miRNA inhibitors (typically small single-stranded RNA) are designed to have near perfect complementarity against a miRNA and are chemically modified to improve their stability. Nevertheless, despite reaching unphysiologically high concentration levels in the cell, artificial miRNA sponges are only capable of partial inhibition [49], which in the case of highly expressed miRNAs does not exceed 50% [50].

1.2.4 Natural miRNA sponges

The first evidence for natural miRNA sponges was discovered in plants where the non-coding gene IPS1 (INDUCED BY PHOSPHATE STARVATION1) from *Arabidopsis thaliana* was observed to sequester miR-399 [51]. IPS1 contains a motif with sequence complementarity to the phosphate starvation–induced miRNA miR-399, but the pairing is interrupted by a mismatched loop at the expected miRNA cleavage site. IPS1overexpression leads to increased accumulation of the miR-399 mRNA target PHO2. This mechanism of miRNA inhibition was termed as "target mimicry".

A similar mechanism was observed later in mammals. Mouse cell lines infected with murine cytomegalovirus, exhibit rapid post-transcriptional down-regulation of miR-27a leading to the hypothesis that miR-27a is inhibited due to the production of a viral or endogenous miRNA sponge [52].

In 2010 a mammalian cellular non-coding RNA was proposed as a miRNA sponge [53]. *PTENP1* is a pseudogene of the tumor suppressor PTEN gene and shares conserved miRNA seed target sites with *PTEN* for the miR-17, miR-21, miR-214, miR-19 and miR-26 miRNA families in its 3'UTR [53]. In this study they showed that retroviral overexpression of the *PTENP1* 3′ UTR increased expression of PTEN (by 50%) in a Dicer-dependent manner and acted as a tumor suppressor. Additionally, knockdown of endogenous PTENP1 in prostate cancer cells resulted in a decrease in *PTEN* mRNA and protein levels. In a similar way as PTENP1 but with different outcome in tumor development, overexpression of the 3’UTR of the pseudogene KRAS1P resulted in increased abundance of the KRAS mRNA and accelerated cell growth [53].

1.2.5 The ceRNA hypothesis

Experimental evidence for miRNA inhibition through sponging effects of artificial RNAs were present since 2007 [54]. Despite reaching unphysiologically high concentration levels these RNAs had mild results in miRNA inhibition [49]. The same year a natural (endogenous) non-coding RNA in plants was reported to invoke miRNA inhibition and upregulation of its mRNA target [51]. Three years later, in 2010, a mammalian pseudogene was shown to upregulate its gene of origin with which it shared miRNA seed target sites. Following these findings, the competing endogenous RNA (ceRNA) hypothesis was formed [55]. The ceRNA hypothesis proposes a layer of gene regulation mediated by transcripts with shared miRNA binding sites where RNAs can impair miRNA activity through sequestration, thereby upregulating miRNA target gene expression. ceRNAs (RNAs targeted by the same miRNA) exhibit indirect positively correlated expression. As one ceRNA increases, it titrates away miRNA from repressing other ceRNAs, and increases expression of all ceRNAs in the network [56]. Several classes of RNAs have been reported to act as ceRNAs including protein coding mRNAs, lncRNAs, pseudogenes, and circular RNAs (circRNAs) as reviewed in [39] (Figure 2) .

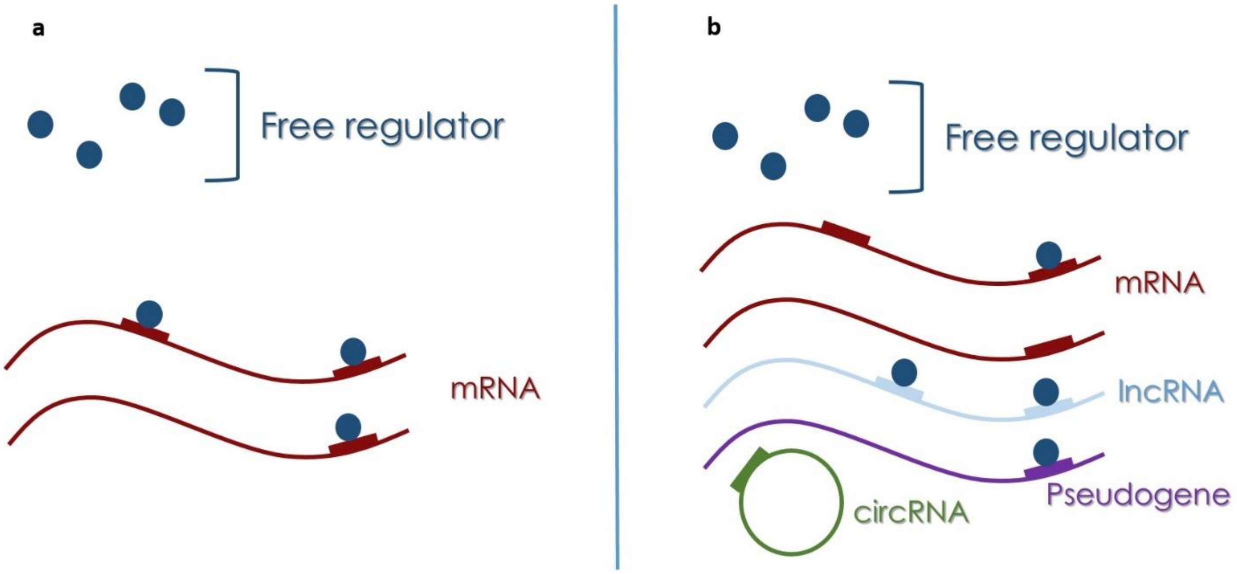

Figure 2: Competition for miRNA binding. In the scenario where only a limited amount of mRNAs is available for miRNA binding (a), the amount of free regulator (miRNA) is enough to bind all sites. Upon expression of additional sites (b) that may belong to mRNAs, lncRNAs, pseudogenes, or circRNAs, the free miRNA is sequestered and cannot bind to additional targets. Unbound targets are free from miRNA mediated repression and thus upregulated.

1.2.6 lncRNAs as competing endogenous RNAs

lncRNAs have been proposed to be part of the ceRNA(competing endogenous RNA) network, titrating miRNAs away from their other mRNA transcripts and are implicated in various dieseases and developmental processes (reviewed in [39]).

1.2.6.1 ceRNAs in Development

linc-MD1: The first evidence of lncRNAs implication in development through sponging effects was demonstrated in mouse and human myoblasts where the muscle-specific long noncoding RNA, linc-MD1 was observed to govern the time of muscle differentiation by acting as a ceRNA for MAML1 and MEF2C titrating away miR-133 [57].

linc-RoR: Another lncRNA, linc-RoR was identified as a regulator of human embryonic stem cell differentiation [58]. Specifically, a direct competition for miR-145 binding occurs between *linc-RoR* and the mRNAs encoding the core TFs transcriptional factors (TFs) *Oct4, Nanog*, and *Sox2*.

H19: lncRNA H19 contains both canonical and non-canonical binding sites for the let-7 miRNA family, and has been reported to act as an effective ceRNA for the abundant let-7 miRNA, hence modulating the expression of other let-7 target transcripts including Dicer and Hmga2 [59]. In addition to let – 7 miRNA, H19 has been reported to act a sponge for miRNAs of the miR-17-5p family during myoblast differentiation [60].

1.2.6.2 ceRNAs in Disease

HULC: Long non-coding RNA HULC, is highly up-regulated in hepatocellular carcinoma and plays an important role in tumorigenesis by sponging miR-372 [61]. miR-372 inhibition by HULC reduces translational repression of its target transcript PRKACB, translational repression which through a series of downstream events leads to the upregulation of the HULC RNA.

PTCSC3**:** lncRNA PTCSC3 (Papillary thyroid carcinoma susceptibility candidate 3) is heavily downregulated in thyroid cancers and its transfection has been shown to lead to a significant decrease in expression levels of the oncogenic miR-574-5p. The significant inverse correlation between PTCSC3 and miR-574-5p suggests that PTCSC3 acts as a competing endogenous RNA to target miRNAs and in turn regulate cell growth and apoptosis in thyroid cancer [62].

1.3 Predicting ceRNA interactions

1.3.1 In silico versus AGO-CLIP guided algorithms for identifying ceRNA ineractions

Competition between RNAs depends on the sharing of MREs for the same miRNA. Hence, identification of MREs on transcripts is of crucial importance for ceRNA interactions.

In silico target prediction algorithms such as TargetScan [63], miRanda [64], PicTar [65] and DIANA microTCDS [66] are widely used and consider a variety of features to identify potential miRNA targets, such as sequence complementarity between the miRNA seed and the target, position of the MRE on the target, free energy and site accessibility, conservation etc.

As an alternative to *in silico* prediction strategies, developed high-throughput biochemical techniques, which identify endogenous miRNA–target interactions can be used. Examples of these experimental methods include high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation (HITS-CLIP) [67] and photoactivatableribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) [4]. In this context, several computational approaches have been developed to analyze AGO-CLIP-Seq datasets and identify miRNA targets (microMUMMIE [68], PARma [69], microCLIP [10]).

Since expression patterns of RNAs are cell type-, tissue type-, developmental stage and disease- specific, interactions between miRNAs and their targets are the same. Both *in silico* and experiment based approaches can identify targets in these different conditions. Localization of lncRNAs on the other hand, is not considered by *in silico* prediction models, while AGO-CLIP guided identification provides only relevant interactions. In addition, targeting patterns of miRNAs change based on target RNA concentration and availability (sponging effects) and *in silico* models are not sensitive to these parameters. Hence, computational methods that analyze AGO-CLIP data may be more appropriate when investigating ceRNA interactions.

1.3.2 Methods for identifying ceRNAs and related databases

Several computational methods have been developed to identify ceRNAs. Correlation of expression between the competing RNAs is one characteristic that several approaches take into consideration [70-72]. For example, if two RNAs are targeted by the same miRNA, then overexpression of one would titrate away miRNA molecules from the other resulting in the upregulation of the latter and thus their expression levels would be positively correlated. Another approach is to score the ceRNA pairs based on the number of shared miRNAs or MREs [43, 73]. By this logic, RNAs with many shared miRNAs or MREs would regulate each other the best because overexpression of one would titrate away multiple miRNAs from the other.

These approaches identify possible competing pairs but cannot predict to what extent changes on the concentration of a transcript, affect other ceRNAs and whether this can amount to a regulatory impact. To address this issue quantitative modeling is needed. Several mathematical models that quantify miRNA and target concentration have been employed (reviewed in [74]) including a stochastic model [75], a mass action model [76] and a model of binding site occupancies [77, 78]. In quantitative models determining the number of transcriptomic miRNA-binding sites is crucial for evaluating the potential for ceRNA regulation. Differences in this estimation may lead to different conclusions.

Databases and tools dedicated to ceRNA interactions are described below:

spongeScan [79]
spongeScan proposes an approach based on sequence complementarity to identify patterns across lncRNAs that could be targeted by miRNAs. Each lncRNA – miRNA pair is further scored in order to assess the lncRNAs potential as a ceRNA.

Each lncRNA - miRNA pair is scored based on three parameters:

1. A log-odds score that is indicative of how many MREs can be found in the lncRNA sequence.
   Multiple MREs would facility the sequestration of miRNAs by the lncRNA.
2. A dispersion score that is indicative of how evenly distributed the MREs are along the transcript.
   Equally distributed MREs are observed in some known ceRNAs and would facilitate multiple miRNA binding.
3. A complexity score to filter out low complexity k-mers that would result in unspecific binding.

SomamiR 2.0 [80]
SomamiR 2.0 database contains cancer somatic mutations in miRNAs and their targets. ceRNAs are considered the mRNA, lncRNA and circRNA targets identified in 21 PAR-CLIP and 13 HITS-CLIP experiments from starBase. Somatic mutations for these targets

were downloaded from the COSMIC database. (The somatic mutations in the ceRNAs are only reported. There is no evidence of how the mutation impacts the interaction). Functional impact of somatic mutations are analyzed from miR2GO only for mutations in the miRNA sequence.

starBase v2.0 [43]
starBase ceRNA network for human and mouse includes computationally predicted targets for conserved miRNA families that are supported by CLIP-Seq data (108 datasets were analyzed in total from PAR-CLIP, HITS-CLIP, iCLIP and CLASH experiments). The biotypes included are mRNAs, lncRNAs, pseudogenes and circular RNAs. Each CLIP-supported target is paired with every other target and a hypergeometric test is executed for each pair separately. Pairs with FDR < 0.05 are imported in the database. The hypergeometric is defined by four parameters: i) the total number of miRNAs used to predict targets ii) the number of miRNAs that interact with the chosen gene of interest iii) the number of miRNAs that interact with the candidate ceRNA of the chosen gene and iv) the common miRNA number between these two genes.

ln*Ce*DB [73]
The database includes computationally predicted mRNA targets that are supported by CLIP-Seq data from starBase. Targets in lncRNAs from miRCode. For more recently annotated lncRNAs target sites were computationally predicted with an algorithm similar to the miRanda algorithm and then were intersected with the dataset of Jalali et al.

The cRNA network consists of mRNA-lncRNA pairs. To find out the likelihood of an lncRNA-mRNA pair for actually being ceRNA they use two methods:

1. a ceRNA score is calculated from the ratio of the number of shared MREs between the pair with the total number of MREs of the individual candidate gene
2. by hypergeometric test using the number of shared miRNAs between the ceRNA pair against the number of miRNAs interacting with the individual RNAs (same as starBase)

They support that the number of shared MREs would be more appropriate instead of the number of shared miRNAs between the ceRNA pair. ln*Ce*Db also incorporates lncRNA and mRNA expression data from 22 tissues for viewing the co-expression of each lncRNA – mRNA pair.

miRSponge [81]
A manually curated database for experimentally supported miRNA-sponge interactions and ceRNAs for 11 species. Database classes include endogenously generated molecules including coding genes, pseudogenes, lnc RNAs and circular RNAs, along with exogenously introduced molecules including viral RNAs and artificial engineered sponges.

LncACTdb 2.0 [72]
A database that integrates both manually curated and predicted ceRNA interactions for 23 species and 213 diseases/phenotypes. Database classes include endogenously generated molecules including coding genes, pseudogenes, lnc RNAs and circular RNAs, along with exogenously introduced molecules including viral RNAs and artificial engineered sponges. ceRNA interactions between lncRNAs and mRNAs were predicted for several cancer types. Expression data was retrieved from TCGA. lncRNA targets were computationally predicted with strict thresholds and then compared with the experimentally supported binding sites from 41 AGO-CLIP datasets from starBase. miRNA targets were derived from TarBase (v8) and mirTarBase (v18). Every lncRNA and mRNA interacting with the same miRNA is considered as a candidate ceRNA triplet (lncRNA-miRNA-mRNA). An lncRNA-miRNA-mRNA triplet is considered functional ceRNA in a specific cancer type depending on the correlation of their expression values.

More specifically the following criteria must be met: i) corr(lncRNA,miRNA) < 0 ii) corr(mRNA,miRNA) < 0 iii) corr(lncRNA,mRNA) > 0.

2. METHODS

2.1 Data Collection

To investigate the competition between lncRNAs and mRNAs for miRNA binding with a quantitative model, both miRNA and target expression needs to be taken into account. Small RNA-seq data was used to identify and quantify miRNA expression, photoactivatable ribonucleoside enhanced crosslinking and immunoprecipitation (PAR-CLIP) experiments [4] were analyzed to retrieve targets for the expressed miRNAs and RNA-seq data provided information on the expression of those targets. Analyzed datasets from microCLIP [10] for 4 different human cell types, one liver cell line (HEK293) and 3 lymphoblastoid EBV infected cell lines (LCL-BAC, LCL-BACD1 and LCL-BACD3).

2.1.1 PAR-CLIP data

A total of 7 PAR-CLIP datasets (4 HEK293 and 3 lymphoblastoid) were retrieved (Table 1). They were previously preprocessed and aligned to hg19.

Table 1: PAR-CLIP datasets

SRA	Authors	Cell type/tissue	Condition	Description
SRR592687	Skalsky et al. [82]	LCL-BAC	NA	LCL infected with an EBV B95-8 BACmid
SRR592688		LCL-BAC-D1	miR-BHRF1-1 mutant virus	LCL infected with an EBV B95-8 BACmid
SRR592689		LCL-BAC-D3	miR-BHRF1-3 mutant virus	LCL infected with an EBV B95-8 BACmid
SRR189784	Kishore et al. [83] .	HEK293	T1RNase	Embryonic Kidney Cells
SRR189785		HEK293	T1RNase	Embryonic Kidney Cells
SRR189786		HEK293	mildMNase	Embryonic Kidney Cells
SRR189787		HEK293	mildMNase	Embryonic Kidney Cells

2.1.2 RNA-seq data

RNA seq data had been previously preprocessed, aligned to hg38 and hg19 for the HEK293 and lymphoblastoid cell lines respectively, quantified and annotated. Quantified transcripts were annotated with ensembl 82 for the HEK293 cell line and ensembl 75 for the lymphoblastoid cell lines.

Table 2: RNA-seq datasets

SRA	Authors	Cell type/tissue	Condition	Description
SRR837794	Majoros et al. [68]	LCL-BAC	NA	LCL infected with an EBV B95-8 BACmid
SRR837795		LCL-BAC-D1	miR-BHRF1-1 mutant virus	LCL infected with an EBV B95-8 BACmid
SRR837796		LCL-BAC-D1	miR-BHRF1-1 mutant virus	LCL infected with an EBV B95-8 BACmid
SRR837798		LCL-BAC-D3	miR-BHRF1-3 mutant virus	LCL infected with an EBV B95-8 BACmid
SRR1240811	Conrad et al. [84]	HEK293	NA	Embryonic Kidney Cells

2.1.3 Small RNA-seq data

Small RNA seq data had been previously pre-processed and aligned. Annotation was retrieved from miRBase Release 22 [85] and miRBase Release 20 [86] for the HEK293 and lymphoblastoid cell lines respectively.

Table 3: small RNA-seq datasets

SRA	Authors	Cell type/tissue	Condition	Description
SRR592692	Skalsky et al. [82]	LCL-BAC	NA	LCL infected with an EBV B95-8 BACmid
SRR592693		LCL-BAC-D1	miR-BHRF1-1 mutant virus	LCL infected with an EBV B95-8 BACmid
SRR592694		LCL-BAC-D3	miR-BHRF1-3 mutant virus	LCL infected with an EBV B95-8 BACmid
SRR1240816	Conrad et al. [84]	HEK293	NA	Embryonic Kidney Cells
SRR1240817		HEK293	NA	Embryonic Kidney Cells

2.1.4 Target identification

In order to examine an adequate number of miRNAs with different expression patterns and target pools, the 200 top expressed miRNAs were considered in each dataset. For those, targets were retrieved with microCLIP [10], a framework that combines deep learning classifiers to identify miRNA – target interactions from PAR-CLIP experiments. The algorithm takes into account and analyzes non T-to-C clusters in addition to the regular clusters with T-to-C transitions and reports canonical and non-canonical bindings. After target identification the four HEK293 libraries were combined into one, keeping all the interactions reported in the deepest library and adding the unique interactions from the rest.

2.1.5 Target annotation

Protein coding transcripts were considered those with “protein coding” gene and transcript biotype.

lncRNA transcripts were considered those with gene and transcript biotype in: “3prime overlapping ncrna”, “antisense”, “lincRNA”, “ retained intron”, “macro lncRNA”, “processed transcript”, “sense intronic”, “sense overlapping”.

Pseudogene transcripts were considered those with gene and transcript biotype in: “IG_C_pseudogene”, “IG_J_pseudogene”, “IG_V_pseudogene”, “processed_pseudogene”, “pseudogene”, “transcribed processed_pseudogene”, “transcribed unprocessed pseudogene”, “transcribed unitary pseudogene”, “translated unprocessed pseudogene”, “TR_J_pseudogene”, “TR_V_pseudogene”, “unitary pseudogene”, “unprocessed pseudogene”.

2.1.6 Target quantification

Targets identified with microCLIP [10] were intersected with the RNA-seq quantified transcripts to gain expression information using BEDTools v2.17.0 [87]. If the same MRE overlapped two different elements then it was assigned to the one with the highest priority. Priorities (in descending order) were considered as : “CDS”, “UTR3”, “UTR5”, “lincRNA”, “sense_intronic” = “sense_overlapping”, “antisense”, “retained_intron”, “3prime_overlapping_ncRNA”, “processed_transcript”, “pseudogene”. Since HEK293 RNA-seq reads were aligned to hg38 while PAR-CLIP (target) reads were aligned to hg19, coordinate liftover from hg19 to hg38 was performed for the reported targets using the ensembl liftover tool. This procedure was not necessary for the lymphoblastoid datasets that were aligned to hg19.

2.2 Mathematical model

2.2.1 Model of binding site occupancies

The ceRNA hypothesis states that sites that are targeted by the same miRNA compete with each other for binding by this miRNA. This way, every binding site reduces the amount of free miRNA available to others. Whether this can amount to a regulatory impact needs further investigation. miRNAs exert their regulatory role through binding on target MREs. For a protein coding transcript miRNA binding may lead to translational repression and thus, the number of transcripts that are free from miRNA regulation are important for determining mRNA activity. Quantitative models of binding site competition predict the changes in binding site occupancy (that is fraction of sites bound by a miRNA) in response to changes in target or miRNA concentration.

The model used to investigate the extent of competition between lncRNAs and mRNAs is the one described by Jens and Rajewsky [77].Their source code (freely available on doRiNA — RNA competition **[88]**), was downloaded and executed with minor modifications concerning the input and output files.

In this model target site occupancies (that is fraction of sites that are bound by a miRNA) change in response to changes in target concentration. Occupancy is measured by the binding equation:

Equation 1

$$\Theta_i = \frac{F}{K_i + F}$$

Θ_i is the fraction of the sites that are bound, F is the concentration of the free regulator and K_i is the dissociation constant which quantifies the strength of interaction between the binding site and the ligand. At equilibrium, all sites of a given *Ki* have the same occupancy (Θ*i*), which is determined by the amount of free regulator (*F*) and the binding Equation 1.

Equation 2

$$F = Total - \sum c_i \Theta_i$$

$$F = Total - Bound$$

Equation 2 expresses that all regulator (*T*) is either free or bound (that is, the sum of all binding sites, each weighted with its occupancy and concentration. The free regulator is equal to the total regulator reduced by the fraction that is bound.

2.2.2 Target site concentration

Each binding site is as abundant as its harbouring RNA. RNA concentration was estimated as follows:

$$Concentration = \frac{Moles}{Volume} = \frac{\frac{Copies}{Avogadro\ number}}{Volume}$$

Cell volume was approximated by a sphere of 6.5 μm radius (r) and is calculated by the equation:

$$Volume = \frac{4}{3}\pi r^3$$

To estimate absolute copy numbers of mRNAs, the assumption that 250.000 mRNA molecules are present in the cell was made. The same assumption was made by [77] and was based on reported measurements [89]. The 250.000 molecules correspond to the total TPM values (or any other type of normalization value) of protein coding transcripts as they derive from the quantification of the RNA-seq experiment. Hence, transcript copies equal its TPM value multiplied by the scaling factor:

$$Trancript\ copies = Transcript\ TPM * \frac{250.000}{Total\ mRNA\ TPM}$$

2.2.3 miRNA concentration

miRNA concentration was calculated with the same method, under the assumptions that miRNA read count is proportional to abundance, all miRNAs are loaded and the total amount of AGO/RISC complexes is constant and known (150000/cell) [77, 90]. Hence, miRNA copies equal its count multiplied by the scaling factor:

$$miRNA\ copies = miRNA\ counts * \frac{150.000}{Total\ miRNA\ counts}$$

2.2.4 Dissociation constant (K_d) calculation

Dissociation constants (K_d) that quantify the strength of binding between the miRNA and its target were calculated by the approximate energy model for mammalian AGO binding as proposed by Jens and Rajewsky [77]. In brief, each nucleotide contributes to the stability of the miRNA-target complex and K_d is calculated by taking into account the base-pairing between the miRNA and the target. The model is described below.

Dissociation constant (K_d) measures the propensity of a complex to dissociate in smaller components. For the general reaction:

$$A + B \leftrightarrows AB$$

In which A and B come together to form complex AB and reversibly AB breaks down to A and B, the dissociation constant is defined:

$$K_d = \frac{[A][B]}{[AB]}$$

Where [A] is the concentration of component A, [B] is the concentration of component B and [AB] is the concentration of complex AB. Small K_d value means that the complex is favored in comparison to the separate components. Thus, the strength of binding between the miRNA and its target (miRNA- target complex) can be estimated by calculating the K_d.

The binding energy directly determines the dissociation constant (K_d), and vice versa:

$$K_d = e^{E/K_B T}$$

Where E is the binding energy of the regulator bound to the site, KB is the Boltzmann constant and T is the temperature in Kelvin.

2.2.5 An approximate energy model for AGO binding

Jens and Rajewsky [77] built an approximate energy model for mouse AGO binding where total binding energy is calculated by the base pairing between the guide miRNA and the target. Each position of the guide miRNA contributes to the total energy of the binding (Figure 3, Figure 4).

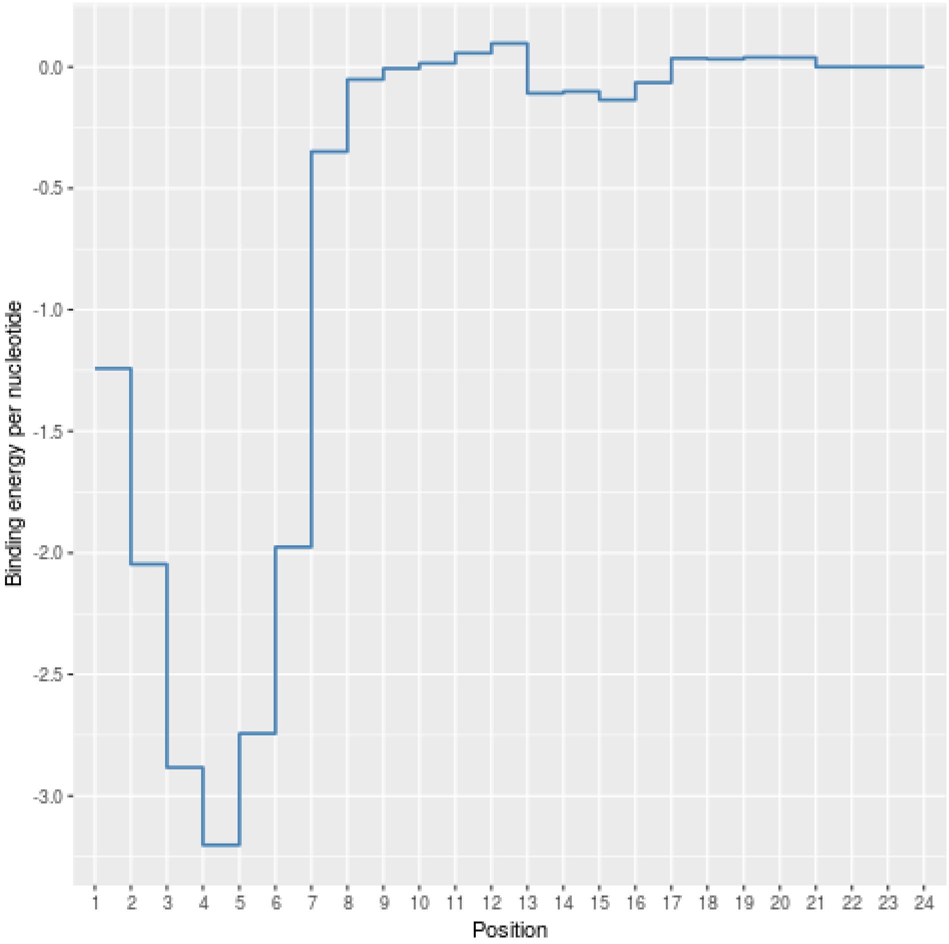

Figure 3. Each nucleotide contributes to the total binding energy independently. The binding energy directly determines the dissociation constant, and vice versa. In this mammalian AGO binding model binding to the seed region (miRNA nucleotides positions 2-7) contributes the most to the stability of the miRNA – target complex (smaller energy means greater stability).

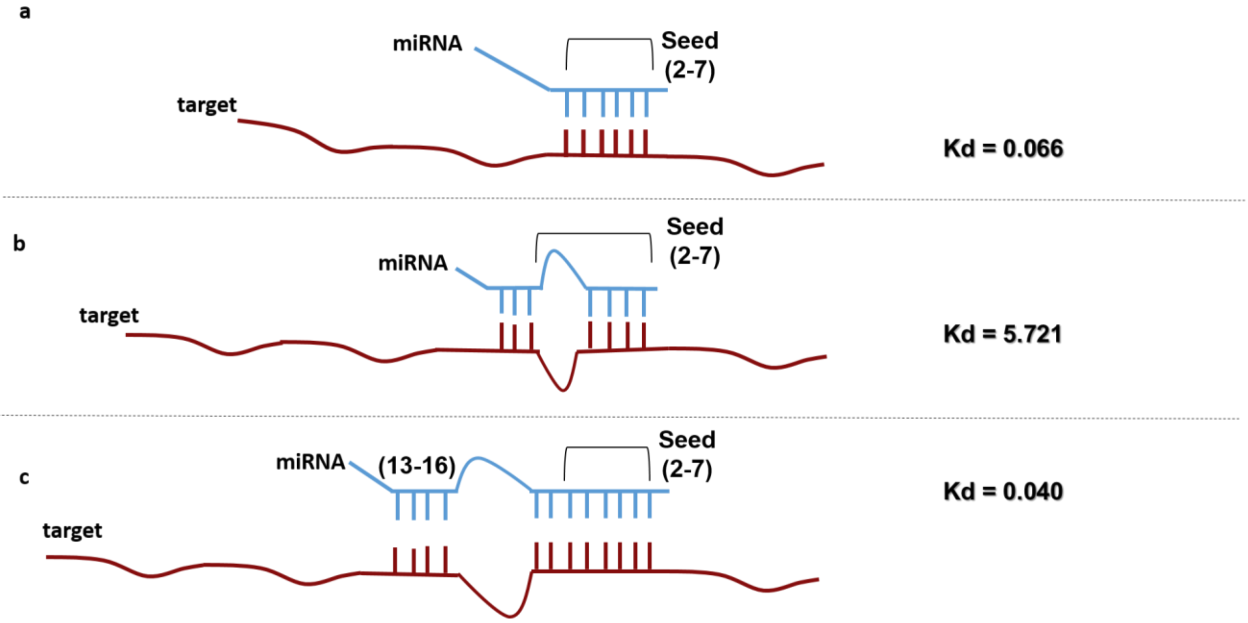

Figure 4: Kd values calculated by the approximate energy model of Jens and Rajewsky. a) Binding to the seed region (position 2-7) leads to stable binding. b) Mismatch in the seed region reduces dramatically the strength of binding. c) Pairing that extends the seed match up to position 9 of the

miRNA and supplementary base-pairing of the 3´ part of the miRNA (around positions 13–16) can further stabilize binding.

To build this model they used data published by [88]. In this study they systematically altered the sequence that corresponds to the let-7 miRNA, in order to determine how the pairing affects fly-AGO function. For each position in the guide RNA they measured the Michaelis-Menten constant (Km) in two cases:

A) If the nucleotide (or dinucleotide) in this position was matched perfectly to the target
B) If the nucleotide (or dinucleotide) in this position was mismatched (after sequence alteration)

Jens and Rajewsky reasoned that the log-ratios of perfect and mismatched K*m* should be proportional to the change in binding energy introduced by the mutation.

The steps that they followed were:

1. **Take perfect and mismatched Km values experimentally measured** for fly AGO

2. **Calculate energy in each position of the guide RNA** (by calculating the log ratio of the perfect and mismatched Km for each nucleotide or dinucleotide)
    a. Assign to each nucleotide (ex. g1) or dinucleotide (ex. g1-g2) the log ratio of the km values($-\log(\frac{\text{Km match}}{\text{Km mismatch}})$)
    b. Evaluate the energy (meaning the calculated ratio) in each position by averaging the available data (ex. For the position g2 there is data for the dinuclotide g1-g2 and the dinucleotide g2-g3. So the energy for position g2 is calculated as (Energy[g1-g2]/2 + Energy[g2-g3]/2)/2)
    c. Scale these energies such that their sum corresponds to the best reported binding (3.7 pM at 25 °C). New energy = old energy *(best energy/sum(old energies))

3. **Estimate mouse AGO energies by scaling appropriately the fly AGO energies** For mammalian (mouse) Argonaute (AGO), three distinct modes of binding were measured (best match , seed match, perfect match) [91]. These data were used to scale the fly Ago energies in each of the three described guide RNA segments to arrive at an approximate energy model for mammalian AGO.

    Scaling:

    a. Calculate energies from $\mathrm{K_d}$ values for each of the 3 categories. This results to energies for segments, but not for each position separately like before.
    b. Each position must be scaled according to the segment that it belongs.

4. **Estimate total energy of the base-pairing between a miRNA and its target** Having estimated the energy each position contributes calculate total energy of the predicted base-pairing between miRNA and target by summing all energies in the matched positions.

5. **Calculate K_d value from the total energy value.** $K_d = e^{E/K_BT}$

2.2.6 Sponge identification

2.2.6.1 Brief overview of the procedure

The steps followed in order to measure the ability of each lncRNA to act a sponge and reduce the amount of miRNA available to other targets are presented in brief below:

1. Target identification and quantification of miRNAs and reported targets
2. Employment of the mathematical model to predict protein coding site occupancies for each miRNA in two cases:
    a. All lncRNAs that are co-targeted are excluded from the target pool
    b. Target pool includes all protein coding and lncRNA sites
3. A suitable threshold is applied for the observed changes in site occupancies
4. Employment of the mathematical model again (only for the miRNAs passing the aforementioned threshold) to predict protein coding site occupancies in two cases:
    a. Each and every lncRNA is excluded by turn from the target pool
    b. Target pool includes all protein coding and lncRNA sites

    This step is needed to estimate if the observed changes in occupancies from step 2 are due to the exclusion of all lcRNAs or if there is an lncRNA more responsible than others
5. The same threshold as before is applied. If the changes in the occupancies of protein coding sites pass the threshold, then the lncRNA is considered a sponge for the miRNA.

2.2.6.2 Threshold for changes in site occupancies

The threshold for assessing the importance of site occupancy changes upon lncRNA exclusion includes 2 conditions:

1. The most affected targets (those with the largest difference in occupancies before and after lncRNA exclusion) should have at least a 5% change in their occupancies.
2. Mean occupancy of targets when the whole target pool is considered (protein coding and lncRNA) should be less than 95%.

The first condition ensures that the targets will be occupied more than 5%. Smaller occupancy percentage would probably mean that the target is not regulated by the miRNA. For example, even if the occupancy of a site increases from 0,001% to 1% in the absence of lncRNAs, the consequences of increased miRNA binding will be negligible. This first condition also ensures that the change will be somewhat measurable and avoids reporting very small differences.

The second condition ensures that the miRNA is not expressed at levels that would allow it to highly occupy all of its targets, since this would mean that it is not effectively sponged by a single target.

3. RESULTS

3.1 miRNA expression

For all cell types, absolute quantification of the 200 most expressed miRNAs was performed by transforming their reported counts to copy numbers. The most expressed miRNAs (regardless of the dataset) were found at approximately 20,000 copies, while the least expressed were as low as 6 to 12 copies (**Table 4**)

More than 75% of miRNAs were expressed at levels below 250 copies (this was observed for all 4 datasets) and although, the 3 datasets from lymphoblastoid cell lines seemed to have very similar miRNA expression patterns, the differences between them were statistically important (Figure 5, **Table 5**).

Table 4: miRNA abundance in each cell type. Copy numbers estimated for the 200 miRNAs, vary greatly from tens of thousands to less than ten. Copies were estimated from small RNA-seq data assuming that 150000 AGO/RISC complexes are present in the cell.

Cell type	Max copy number	Min copy number	Mean copy number
HEK293	26280	6	747
LCL-BAC	20060	12	745
LCL-BACD1	17530	10	746
LCL-BACD3	20330	11	746

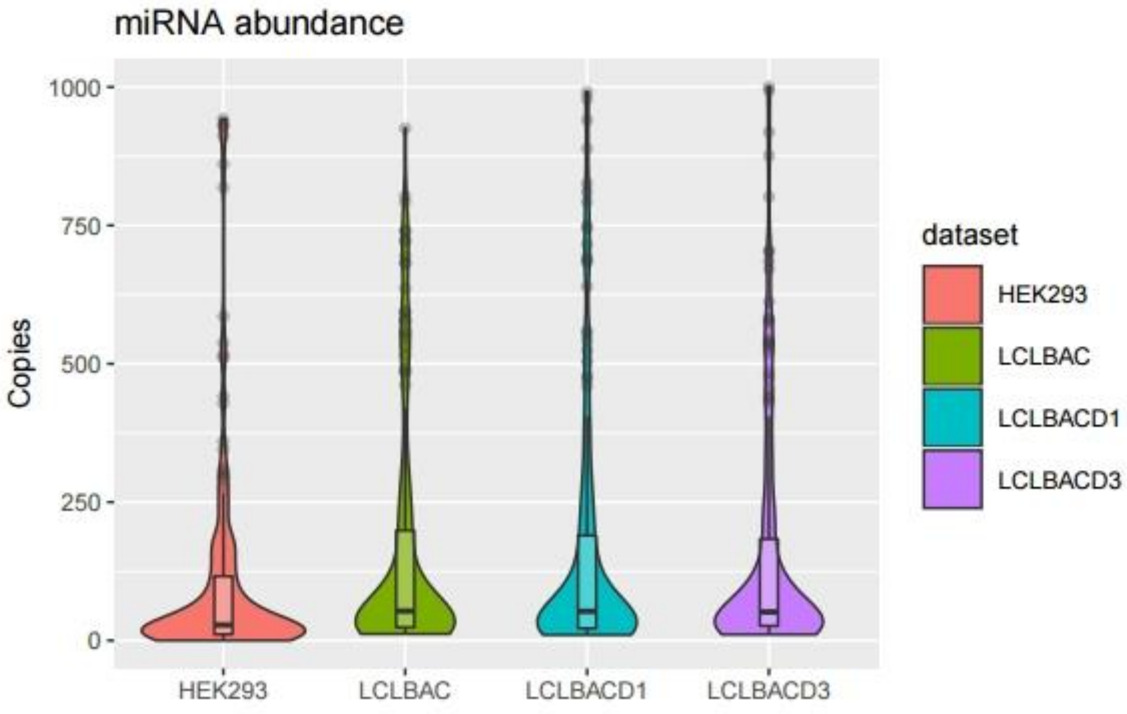

Figure 5: miRNA expression patterns per dataset (miRNAs with more than 1000 copies are not shown in this plot). Copies were estimated from small RNA-seq data assuming that 150000 AGO/RISC complexes are present in the cell. For all datasets, 75% of miRNAs are expressed at levels below 250 copies.

Table 5: Statistical importance of miRNA expression differences between cell types. Kruskal-Wallis test and Dunn's post-hoc test for multiple comparisons were performed. Bonferroni adjusted p-values are presented.

	HEK293	LCLBAC	LCLBACD1	LCLBACD3
HEK293	_	_	_	_
LCLBAC	0	_	_	_
LCLBACD1	0	5.1625E-021	_	_
LCLBACD3	0	0.0011	2.2941E-007	_

3.2 Identified targets for the most expressed miRNAs

After the identification of the most expressed miRNAs in each dataset, their targets across Ensembl annotated 3'UTR, CDS, 5'UTR, lncRNAd and pseudogenes were retrieved from PAR-CLIP data with microCLIP. More extensive targeting was observed in the HEK293 cell line where approximately 7000 genes were reported as targets (Table 6, Figure 6).

Table 6: Number of targeted MREs and corresponding transcripts and genes in each cell type. MREs were identified with microCLIP from PAR-CLIP data.

Cell type	MREs	Transcripts	Genes
HEK293	167883	13687	7132
LCL-BAC	34566	5234	2704
LCL-BACD1	52306	8240	3918
LCL-BACD3	27268	4527	2429

The majority of MREs were located in the 3'UTR of protein coding transcripts, while lncRNA MREs make up approximately 2% of the total. Pseudogene targets were almost non-existent in the lymphoblastoid cell lines and make up only 0.3% of the total MREs in the HEK293 cell line (Figure 6, Figure 7).

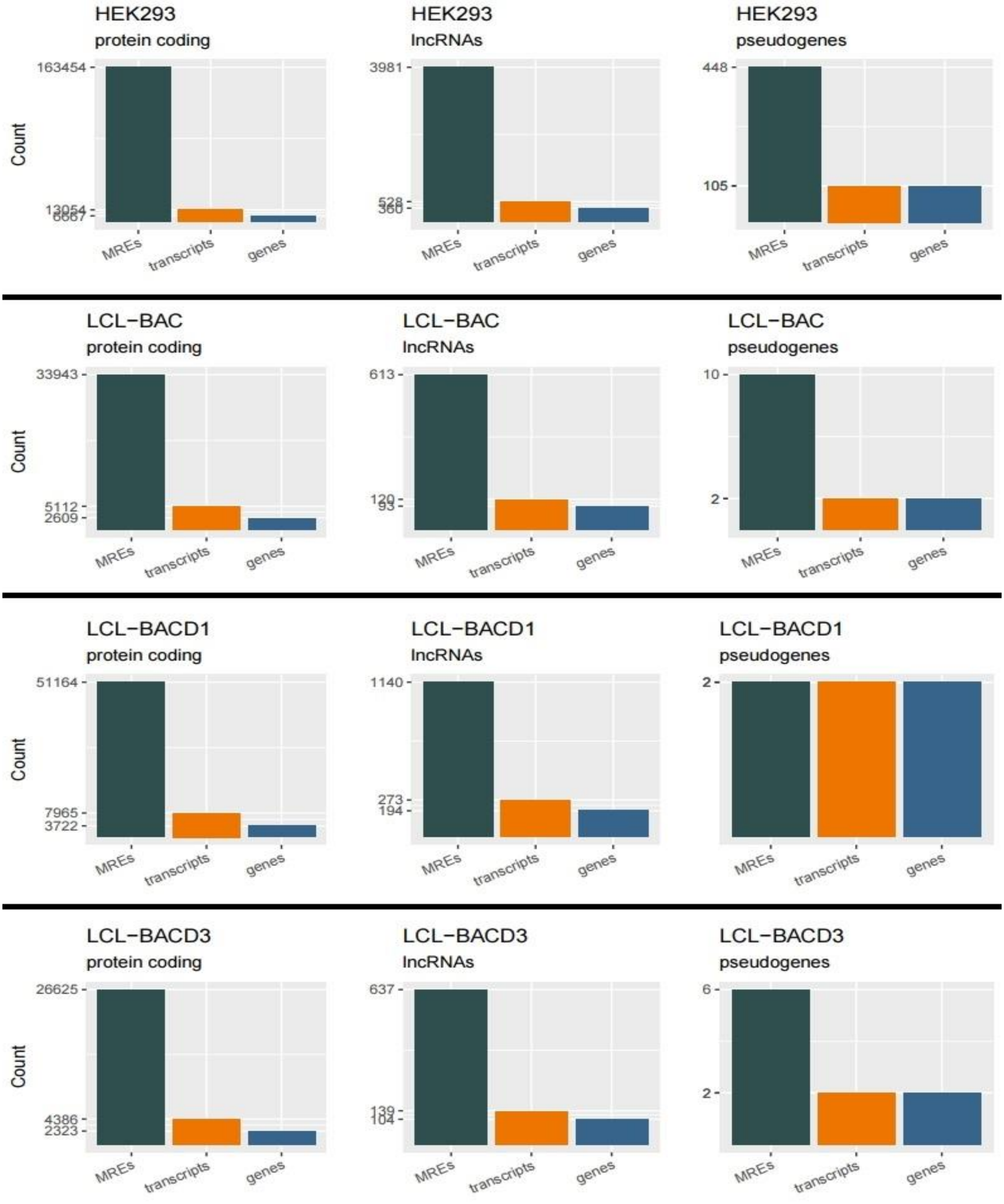

Figure 6: Number of MREs, transcripts and genes reported in each cell type, grouped according to their biotype. More targets are identified in the HEK293 cell line compared to the lymphoblastoid cell lines. The majority of MREs are located in protein coding transcripts.

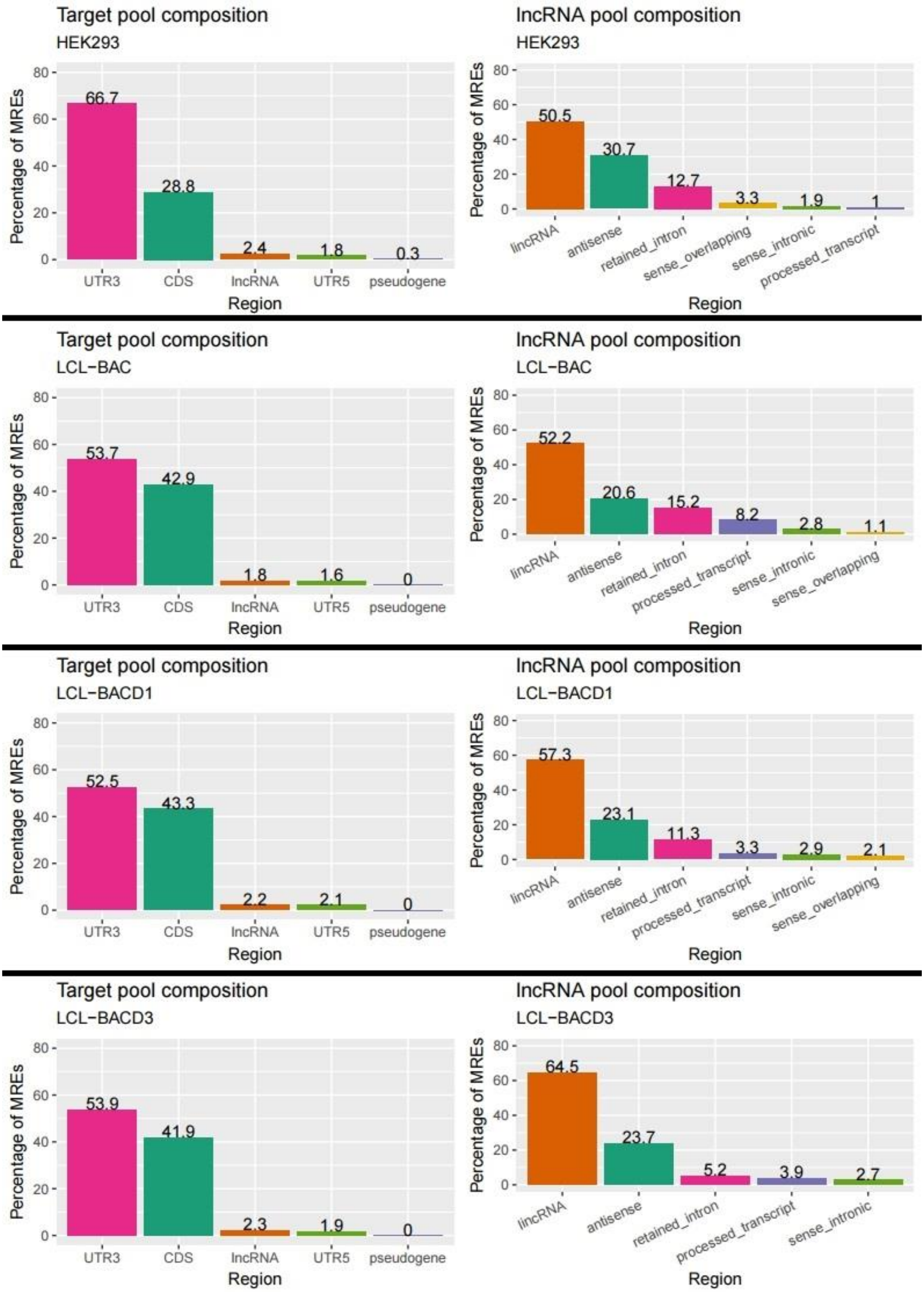

Figure 7: Target pool composition. Figure shows the percentage of MREs located in 3'UTR, CDS, 5'UTR, lncRNAs and pseudogenes for every dataset (column 1). For the targeted lncRNAs, the percentage of MREs per biotype is also shown (column 2).

3.3 Target expression

Absolute quantification was also performed for the reported targets, by transforming their TPM values to copies per cell. Protein coding targets were more abundant and more highly expressed in every dataset (Figure 8, Table 7).

For every miRNA, the expression (in number of copies) of each transcript they target was added together to form the total target pool. This resulted in target pool estimates varying from more than 70000 copies to less than 20 (Table 8). Larger target pools were observed for miRNAs in the HEK293 cell line, in accordance with the higher number of targeted transcripts in this dataset, while in the lymphoblastoid cell lines LCL-BAC and LCL-BACD3 target pools were of similar magnitude (Table 8, Figure 9, Table 9).

The majority of miRNAs were expressed at levels lower than their target pool, leading to miRNA:target ratios < 1. This phenomenon was observed even for highly expressed miRNAs like miR-101-3p, the third most expressed miRNA in the HEK293 cell line (Table 10).

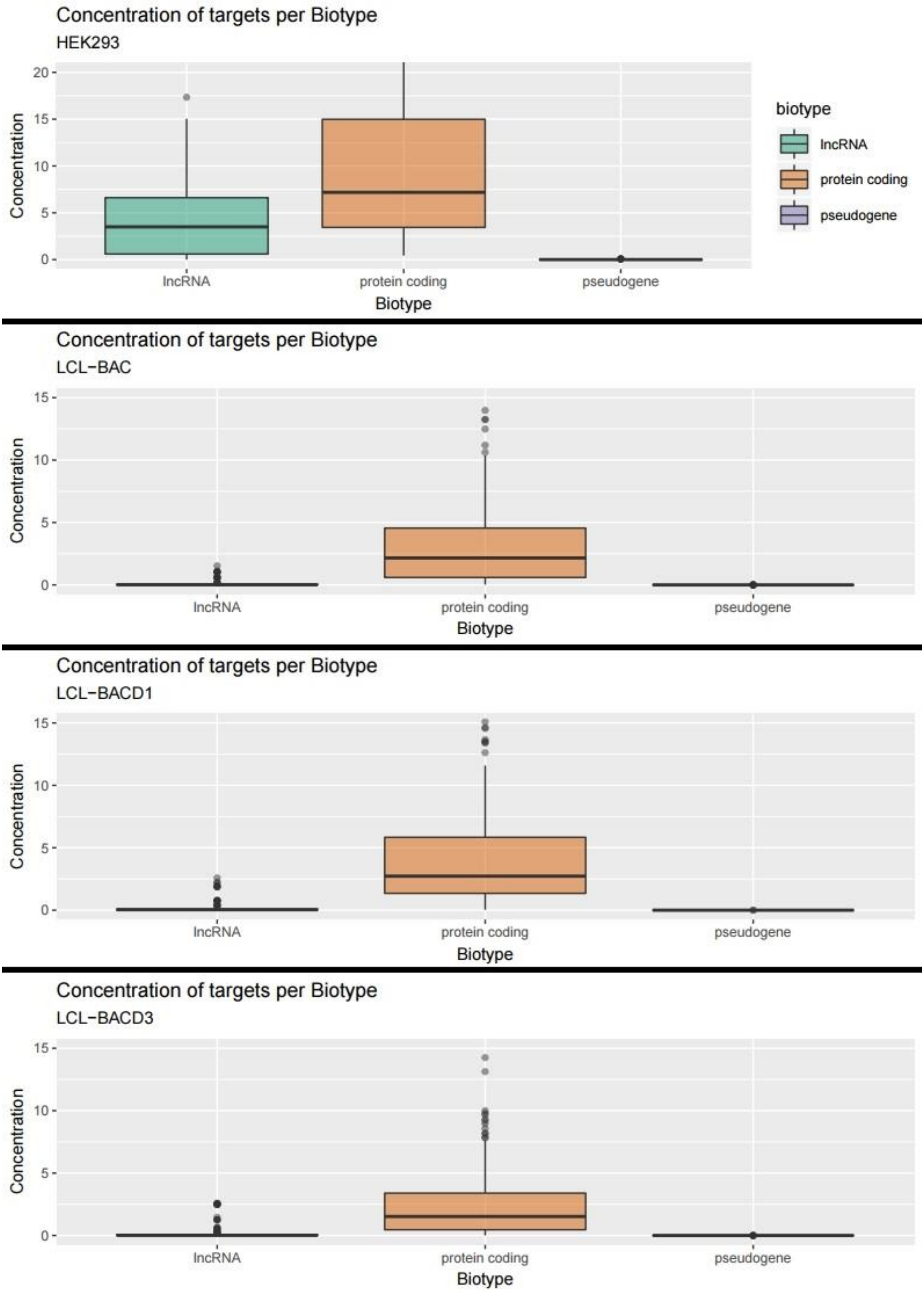

Figure 8: Concentration of targets per biotype. Concentration is calculated as ((copies/Avogadro number)/cell volume). Cell volume is approximated by a sphere of 6.5 μm radius. In all datasets protein coding targets have higher concentration meaning they are more expressed/abundant.

Table 7: Statistical importance of expression differences between biotypes for every cell type. Kruskal-Wallis test and Dunn's post-hoc test for multiple comparisons were performed for each cell type. Bonferroni adjusted p-values are presented.

		protein coding	**lncRNA**	**pseudogene**
HEK293	**protein coding**	_	_	_
	lncRNA	2.99E-06	_	_
	pseudogene	6.85E-82	8.32E-48	_
		protein coding	**lncRNA**	**pseudogene**
LCLBAC	**protein coding**	_	_	_
	lncRNA	2.67E-35	_	_
	pseudogene	4.08E-108	1.44E-21	_
		protein coding	**lncRNA**	**pseudogene**
LCLBACD1	**protein coding**	_	_	_
	lncRNA	1.97E-31	_	_
	pseudogene	2.31E-113	2.78E-27	_
		protein coding	**lncRNA**	**pseudogene**
LCLBACD3	**protein coding**	_	_	_
	lncRNA	1.25E-29	_	_
	pseudogene	1.31E-104	1.01E-24	_

Table 8: Target pool size. For every miRNA, the expression (in number of copies) of each transcript they target was added together to form the total target pool. The table presents the larger, smaller and average target pool per dataset.

Dataset	**Max target pool size (in copies)**	**Min target pool size (in copies)**	**Mean target pool size (in copies)**
HEK293	73370	349	16730
LCL-BAC	12550	14	2294
LCL-BACD1	17120	24	3124
LCL-BACD3	12330	19	1896

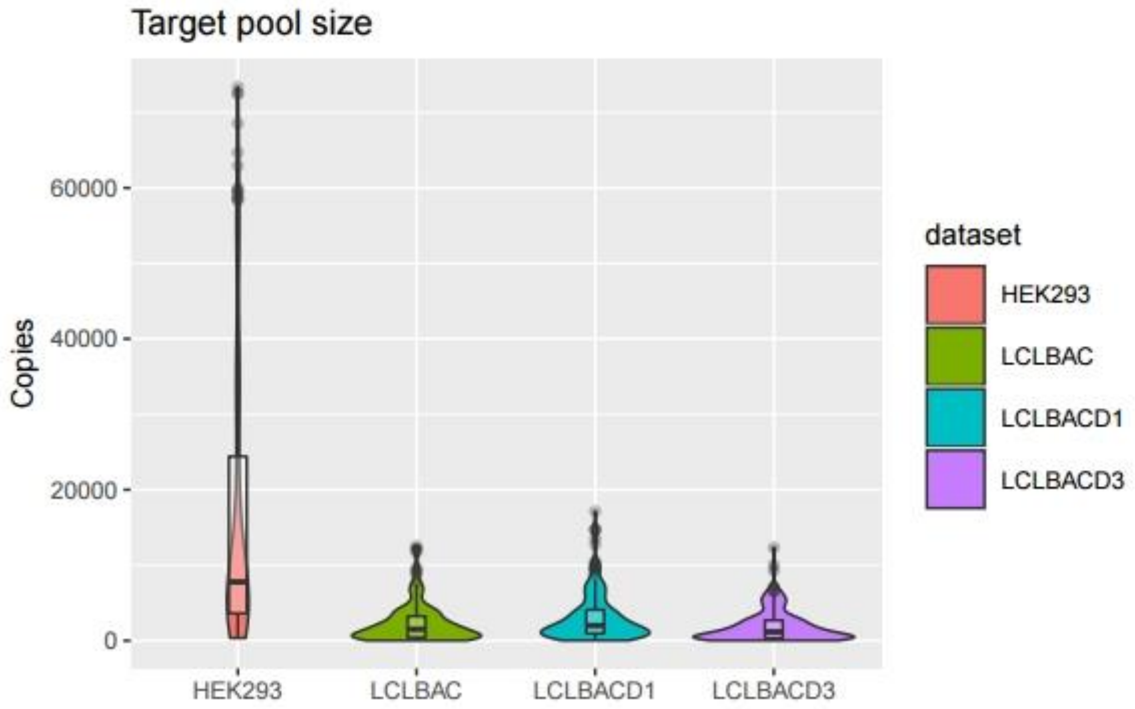

Figure 9: Target pool size for the 200 miRNAs tested per dataset. For every miRNA, the expression (in number of copies) of each transcript they target was added together to form the total target pool.

Table 9: Statistical importance of target pool expression differences across cell types. Kruskal-Wallis test and Dunn's post-hoc test for multiple comparisons were performed. Bonferroni adjusted p-values are presented. Differences that are not significant (>0.05) are highlighted.

	HEK293	LCLBAC	LCLBACD1	LCLBACD3
HEK293	_	_	_	_
LCLBAC	9.96E-38	_	_	_
LCLBACD1	6.47E-24	3.95E-02	_	_
LCLBACD3	2.67E-46	9.10E-01	1.98E-04	_

Table 10: Top 5 expressed miRNAs in each dataset. Calculated copies per cell for the miRNAs and their corresponding target pools are shown. miRNA:target ratio is calculated by dividing miRNA copies by total target copies. Total target copies (also referred to as target pool size) is estimated for every miRNA by adding together the expression (in number of copies) of each transcript they target. Even some highly expressed miRNAs are expressed at levels below their target pool, resulting in miRNA:target ratios < 1.

Dataset	miRNA	miRNA Copies	Total Target Copies	miRNA:target ratio
HEK293	hsa-miR-30a-5p	26281	15343	1.71289839
	hsa-miR-192-5p	22977	2395	9.593736952
	hsa-miR-101-3p	10500	16750	0.626865672
	hsa-let-7a-5p	8876	37869	0.234386966

	hsa-let-7f-5p	8842	30583	0.289114868
LCL-BAC	hsa-miR-21-5p	20059	2840	7.063028169
	hsa-miR-155-5p	17372	6875	2.526836364
	hsa-miR-142-3p	11338	6501	1.744039379
	hsa-miR-103a-3p	6413	4464	1.436603943
	hsa-let-7a-5p	5868	11649	0.503734226
LCL-BACD1	hsa-miR-21-5p	17525	2893	6.057725544
	hsa-miR-155-5p	13768	7727	1.781804064
	hsa-miR-16-5p	12145	4129	2.941390167
	hsa-miR-103a-3p	10784	5029	2.144362696
	hsa-miR-142-3p	9451	7153	1.321263805
LCL-BACD3	hsa-miR-21-5p	20333	4687	4.338169405
	hsa-miR-142-3p	14440	6004	2.405063291
	hsa-miR-155-5p	14185	6041	2.348121172
	hsa-miR-103a-3p	10777	3820	2.821204188
	hsa-miR-16-5p	8437	2550	3.308627451

3.4 Target binding types

microCLIP reports a wide variety of binding types based on the base pairing between the miRNA and the MRE. The most common binding type seems to be “8mer nonCanonical” which is described as base pairing in positions 1-9 with mismatch or miRNA bulge and/or a target bulge and/or a GU wobble pair (Figure 10). The main classes that include those binding types are the canonical class (bindings with perfect complementarity with the miRNA seed region) and the non-canonical class (bindings with imperfect complementarity with the miRNA seed region). The vast majority of sites in all datasets represent non canonical bindings, regardless of their biotype (Figure 11). To further investigate the abundance of canonical and non-canonical bindings, each site was weighted by the expression of its harbouring transcript. Non canonical sites seem to be more abundant than canonical sites, meaning that the transcripts with non-canonical sites are more highly expressed (Figure 12, Table 11).

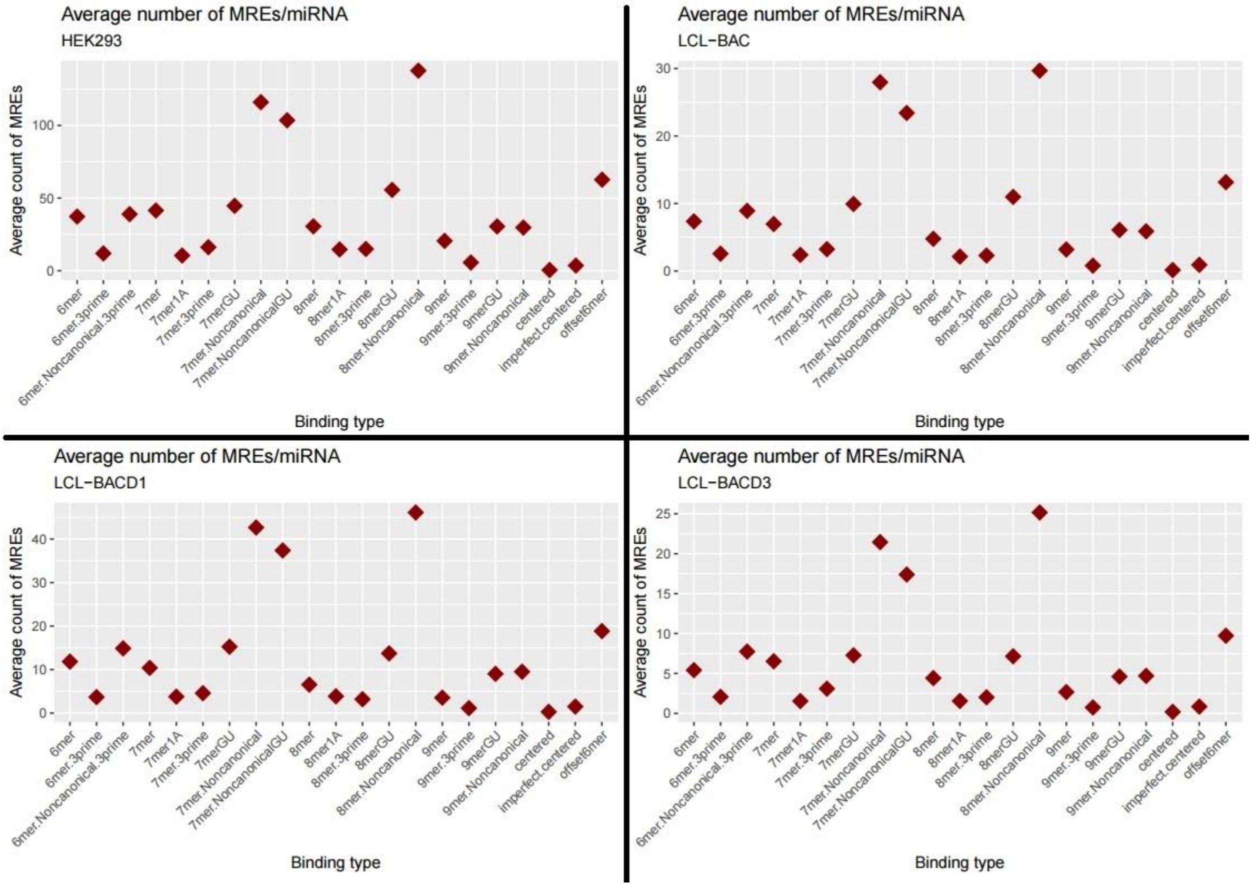

Figure 10: Figure shows the average number of MREs with a specific binding type per miRNA. For example, miRNAs in the LCL-BACD3 cell line have on average five 6mer MREs, two 6mer.3prime MREs, twenty five 8mer.Noncanonical MREs etc.

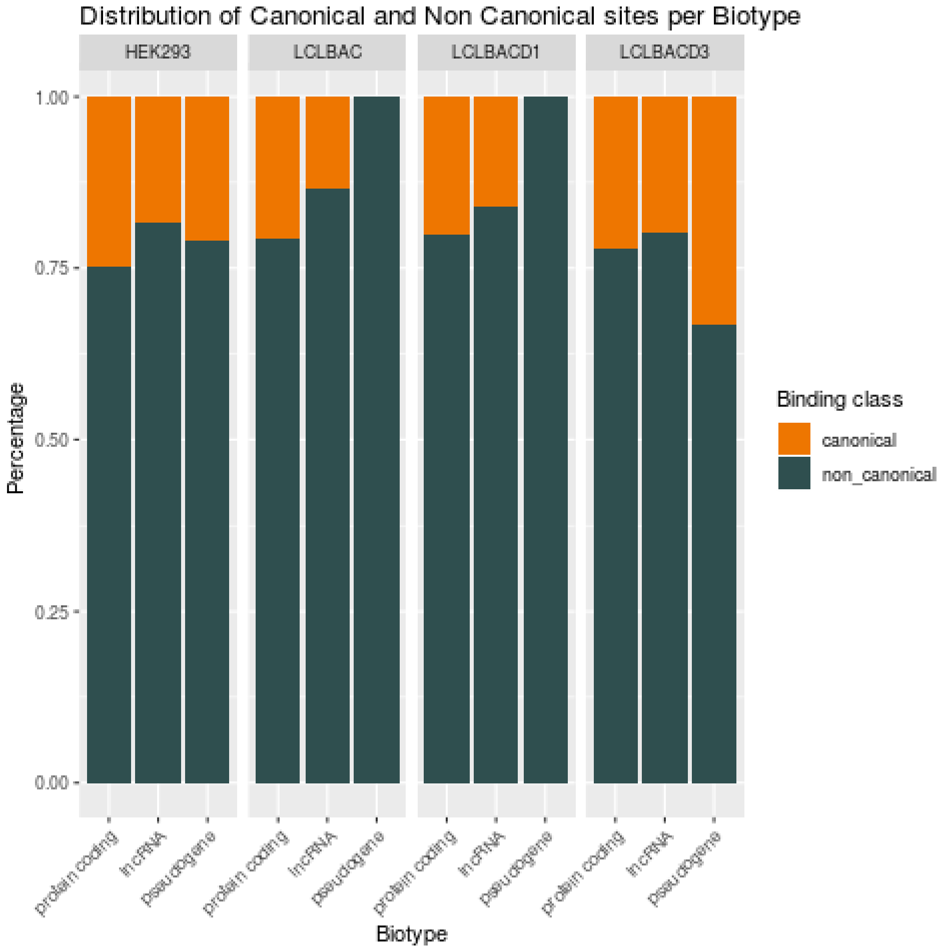

Figure 11: Distribution of canonical and non-canonical sites (MREs) per biotype. The highest percentage of sites is bound non-canonically regardless of their biotype.

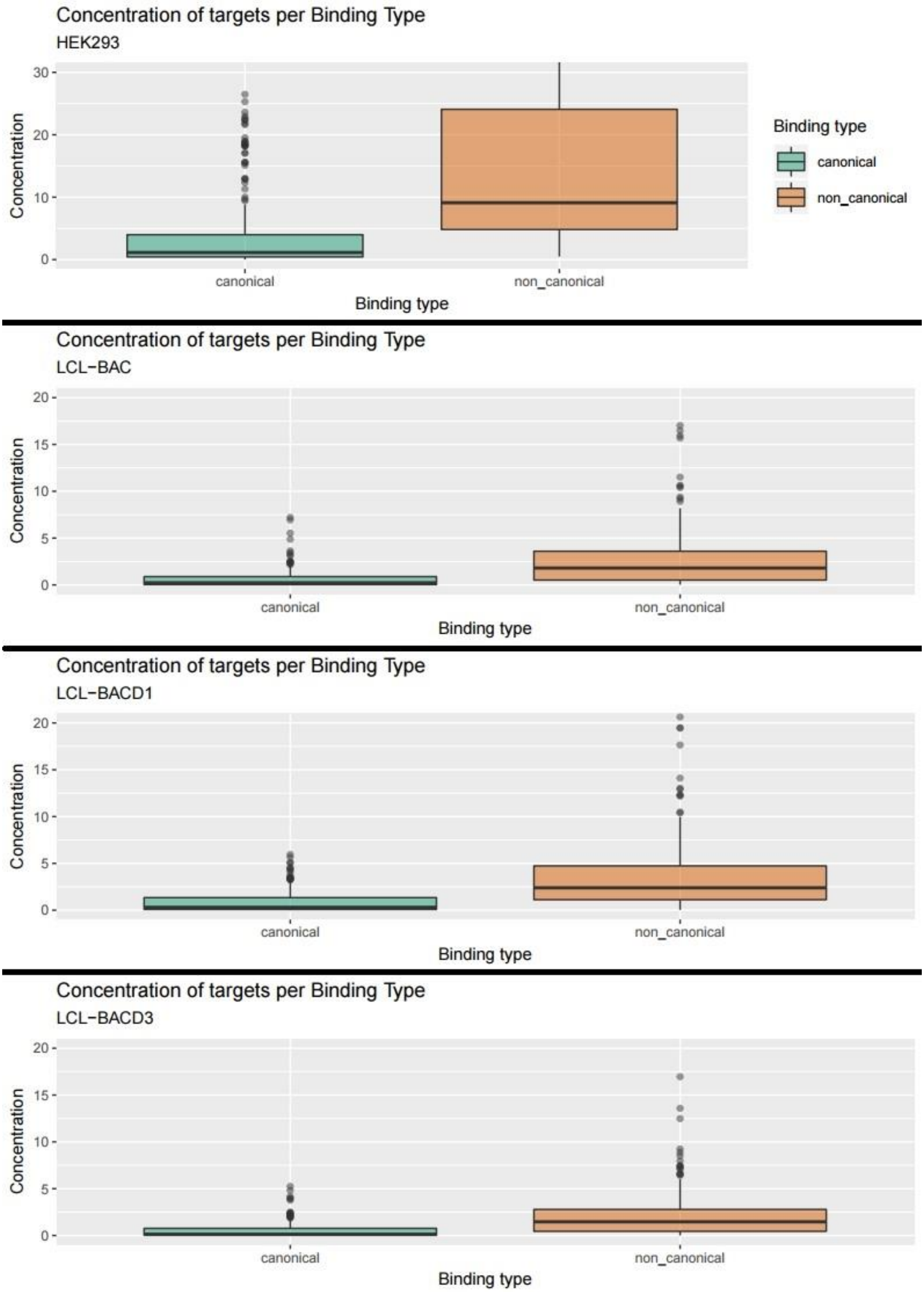

Figure 12: Concentration of targets per binding type. Concentration is calculated as ((copies/Avogadro number)/cell volume). Cell volume is approximated by a sphere of 6.5 μm radius. The concentration of every site (canonical or non-canonical) equals the concentration of its harbouring transcript. Hence, the differences between canonical and non-canonical sites

depict differences in the concentration of the transcripts. Transcripts with non-canonical sites seem to be more highly expressed.

Table 11. Statistical importance of expression differences between canonical and non-canonical sites for the different cell types. Wilcoxon rank sum test was performed.

	canonical - non canonical
HEK293	2.20E-16
LCLBAC	2.20E-16
LCLBACD1	2.20E-16
LCLBACD3	2.20E-16

3.5 Target binding strength

Strength of binding between the miRNA and its target can be measured in terms of K_d value (dissociation constant). Jens and Rajewsky built an approximate energy model to calculate mouse K_d values based on experimental measurements of fly AGO binding strength reported by [92]. Since the AGO protein is conserved in mammals, this approximate model is expected to give reasonable K_d estimates for human miRNA-target interactions. The distribution of K_d values for canonical and non-canonical sites is shown in Figure 13. As expected, canonical sites have mostly small K_d values (increased binding strength). Non canonical sites have a comparable distribution to that of canonical but large K_d values characterize only non-canonical bindings. Overall, these observations do not provide a reason to reject the approximate energy model for human miRNA-target interactions. The distribution of K_d values across the different transcript types (protein coding, lncRNAs and pseudogenes) suggests that binding strength is independent of the biotype (Figure 13). K_d distribution seems to differ between the 4 cell types (Table 12).

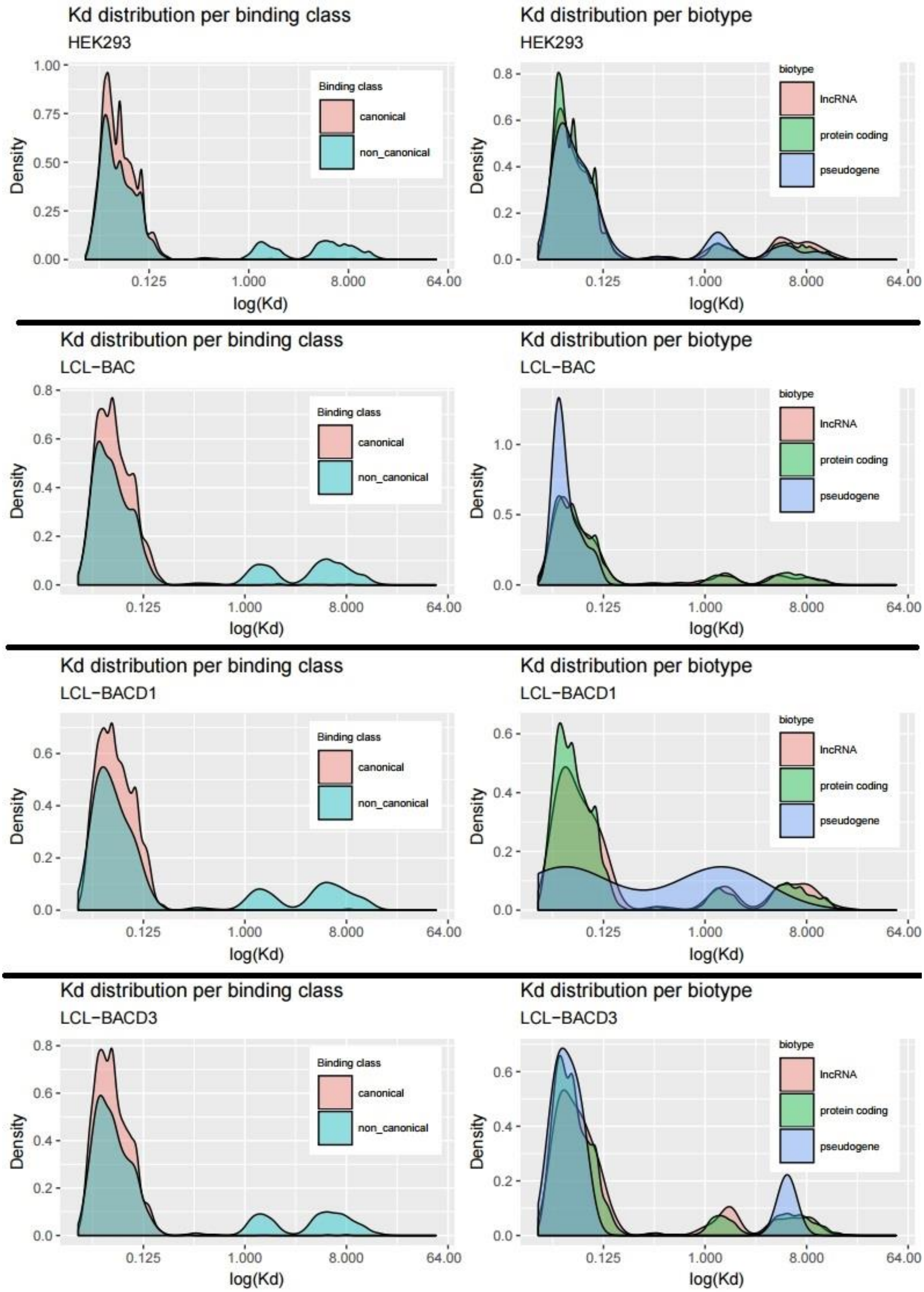

Figure 13: Kd (dissociation constant) distribution of MREs per binding class (column 1) and biotype (column 2). Kd depicts the strength of binding, with smaller values corresponding to stronger binding.

Table 12: Statistical importance of difference between the distributions of Kd values of all cell types. Two-sided Kolmogorov-Smirnov test was performed. P-values were FDR corrected.

	HEK293	**LCLBAC**	**LCLBACD1**	**LCLBACD3**
HEK293	_	_	_	_
LCLBAC	0.00E+00	_	_	_
LCLBACD1	0.00E+00	1.34E-05	_	_
LCLBACD3	0.00E+00	2.69E-03	2.20E-12	_

3.6 Employment of the mathematical model of binding site competition to identify miRNA sponges

Having estimated the abundance and affinity of all target sites for every miRNA (both canonical and non-canonical sites were included), the quantitative model was employed to predict the occupancy of protein coding targets in two cases:

1. When the whole target pool was taken into account (protein coding and lncRNA targets)
2. When all lncRNA sites were excluded

This analysis aimed to measure in what extent lncRNA sites titrate miRNAs away from the other protein coding targets.

3.6.1 Changes in mRNA occupancies upon lncRNA exclusion

After excluding lncRNA sites from the target pool of every miRNA, protein coding sites showed a variety of changes for different miRNAs (Figure 14, Figure 15). Targets of miR-200a-3p did not seem to be affected and were bound at similar levels as before. On the other hand, miR-199a-5p targets were occupied more highly after lncRNA exclusion. The same pattern was observed for let-7d-5p and its targets. The difference between miR-199a and let-7d is the absolute value of site occupancy after the exclusion of lncRNAs. For miR-199a even though the differences were profound, final target site occupancy did not exceed 0.02 (2%), meaning that only 2% of the most affected site was occupied. For let-7d, the differences were observed in a more relevant window with occupancies shifting from 20% to 35% for some sites. Changes in the regulatory effect of miRNAs are more likely to be observed for targets occupied in a relevant window from 5% to 95%. Thus, in order for lncRNAs to effectively alter the fate of other competing mRNAs, occupancy changes of protein coding sites should be in that relevant window. Following this reasoning, an appropriate threshold was defined to distinguish between miRNAs with potential lncRNA sponges and miRNAs where lncRNA expression is not able to meaningfully affect other targets (see 2.2.6.2).

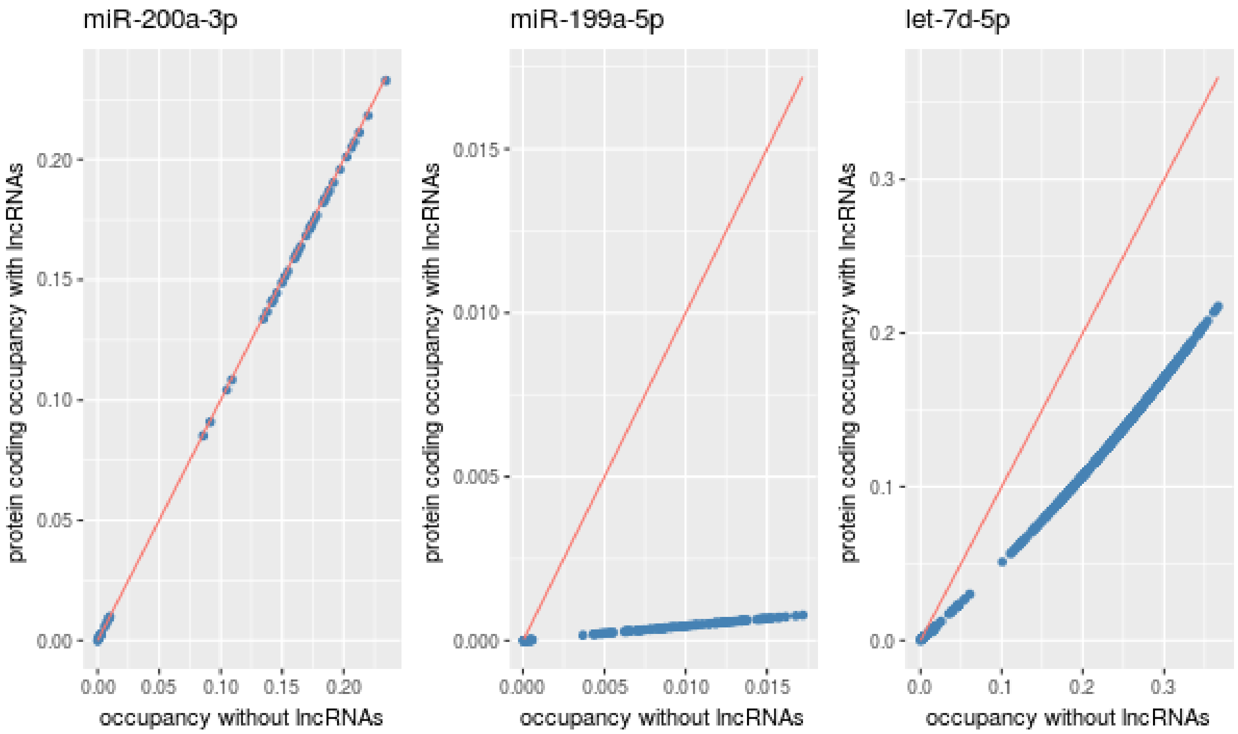

Figure 14: Altered occupancies of protein coding sites when all lncRNA targets are excluded from the target pool. The figure depicts the occupancy of every protein coding site (that is the fraction of the site bound by the miRNA), when lncRNAs are considered as part of the target pool and when they are excluded. Red line is a reference line representing the occupancies if there was no change. For miR-200a, exclusion of lncRNAs does not affect the occupancy of protein coding sites. On the other hand, exclusion of lncRNAs from the miR-199-5p and let-7d-5p target pools leads to increased binding of protein coding sites. Data from the HEK293 dataset. Differences for all three miRNAs were statistically important (Wilcoxon signed rank test <2.2e-11)

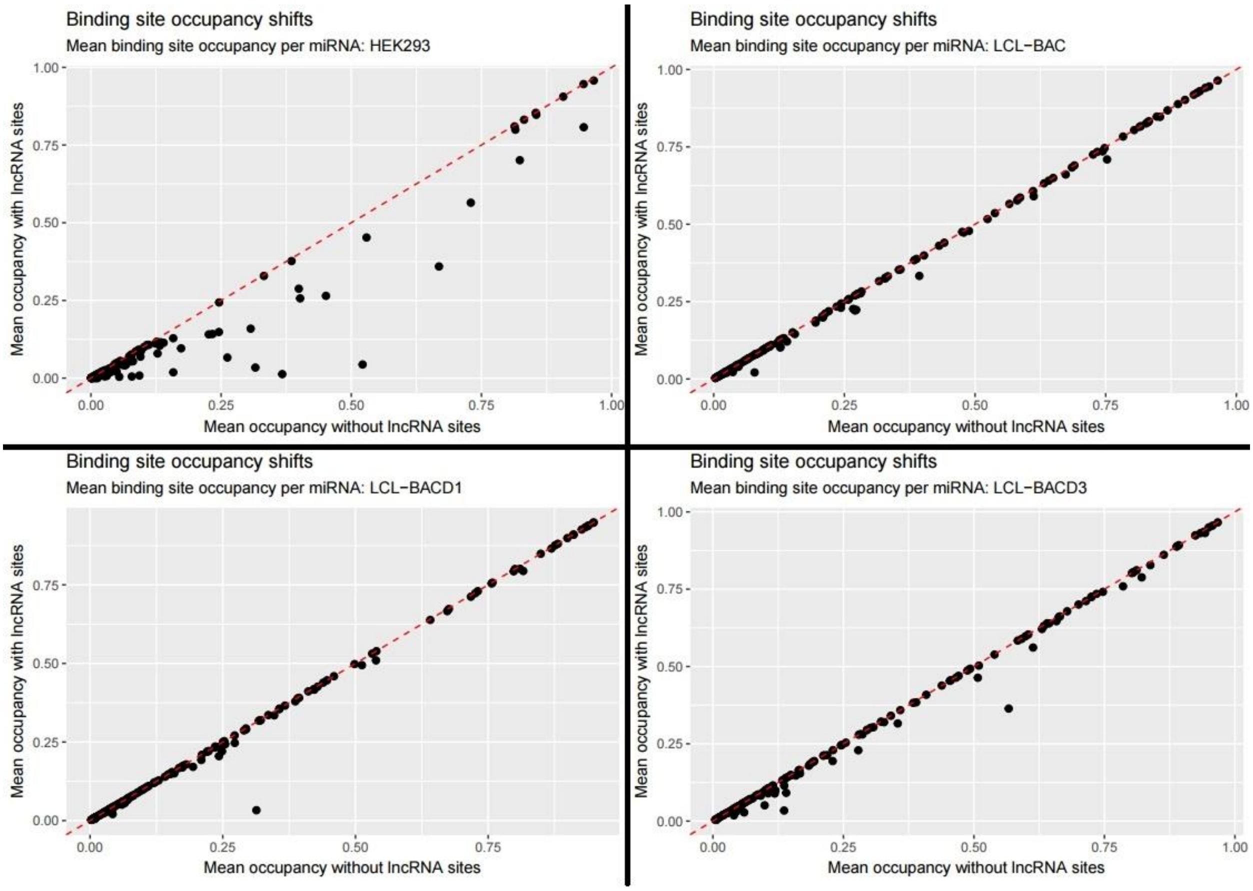

Figure 15: Every miRNA occupies to some extent its targets. For every miRNA, mean occupancy of all targets is plotted in the absence and presence of lncRNAs. Red dotted line is a reference line representing the mean occupancy of the miRNA targets if there is no change. miRNAs in the HEK293 dataset seem to be more affected by the lncRNA exclusion and some of them occupy their targets more upon lncRNA loss. Differences for all cell types were statistically important (Wilcoxon signed rank test <2.2e-16)

3.6.2 lncRNA sponges

For miRNAs with altered protein coding site occupancies upon exclusion of the whole lncRNA target pool, the potential of each and every lncRNA to contribute to these changes was evaluated separately. Occupancies of protein coding sites were calculated first by taking into account the whole target pool (all lncRNAs and mRNAs, same as before) and then by taking into account the whole target pool (again lncRNAs and mRNAs) except sites of the lncRNA gene of interest. Upon exclusion of a specific lncRNA, protein coding sites were either unaffected or more highly occupied (Figure 16). By applying the same threshold as before, following the reasoning that occupancy changes should be in a relevant window between 5% and 95% to be measurable and meaningful, lncRNAs that caused changes of that magnitude were deemed as miRNA sponges.

This analysis resulted in the identification of 8 lncRNAs that can act as sponges for a total of 38 miRNAs (Figure 17, Figure 18). Among the reported sponges, XIST and MALAT1 are two well-studied lncRNAs with implications in chromosome X inactivation [16] and splicing [33] respectively. Their identification as miRNA sponges is somewhat controversial due to their localization on the nucleus. Regulation through competition for miRNA binding presupposes co-localization of the competitors. Nevertheless, these findings are supported by recent reports on the sponging capabilities of both XIST and

MALAT1 (Table 13). Overall, 6 out of the 8 lncRNAs have already been reported as sponges (Table 13).

XIST, MALAT1 and RMRP can sponge multiple miRNAs from different miRNA families (Figure 18). In the case of XIST, 12 miRNAs are sponged corresponding to miR-101, let-7, miR-191, miR-26, miR-30 and miR-23 families in the HEK293 cell line.

The majority of lncRNAs were unable to alter the fraction of protein coding sites that were bound by their shared miRNA and only highly expressed lncRNAs could function as sponges (Figure 19).

Highly expressed miRNAs like miR-101-3p and miRNAs of the let-7 family were susceptible to sponging effects by one or more lncRNAs, indicating that high expression is not enough for stable control over their targets. On the other hand, the relative abundance between the miRNA and its target pool (miRNA:target ratio) seems to be a better predictor for miRNA resistance to sponging effects, since none of the miRNAs that were expressed at levels above their target pool has lncRNA sponges (Figure 20).

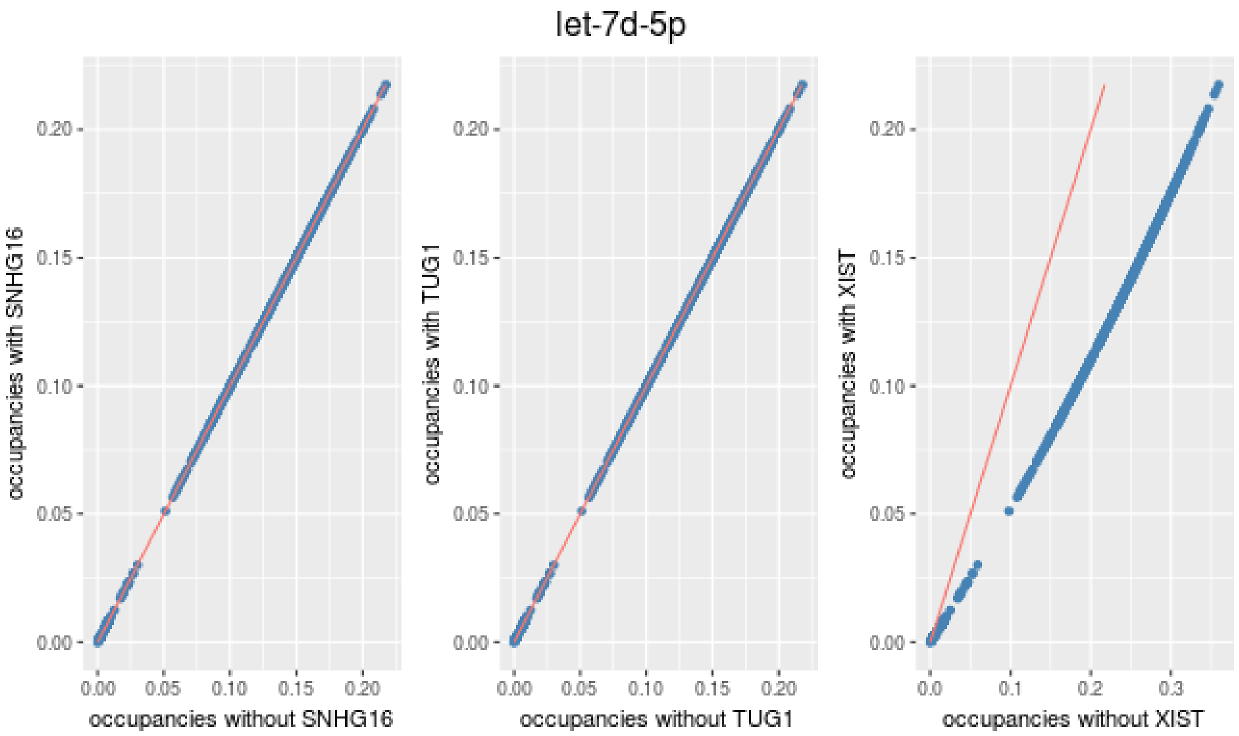

Figure 16: Altered occupancies of protein coding sites when only one lncRNA gene is excluded from the target pool. The figure depicts the occupancy of every protein coding site (that is the fraction of the site bound by the miRNA), when all sites of an lncRNA gene (meaning sites in all transcripts) are considered as part of the target pool and when they are excluded. Red line is a reference line representing the occupancies if there was no change. For let-7d-5p, exclusion of SNHG16 or TUG1 sites does not lead to increased binding of the protein coding sites but exclusion of XIST does. Data from HEK293 dataset. Differences for the three miRNAs were statistically important (Wilcoxon signed rank test < 2.2e-16)

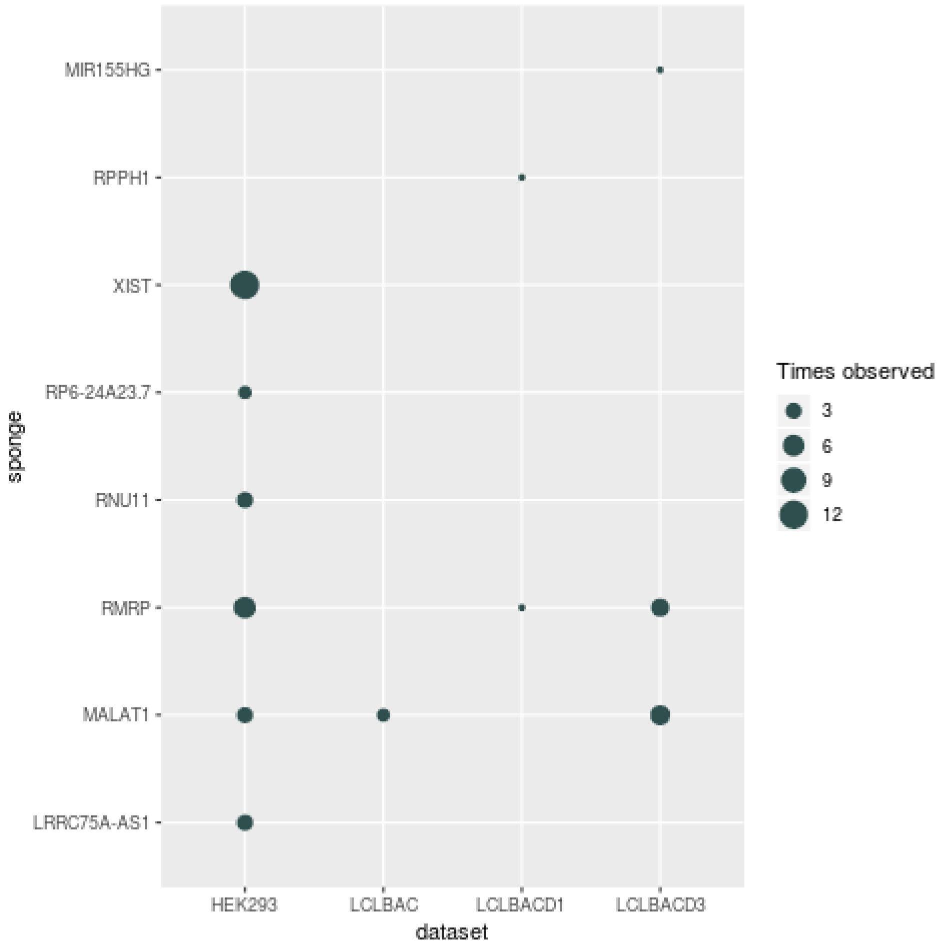

Figure 17: Identified sponges. Some of them are observed as sponges for multiple miRNAs.

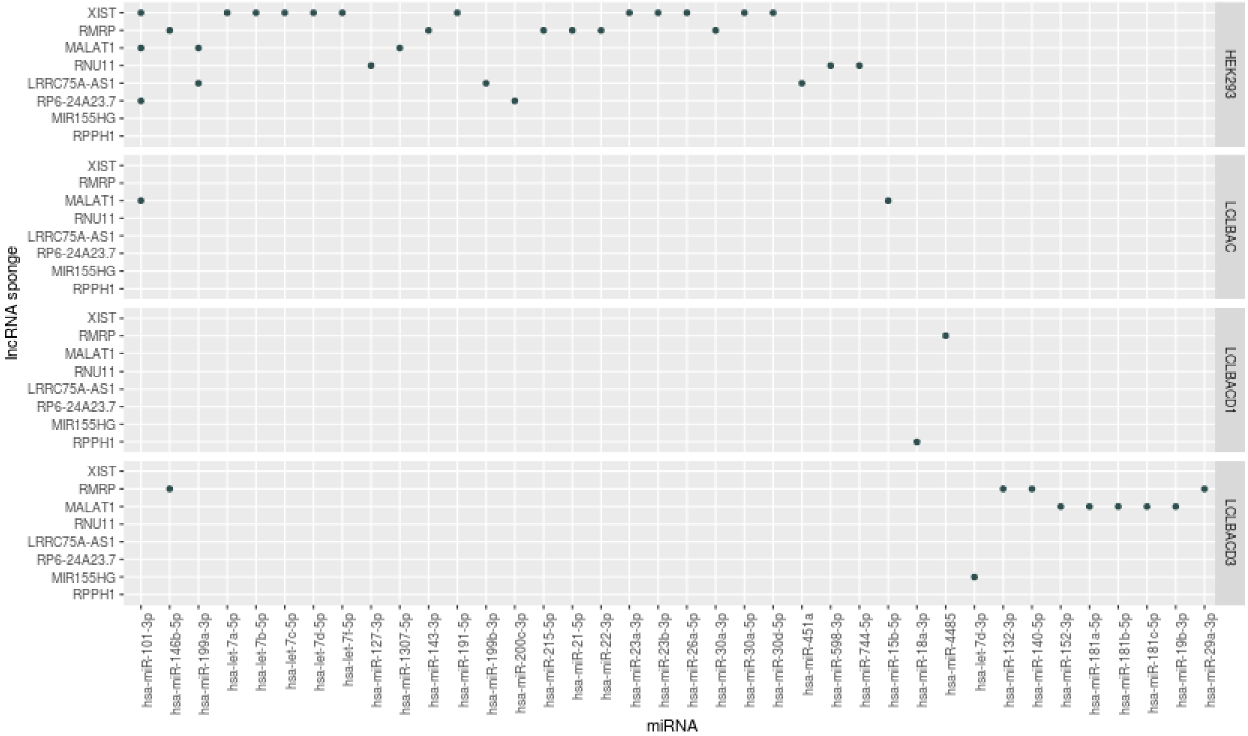

Figure 18: Identified sponges and their corresponding miRNAs. Some lncRNAs may act as sponges for multiple miRNAs and some miRNAs may have multiple sponges.

Table 13: 6 out of the 8 identified sponges have been previously reported as miRNA sponges in different disease settings. Table shows the lncRNA sponge, the miRNA that it titrates (if it is specified) and the publication that reports the interaction.

lncRNA	miRNA	publication
RPPH1	miR-326-3p	[93]
	miR-330-5p	
miR-155HG	miR-185	[94]
LRRC75A-AS1		[95]
XIST	miR-181a	[96]
	miR-133a	[97]
	miR-137	[98]
	miR-194-5p	[99]
MALAT1	miR-34a	[100]
	miR-200c	[101]
		[102]
	miR-211	[103]
	miR-30a-5p	[104]
RMRP	miR-206	[105]

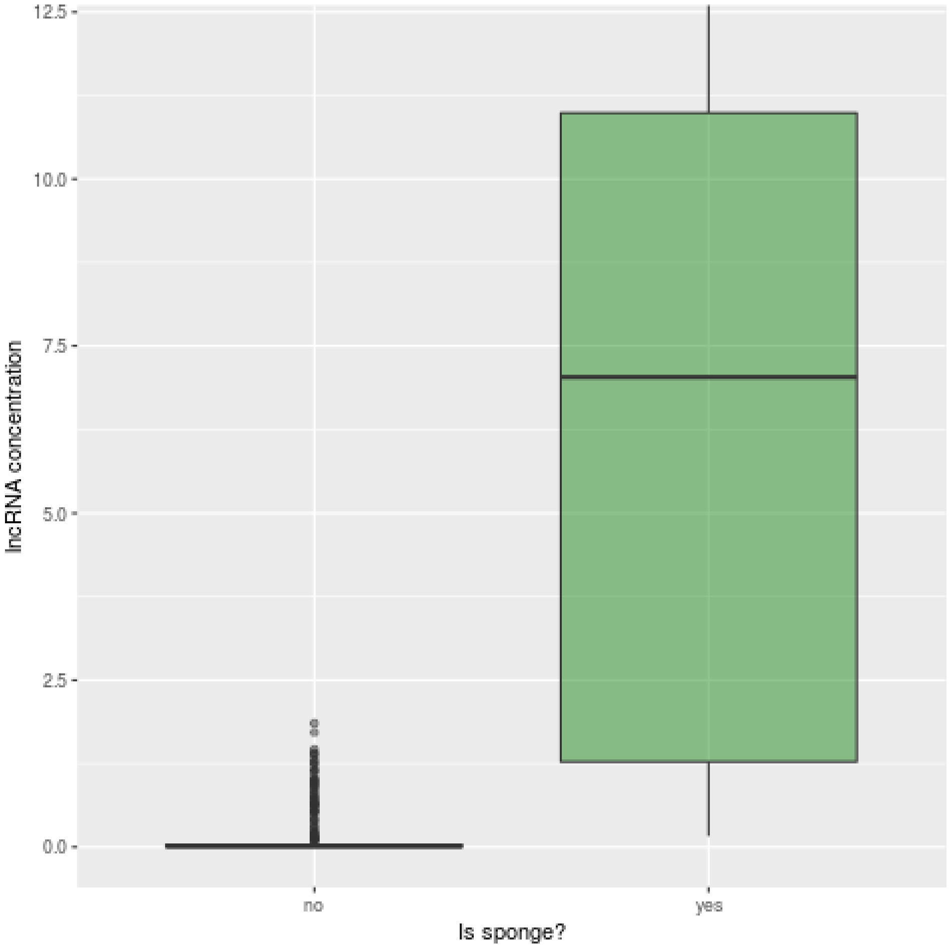

Figure 19: Concentration of all lncRNAs identified as sponges compared to the concentration of the rest targeted lncRNAs. Sponges are more highly expressed. Wilcoxon rank sum test p-value = 8.199e-16.

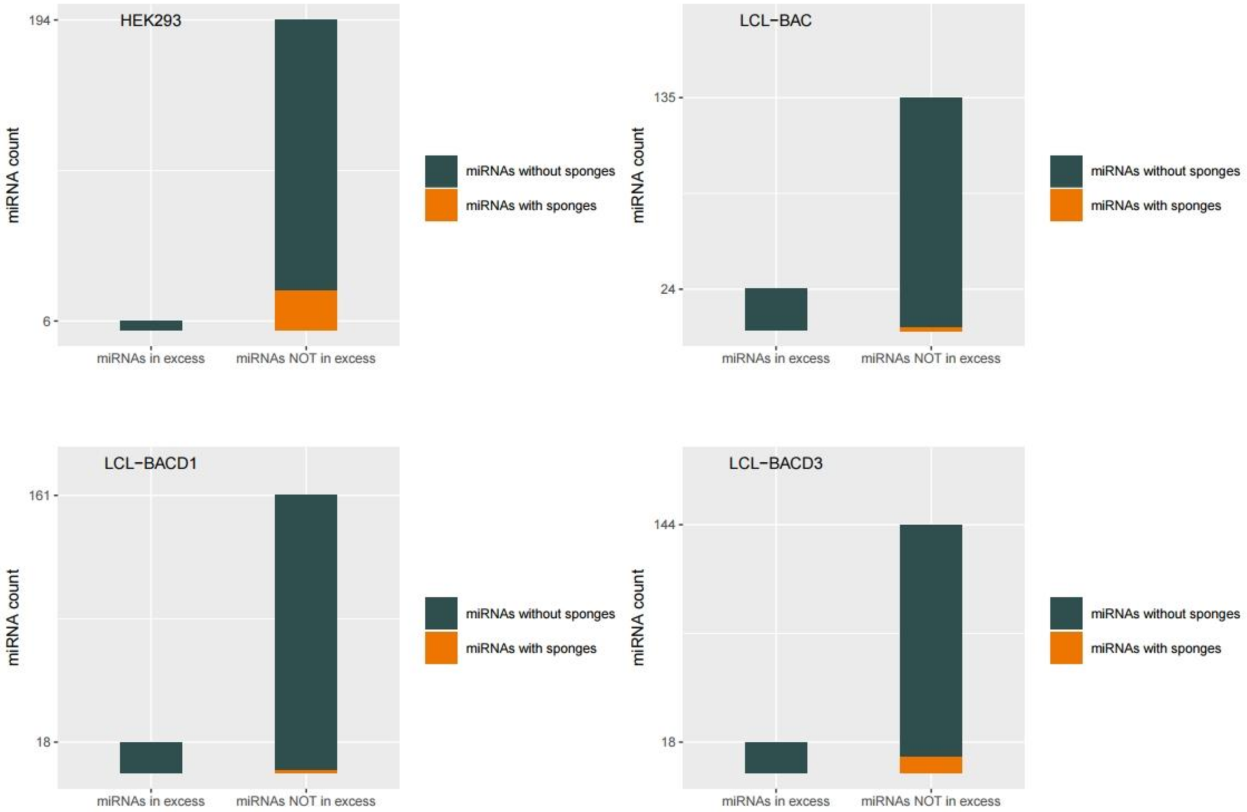

Figure 20: miRNAs with lncRNA targets are devided in two categories based on whether they are expressed above their target pool (are in excess of their targets) or not. Expressed above their target pool means that the miRNA copies are more than the total copies of all targeted transcripts. Those categories are subdivided based on whether they contain a miRNA with a sponge. miRNAs that are in excess of their targets do not have sponges.

3.7 Repeating the analysis for HEK293 cell line by considering only canonical sites

When modeling competition, several teams have relied only on canonical bindings to estimate target pools or have restricted their targets to computationally predicted, conserved sites on 3'UTRS [77, 78, 106]. Non canonical interactions would not generally be identified by target prediction programs, which are biased toward canonical seed interactions [107]. On the other hand, microCLIP [10] provides a wide range of non-canonical bindings which constitute the majority of targeted MREs and were all included in the estimation of miRNA target pools. To investigate whether target pools should include non-canonical sites or canonical sites are enough to model competition, the analysis for sponge identification was repeated using only the reported canonical sites for the HEK293 dataset.

This resulted in extensive changes on the reported pairs of miRNA – lncRNA sponge. 58 miRNA-sponge pairs were identified when only canonical sites where considered compared to 28 when mixed targets were included in the target pool (Figure 21). From those pairs only 6 were common and 15 additional lncRNAs were reported as sponges. Those changes may be attributed to the increased concentration ratio of specific lncRNAs and the rest of the targets. For example, if an lncRNA constitutes 5% of the total target pool in terms of expressed copies, and after excluding all non-canonical sites it constitutes

25% of the total target pool, then it is more likely to cause changes in the binding of other sites when excluded (Figure 22).

Overall, the model becomes more sensitive to sponging effects when only canonical sites are included. In the context of a biological network, this scenario is less likely to occur. Regulation by the sponging activity of a single gene is not observed in that extent in nature.

Based on these findings, an appealing emerging hypothesis is that non-canonical sites contribute to the robustness of the regulatory network by reducing the ability of a single gene to act as sponge and alter the occupancy of other targets.

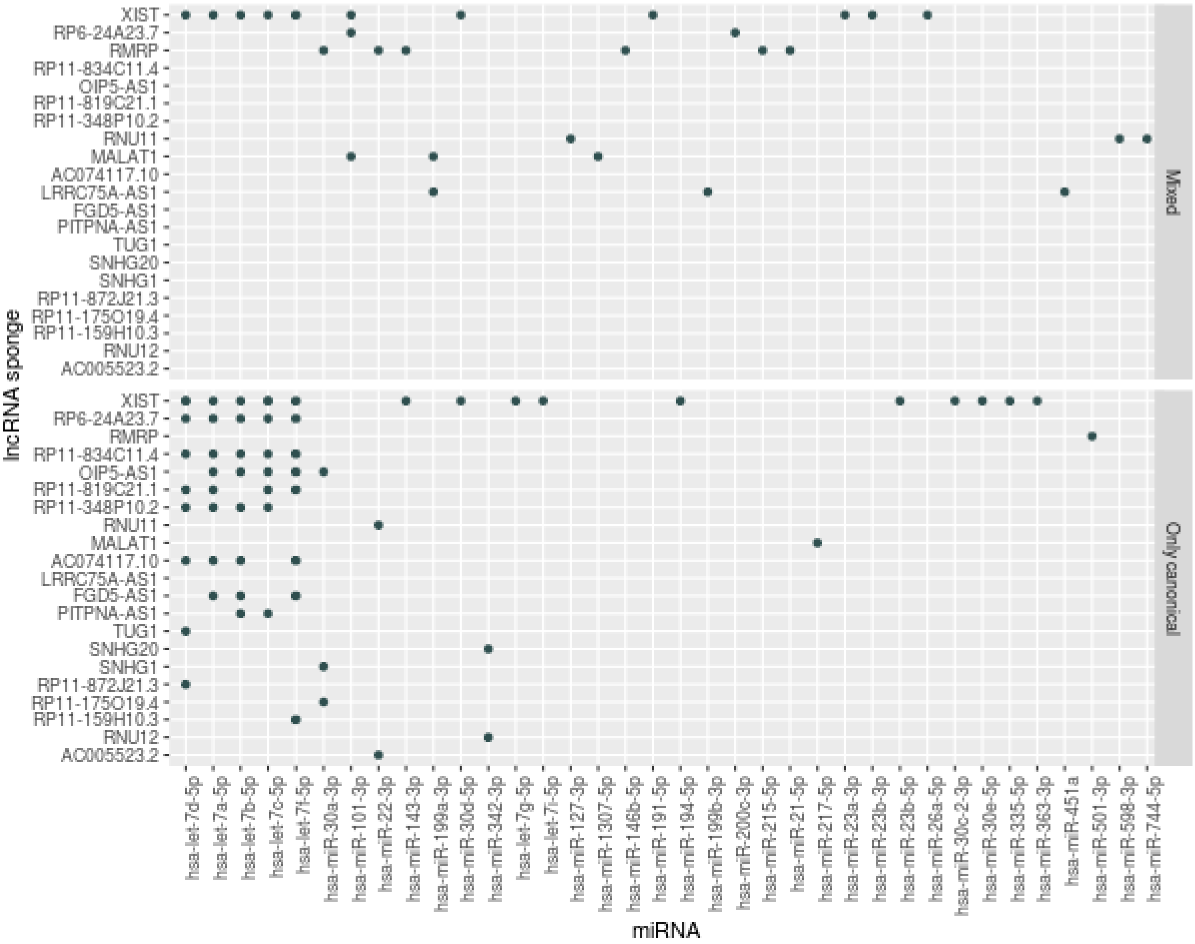

Figure 21: Identified miRNA-sponge pairs when all sites (canonical and non-canonical) are considered to be part of the target pool compared to identified miRNA-sponge pairs when only canonical sites are considered.

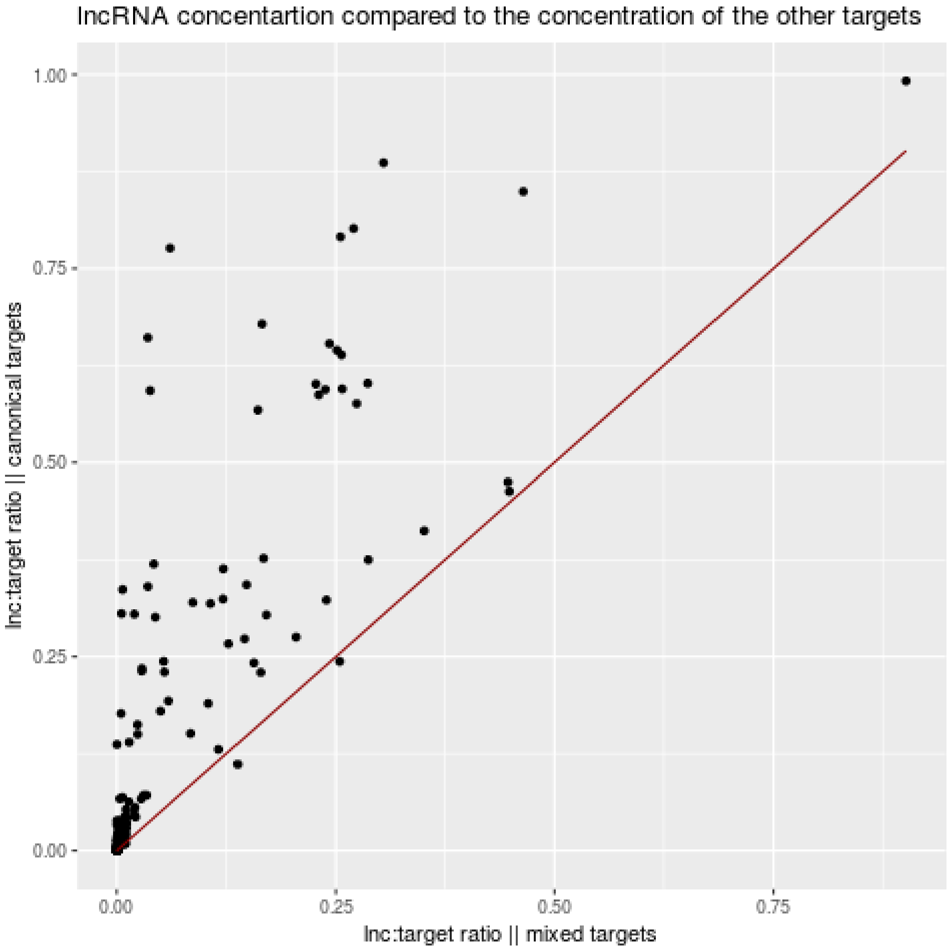

Figure 22: For every lncRNA that is targeted by a specific miRNA, the % ratio of its concentration to the concentration of the other targets is plotted in two cases: first when the target pool includes both canonical and non-canonical sites (mixed targets) and second when the target pool includes only canonical sites. When only canonical sites are considered, the majority of lncRNAs seem to have increased concentration ratios, meaning they make up a bigger part of the target pool than before.

4. DISCUSSION

Computational methodologies that rely on sequence characteristics or number of shared miRNAs and MREs to score the interaction between a potential sponge and an affected transcript, cannot account for the cell type-, tissue type-, developmental stage and disease-specific changes in expression patterns of competing RNAs. Thus, they are not dynamic and cannot predict in what extent changes in the concentration of a transcript can affect other targets. The same problem arises in approaches that rely on correlation of expression between the potential sponge and the other co-targeted transcripts. An additional disadvantage of those approaches is that correlation of expression is an indirect way to assess the sponging potential of a transcript since it does not prove that the outcome is due to increased binding of the miRNA to the sponge.

Quantitative models that predict the fraction of sites that are bound by a specific miRNA can address these issues and have been used to explore the competition between targets. However they have not been used to identify endogenously expressed sponges.

Here, a computational approach for identifying lncRNAs that act as miRNA sponges was presented. By employing a mathematical model of target site occupancies and integrating small RNA-seq, RNA-seq and PAR-CLIP data the ability of lncRNAs to titrate miRNAs away from other protein coding targets was assessed. This analysis was performed for 4 different cell lines (HEK293, LCL-BACD1, LCL-BACD2, LCL-BACD3) and 200 miRNAs in each were examined.

Overall, 8 lncRNAs were identified as sponges. 6 out of 8 had been previously reported as sponges in different disease settings (Table 13). Nuclear lncRNAs, XIST and MALAT1 were among the identified sponges and even though their sponging potential has been observed, further investigation is needed.

A previous study that used quantitative measurements of target and miRNA abundance to assess the ceRNA hypothesis, concluded that highly expressed miRNAs are not affected by physiological changes in ceRNA expression [106]. Here, it was shown that even for highly expressed miRNAs like miR-101-3p and miRNAs of the let-7 family, expression of an lncRNA at its endogenous levels was sufficient to cause measurable changes on the occupancies of other protein coding targets (Table 10, Figure 18). Thus, it can be concluded that it is not only a matter of how highly a miRNA is expressed but also a matter of its abundance compared to that of its targets (miRNA:target ratio). The susceptibility of miRNAs to competition effects (like sponging) depending on miRNA:target ratios has been demonstrated by another study where they used single cell measurment of miRNA activity for different target abundances [78].

Although some lncRNAs displayed sponge functionality, the abundance of most individual targets was insufficient to alter free miRNA levels and only highly expressed lncRNAs were identified as sponges. This observation is in line with the predictions of other quantitative models [77, 106].

Several teams estimated target pools by taking into account only canonical or computationally predicted, conserved sites [77, 78, 106]. Reported interactions from microCLIP provide evidence that the majority of binding sites are non-canonical (Figure 11). Widespread non-canonical bindings have been reported for miR-155 in T-cells where 40% of Argonaute binding occurs at sites without perfect seed matches. These non-canonical sites were estimated to confer regulation of gene expression, however less potently than canonical sites [108]. Analysis of CLASH data also reported extensive non canonical binding (around 60% of seed interactions were non-canonical) [107], further supporting the idea that target pools without non-canonical sites are underestimated and not very representative. When the quantitative, mathematical model was employed with target pool estimates relying only on canonical bindings, a large number of additional sponges was reported. The increased sensitivity of the model to sponging effects is less likely to be biologically relevant, since such extensive regulation by sponges has not been observed (at least not yet) in nature. These observations lead to the hypothesis that non-canonical sites contribute to the robustness of the regulatory network and thus they should be taken into account when estimating miRNA target pools.

4.1 Future directions

1. **Separate analysis of nuclear and cytoplasmic RNAs** is needed to accurately estimate interacting miRNA and target pool abundance.

2. **Model extension**. The same way each miRNA target reduces the amount of the miRNA available to other sites, each miRNA reduces the amount of free target RNA available to other miRNAs. Currently, the model examines one miRNA species at a time and reduction of targeted RNAs due to miRNA –miRNA competition is not taken into account. This may lead to overestimated target abundance and subsequent underestimation of the miRNA:target ratio and miRNA regulatory capacity. Interactions between miRNAs and RNAs form a complex regulatory network where each miRNA regulates up to hundreds of targets and each target can be regulated by multiple miRNAs. To capture this complexity, the current model should be expanded to account for miRNA competition too. Figure 23 and Figure 24 demonstrate how the model “sees” the network currently and how it should treat it after the extension.

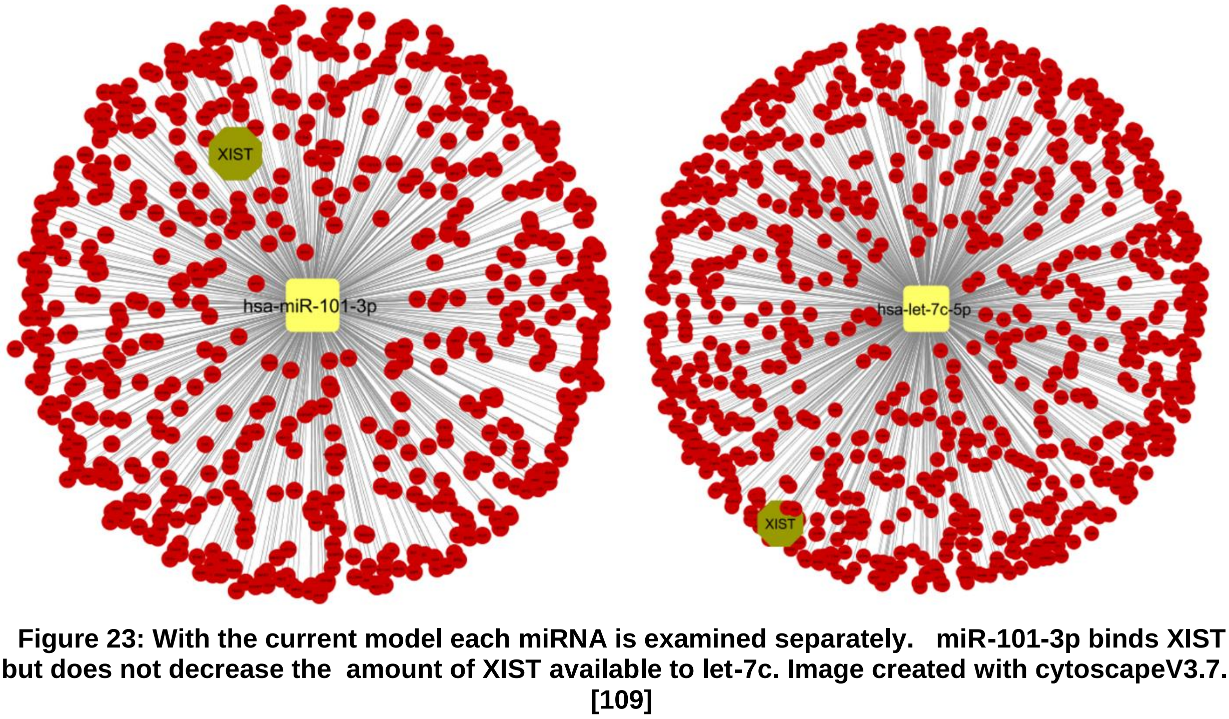

Figure 23: With the current model each miRNA is examined separately. miR-101-3p binds XIST but does not decrease the amount of XIST available to let-7c. Image created with cytoscapeV3.7.1 [109]

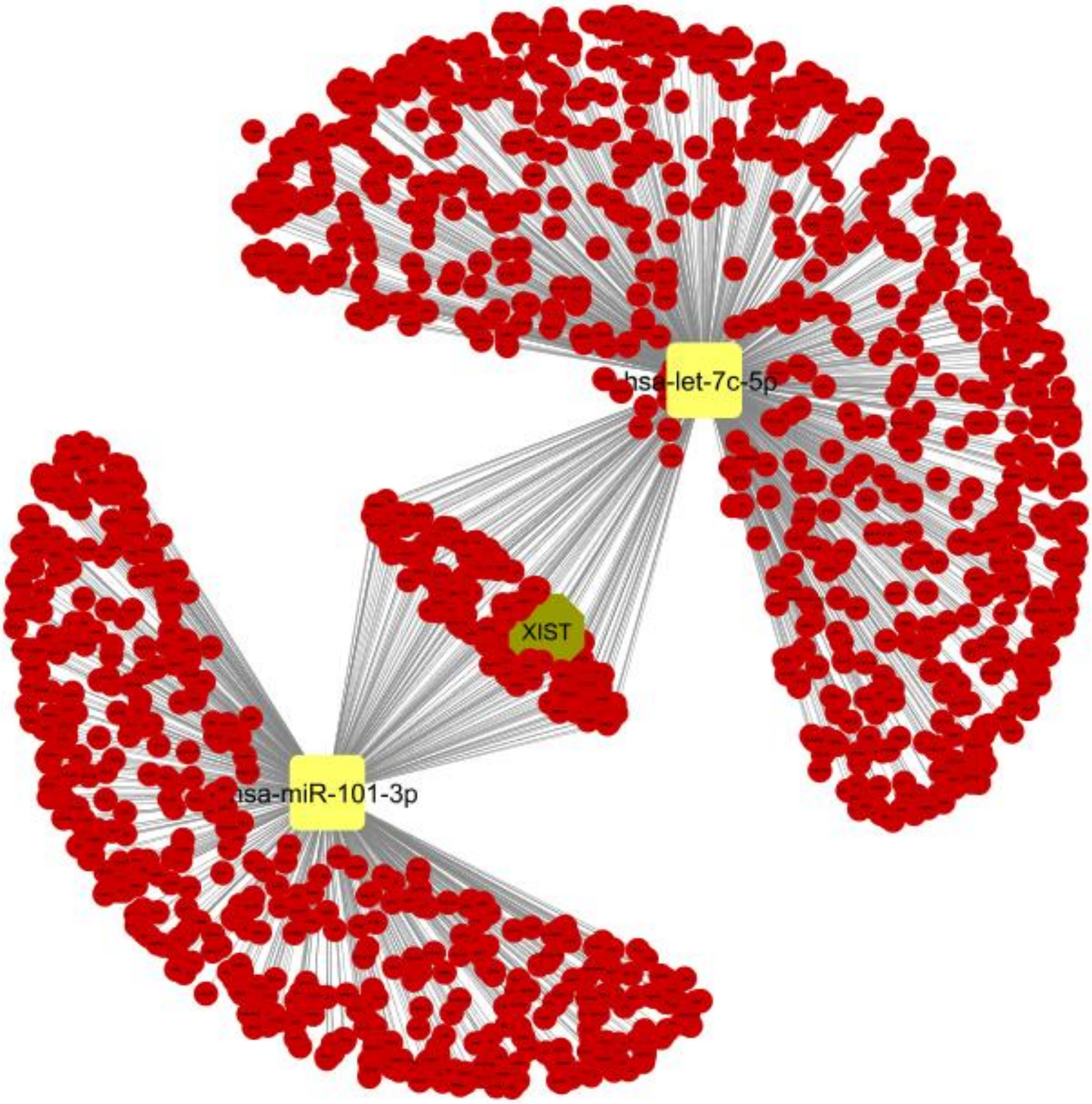

Figure 24: miRNAs and their targets form a network. Each miRNA reduces the amount of shared targets available to the other. Image created with cytoscape V3.7.1

5. REFERENCES

[1] N. Bushati and S. M. Cohen, "microRNA functions," (in eng), no. 1081-0706 (Print).

[2] R. W. Carthew and E. J. Sontheimer, "Origins and Mechanisms of miRNAs and siRNAs," (in eng), no. 1097-4172 (Electronic).

[3] L. He and G. J. Hannon, "MicroRNAs: small RNAs with a big role in gene regulation," (in eng), no. 1471-0056 (Print).

[4] M. Hafner *et al.*, "Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP," (in eng), no. 1097-4172 (Electronic).

[5] E. van Rooij, X. Sutherland Lb Fau - Qi, J. A. Qi X Fau - Richardson, J. Richardson Ja Fau - Hill, E. N. Hill J Fau - Olson, and E. N. Olson, "Control of stress-dependent cardiac growth and gene expression by a microRNA," (in eng), no. 1095-9203 (Electronic).

[6] A. K. Leung and P. A. Sharp, "MicroRNA functions in stress responses," (in eng), no. 1097-4164 (Electronic).

[7] H. Dvinge *et al.*, "The shaping and functional consequences of the microRNA landscape in breast cancer," (in eng), no. 1476-4687 (Electronic).

[8] C. Guay, V. Roggli E Fau - Nesca, C. Nesca V Fau - Jacovetti, R. Jacovetti C Fau - Regazzi, and R. Regazzi, "Diabetes mellitus, a microRNA-related disease?," (in eng), no. 1878-1810 (Electronic).

[9] A. M. Gurtan and P. A. Sharp, "The role of miRNAs in regulating gene expression networks," (in eng), no. 1089-8638 (Electronic).

[10] M. D. Paraskevopoulou, D. Karagkouni, I. A.-O. X. Vlachos, S. A.-O. Tastsoglou, and A. G. Hatzigeorgiou, "microCLIP super learning framework uncovers functional transcriptome-wide miRNA interactions," (in eng), no. 2041-1723 (Electronic).

[11] D. M. Anderson *et al.*, "A micropeptide encoded by a putative long noncoding RNA regulates muscle performance," (in eng), *Cell,* vol. 160, no. 4, pp. 595-606, Feb 12 2015, doi: 10.1016/j.cell.2015.01.009.

[12] A. Matsumoto, J. G. Clohessy, and P. P. Pandolfi, "SPAR, a lncRNA encoded mTORC1 inhibitor," (in eng), *Cell Cycle,* vol. 16, no. 9, pp. 815-816, May 3 2017, doi: 10.1080/15384101.2017.1304735.

[13] T. Derrien *et al.*, "The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression," (in eng), *Genome Res,* vol. 22, no. 9, pp. 1775-89, Sep 2012, doi: 10.1101/gr.132159.111.

[14] T. R. Mercer, M. E. Dinger, S. M. Sunkin, M. F. Mehler, and J. S. Mattick, "Specific expression of long noncoding RNAs in the mouse brain," (in eng), *Proc Natl Acad Sci U S A,* vol. 105, no. 2, pp. 716-21, Jan 15 2008, doi: 10.1073/pnas.0706729105.

[15] M. N. Cabili *et al.*, "Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution," (in eng), *Genome Biol,* vol. 16, p. 20, Jan 29 2015, doi: 10.1186/s13059-015-0586-4.

[16] L. P. Benoit Bouvrette *et al.*, "CeFra-seq reveals broad asymmetric mRNA and noncoding RNA distribution profiles in Drosophila and human cells," (in eng), *Rna,* vol. 24, no. 1, pp. 98-113, Jan 2018, doi: 10.1261/rna.063172.117.

[17] C. M. Clemson, J. A. McNeil, H. F. Willard, and J. B. Lawrence, "XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure," (in eng), *J Cell Biol,* vol. 132, no. 3, pp. 259-75, Feb 1996, doi: 10.1083/jcb.132.3.259.

[18] J. M. Engreitz *et al.*, "The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome," (in eng), *Science,* vol. 341, no. 6147, p. 1237973, Aug 16 2013, doi: 10.1126/science.1237973.

[19] T. Kino, D. E. Hurt, T. Ichijo, N. Nader, and G. P. Chrousos, "Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor," (in eng), *Sci Signal,* vol. 3, no. 107, p. ra8, Feb 2 2010, doi: 10.1126/scisignal.2000568.

[20] A. M. Khalil *et al.*, "Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression," (in eng), *Proc Natl Acad Sci U S A,* vol. 106, no. 28, pp. 11667-72, Jul 14 2009, doi: 10.1073/pnas.0904715106.

[21] L. L. Chen and G. G. Carmichael, "Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA," (in eng), *Mol Cell,* vol. 35, no. 4, pp. 467-78, Aug 28 2009, doi: 10.1016/j.molcel.2009.06.027.

[22] T. R. Mercer and J. S. Mattick, "Structure and function of long noncoding RNAs in epigenetic regulation," (in eng), *Nat Struct Mol Biol,* vol. 20, no. 3, pp. 300-7, Mar 2013, doi: 10.1038/nsmb.2480.

[23] D. Mas-Ponte, J. Carlevaro-Fita, E. Palumbo, T. Hermoso Pulido, R. Guigo, and R. Johnson, "LncATLAS database for subcellular localization of long noncoding RNAs," (in eng), *Rna,* vol. 23, no. 7, pp. 1080-1087, Jul 2017, doi: 10.1261/rna.060814.117.

[24] J. Carlevaro-Fita, A. Rahim, R. Guigo, L. A. Vardy, and R. Johnson, "Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells," (in eng), *Rna,* vol. 22, no. 6, pp. 867-82, Jun 2016, doi: 10.1261/rna.053561.115.

[25] S. van Heesch *et al.*, "Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes," (in eng), *Genome Biol,* vol. 15, no. 1, p. R6, Jan 7 2014, doi: 10.1186/gb-2014-15-1-r6.

[26] F. Kopp and J. T. Mendell, "Functional Classification and Experimental Dissection of Long Noncoding RNAs," (in eng), *Cell,* vol. 172, no. 3, pp. 393-407, Jan 25 2018, doi: 10.1016/j.cell.2018.01.011.

[27] P. A. Latos *et al.*, "Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing," (in eng), *Science,* vol. 338, no. 6113, pp. 1469-72, Dec 14 2012, doi: 10.1126/science.1228110.

[28] K. M. Anderson, D. M. Anderson, J. R. McAnally, J. M. Shelton, R. Bassel-Duby, and E. N. Olson, "Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development," (in eng), *Nature,* vol. 539, no. 7629, pp. 433-436, Nov 17 2016, doi: 10.1038/nature20128.

[29] J. M. Engreitz *et al.*, "Local regulation of gene expression by lncRNA promoters, transcription and splicing," (in eng), *Nature,* vol. 539, no. 7629, pp. 452-455, Nov 17 2016, doi: 10.1038/nature20149.

[30] N. Dimitrova *et al.*, "LincRNA-p21 activates p21 in cis to promote Polycomb target gene expression and to enforce the G1/S checkpoint," (in eng), *Mol Cell,* vol. 54, no. 5, pp. 777-90, Jun 5 2014, doi: 10.1016/j.molcel.2014.04.025.

[31] A. F. Groff *et al.*, "In Vivo Characterization of Linc-p21 Reveals Functional cis-Regulatory DNA Elements," (in eng), *Cell Rep,* vol. 16, no. 8, pp. 2178-2186, Aug 23 2016, doi: 10.1016/j.celrep.2016.07.050.

[32] J. L. Rinn *et al.*, "Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs," (in eng), *Cell,* vol. 129, no. 7, pp. 1311-23, Jun 29 2007, doi: 10.1016/j.cell.2007.05.022.

[33] J. M. Engreitz *et al.*, "RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites," (in eng), *Cell,* vol. 159, no. 1, pp. 188-199, Sep 25 2014, doi: 10.1016/j.cell.2014.08.018.

[34] D. L. Spector and A. I. Lamond, "Nuclear speckles," (in eng), *Cold Spring Harb Perspect Biol,* vol. 3, no. 2, Feb 1 2011, doi: 10.1101/cshperspect.a000646.

[35] B. Zhang *et al.*, "The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult," (in eng), *Cell Rep,* vol. 2, no. 1, pp. 111-23, Jul 26 2012, doi: 10.1016/j.celrep.2012.06.003.

[36] M. A. Miller and W. M. Olivas, "Roles of Puf proteins in mRNA degradation and translation," (in eng), *Wiley Interdiscip Rev RNA,* vol. 2, no. 4, pp. 471-92, Jul-Aug 2011, doi: 10.1002/wrna.69.

[37] S. Lee *et al.*, "Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins," (in eng), *Cell,* vol. 164, no. 1-2, pp. 69-80, Jan 14 2016, doi: 10.1016/j.cell.2015.12.017.

[38] A. Tichon *et al.*, "A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells," (in eng), *Nat Commun,* vol. 7, p. 12209, Jul 13 2016, doi: 10.1038/ncomms12209.

[39] Y. Tay, J. Rinn, and P. P. Pandolfi, "The multilayered complexity of ceRNA crosstalk and competition," (in eng), *Nature,* vol. 505, no. 7483, pp. 344-52, Jan 16 2014, doi: 10.1038/nature12986.

[40] E. Leucci *et al.*, "microRNA-9 targets the long non-coding RNA MALAT1 for degradation in the nucleus," (in eng), *Sci Rep,* vol. 3, p. 2535, 2013, doi: 10.1038/srep02535.

[41] Y. Ge *et al.*, "MiRNA-192 [corrected] and miRNA-204 Directly Suppress lncRNA HOTTIP and Interrupt GLS1-Mediated Glutaminolysis in Hepatocellular Carcinoma," (in eng), *PLoS Genet,* vol. 11, no. 12, p. e1005726, Dec 2015, doi: 10.1371/journal.pgen.1005726.

[42] M. D. Paraskevopoulou *et al.*, "DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts," (in eng), *Nucleic Acids Res,* vol. 44, no. D1, pp. D231-8, Jan 4 2016, doi: 10.1093/nar/gkv1270.

[43] J. H. Li, S. Liu, H. Zhou, L. H. Qu, and J. H. Yang, "starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data," (in eng), *Nucleic Acids Res,* vol. 42, no. Database issue, pp. D92-7, Jan 2014, doi: 10.1093/nar/gkt1248.

[44] Y. Hao *et al.*, "NPInter v3.0: an upgraded database of noncoding RNA-associated interactions," (in eng), *Database (Oxford),* vol. 2016, 2016, doi: 10.1093/database/baw057.

[45] G. Meister, M. Landthaler, Y. Dorsett, and T. Tuschl, "Sequence-specific inhibition of microRNA- and siRNA-induced RNA silencing," (in eng), *Rna,* vol. 10, no. 3, pp. 544-50, Mar 2004.

[46] J. Krutzfeldt *et al.*, "Silencing of microRNAs in vivo with 'antagomirs'," (in eng), *Nature,* vol. 438, no. 7068, pp. 685-9, Dec 1 2005, doi: 10.1038/nature04303.

[47] B. D. Brown *et al.*, "Endogenous microRNA can be broadly exploited to regulate transgene expression according to tissue, lineage and differentiation state," (in eng), *Nat Biotechnol,* vol. 25, no. 12, pp. 1457-67, Dec 2007, doi: 10.1038/nbt1372.

[48] A. A. Farooqi, Z. U. Rehman, and J. Muntane, "Antisense therapeutics in oncology: current status," (in eng), *Onco Targets Ther,* vol. 7, pp. 2035-42, 2014, doi: 10.2147/ott.s49652.

[49] M. S. Ebert and P. A. Sharp, "MicroRNA sponges: progress and possibilities," (in eng), *Rna,* vol. 16, no. 11, pp. 2043-50, Nov 2010, doi: 10.1261/rna.2414110.

[50] S. Davis, B. Lollo, S. Freier, and C. Esau, "Improved targeting of miRNA with antisense oligonucleotides," (in eng), *Nucleic Acids Res,* vol. 34, no. 8, pp. 2294-304, 2006, doi: 10.1093/nar/gkl183.

[51] J. M. Franco-Zorrilla *et al.*, "Target mimicry provides a new mechanism for regulation of microRNA activity," (in eng), *Nat Genet,* vol. 39, no. 8, pp. 1033-7, Aug 2007, doi: 10.1038/ng2079.

[52] A. H. Buck *et al.*, "Post-transcriptional regulation of miR-27 in murine cytomegalovirus infection," (in eng), *Rna,* vol. 16, no. 2, pp. 307-15, Feb 2010, doi: 10.1261/rna.1819210.

[53] L. Poliseno, L. Salmena, J. Zhang, B. Carver, W. J. Haveman, and P. P. Pandolfi, "A coding-independent function of gene and pseudogene mRNAs regulates tumour biology," (in eng), *Nature,* vol. 465, no. 7301, pp. 1033-8, Jun 24 2010, doi: 10.1038/nature09144.

[54] M. S. Ebert, J. R. Neilson, and P. A. Sharp, "MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells," (in eng), *Nat Methods,* vol. 4, no. 9, pp. 721-6, Sep 2007, doi: 10.1038/nmeth1079.

[55] L. Salmena, L. Poliseno, Y. Tay, L. Kats, and P. P. Pandolfi, "A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language?," (in eng), *Cell,* vol. 146, no. 3, pp. 353-8, Aug 5 2011, doi: 10.1016/j.cell.2011.07.014.

[56] P. Sumazin *et al.*, "An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma," (in eng), *Cell,* vol. 147, no. 2, pp. 370-81, Oct 14 2011, doi: 10.1016/j.cell.2011.09.041.

[57] M. Cesana *et al.*, "A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA," (in eng), *Cell,* vol. 147, no. 2, pp. 358-69, Oct 14 2011, doi: 10.1016/j.cell.2011.09.028.

[58] Y. Wang *et al.*, "Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal," (in eng), *Dev Cell,* vol. 25, no. 1, pp. 69-80, Apr 15 2013, doi: 10.1016/j.devcel.2013.03.002.

[59] A. N. Kallen *et al.*, "The imprinted H19 lncRNA antagonizes let-7 microRNAs," (in eng), *Mol Cell,* vol. 52, no. 1, pp. 101-12, Oct 10 2013, doi: 10.1016/j.molcel.2013.08.027.

[60] J. Imig *et al.*, "miR-CLIP capture of a miRNA targetome uncovers a lincRNA H19-miR-106a interaction," (in eng), *Nat Chem Biol,* vol. 11, no. 2, pp. 107-14, Feb 2015, doi: 10.1038/nchembio.1713.

[61] J. Wang *et al.*, "CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer," (in eng), *Nucleic Acids Res,* vol. 38, no. 16, pp. 5366-83, Sep 2010, doi: 10.1093/nar/gkq285.

[62] M. Fan, X. Li, W. Jiang, Y. Huang, J. Li, and Z. Wang, "A long non-coding RNA, PTCSC3, as a tumor suppressor and a target of miRNAs in thyroid cancer cells," (in eng), *Exp Ther Med,* vol. 5, no. 4, pp. 1143-1146, Apr 2013, doi: 10.3892/etm.2013.933.

[63] V. Agarwal, G. W. Bell, J. W. Nam, and D. P. Bartel, "Predicting effective microRNA target sites in mammalian mRNAs," (in eng), *Elife,* vol. 4, Aug 12 2015, doi: 10.7554/eLife.05005.

[64] K. C. Miranda *et al.*, "A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes," (in eng), *Cell,* vol. 126, no. 6, pp. 1203-17, Sep 22 2006, doi: 10.1016/j.cell.2006.07.031.

[65] A. Krek *et al.*, "Combinatorial microRNA target predictions," (in eng), *Nat Genet,* vol. 37, no. 5, pp. 495-500, May 2005, doi: 10.1038/ng1536.

[66] M. Reczko, M. Maragkakis, P. Alexiou, I. Grosse, and A. G. Hatzigeorgiou, "Functional microRNA targets in protein coding sequences," (in eng), *Bioinformatics,* vol. 28, no. 6, pp. 771-6, Mar 15 2012, doi: 10.1093/bioinformatics/bts043.

[67] S. W. Chi, J. B. Zang, A. Mele, and R. B. Darnell, "Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps," (in eng), *Nature,* vol. 460, no. 7254, pp. 479-86, Jul 23 2009, doi: 10.1038/nature08170.

[68] W. H. Majoros *et al.*, "MicroRNA target site identification by integrating sequence and binding information," (in eng), *Nat Methods,* vol. 10, no. 7, pp. 630-3, Jul 2013, doi: 10.1038/nmeth.2489.

[69] F. Erhard, L. Dolken, L. Jaskiewicz, and R. Zimmer, "PARma: identification of microRNA target sites in AGO-PAR-CLIP data," (in eng), *Genome Biol,* vol. 14, no. 7, p. R79, Jul 29 2013, doi: 10.1186/gb-2013-14-7-r79.

[70] J. Xu *et al.*, "The mRNA related ceRNA-ceRNA landscape and significance across 20 major cancer types," (in eng), *Nucleic Acids Res,* vol. 43, no. 17, pp. 8169-82, Sep 30 2015, doi: 10.1093/nar/gkv853.

[71] X. Zhou, J. Liu, and W. Wang, "Construction and investigation of breast-cancer-specific ceRNA network based on the mRNA and miRNA expression data," (in eng), *IET Syst Biol,* vol. 8, no. 3, pp. 96-103, Jun 2014, doi: 10.1049/iet-syb.2013.0025.

[72] P. Wang *et al.*, "LncACTdb 2.0: an updated database of experimentally supported ceRNA interactions curated from low- and high-throughput experiments," (in eng), *Nucleic Acids Res,* vol. 47, no. D1, pp. D121-d127, Jan 8 2019, doi: 10.1093/nar/gky1144.

[73] S. Das, S. Ghosal, R. Sen, and J. Chakrabarti, "lnCeDB: database of human long noncoding RNA acting as competing endogenous RNA," (in eng), *PLoS One,* vol. 9, no. 6, p. e98965, 2014, doi: 10.1371/journal.pone.0098965.

[74] T. D. Le, J. Zhang, L. Liu, and J. Li, "Computational methods for identifying miRNA sponge interactions," (in eng), *Brief Bioinform,* vol. 18, no. 4, pp. 577-590, Jul 1 2017, doi: 10.1093/bib/bbw042.

[75] C. Bosia, A. Pagnani, and R. Zecchina, "Modelling Competing Endogenous RNA Networks," (in eng), *PLoS One,* vol. 8, no. 6, p. e66609, 2013, doi: 10.1371/journal.pone.0066609.

[76] U. Ala *et al.*, "Integrated transcriptional and competitive endogenous RNA networks are cross-regulated in permissive molecular environments," (in eng), *Proc Natl Acad Sci U S A,* vol. 110, no. 18, pp. 7154-9, Apr 30 2013, doi: 10.1073/pnas.1222509110.

[77] M. Jens and N. Rajewsky, "Competition between target sites of regulators shapes post-transcriptional gene regulation," (in eng), no. 1471-0064 (Electronic).

[78] A. D. Bosson, J. R. Zamudio, and P. A. Sharp, "Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition," (in eng), *Mol Cell,* vol. 56, no. 3, pp. 347-59, Nov 6 2014, doi: 10.1016/j.molcel.2014.09.018.

[79] P. Furio-Tari, S. Tarazona, T. Gabaldon, A. J. Enright, and A. Conesa, "spongeScan: A web for detecting microRNA binding elements in lncRNA sequences," (in eng), *Nucleic Acids Res,* vol. 44, no. W1, pp. W176-80, Jul 8 2016, doi: 10.1093/nar/gkw443.

[80] A. Bhattacharya and Y. Cui, "SomamiR 2.0: a database of cancer somatic mutations altering microRNA-ceRNA interactions," (in eng), *Nucleic Acids Res,* vol. 44, no. D1, pp. D1005-10, Jan 4 2016, doi: 10.1093/nar/gkv1220.

[81] P. Wang *et al.*, "miRSponge: a manually curated database for experimentally supported miRNA sponges and ceRNAs," (in eng), *Database (Oxford),* vol. 2015, 2015, doi: 10.1093/database/bav098.

[82] R. L. Skalsky *et al.*, "The viral and cellular microRNA targetome in lymphoblastoid cell lines," (in eng), *PLoS Pathog,* vol. 8, no. 1, p. e1002484, Jan 2012, doi: 10.1371/journal.ppat.1002484.

[83] S. Kishore, L. Jaskiewicz, L. Burger, J. Hausser, M. Khorshid, and M. Zavolan, "A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins," (in eng), *Nat Methods,* vol. 8, no. 7, pp. 559-64, May 15 2011, doi: 10.1038/nmeth.1608.

[84] T. Conrad, A. Marsico, M. Gehre, and U. A. Orom, "Microprocessor activity controls differential miRNA biogenesis In Vivo," (in eng), *Cell Rep,* vol. 9, no. 2, pp. 542-54, Oct 23 2014, doi: 10.1016/j.celrep.2014.09.007.

[85] A. Kozomara, M. Birgaoanu, and S. Griffiths-Jones, "miRBase: from microRNA sequences to function," (in eng), no. 1362-4962 (Electronic).

[86] A. Kozomara and S. Griffiths-Jones, "miRBase: annotating high confidence microRNAs using deep sequencing data," (in eng), no. 1362-4962 (Electronic).

[87] A. R. Quinlan and I. M. Hall, "BEDTools: a flexible suite of utilities for comparing genomic features," (in eng), no. 1367-4811 (Electronic).

[88] G. Anders *et al.*, "doRiNA: a database of RNA interactions in post-transcriptional regulation," (in eng), no. 1362-4962 (Electronic).

[89] G. K. Marinov *et al.*, "From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing," (in eng), *Genome Res,* vol. 24, no. 3, pp. 496-510, Mar 2014, doi: 10.1101/gr.161034.113.

[90] D. Wang *et al.*, "Quantitative functions of Argonaute proteins in mammalian development," (in eng), *Genes Dev,* vol. 26, no. 7, pp. 693-704, Apr 1 2012, doi: 10.1101/gad.182758.111.

[91] J. Konig *et al.*, "iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution," (in eng), *Nat Struct Mol Biol,* vol. 17, no. 7, pp. 909-15, Jul 2010, doi: 10.1038/nsmb.1838.

[92] L. M. Wee, W. E. Flores-Jasso Cf Fau - Salomon, P. D. Salomon We Fau - Zamore, and P. D. Zamore, "Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties," (in eng), no. 1097-4172 (Electronic).

[93] Y. Cai *et al.*, "Rpph1 Upregulates CDC42 Expression and Promotes Hippocampal Neuron Dendritic Spine Formation by Competing with miR-330-5p," (in eng), *Front Mol Neurosci,* vol. 10, p. 27, 2017, doi: 10.3389/fnmol.2017.00027.

[94] W. Wu, T. Yu, Y. Wu, W. Tian, J. Zhang, and Y. Wang, "The miR155HG/miR-185/ANXA2 loop contributes to glioblastoma growth and progression," (in eng), *J Exp Clin Cancer Res,* vol. 38, no. 1, p. 133, Mar 21 2019, doi: 10.1186/s13046-019-1132-0.

[95] F. Wang *et al.*, "A threelncRNA signature for prognosis prediction of acute myeloid leukemia in patients," (in eng), *Mol Med Rep,* vol. 18, no. 2, pp. 1473-1484, Aug 2018, doi: 10.3892/mmr.2018.9139.

[96] S. Chang, B. Chen, X. Wang, K. Wu, and Y. Sun, "Long non-coding RNA XIST regulates PTEN expression by sponging miR-181a and promotes hepatocellular carcinoma progression," (in eng), *BMC Cancer,* vol. 17, no. 1, p. 248, Apr 7 2017, doi: 10.1186/s12885-017-3216-6.

[97] W. Wei, Y. Liu, Y. Lu, B. Yang, and L. Tang, "LncRNA XIST Promotes Pancreatic Cancer Proliferation Through miR-133a/EGFR," (in eng), *J Cell Biochem,* vol. 118, no. 10, pp. 3349-3358, Oct 2017, doi: 10.1002/jcb.25988.

[98] H. Yu *et al.*, "Knockdown of long non-coding RNA XIST increases blood-tumor barrier permeability and inhibits glioma angiogenesis by targeting miR-137," (in eng), *Oncogenesis,* vol. 6, no. 3, p. e303, Mar 13 2017, doi: 10.1038/oncsis.2017.7.

[99] Q. Kong *et al.*, "LncRNA XIST functions as a molecular sponge of miR-194-5p to regulate MAPK1 expression in hepatocellular carcinoma cell," (in eng), *J Cell Biochem,* vol. 119, no. 6, pp. 4458-4468, Jun 2018, doi: 10.1002/jcb.26540.

[100] F. Li, X. Li, L. Qiao, W. Liu, C. Xu, and X. Wang, "MALAT1 regulates miR-34a expression in melanoma cells," (in eng), *Cell Death Dis,* vol. 10, no. 6, p. 389, May 17 2019, doi: 10.1038/s41419-019-1620-3.

[101] Q. Li *et al.*, "Disrupting MALAT1/miR-200c sponge decreases invasion and migration in endometrioid endometrial carcinoma," (in eng), *Cancer Lett,* vol. 383, no. 1, pp. 28-40, Dec 1 2016, doi: 10.1016/j.canlet.2016.09.019.

[102] M. Pa, G. Naizaer, A. Seyiti, and G. Kuerbang, "Long Noncoding RNA MALAT1 Functions as a Sponge of MiR-200c in Ovarian Cancer," (in eng), *Oncol Res,* Sep 11 2017, doi: 10.3727/096504017x15049198963076.

[103] F. Tao, X. Tian, S. Ruan, M. Shen, and Z. Zhang, "miR-211 sponges lncRNA MALAT1 to suppress tumor growth and progression through inhibiting PHF19 in ovarian carcinoma," (in eng), *Faseb j,* p. fj201800495RR, Jun 6 2018, doi: 10.1096/fj.201800495RR.

[104] Y. Pan *et al.*, "Long Non-Coding MALAT1 Functions as a Competing Endogenous RNA to Regulate Vimentin Expression by Sponging miR-30a-5p in Hepatocellular Carcinoma," (in eng), *Cell Physiol Biochem,* vol. 50, no. 1, pp. 108-120, 2018, doi: 10.1159/000493962.

[105] Y. Shao *et al.*, "LncRNA-RMRP promotes carcinogenesis by acting as a miR-206 sponge and is used as a novel biomarker for gastric cancer," (in eng), *Oncotarget,* vol. 7, no. 25, pp. 37812-37824, Jun 21 2016, doi: 10.18632/oncotarget.9336.

[106] R. Denzler, S. E. McGeary, A. C. Title, V. Agarwal, D. P. Bartel, and M. Stoffel, "Impact of MicroRNA Levels, Target-Site Complementarity, and Cooperativity on Competing Endogenous RNA-Regulated Gene Expression," (in eng), *Mol Cell,* vol. 64, no. 3, pp. 565-579, Nov 3 2016, doi: 10.1016/j.molcel.2016.09.027.

[107] A. Helwak, G. Kudla, T. Dudnakova, and D. Tollervey, "Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding," (in eng), *Cell,* vol. 153, no. 3, pp. 654-65, Apr 25 2013, doi: 10.1016/j.cell.2013.03.043.

[108] G. B. Loeb *et al.*, "Transcriptome-wide miR-155 binding map reveals widespread noncanonical microRNA targeting," (in eng), *Mol Cell,* vol. 48, no. 5, pp. 760-70, Dec 14 2012, doi: 10.1016/j.molcel.2012.10.002.

[109] P. Shannon *et al.*, "Cytoscape: a software environment for integrated models of biomolecular interaction networks," (in eng), *Genome Res,* vol. 13, no. 11, pp. 2498-504, Nov 2003, doi: 10.1101/gr.1239303.