UNIVERZA V LJUBLJANI

FAKULTETA ZA FARMACIJO

MAŠA KENDA

MAGISTRSKA NALOGA

ENOVITI MAGISTRSKI ŠTUDIJ FARMACIJE

Ljubljana, 2014

UNIVERSITY OF LJUBLJANA

FACULTY OF PHARMACY

MAŠA KENDA

ALTERNATIVE MICRORNA PROMOTERS

ALTERNATIVNI PROMOTORJI GENOV ZA MIKRO RNA

UNIFORM MASTER'S STUDY PROGRAMME PHARMACY

Ljubljana, 2014

The research work was carried out at the Max Planck Institute for Molecular Genetics in Berlin, Germany, under supervision of my co-mentor Ulf Andersson Ørom, PhD, and home mentorship of Assist. Prof. Dr. Tomaž Bratkovič. Sequential analysis was performed by GATC Biotech in Köln, Germany.

Acknowledgements

First and foremost, I would like to thank my co-mentor Ulf Ørom, PhD, for giving me the opportunity to do my master's thesis research work in his group and for thoughtful guidance throughout the process. His support and advice are much appreciated. I am extremely grateful to members of the lab for making me feel as an equal part of the team, for the encouragement, excellent advice and all the fun moments. My time at the MPI was truly enjoyable, thank you.

Zahvaljujem se svojemu mentorju doc. dr. Tomažu Bratkoviču za hitre odgovore na vsa vprašanja, nasvete in temeljit pregled teze. Hvala družini za pomoč in podporo ter prijateljem, še posebej Ani za vse spodbudne besede in sošolcem za zabavna študijska leta.

Statement

I hereby declare that I carried out my master's thesis work independently under the mentorship of Assist. Prof. Dr. Tomaž Bratkovič and co-mentorship of Ulf Andersson Ørom, PhD.

Ljubljana, 2014

Graduation commission president: Prof. Dr. Julijana Kristl

Graduation commission member: Assist. Prof. Rok Frlan

Contents

Abstract

MicroRNAs (miRNAs) are small noncoding RNAs that regulate the expression of protein-coding genes post-transcriptionally. They are important for normal cellular functions and have been found missexpressed in many diseases. Little is known about how the expression of miRNAs is transcriptionally regulated. In this survey we addressed the question of their biogenesis from alternative promoters and the role of these promoters as regulatory elements in feedback loops using a miRNA family miR-200b-200a-429 and its involvement with transcription factor ZEB2 and E-cadherin protein as our study model. By analyzing bioinformatic data we recognized two promoter regions – 4.2 kb and 2.0 kb upstream of the miR-200b-200a-429 cluster, that are differentially expressed in a set of cell lines, depending on presence of proteins ZEB2 and E-cadherin. We observed that when E-cadherin is weakly expressed primary miR-200b-200a-429 transcript is generated only from -2.0-kb promoter in low quantity, suggesting E-cadherin could repress miR-200b-200a-429 promoters by acting mainly on -4.2-kb promoter. With the luciferase reporter assay we experimentally confirmed promoter activity of the two predicted promoter regions, and found that they are differentially active in two cell lines. In response to knockdown of ZEB2 and E-cadherin we observed an increase in primary miR-200b-200a-429 transcript level, thus implying that both act as transcription factors repressing transcription of subjected miRNAs. To identify how ZEB2 and E-cadherin affect the activity of each alternative promoter we coupled knockdown with reporter luciferase assay, but could not produce useful results from that experiment. Alternative miRNA promoters could provide a loophole in regulation of expression by feedback circuits. Their exact role should be well described before therapeuticals that regulate miRNA transcription can be designed.

Povzetek

Mikro RNA (miRNA) so približno 22 nukleotidov dolge enoverižne nekodirajoče RNA, ki posttranskripcijsko regulirajo izražanje genov, ki kodirajo proteine. Delujejo tako, da se preko komplementarnih baznih parov vežejo na 3'-neprevedljivo regijo tarčnih informacijskih RNA (mRNA). To vodi do represije translacije, destabilizacije ali cepitve tarčne mRNA – tako miRNA utišajo tarčne gene, saj je iz njih onemogočen nastanek proteina. Čeprav miRNA predstavljajo le majhen delež (približno 3 %) človeškega genoma, bi lahko znižale izražanje kar polovice vseh kodirajočih genov. MiRNA imajo vlogo v fizioloških in patoloških procesih, npr. pri diferenciaciji, proliferaciji, celični smrti, kontroli metabolizma in razvoju raka. Precej je znanega o mehanizmu delovanja in vpletenosti miRNA v te procese, medtem ko njihova biogeneza, predvsem regulacija transkripcije, še ni dobro pojasnjena.

Kanonična biogeneza miRNA se začne v jedru s transkripcijo iz miRNA-gena, ki jo vodi polimeraza II, izjemoma tudi polimeraza III. Nastanejo primarni transkripti (pri-miRNA), dolgi več kilobaznih parov. Pri-miRNA zaradi delne komplementarnosti baz tvorijo sekundarne strukture v obliki lasnične zanke s približno 33 nukleotidov dolgim steblom in terminalnim nesparjenim zavojem. To sekundarno strukturo nato prepozna protein DGCR8, ki usmeri encim RNazo Drosha, da cepi zanko približno 11 baznih parov od začetka stebla. Drosha tako ustvari prekurzorsko miRNA (pre-miRNA) – približno 65 nukleotidov dolgo zanko z dvonukleotidnim podaljškom na 3'-koncu, ki ga prepozna prenašalni protein eksportin 5. Ta skozi jedrne pore prenese pre-miRNA v citoplazmo. Tam terminalno zanko pre-miRNA odcepi protein Dicer in tako ustvari 22 nukleotidov dolg miRNA-dupleks. Alternativno kanonični biogenezi obstajajo tudi miRNA, ki generirajo dupleks brez Droshine ali Dicerjeve cepiteve. Proteini Dicer, TRBP in argonavt (Ago) sestavijo z RNA-inducirani utiševalni kompleks (RISC), v katerega se vgradi ena od verig (funkcionalna veriga) miRNA-dupleksa, komplementarna veriga pa se razgradi. Funkcionalna veriga (zrela miRNA) ostane vgrajena v protein Ago kompleksa RISC in ga vodi do tarčne mRNA. MiRNA lahko povzročijo represijo translacije preko upočasnitve oz. zaustavitve ribosomov, nanizanih na tarčni mRNA, ali z razgradnjo na novo sintetiziranega peptida. MiRNA preprečijo izražanje proteina iz tarčne mRNA tudi tako, da odstranijo njeno 5'-kapo ali poliA-rep; oboje vodi do destabilizacije in razkroja tarčne mRNA. Zaradi znanja o biogenezi in mehanizmu delovanja miRNA-genov, smo zmožni

znižati izražanje specifičnega gena tudi z eksogeno vnesenimi molekulami (siRNA ali shRNA), ki delujejo podobno kot miRNA in za aktivacijo potrebujejo encime, vpletene v biogenezo miRNA. Temu postopku pravimo RNA-interferenca, njenemu učinku na tarčni gen pa utišanje.

Regulacija biogeneze miRNA je zelo pomemben proces za vzdrževanje normalnih celičnih funkcij, saj neobičajne količine miRNA v celici vodijo do napačnega sorazmerja izraženih proteinov – to pa je lahko vzrok za patološke procese. Regulacija biogeneze miRNA je možna na transkripcijski (s transkripcijskimi dejavniki ali epigenetsko), posttranskripcijski (preko regulacije procesiranja Droshe, hitrosti nastajanja in razgradnje miRNA ali urejanja baz v miRNA) in na ravni povratnih povezav (enojne in dvojne negativne povratne zanke). Primer dvojne negativne povratne zanke je družina miR-200, ki posttranskripcijsko zavre izražanje transkripcijskih dejavnikov ZEB1 in ZEB2 – ta pa zavreta transkripcijo družine miR-200 in tudi proteina kadherina E. Ker je kadherin E celični adhezijski protein, izguba njegovega izražanja pomeni, da bodo celice prešle iz epitelijskega v mezenhimsko stanje, torej da bodo migrirale iz tumorja in metastazirale, če gre za rakave celice. V tej raziskavi smo se ukvarjali z biogenezo miRNA iz alternativnih promotorjev in vlogo teh promotorjev kot regulatornih elementov v povratnih zankah z delom na primeru miRNA družine miR-200b-200a-429 in njeno povezanostjo s transkripcijskim dejavnikom ZEB2 in kadherinom E.

Z analizo bioinformatičnih podatkov (evolucijske ohranjenosti, hipersenzitivnosti (tj. povečane občutljivosti DNA za vezavo transkripcijskih dejavnikov ali encimov), histonskih modifikacij, vezave polimeraze II, sekvenciranja RNA in sekvenciranja 5'-koncev primarnih transkriptov) smo prepoznali dva promotorja oz. mesti začetka transkripcije – 4.2 kb in 2.0 kb navzgor od gena za miR-200b-200a-429. Analiza RNA sekvenciranja v 11 celičnih linijah je pokazala, da sta transkripta s teh dveh alternativnih mest začetka transkripcije diferencialno izražena v različnih celičnih linijah, in sicer odvisno od prisotnosti in količine izraženih proteinov ZEB2 in E kadherina. Kadar je ZEB2 dobro izražen in E kadherin ni, miR-200b-200a-429 niso izražene. Kadar ZEB2 ni prisoten, E kadherin pa je v večji količini, se primarni transkript miR-200b-200a-429 začne prepisovati iz -4.2-kb promotorja. Obstaja še tretja možnost, ko je ZEB2 prisoten v majhni količini ali sploh ni prisoten in kadherin E izražen v majhni količini. Takrat se miR-200b-200a-429 prepisuje v manjši količini iz -2.0-kb promotorja, kar bi lahko pomenilo, da poleg ZEB2 tudi kadherin E

transkripcijsko zavre promotorja miR-200b-200a-429 – popolnoma -4.2-kb promotor in deloma -2.0-kb promotor.

Da predvideni promotorski regiji zares posedujeta promotorsko aktivnost, in da sta alternativno aktivni v različnih celičnih linijah, smo potrdili z luciferaznim poročevalskim testom. V celični liniji MCF-7 sta aktivna oba promotorja, v celični liniji A549 pa samo -2.0-kb promotor, kar je skladno z ugotovitvami analize sekvenciranja RNA.

Po utišanju Droshe, transkripcijskega dejavnika ZEB2 in kadherina E smo opazili povečano izražanje primarnega transkripta miR-200b-200a-429, kar kaže na to, da poteka biogeneza miR-200b-200a-429 po kanonični poti biogeneze, ki vključuje Droshino cepitev primarnega transkripta; da ZEB2, kot že znano iz preteklih študij, zavre izražanje primarnega transkripta miR-200b-200a-429; da bi kadherin E tudi lahko deloval kot transkripcijski represor na miR-200b-200a-429 (zaradi velike standardne deviacije pri tem eksperimentu tega ne moremo z gotovostjo trditi, je pa to verjetna možnost, saj so za kadherin E že dokazali vpliv na transkripcijo genov in njegovo jedrno lokalizacijo pri več vrstah tumorjev).

Da bi ugotovili, kako ZEB2 in kadherin E vplivata na vsakega od alternativnih promotorjev (ali povečata ali zmanjšata aktivnost promotorskih regij), smo združili njuno utišanje in luciferazni poročevalski test. Za primerjavo s posttranskripcijsko regulacijo izražanja miRNA smo vključili tudi utišanje Droshe. Eksperiment ni dal veljavnih rezultatov, saj se kontrole niso odzivale kot pričakovano. Problematičen je bil prevelik signal pri negativni kontroli za luciferazni test, prevelike razlike med pozitivnimi kontrolami za luciferazni test in različen odziv dveh negativnih kontrol za utišanje. Kot alternativo smo predlagali eksperiment z imunoprecipitacijo kromatina in sekvenciranjem regij, kamor se vežeta ZEB2 in kadherin E.

Vloga alternativnih promotorjev v povratnih zankah še ni pojasnjena. Ena od možnosti je, da delujejo kot stikalo, ki omogoči izhod iz povratnih zank. Točna vloga alternativnih promotorjev za miRNA-gene mora biti dobro raziskana pred razvojem zdravil, s katerimi bi vplivali na regulacijo transkripcije miRNA.

Key words

MicroRNA; promoter recognition; alternative promoters; RNA interference

List of abbreviations

ADAR – adenosine deaminase acting on RNA

Ago – Argonaute

Bcl2 – B-cell lymphoma 2

bp – base pair

CDH1 – E-cadherin

Cdk4 – cyclin-dependent kinase 4

Cdk6 – cyclin-dependent kinase 6

ChIP-seq – chromatin immunoprecipitation followed by sequencnig

ddH_2O – double destilled H_2O

DGCR8 - DiGeorge syndrom critical region gene 8 protein

dsDNA – double stranded DNA

dsRNA – double stranded RNA

EMT – epithelial-mesenchymal transition

GTP – guanosine-5'-triphosphate

kb – kilo base pair

KLF4 – Kruppel-like factor 4

MET – mesenchymal-epithelial transition

MID – middle domain

miRNA – microRNA

MYOD1 – myoblast determination 1

nt – nucleotide

OCT4 – octamer-binding transcription factor 4

PACT – protein activator of the interferon induced protein kinase

PAZ – Piwi-Argonaute-Zwille

PIWI – P-element induced wimpy testis

Pol II – Polymerase II

Pol III – Polymerase III

pre-miRNA – precursor microRNA

pri-miRNA – primary microRNA

Ran – Ras-related nuclear protein

RISC – RNA-induced silencing complex

RLC – RISC loading complex

RNAi – RNA interference

RNA-seq – RNA sequencing

RT-qPCR – quantitative reverse transcription Polymerase chain reaction

shRNA – short hairpin RNA

siRNA – small interfering RNA

SOX2 – SRY (sex determining region Y)-box 2

ssRNA – single-stranded RNA

TF – transcription factor

TGFβ – transforming growth factor beta

TRBP – TAR RNA-binding protein

tRNA – transfer RNA

TSS – transcription start site

UCSC – University of California Santa Cruz

UTR – untranslated region

ZEB1 – zinc finger E-box binding homoprotein 1 transcription factor

ZEB2 – zinc finger E-box binding homoprotein 2 transcription factor

I. Introduction

MicroRNAs (miRNAs) are non-coding RNAs whose genes comprise 3% of all predicted genes in the human genome, but affect up to one half of protein-coding genes (1,2). They function via gene silencing (i.e., downregulation of gene expression) and their roles span from regulating physiological processes (e.g., developmental timing, cell differentiation, cell proliferation, cell death and metabolic control) to development of different pathologies like cancer (3). A single miRNA can target hundreds of genes and likewise – abundance of an individual miRNA can be influenced by many factors (1,4). Consequently, miRNAs form complex regulatory networks that differ in specific conditions and from cell type to cell type as they are expressed only in some tissues, developmental stages or under specific circumstances (2,3).

MiRNAs are ~22 nucleotides (nts) long single-stranded RNAs (ssRNAs); important post-transcriptional gene regulators that fine-tune the output of actively transcribed protein-coding genes. They most commonly act by base-pairing in the 3' untranslated region (UTR) of target messenger RNAs (mRNAs), leading to repression of translation and destabilization or mRNA cleavage (Fig. 1), thereby silencing target genes (3,5).

With progress in miRNA research numerous miRNAs have been identified (currently the microRNA database miRBase contains 1872 miRNAs), along with genes regulated by them, and entities affecting them. This contributed to a deeper insight into their biogenesis, mechanism of action and function (6). Altogether, new discoveries denote the complexity and diversity of miRNA biological regulation. For example, miRNAs do not act as silencers by default, as some can also activate transcription by binding to gene regulatory sequences (7). There is not just one universal biogenesis pathway as some alternatives that give rise to functional miRNAs occur. Regulation of biogenesis can happen at many levels on the road to maturity, altering the entire transcriptome in a cell (8).

The transcriptional characteristics of miRNAs may be equally intricate, though these are yet to be fully delineated. It is important to understand what drives the expression of genes that influence so many others. In this thesis, we set out to explore the possibility of miRNA transcriptional regulation by alternative promoters.

1.1 Biogenesis of mammalian miRNAs

Canonical miRNA biogenesis (overviewed in Fig. 1) starts in the nucleus with transcription from miRNA gene by Polymerase II or III, generating primary transcripts (pri-miRNAs). Due to partial base pair complementarity, pri-miRNAs fold back on themselves and form a secondary structure with a stem and a non-base pairing string that forms a stem loop. Primary tanscripts are then cleaved at the base of the stem by the Microprecessor complex, which consists of two proteins, Drosha and DiGeorge syndrome critical region gene 8 (DRCR8), to form ~65 nt precursor transcripts (pre-miRNAs). Nuclear transporter Exportin 5 mediates pre-miRNA transport through the nuclear pores into the cytoplasm. There, the hairpin is cleaved off by cytoplasmic protein Dicer producing core miRNA duplexes. Dicer, TRBP (TAR RNA-binding protein) and Argonaute (Ago) proteins form RNA-induced silencing complex (RISC) where one strand is chosen to be the functional component and its complementary strand is degraded. Mature miRNA remains loaded on the Ago protein and then exerts its effect on target messenger RNA (4). Noncanonical pathways also yield some mature miRNAs, such as pri-miRNAs that bypass Drosha cleavage (for example mirtrons) or pre-miRNAs that bypass Dicer cleavage (9).

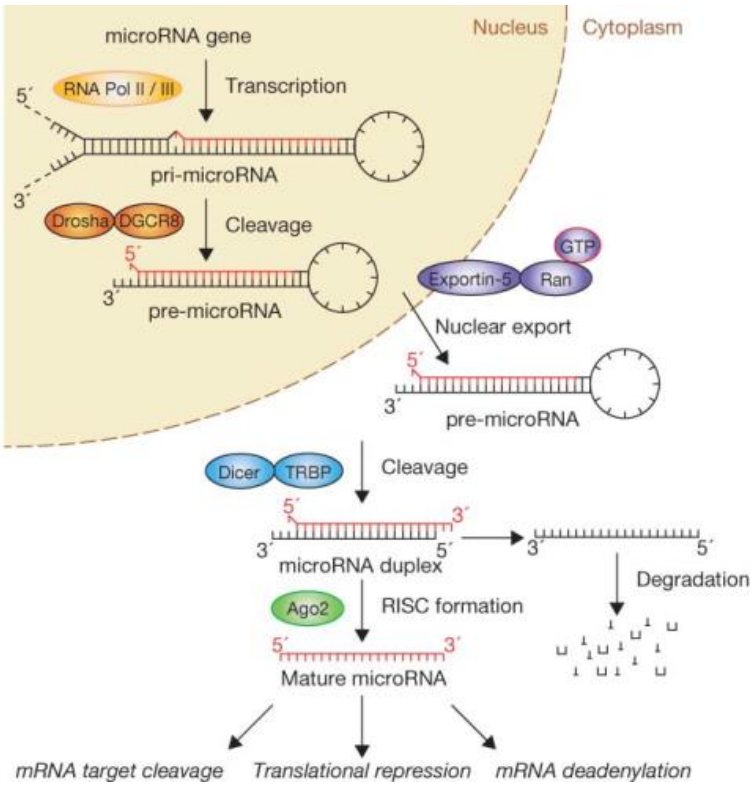

Figure 1: The canonical pathway of miRNA biogenesis (from (4)). Pol II / III – Polymerase II / III, DGCR8 – DiGeorge syndrom critical region gene 8 protein, TRBP – TAR RNA-binding protein, Ago2 – Argonaute protein 2, RISC – RNA-induced silencing complex.

1.1.1 Properties of miRNA genes, their transcription and promoter features

Before the regulatory effect of miRNA genes was known, high evolutionary conservation across species already pointed out their significance (3). Mammalian miRNAs are well conserved in the genome and can often be found as paralogs (derived from the same ancestral gene) with identical or almost identical seed regions, i.e., sequence of 6 – 8 nucleotides usually located at positions 2 – 7 counting from the 5' end of the miRNA (10). The seed region is the main element responsible for complementary base-pairing with target mRNA. A higher conservation score of seed region complementary sites in the 3’ UTR of target mRNA compared to adjacent nucleotides implies the importance of seed region pairing in determining miRNA targets (10). MiRNAs with identical seed region sequences are termed an miRNA family and would be expected to share the same targets and have similar roles. Due to the influence of other possible interactions (supplementary and compensatory pairing at the 3’ end of the miRNA) this is not always the case (3).

Members of a miRNA family are often found as clusters in the genome; two to six miRNAs in close proximity are transcribed in the same orientation, are not separated by another transcription unit or miRNA transcribed in the opposite direction. Clustered miRNAs share promoters of a single polycistronic transcription unit (11). The expression of a cluster of miRNAs is driven from the same promoter producing a single few kilobases (kb) long primary transcript on which several miRNAs are beaded (3,5). However, there is also a possibility of separate transcription and regulation of each miRNA in a cluster (4).

MiRNAs are encoded in regions between annotated genes (intergenic miRNAs) or are embedded in introns or exons of protein-coding transcripts (intragenic miRNAs) (5). Protein-coding genes that contain an intragenic miRNA are termed miRNA host genes. Roughly two thirds of miRNAs are intragenic, the remainder being intergenic (2). It has been proposed that intragenic miRNAs are usually co-transcribed with their host gene by RNA Polymerase II (Pol II), while intergenic miRNAs have their own promoter and are transcribed by Pol II or RNA Polymerase III (Pol III) (5). Pol II indeed confers transcription of most miRNAs - hallmarks of Pol II transcription are primary transcripts longer than 1 kb (while Pol III products are shorter), 3' polyadenilated and 5' capped primary transcripts that are also rich in uridine residues, that would terminate Pol III transcription. Moreover, differential expression during development is solely observed for Pol II products (4,12).

Promoter is a region upstream of a gene where transcription machinery binds. It is directly followed by the transcription start site (TSS) – a location on the DNA, specified to exact nucleotide, where 5' end of RNA is generated (13). DNA is wound up around nucleosomes that constitute of 8 histone proteins. This entanglement of DNA and proteins is called chromatin. There are certain chromatin signatures that are useful in promoter prediction methods, namely covalent histone modifications like trimethylation of Lys 4 of histone 3 (H3K4me3) and acetylation of Lys 9/14 of histone 3 (H3K9/14Ac). H3K4me3 is a characteristic of TSS and is limited to a transcription initiation region. Expression of genes is under the control of transcription factors that bind proximally to TSS. 100 – 130 bp region of the TSS of transcriptionally active genes is nucleosome depleted, which makes the DNA more accessible for transcription factor and enzyme binding (5). These regions are also termed hypersensitive sites (14). Other structural features that contribute to promoter identification are CpG islands, vertebrate (evolutionary) conservation, transcription factor binding sites and proximity of the mature miRNA (2). These genomic landmarks are used among others in next-generation bioinformatics to predict putative promoters.

Recent study reveals that about 60% of intragenic miRNAs that reside in introns have their own dedicated promoters allowing for independent regulation from the host gene. Such intronic promoters are not the sole promoters of a particular miRNA, as about 60% of intragenic miRNAs are still coexpressed from the host gene promoter. Duducing from these findings, a single miRNA can have more than one promoter. In fact, on average 4.7 alternative TSSs per miRNA have been computationally predicted. The same study estimates that 84% of all miRNA have more than one promoter (2).

1.1.2 Primary transcripts are processed by Drosha

Upon transcription from DNA pri-miRNA forms local stem-loop structures with miRNA encoded in the stem. These secondary structures typically have a stem of ~33 bp, a terminal loop and flanking ssRNA segments (5). The stem is formed due to Watson-Crick pairing (G-C and U-A) and non-Watson-Crick interactions, also referred as wobble base pairs (e.g. G-U or G-G) (9). *En route* to the effective mature miRNA, the stem loop is cleaved at the base of the stem in the nucleus by Drosha. Drosha is a nuclear RNase III protein of 130-160 kDa. It has two domains, each containing one catalytic site, that crop 3' strand and 5' strand of double stranded RNA (dsRNA) separately, creating a 2 nt 3' overhang on the released short hairpin product called precursor miRNA (Fig. 2). To function, Drosha requires assembly of Microprocessor complex together with a cofactor, DiGeorge syndrom critical region 8, a protein of ~120 kDa that is responsible for the recognition and binding to the ssRNA-dsRNA junction at the base of the stem (5,8). The other role of DGCR8 is to position the two catalytic sites of Drosha to precise distance on the stem (4). The cleavage happens approximately two helical turns from the ssRNA-dsRNA junction, 11 bp into the stem. Apart from those two helical turns flanked by at least nine unstructured nts (i.e., nucleotides with low base pairing probability that are therefore not prone to formation of secondary structures), important determinants enhancing processing by the Microprocessor complex are also the UG motif at the base of the stem, upstream of the 5' Drosha cleavage site, the CNNC motif 17-18 nts downstream of the 3' Drosha cleavage site and the UGUG in the stem loop (Fig. 2) (9). Drosha processing happens practically co-transcriptionally, meaning that the primary miRNA transcripts are present in the cell for a very short amount of time. Drosha cleavage also affects the host gene, fractionating the primary transcript before splicing (i.e., removing the miRNA-harboring introns), yet it does not impair this process as an intact intron is not absolutely necessary for splicing to occur (4).

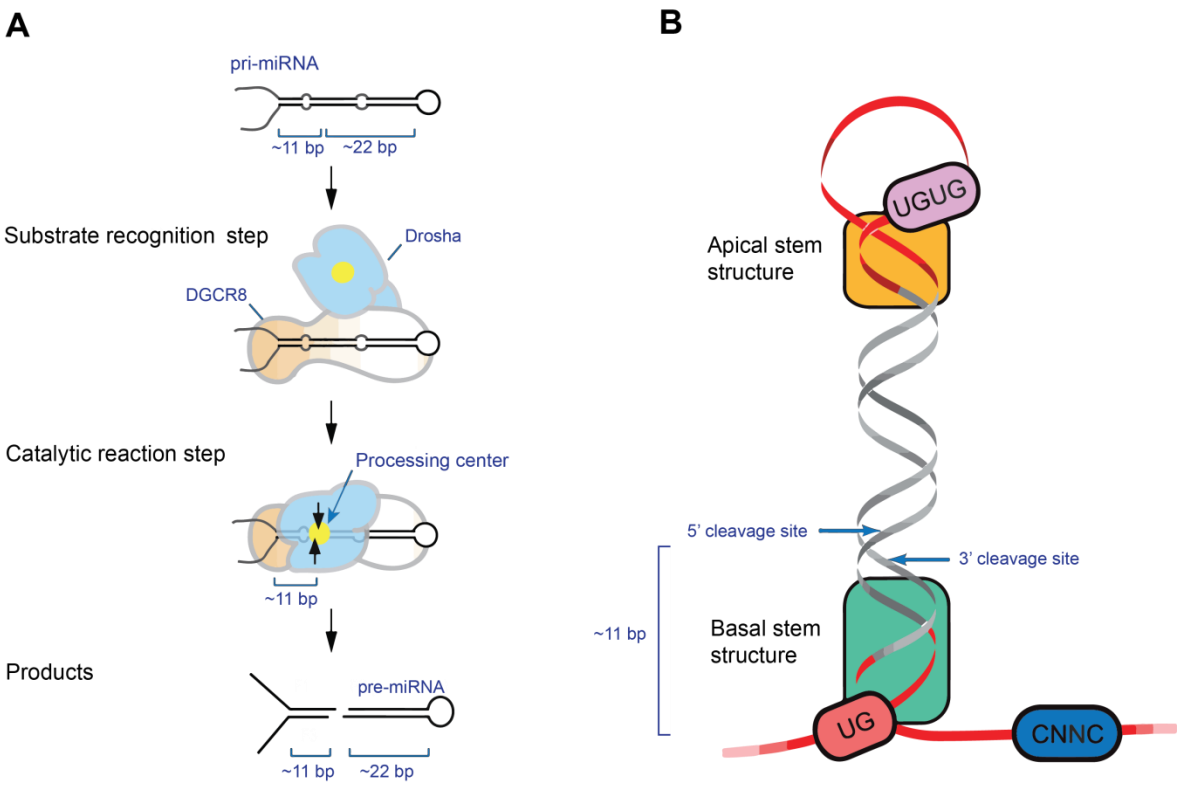

Figure 2: (A) Drosha cleavage of the pri-miRNA (from (15)). 33 nt stem of pri-miRNA is cleaved by Drosha 11 bp from ssRNa-dsRNA junction, which has to first be recognized by DGCR8. Releasing pre-miRNA has a 2 nt 3' overhang – hallmark of RNase III type enzyme action. DGCR8 – DiGeorge syndrom critical region gene 8 protein **(B)** Important primary-sequence motifs for Drosha cleavage (from (9)).

Drosha cleavage is not always required for generation of miRNAs. Mirtrons are splicing-derived miRNAs from introns of pre-mRNA or pri-miRNA transcripts. After splicing and debranching excised introns fold and form secondary structures that resemble pre-miRNAs and therefore enter the canonical biogenesis pathway without Drosha cleavage at the stage of dicing off the stem loop (Fig. 3) (16).

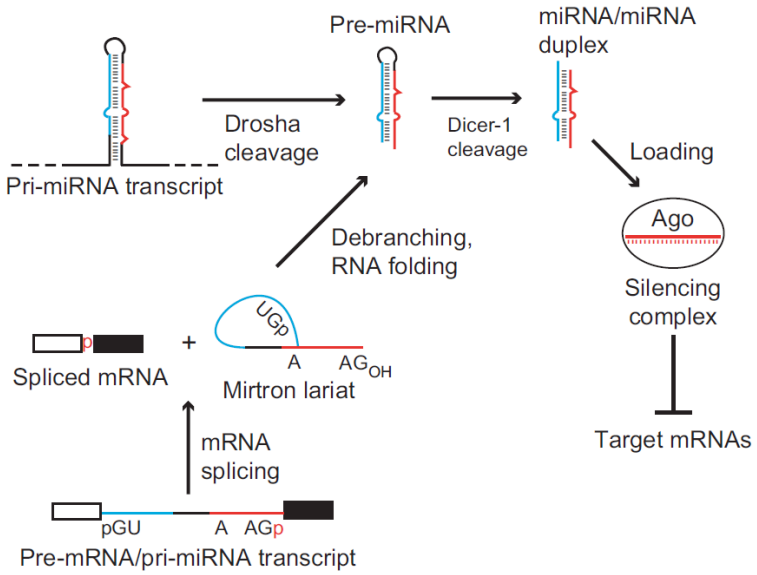

Figure 3: Mirtrons are spliced-out introns that form pre-miRNAs (from (16)).

1.1.3 Export to the cytoplasm is mediated by Exportin 5

Following Drosha cleavage, nucleocytoplasmic export of pre-miRNAs is executed by nuclear transport receptor protein Exportin 5. Its main cargo are the pre-miRNAs, but it is also a less important transporter for tRNAs (3). Exportin 5 assures that only correctly processed pre-miRNAs are exported to the cytoplasm. This is enabled by specific recognition of pre-miRNAs – determinants that interact with small pocket of the binding groove are 1 – 8 nt 3' overhang, a consequence of Drosha cleavage, and at least 14 bp stem (3,4). Due to strongly basic nature of the binding groove it attracts entities that have a strong net negative charge, namely pre-miRNA's double helical stem structure, whereas the miRNA terminal loop protrudes from the groove. In this binding manner co-transport of proteins bound to it might be possible (8). Exportin 5 is RanGTP-dependent. Prior to export it binds to a cofactor Ran which is in GTP-bound form in the nucleus and is hydrolyzed upon cargo release in the cytoplasm (3). Exportin 5 has also been shown to protect pre-miRNAs from digestion in the nucleus (4).

1.1.4 Dicer cleaves the terminal loop

In the cytoplasm RNase Dicer chops off the terminal loop and produces approximately 22 nt miRNA duplexes. Dicer binds 3' overhang of pre-miRNA with its PAZ domain and positions two catalytic domains to cleave the stem 22 nts upstream from its base, resulting in removal of the terminal loop (8). Like Drosha, Dicer is a ribonuclease type III enzyme that cuts dsRNA in a staggering fashion, creating another 2 nt 3' overhang. Dicer

collaborates with two proteins, TRBP and/or PACT, which both have dsRNA-binding domains and help generate RISC loading complex (RLC) (4).

1.1.5 Loading into Argonaute proteins

Dicer cleaved miRNA duplex is loaded on an Argonaute protein in RISC. RISC is an assembly of proteins that are required for miRNA maturation and silencing effect. It is comprised of Dicer, TRBP and/or PACT and one of the four Ago proteins. Upon Dicer cleavage, the two strands of miRNA duplex are separated. The strand with less thermodynamically stable 5' end is used as the functional strand (guide strand), while the other (passenger strand) is degraded, although sometimes both might be functional (1). Stability depends on whether Watson-Crick (G-C and U-A) or wobble base pairing (e.g. G-U or G-G) is present, where the canonical, Watson-Crick, base pairs interact more stably compared with wobble interactions (3). The decision making element for determining base pair stability at 5' ends is the helicase domain of Dicer. This domain also relocates guide strand to an Ago protein in RISC (8). There are four Ago proteins in humans; miRNAs load into either one of them with comparable frequencies (3). Mature miRNA positioned on effector Ago protein in active RISC is then carried to target mRNA genes to exhibit its regulatory mode of action.

1.2 Regulation of miRNA biogenesis

Dysregulation of miRNA levels leads to an imbalance of synthesized proteins – a condition within a cell that can cause disease (3). For example, impaired levels of miRNAs have been linked to many human cancers (8). Because regulation of miRNA abundance and activity is so important to maintain normal cellular functions, many mechanisms have evolved to govern miRNA biogenesis. Regulation of biogenesis is possible on transcriptional, post-transcriptional and a somewhat more complex feedback circuit level (3).

1.2.1 Transcriptional regulation

Transcription of miRNA genes can be positively or negatively regulated. Expression is under control of several transcription factors that can induce or repress it. Some examples are Pol II-associated transcription factors (myogenin, MYOD1), tumor-suppressive (tumor suppressor p53) or oncogenic (oncogenic protein MYC) transcription factors (3). Tumor suppressor p53 acts via induction of the miR-34 family, which consecutively targets Bcl2, Cdk4 and Cdk6 – factors that promote cell proliferation and survival (8). Epigenetic

changes – histone modifications and DNA methylation – can alter transcription profile of miRNA genes, especially changes at promoter sequences (3). Environmental conditions can also affect miRNA expression. For instance, hypoxic stress reflects in upregulation of miR-210 mediated by Hypoxia-inducible factors (5). MiR-210 targets factors involved in inhibition of angiogenesis and induction of apoptosis (17). This is thought to help cells adapt to hypoxic conditions. Since tumor cells are exposed to such environment, this regulation is promoting the survival of tumor cells (5).

1.2.2 Post-transcriptional regulation

There are several possibilities of miRNA biogenesis regulation post-transcriptionally. One of them is interference with biogenesis pathway proteins Drosha or Dicer (3). Transforming growth factor-β (TGFβ) induces maturation of some miRNAs, which causes more pri-miRNAs being recruited to the Microprocessor complex so Drosha can cleave them more efficiently (8). To maintain a pluripotent state in embryonic stem cells during development Lin-28 represses miRNA let-7 by inducing pre-let-7 terminal polyuridylation, which in turn blocks Dicer processing (3,8). Another mechanism of post-transcriptional regulation is RNA editing. This is alteration (deamination) of adenosines to inosines by ADARs (adenosine deaminases acting on RNA) in primary transcripts. Such modified miRNAs become poor substrates for RNase III proteins Drosha and Dicer. A-to-I editing can also impair target specificity due to guanosine-like base pairing properties of inosine (3,4). Little is known about miRNA turnover rate, but there is some evidence that RNA decay enzymes are directed towards pri- and pre-miRNAs in addition to targeting mature miRNAs (3). Retention of pre-miRNAs in the nucleus disables production of mature miRNAs and was observed in cancerous cells with mutated Exportin 5 (3,8).

1.2.3 Feedback circuits

Feedback circuits include interactions of transcription factors, biogenesis enzymes, miRNAs and their targets. Single- and double-negative feedback loops are engaged in miRNA biogenesis. The expression profile of an entity pursued in single-negative feedback usually fluctuates, while in double-negative feedback loops two elements repress each other (3). An example of a single-negative feedback is self-regulation of the Microprocessor complex. DGCR8 stabilizes Drosha protein and Drosha cleaves DGCR8 mRNA when activity of Microprocessor complex is adequate. In this way, unnecessary synthesis of DGCR8 protein is spared (4). A well-known case of double-negative feedback regulation is miR-200 family and transcriptional repressors ZEB1 and ZEB2 (zinc finger E-box-binding homeobox 1 and 2), which together comprise a regulatory loop involved in

epithelial to mesenchymal transition (Fig. 4) (3). High expression of the miR-200 family maintains a stable epithelial state through post-transcriptional inhibition of transcription factors ZEB1 and ZEB2. If expressed, ZEB1 and ZEB2 would transcriptionally repress miR-200 family and E-cadherin. This situation happens if cells undergo transition to mesenchymal state inducible by TGFβ (18). E-cadherin is a cell-cell adhesion protein therefore its loss causes invasive migratory cell behavior, metastases and tumor progression (18,19,20).

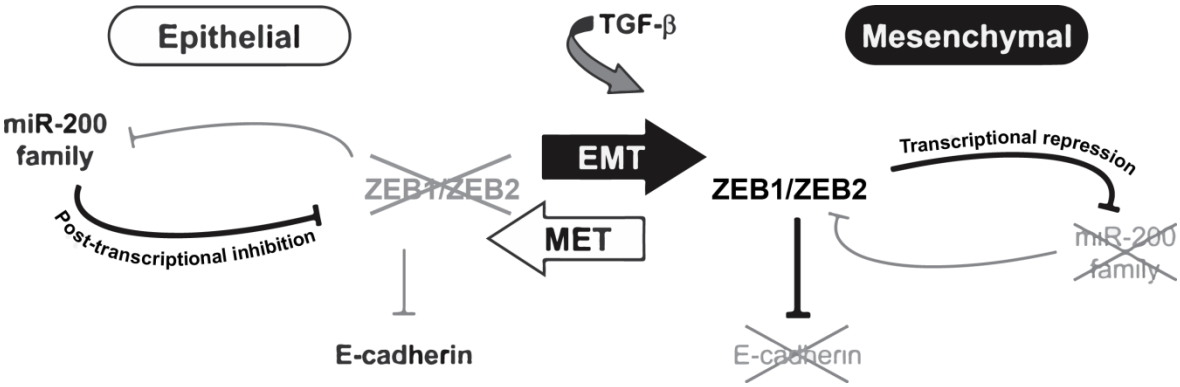

Figure 4: A double-negative feedback loop between miR-200 family and transcription factors ZEB1 and ZEB2 regulates epithelial to mesenchymal transition (from (18)). TGFβ – tumor growth factor β, ZEB1/2 – zinc finger E-box-binding homeobox 1/2 transcription factors, EMT – epithelial-mesenchymal transition, MET – mesenchymal-epithelial transition.

Aforementioned let-7-Lin28 regulation is also a double-negative feedback loop (3). Another example is miR-145 and its connection with pluripotency factors OCT4, SOX2, and KLF4 (21). Some positive feedbacks were also suggested, such as cross-activation of miR-145 and Tp53 resulting in apoptosis (5). A single miRNA, such as miR-145, can therefore be involved in more feedback circuits.

1.3 miRNA mechanism of action and RNA interference

MiRNA mechanism of action is mainly gene silencing achieved by targeting mRNAs. On the molecular scale this means mRNA cleavage, decay or repression of translation (3,5). MiRNA machinery is harnessed in a laboratory-based setting, called knockdown or RNA interference (RNAi). RNA interference is a mechanism of silencing specific genes post-transcriptionally. It is achieved by various small RNAs, such as endogenous microRNAs, exogenous small interfering RNAs (siRNAs), or short hairpin RNAs (shRNAs). RNAi results in reduced expression of the targeted gene, known as gene knockdown (22).

1.3.1 MiRNA effects on protein-coding genes

RISC is the effector complex that is guided by the mature miRNA to its target (3). MiRNAs are partially complementary to target mRNAs. Complementarity mostly occurs in the 3' UTR, but there is also evidence of miRNA binding sites in the 5' UTR and open reading frame. If complementarity between miRNA and mRNA is perfect, target mRNA undergoes cleavage mediated by an Ago protein in RISC. Ago proteins have three domains: PAZ, MID and PIWI. PIWI structure is reminiscent of RNase H. This might be the underlying reason for its endonuclease activity – PIWI domain cleaves mRNA. Out of four Ago proteins in mammals, only Ago 2 is cleavage competent. MRNA cleavage mainly happens with exogenously introduced small RNAs in RNAi, while the usual phenomenon with endogenous miRNAs in mammals is not perfect but partial complementarity, leading to repression of translation (15%) or mRNA degradation (85%) rather than mRNA cleavage. Particularly important for target base pairing is the seed region (i.e., nucleotides at positions 2 – 7 on the miRNA counting from the 5' end) (8). In RISC, 5' monophosphate of the miRNA is positioned between MID and PIWI domains, while 2 nt 3' overhang is positioned in PAZ domain of Ago2. This orientation of mature miRNA on Ago protein exposes its seed region and makes it accessible for complementary base pairing with target mRNA. Additional complementarity at the 3' end of miRNA, called supplementary and compensatory pairing, usually results in mRNA cleavage (10). Pairing 2 or 3 nts downstream of the seed region towards 3' end makes phosphate bond between nts 10 and 11 feasible for PIWI-mediated cleavage (22). Complementarity solely at the seed region brings about translational repression or mRNA degradation by various possible mechanisms (10,22). Repression of translation is caused by slowed or halted ribosomes on the mRNA that inhibit translational initiation, elongation or end translation abortively, and specific degradation of a newly synthesized peptide (1,22). MiRNA-mediated removal of 5' cap or polyadenylated 3' tail (features of mature mRNAs) leads to mRNA destabilization and thus its decay (1).

1.3.2 RNA interference as a tool for artificial gene-silencing

Gene knockdown can be achieved by introducing siRNA or shRNA into a cell. SiRNAs are 21 nts long synthesized dsRNA, phosphorylated at 5' end and with a 2 nt hydrohylated 3' overhang. They are directly incorporated into RISC. A single siRNA molecule is used more than once – it affects more than one target mRNA molecule (22). However, due to dilution upon cell division and growth, siRNAs have a short-half life, thus exhibiting only a transient silencing effect (22,23). ShRNAs are short dsDNA molecules containing a

hairpin loop. They are introduced into the cell by an external expression vector where they occupy miRNA machinery starting with transcription by Pol II or III. Compared with siRNAs they bring about longer lasting gene silencing since the effector molecule is constantly produced in the cell. Both siRNAs and shRNAs are designed to perfectly match complementary mRNA, probably leading to its cleavage (22).

RNAi is a very effective and useful way of gene silencing *in vitro* for experimental and *in vivo* for therapeutic purposes. However, there are some challenges yet to be resolved: short-half life, interruption of endogenous miRNA biogenesis and function, nonspecific off-target effects, immunogenicity, nuclease degradation and delivery problems, to name a few (22,24). Risk of nuclease degradation demands that at all times handling with si- or shRNAs RNase free protocol is engaged. This is also an issue upon systemic delivery of RNAi therapeuticals. To promote their stability chemically modified siRNAs are routinely used nowadays. Modifications on sugar component (2'-fluoro, 2'-O-methyl, 2'-halogen, 2'-amine, 2'-deoxy or linking 2'- and 4'-positions with -O-CH_2) stabilize siRNAs and have the ability to reduce activation of the immune system. Backbone modifications, such as phosphothioate (P=S) or boranophosphate (P=B) can enhance cellular uptake and resistance against nucleases. Nucleobase modifications on the passenger strand (5-iodouridine, N-3-Me-uridine, 2,6-diaminopurine residues) improve stability. Modification of siRNA termini (conjugation to cholesterol, folate, various peptides, aptamers) can enhance cell entry and provide means for delivery of RNAi pharmaceuticals to specific cells in the body (22).

In addition to the risk of saturation of miRNA machinery, which leads to imbalance of endogenous miRNA biogenesis and function, the greatest hinderance to research in the past few years has been how to overcome cellular hurdles (22,24). Small RNAs act in the cytoplasm, but they do not easily cross cellular membrane, because they are rather large molecules and negatively charged. Though the challenge remains, a few delivery circumventions have shown some progress. **Viral vectors** are mainly used with shRNAs delivery and some are capable of integrating genetic information into the cell genome, thereby producing prolonged silencing effect. Retrovirus and lentivirus vectors can be used for this intent, while adenovirus vectors and adenovirus-associated vectors also provide for successful shRNA delivery but lack the ability to integrate into the cell genome. **Liposomes** are generated from synthetic lipids that form a globularly shaped phospholipid bilayer, trapping siRNA molecules in the aqueous phase inside the vehicle. They are easy to use in a laboratory-based experiment, because their preparation only requires mixing and incubation of ingredients. Neutral or cationic lipids can be used for phospholipid

bilayer formation, but transfection with neutral liposomes is poorly efficient. On the other hand, positive surface charges of cationic liposomes make them feasible for cellular uptake, though this can potentially interact with negatively charged proteins and evokes an immune response. A representative reagent of cationic lipids is lipofectamine. **Nanoparticles** range from 10 to 1000 nm in size and consist of solid polymers that encapsulate siRNAs. In this way, siRNAs are protected from nuclease degradation. The drawback is low packing volume to surface ratio. Large numbers of nanoparticles are needed to deliver enough siRNA and their vast surface is potential area for immunogenicity inducing interactions with endogenous molecules (22).

II. Research aim

MiRNAs are endogenous small non-coding RNAs that control the expression of protein-coding genes by gene-silencing mechanisms (5). Their expression was found to be regulated at numerous levels but little is known about their transcriptional regulation (8). Bioinformatically, an average of 4.7 transcriptional start sites have been predicted per miRNA and 84% of miRNAs were estimated to have multiple promoters (2). What is the role of such promoters? One possibility could be that they are alternatively activated to enable escape from feedback regulation of the miRNA transcription. Here, we will address this question.

Our study model will be the miR-200 family and regulatory loop with transcription factors ZEB1 and ZEB2 it is involved in. They comprise a double-negative feedback loop; miR-200 family post-transcriptionally inhibits ZEB1 and ZEB2 which in turn transcriptionally repress miR-200 family and cell adhesion protein E-cadherin (Fig. 5) (18).

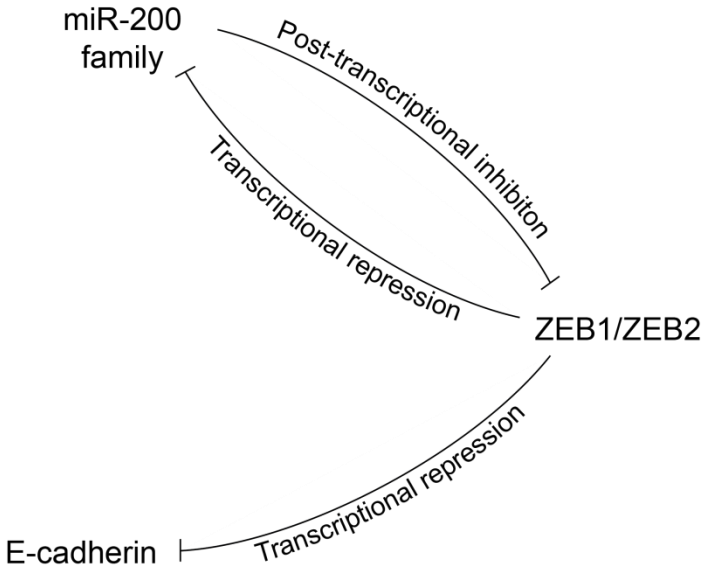

Figure 5: ZEB/miR-200 double-negative feedback loop.

MiR-200 family has five members encoded in two locations in the genome: miR-200b, miR-200a and miR-429 are found together as a cluster, and miR-200c and miR-141 together as a separate cluster (25). MiR-200b-200a-429 cluster is detectable by nuclear RNA sequencing from ENCODE project. Different start sites of its primary transcripts in different cell lines indicate that alternative promoters for this cluster could exist. Thus, our

study will concentrate on miR-200b-200a-429 cluster and ZEB2 transcription factor, since miR-200b was shown to repress it (20).

First, we will identify potential alternative miR-200b-200a-429 promoter regions and expression levels of entities involved in the above specified regulatory loop by the means of bioinformatics. Bioinformatic analysis will include searching for features that characterize promoters upstream of miR-200b in the UCSC genome browser and comparing expression levels of recognized promoter regions, E-cadherin and ZEB2 protein in various cell lines based on RNA-sequencing data. Three cell lines with differential expression of promoter regions will be chosen for further experiments.

The second aim is to validate recognized promoter regions using molecular biology techniques. We will clone recognized promoter regions of miR-200b-200a-429 cluster and E-cadherin promoter into a plasmid pGL3-basic as the vector system. E-cadherin, like miR-200 family, is targeted by ZEB2 and its promoter will serve as a positive control in later experiment that will include observing the effect of absence of ZEB2 on promoter regions. Plasmid pGL3-basic has a built-in luciferase gene which is under control of inserted sequence, thus by measuring luminescence signal in a luciferase reporter assay we will be able to assess the strength of promoters (26). We will transfect three cell lines that have differential expression of miR-200b-200a-429 promoter regions as deduced from RNA-sequencing data. Then, we will compare promoter strength to expression level of respective regions as seen in analysis of RNA-sequencing data.

To see how elements of studied feedback loop affect expression of miR-200b-200a-429 we will carry out siRNA-mediated knockdowns of ZEB2, E-cadherin and pri-miRNA processing protein Drosha. Drosha is an example of post-transcriptional regulation of miRNA expression as opposed to transcription factor activity of ZEB2 and possibly E-cadherin. Upon knockdown, we will measure its efficiency and the expression of related elements by quantitative reverse transcription PCR (RT-qPCR). ZEB2 knockdown will be followed by measurements of ZEB2, miR-200b-200a-429 primary transcript and E-cadherin expression as a positive control since ZEB2 has been shown to transcriptionally repress E-cadherin (18). E-cadherin knockdown will be followed by measurements of E-cadherin and miR-200b-200a-429 primary transcript expression. Drosha knockdown will be followed by Drosha and miR-200b-200a-429 primary transcript expression.

Next, cell lines will be cotransfected with siRNAs targeting ZEB2, E-cadherin or Drosha, and the promoter luciferase constructs. This experiment will give us insight into how alternative promoters are regulated: on which promoter ZEB2 acts and in what way;

whether other elements included or influenced in miR-200-ZEB loop (like E-cadherin) also regulate promoter activity; and whether interference with post-transcriptional processing (like the absence of Drosha) affects expression from alternative promoters. Again, activity of promoters will be detected by measuring luminescence levels in a luciferase assay.

Workflow of experiments is presented in Fig. 6.

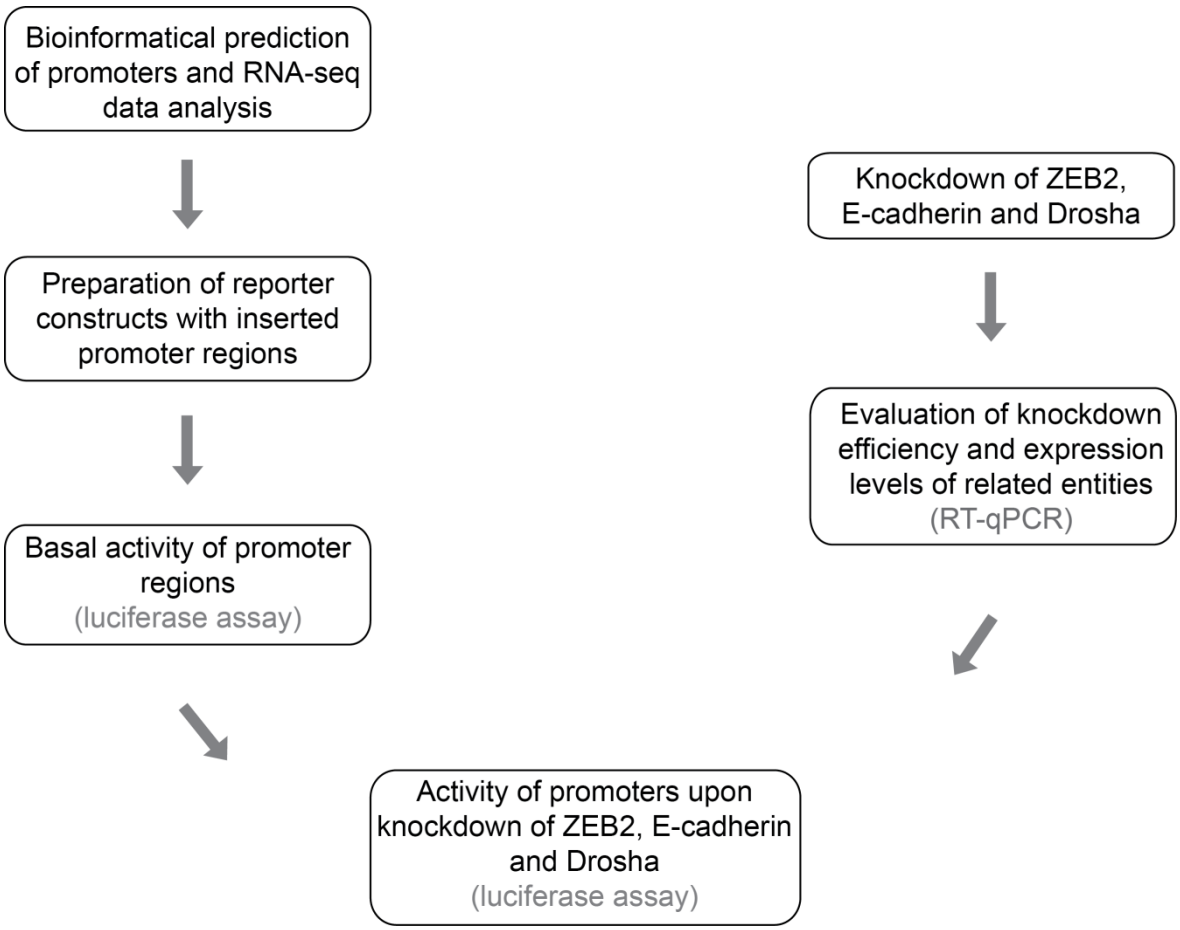

Figure 6: Schematic of the experimental workflow.

III. Materials and methods

3.1 Materials

3.1.1 Chemicals

Chemical	Manufacturer
2-Propanol for analysis ≥99.8%	Merck KGaA (Darmstadt, Germany)
Agarose	Biozym, (Oldendorf, Germany)
Ampicillin sodium salt	AppliChem (Darmstadt, Germany)
Bacto™ Agar	Dickinson and Company (Le Pont de Claix, France)
Bacto™ Tryptone	Dickinson and Company (Le Pont de Claix, France)
Bacto™ Yeast Extract Becton	Dickinson and Company (Le Pont de Claix, France)
Bovine serum albumin, REF: B9000S	New England Biolabs (Frankfurt am Main, Germany)
Bromphenol Blue	Alfa Aesar (Karlsruhe, Germany)
Chloroform	Merck KGaA (Darmstadt, Germany)
Deoxynucleotide Triphosphates (dNTPs)	Invitrogen (Karlsruhe, Germany)
Diethylpyrocarbonate ≥ 97%	Sigma-Aldrich (Steinheim, Germany)
EDTA pH 8.0, 0.5 M	AppliChem (Darmstadt, Germany)
Ethanol absolute for analysis	Merck KGaA (Darmstadt, Germany)
Glycerol 85% for analysis	Merck KGaA (Darmstadt, Germany)
Lipofectamine® 2000 Transfection Reagent, REF: 11668-019	Invitrogen (Karlsruhe, Germany)
Penicillin-Streptomycin (10,000 U/mL), REF: 15140-122	Invitrogen (Karlsruhe, Germany)
Sodium chloride	Sigma-Aldrich (Steinheim, Germany)
SYBR® Safe DNA gel stain	Invitrogen (Karlsruhe, Germany)
TRIzol® Reagent	Invitrogen (Karlsruhe, Germany)

3.1.2 Solutions, buffers and media

Solution, buffer or medium	Concentration, composition or reference	Manufacturer
Ampicillin (1000x)	100 mg/mL	
DNA loading buffer, 6x	0.25% (w/v) bromphenol blue 30% (v/v) glycerol	
Dulbecco's Modified Eagle Medium (DMEM)	REF: 41966-029	Invitrogen (Karlsruhe, Germany)
Fetal Bovine Serum	REF: ECS0180L	EUROCLONE S.p.A. (Siziano, Italy)
Kanamycin (1000x)	20 mg/mL	
LB-agar, pH 7.4	10 g/L NaCl 10 g/L tryptone 5 g/L yeast extract 15 g/L agar	
LB-medium, pH 7.4	10 g/L NaCl 10 g/L tryptone 5 g/L yeast extract	
Nuclease-free Duplex Buffer		Integrated DNA Technologies, BVBA (Munich, Germany)
Opti-MEM® I Reduced Serum Media	REF: 31985-062	Invitrogen (Karlsruhe, Germany)
PBS buffer	REF: 10010-023	Invitrogen (Karlsruhe, Germany)
Penicillin-Streptomycin (10,000 U/mL)	REF: 15140-122	Invitrogen (Karlsruhe, Germany)
TBE buffer (10x)	REF: 15581-028	Invitrogen (Karlsruhe, Germany)
Trypsin-EDTA	REF: 25300-054)	Invitrogen (Karlsruhe, Germany)

3.1.3 Biological material

3.1.3.1 Plasmid vectors

Plasmid vector	Manufacturer
pGL3-Basic Vector	Promega (Mannheim, Germany)
pRL-TK Vector	Promega (Mannheim, Germany)

3.1.3.2 Bacterial strain

DH5α Z-Competent™ *E. coli* StrainGenotype: F⁻φ80dlacZΔM15 Δ(lacZYA-argF)U169 deoR, recA1 endA1 hsdR17(r_k^- m_k^+ phoA supE44 λ^- thi-1 gyrA96 relA1, Zymo Research Europe GmbH, Freiburg, Germany

3.1.3.3 Eukaryotic cell lines

Cell line	Description
A549	adenocarcinomic human alveolar basal epithelial cells
MCF-7	human breast cancer cells
HepG2	human liver carcinoma cells

3.1.4 Primers

All primers were purchased from Integrated DNA Technologies, BVBA (Munich, Germany).

3.1.4.1 Primers for cloning into plasmid

Primer pair label	Forward primer sequence (5'-3'), reverse primer sequence (5'-3')
miR200b4200cloning	TGGCTCGAGACACCCCTCCCTCATTCCCC, TGGAAGCTTCCTGGCACAGGAAGTCAGTT
miR200b2000cloning	TGGGGTACCTTTGCGGTTCTTTTCAGACA, TGGAAGCTTGTGGCCACAGGTCAAGAAAT
CDH1cloning	TGGCTCGAGTAGAGGGTCACCGCGTCTAT, TGGAAGCTTGGGCTGGAGTCTGAACTGAC

3.1.4.2 Primers for sequencing

Primer pair label	Forward primer sequence (5'-3'), reverse primer sequence (5'-3')
pGL3_Pr_seq	CCATCTTCCAGCGGAT, GCTGTCCCCAGTGCAA

3.1.4.3 Primers for RT-qPCR

Primer pair label	Forward primer sequence (5'-3'), reverse primer sequence (5'-3')
qmiR200b5	TACTGAGCTTCCCAGCGAGT, AATGCTGCCCAGTAAGATGG
qZEB2ex1-ex2	CCAATCCCAGGAGGAAAAAC, GTTGGCAATACCGTCATCCT
qCDH1ex3-4	CCTGGGACTCCACCTACAGA, TGTGAGCAATTCTGCTTGGA
qPCRDroshaE1-E2	TGGTACCTGCGGTAGTAGCC, GTTGCGGTTTGGAAACAGAG
qGAPDH	GCTCTCTGCTCCTCCTGTTC, ACGACCAAATCCGTTGACTC
qbetaactin	CGACAGGATGCAGAAGGAG, GTACTTGCGCTCAGGAGGAG

3.1.5 siRNAs

All siRNAs were purchased from Integrated DNA Technologies, BVBA (Munich, Germany).

Dicer-substrate RNA duplex label	Sense strand sequence (5'-3'), antisense strand sequence (5'-3')
siDrosha	ACAAUGAGACGAGAAGUAACGGUGG, CCACCGUUACUUCUCGUCUCAUUGUAU
siCDH1	CCAAAGUCACGCUGAAUACAGUGGG, CCCACUGUAUUCAGCGUGACUUUGGUG
siZEB2	GAGGCGCAAACAAGCCAAUCCCAGG, CCUGGGAUUGGCUUGUUUGCGCCUCUU
siNC1	Sequence kept secret by Integrated DNA Technologies, BVBA (Munich, Germany).

Dicer-substrate RNA duplex label	Sense strand sequence (5'-3'), antisense strand sequence (5'-3')
siNC Rb first ex1	CAAAGAGUCUGGUGGGUGACUGUGG, CCACAGUCACCCACCAGACUCUUUG
siNC Rb first ex3	CUUCCAUUUCCAAGUGUGUCGAUTT, AAAUCGACACACUUGGAAAUGGAAG

3.1.6 Enzymes and kits

Enzyme/kit	Manufacturer
AllPrep DNA/RNA Mini Kit	Qiagen (Hilden, Germany)
Antarctic Phosphatase Reaction Buffer 10x, REF: B0289S	New England Biolabs (Frankfurt am Main, Germany)
Antarctic Phosphatase, REF: M0289S	New England Biolabs (Frankfurt am Main, Germany)
DMSO, REF: B0515A	New England Biolabs (Frankfurt am Main, Germany)
Dual-Glo® Luciferase Assay System, REF: E2940	Promega (Mannheim, Germany)
Fast SYBR®Green Master Mix	Applied Biosystems (Darmstadt, Germany)
High Capacity RNA-to-cDNA Kit	Applied Biosystems (Darmstadt, Germany)
HindIII, REF: R0104L	New England Biolabs (Frankfurt am Main, Germany)
KpnI, REF: R0142S	New England Biolabs (Frankfurt am Main, Germany)
NEBuffer 2, REF: B7002S	New England Biolabs (Frankfurt am Main, Germany)
Phusion® HF Buffer, REF: B0518S	New England Biolabs (Frankfurt am Main, Germany)
Phusion® High-Fidelity DNA Polymerase, REF: M0530L	New England Biolabs (Frankfurt am Main, Germany)
QIAGEN Plasmid Mini Kit	Qiagen (Hilden, Germany)
QIAquick Gel Extraction Kit	Qiagen (Hilden, Germany)
QIAquick PCR Purification Kit	Qiagen (Hilden, Germany)
T4 DNA Ligase Reaction Buffer 10x, REF: B0202S	New England Biolabs (Frankfurt am Main, Germany)

Enzyme/kit	Manufacturer
T4 DNA Ligase, REF: M0202L	New England Biolabs (Frankfurt am Main, Germany)
TOPO® TA Cloning® Kit, REF: 45-0641	Invitrogen (Karlsruhe, Germany)
XhoI, REF: R0146S	New England Biolabs (Frankfurt am Main, Germany)

3.1.7 Molecular weight markers

Molecular weight marker	Manufacturer
100 bp DNA Ladder	Jena Bioscience (Jena, Germany)
500 bp DNA Ladder	Jena Bioscience (Jena, Germany)
Mid Range DNA Ladder	Jena Bioscience (Jena, Germany)

3.1.8 Devices

Device	Manufacturer
Analytical balance, BP 61	Sartorius (Göttingen, Germany)
Aspirator pump hand control VacuuHandControl	VACUUBRAND (Wertheim, Germany)
Bacteriological incubator (E.coli), Innova® 44	Eppendorf (Hamburg. Germany)
Bacteriological incubator Heraeus-Brutschrank B 504	Heraeus (Hanau, Germany)
Centrifuge, 5430	Eppendorf (Hamburg. Germany)
Centrifuge, 5810 R	Eppendorf (Hamburg. Germany)
Centrifuge, MiniSpin	Eppendorf (Hamburg. Germany)
Digital balance, A&D EK-300i	A&D Company, Limited (Griesheim, Germany)
Electrophoresis apparatus Mini-PROTEAN Tetra Cell	Biorad (München, Germany)
Electrophoresis chamber Sub-Cell® GT Cell	Biorad (München, Germany)
Fluid aspiration system BioChem-VacuuCenter BVC 21	VACUUBRAND (Wertheim, Germany)
Fluorescence spectrofluorometer LUMIstar Omega	BMG LABTECH (Ortenberg, Germany)

Device	Manufacturer
Gel imager Bio-Rad Gel Doc 2000	Biorad (München, Germany)
Incubator for A549 and MCF-7 cells, Heracell CO2	Heraeus/Kendro (Hanau, Germany)
Inverted microscope Axiovert 40 CFL	Zeiss (Jena, Germany)
Laminar flow bench, HERAsafe® HSF 12	Heraeus/Kendro (Hanau, Germany)
Microcentrifuge, FRESCO 17	Heraeus (Hanau, Germany)
Microwave SEVERIN 900&Grill	SEVERIN (Sundern, Germany)
PCR cycler peqSTAR 2X Gradient Thermocycler	PEQLAB (Erlangen, Germany)
Pipet controller pipetus®	Hirschmann Laborgeräte (Eberstadt, Germany)
Pipette Multipette ® Xstream	Eppendorf (Hamburg. Germany)
Pipettes PIPETMAN P2, P20, P200, P1000	Gilson (Middleton, USA)
Power supply Power Pac 300	Biorad (München, Germany)
RT-qPCR reader 7900HT Fast Real-Time PCR System	Applied Biosciences (Foster City, USA)
Refrigerator Forma Scientific -86C (-80 °C)	Thermo Scientific (Waltham, MA, United States of America)
Refrigerator Liebherr Comfort (-20 °C)	Liebherr (Bulle, Switzerland)
Refrigerator Liebherr Premium (-20 °C)	Liebherr (Bulle, Switzerland)
Refrigerator Liebherr Premium NoFrost (-20 °C)	Liebherr (Bulle, Switzerland)
Refrigerator Liebherr ProfiLine (4 °C)	Liebherr (Bulle, Switzerland)
Shaker ST5	neoLab (Heidelberg, Germany)
Spinner, mini centrifuge Stuart SCF1	Bibby Scientific (Staffordshire, United Kingdom)
UV-Vis Spectrophotometer NanoDrop 2000	PEQLAB (Erlangen, Germany)
Vortex Heidolph Reax 2000	Heidolph (Schwabach, Germany)
Waterbath Thermomixer® compact	Eppendorf (Hamburg. Germany)
Waterbath, WNE	Memmert (Büchenbach, Germany)

3.2 Methods

3.2.1 Cell culturing

Three cell lines were employed in our experiments: A549, MCF-7 and HepG2. All are adherent type and were handled in the same manner. They were grown in 10 cm tissue culture dishes in 10 mL of cell growth medium (DMEM supplemented with 10% fetal calf serum (FBS), 50 U/mL penicillin, and 50 U/mL streptomycin) in humidified incubator at 37°C and 5% CO_2 in air. The cells were routinely passaged once they reached 90% confluency (usually every three to four days). Cell were detached with trypsin/EDTA. Growth medium was removed, cells were gently washed with PBS and 1 mL trypsin/EDTA was added per 10 cm tissue culture dish. The plate was incubated at 37°C for 4 minutes and then microscopically examined. If detachment was sufficient, cells were dispersed in 9 mL of growth medium and diluted 1:10 with growth medium into a new culture dish.

3.2.2 Bioinformatical prediction of promoters with UCSC genome browser

For prediction of promoter regions we used UCSC genome browser (hg19 assembly, released Feb. 2009), accessed at http://genome.ucsc.edu/. Vast amounts of genomic data that emerged in recent years had become difficult to understand without data visualization (27). University of California Santa Cruz genome browser is an interactive on-line tool for viewing genomic data from various databases in graphical form, displaying a location in the selected genome horizontally and annotated datasets, called 'tracks', vertically (28). Users enter position, gene symbol or search term in the search window or drag an indicator tool to the desired position on the chromosome ideogram to visualize a specific region in a genome. The scale of a region being viewed is always displayed above all tracks and can be adjusted by zooming in and out. Tracks are arranged in a tree-like structure. First by species, then in track groups, e.g. Mapping and Sequencing Tracks, Phenotype and Disease Associations, Genes and Gene Prediction Tracks, mRNA and EST Tracks, Expression, Regulation, Comparative Genomics, etc.. Further on, tracks are clustered in subgroups (called super-tracks), narrowing down to specific tracks for cell lines and cellular localization. Each track group, super-track or specific track can be turned on or off. If turned on, data from the tracks can be viewed in dense, squish, pack or full mode. An example of tracks viewed for a gene in UCSC genome browser is represented in Fig. 7. Users can download data from the browser in various file formats to

use it for further studies, or upload a dataset to view their own data and to contribute to this open-source platform (29).

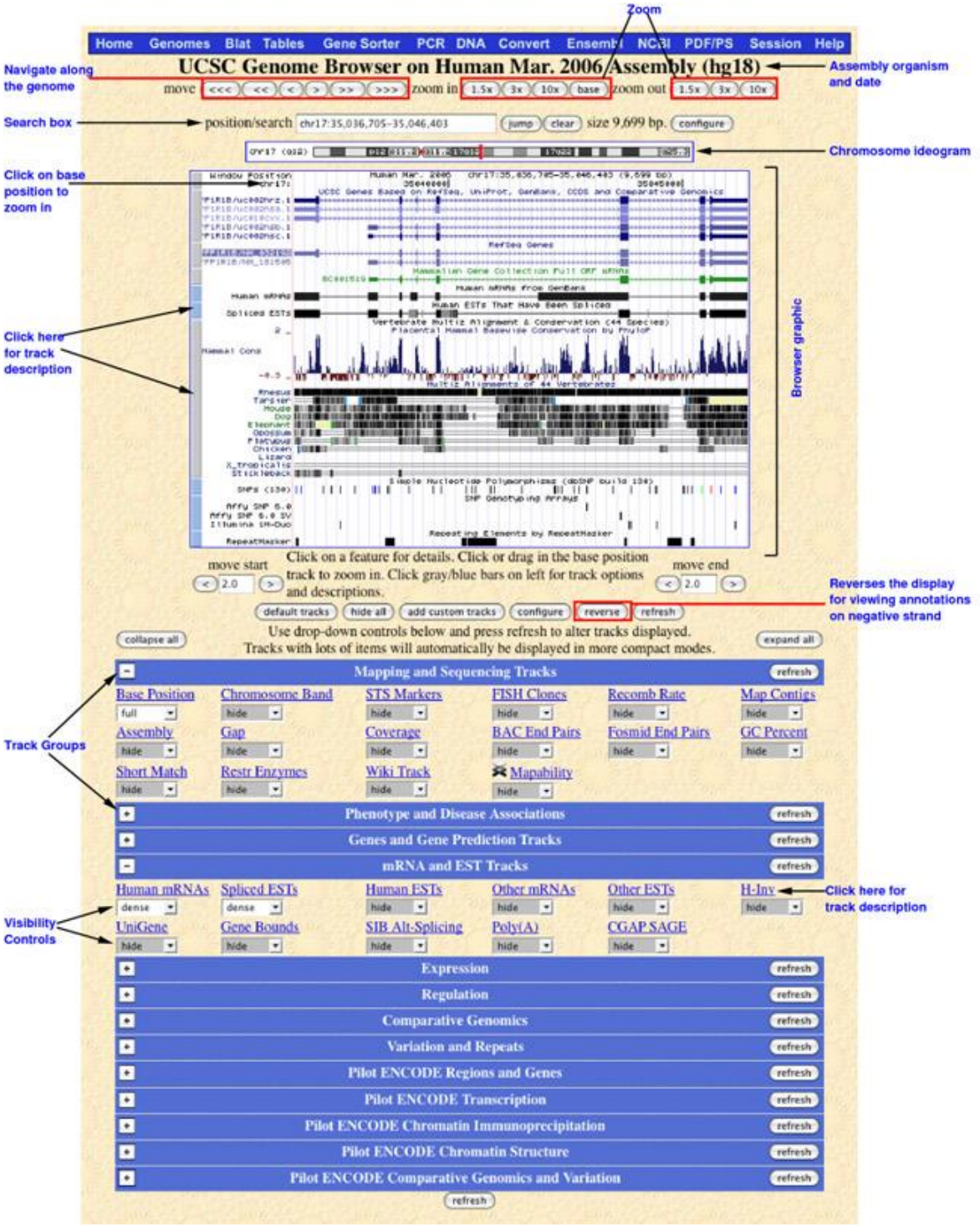

Figure 7: Default tracks for the human hg18 (NCBI Build 36) assembly at the PPP1R1B gene locus. This gene is on the positive strand; using the reverse button, the display can be reversed to observe genes in the 5' to 3' orientation if they are located on the opposite

strand. Copyrighted by 2010 Nature Education. All rights reserved. Used here under terms of use (27).

3.2.2.1 Promoter recognition

For recognizing the promoter regions of miR-200b-200a-429, we used different annotations in several cell lines. We inspected a larger portion of the genome (10 kb) upstream of miR-200b and turned on tracks for mammal conservation, cap analysis of gene expression (CAGE) data, DNase I hypersensitivity clusters, Polymerase II binding, H3K27Ac histone acetylation and H3K4Me3 histone trimethylation marks and RNA sequencing data. Based on that, three cell lines with differential expression of predicted promoter regions were selected to continue wet-lab experiments with.

3.2.2.2 Analysing RNA-seq data from ENCODE/Cold Spring Harbor Lab

To analyze RNA expression of the two predicted promoters of miR-200b-200a-429 we used ENCODE/Cold Spring Harbor Laboratory RNA sequencing annotation. We displayed data for all available cell lines in this dataset in the UCSC genome browser. We zoomed in the area approximately 6 kb upstream of miR-200b and estimated expression levels for each promoter in each cell line with a minus (not expressed), plus (weakly expressed), double plus (moderately expressed) and triple plus (highly expressed). With the same descriptive signs we estimated E-cadherin and ZEB2 expression levels.

3.2.3 Cloning of promoter regions into luciferase reporter vector

Putative promoter regions (denoted as -2.0-kb promoter and -4.2-kb promoter) and E-cadherin promoter region were inserted in luciferase reporter vector system, plasmid pGL3-basic, and were later employed in luciferase assay (see 3.2.9). Plasmid pGL3-basic (Fig. 8) is a tool for quantitative analysis of putative regulatory sequences, e.g. promoters. It encodes an enzyme luciferase, which in presence of a specific substrate catalyzes luminescent reaction. However, pGL3-basic alone lacks promoter sequence that would confer luciferase expression. Numerous restriction sites upstream of the luciferase gene enable putative promoter sequence insertion. Strength of inserted promoter correlates with the yield of emitted light – the more powerful the inserted promoter the more luciferase is transcribed and more luminescence is produced.

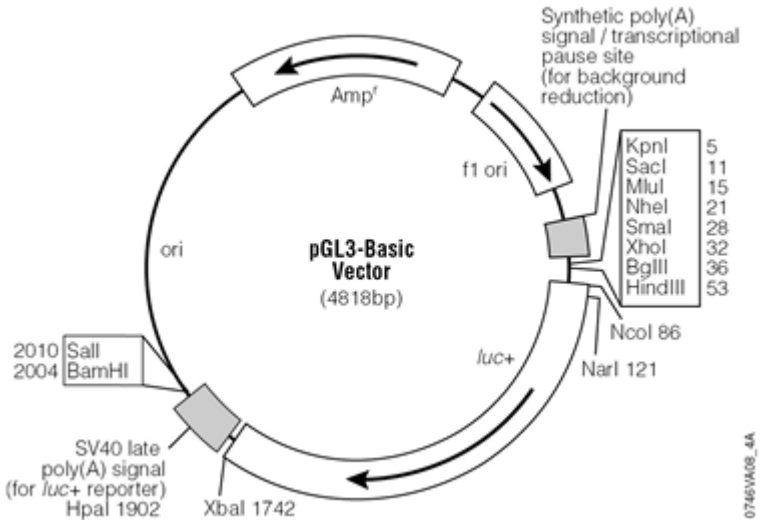

Figure 8: Plasmid pGL3-basic vector circle map (from 26). *luc*+ – cDNA encoding firefly luciferase, Amp – gene conferring ampicillin resistance in E. coli, ori – origin of replication in *E. coli*. Arrows indicate the direction of transcription.

3.2.3.1 Primer design

Input sequence was obtained at UCSC Genome Browser. Primers were designed with Primer3 program (v. 0.4.0) (30,31), accessible online at http://bioinfo.ut.ee/primer3-0.4.0/, in a way that the amplified region was approximately 800 nts long with a 100 nts overlap of the putative transcription start sites. Restriction sites for *Xho*I to cut the amplified region at the 5' end and *Hind*III at the 3' end were included in primers and an additional 5' – TGG – 3' tag at the 5' end to higher the specificity and efficiency of restriction. Primers were checked for presence of more restriction sites for the enzymes used, with NEBcutter V2.0 (32), accessible online at http://tools.neb.com/NEBcutter2/.

3.2.3.2 Genomic DNA isolation

We used AllPrep DNA/RNA Mini Kit from Qiagen (Hilden, Germany) to isolate genomic DNA from A549 cells according to manufacturers' instructions with some modifications. Detailed protocol is described in supplementary information (Supp. 1).

3.2.3.3 Amplification of promoter regions with PCR

To amplify promoter regions we used Polymerase chain reaction (PCR) in a 20 µL reaction volume. Primer mixture was prepared with forward and reverse primers in ratio 1:1.

Reaction mixture	V [µL]
ddH2O	8.8
HF buffer (5x)	4.0
dNTP (10 mM)	0.4
DNA (<250 ng)	5.0
DMSO	0.6
DNA Polymerase	0.2
Primer mixture (10 µM)	1.0

We used the following PCR protocol:

PCR protocol		T [°C]	t
Initial denaturation		98	30''
35 cycles	Denaturation	98	10''
	Annealing	60	10''
	Extension	72	15''
Final extension		72	10'
Hold		8	∞

3.2.3.4 Gel electrophoresis of the PCR product

To determine the size of obtained PCR products and control PCR specificity agarose gel electrophoresis was used. We prepared 50 mL of 2% (w/V) agarose gel in ddH_2O. The mixture was then heated up near boiling point to melt agarose. To visualize PCR product 1 µL of SybrSafe was added. SybrSafe is an in-gel stain that emits strong fluorescent light when bound to dsDNA but not in solution. It is sensitive to heat, therefore it is added to agarose solution when it is cooled down to 60°C. The solution was poured into a mould and left to gel for 30 minutes with a comb to create loading pockets inserted. Gel was then sunk in TBE buffer. Loading dye was added to PCR products, diluted with ddH_2O in 1:6 ratio. Samples were loaded in gel pockets. To determine the size of amplicons, suitable molecular weight marker was added in a gel pocket next to samples. Electrophoresis was run for 25 minutes at 110 V.

3.2.3.5 Gel extraction

We used QIAquick Gel Extraction Kit from Qiagen (Hilden, Germany) to extract DNA from agarose gel according to manufacturer's instructions with some modifications. Detailed protocol is described in supplementary information (Supp. 2).

3.2.3.6 Restriction digestion of PCR-amplified promoter regions and plasmid pGL3-basic

To insert promoter regions into plasmid pGL3-basic we created complementary sticky ends in both, using restriction enzyme digestion in a 20 µL reaction volume. In one case, *Xho*I would cut the fragment we wanted to insert into the plasmid, so we used *Kpn*I instead, for both amplified promoter region and plasmid digest. All restriction enzyme digestion reactions were performed with initial 20 minutes at 37°C for optimal enzyme activity. Subsequently, the temperature was raised to 65°C for 20 minutes to inactivate the enzymes.

Reaction mixture for promoter regions restriction	**V [µL]**
XhoI or KpnI	1
HindIII	1
Buffer II (10x)	2
BSA (10x)	2
Promoter region (from step 0)	14

Reaction mixture for plasmid restriction	**V [µL]**	**m [µg]**
XhoI or KpnI	1	
HindIII	1	
Buffer II (10x)	2	
BSA (10x)	2	
Plasmid pGL3-basic		3
ddH_2O	ad 20	

3.2.3.7 Purification of digested DNA fragment and plasmid

To purify digested DNA fragment we used QIAquick PCR Purification Kit from Qiagen (Hilden, Germany) according to manufacturer's instructions with some modifications. Detailed protocol is described in supplementary information (Supp. 3).

To purify digested plasmid we used gel electrophoresis followed by gel extraction with a kit in the same manner as for PCR product purification (see 3.2.3.4 and 3.2.3.5).

3.2.3.8 Dephosphorylation of plasmid

To dephosphorylate digested plasmid pGL3-basic we used 20 µL reaction volume. This reaction was performed with initial 10 minutes at 37°C for optimal enzyme activity. Then, the temperature was raised to 65°C for 5 minutes to inactivate the enzyme.

Reaction mixture	**V [µL]**
Digested plasmid pGL3-basic	17
10x Antarctic p. Buffer 2	2
Antarctic phosphatase	1

3.2.3.9 Ligation

To ligate promoter regions into plasmid pGL3-basic we used 10 µL reaction volume. This reaction was performed at 16°C for at least 18 hours and up to over-weekend incubation.

Reaction mixture	**V [µL]**
Plasmid GL3-basic	1
PCR product	7
T4 DNA ligation buffer	1
T4 DNA ligase	1

3.2.3.10 Transformation of E. coli

A vial of 50 µL competent E. coli from -80°C refrigerator was left to thaw on ice for 10 minutes. 4 µL of ligation mixture was pipetted into the vial and gently stirred with the tip. Then bacteria with ligation product were put on ice for 10 minutes so bacteria accepted introduced material. The content of the vial was then evenly distributed on LB agar in petri plates. To assure aseptic environment this was done in proximity of a flame. LB agar was earlier supplemented with 0.1% of appropriate antibiotic to provide conditions for selective bacteria growth. Petri dish with bacteria was incubated overnight (up to 16 hours) at 37°C.

3.2.3.11 Miniculturing

To amplify colonies that grew after transforming E. coli, a single colony was picked from the selective plate with a pipette tip and inoculated into 4 mL LB medium supplemented

with 0.1% of appropriate antibiotic. Those preparations were then incubated at 37°C overnight with continuous shaking at 200 rpm.

3.2.3.12 Plasmid extraction

We used QIAGEN Plasmid Mini Kit from Qiagen (Hilden, Germany) to isolate the plasmid from overnight bacteria minicultures according to manufacturer's instructions with some modifications. Detailed protocol is described in supplementary information (Supp. 4).

3.2.3.13 Confirmation of insert presence in the plasmid

To test for the presence of the insert, plasmid DNA was digested with restriction enzymes as in step 0. Products of restriction digestion were run on agarose gel (as in step 0) to confirm the appropriate size of the insert.

3.2.3.14 Sequencing of cloned regions

Sequential analysis was performed by GATC Biotech in Köln, Germany. Samples were prepared according to service provider's request. Results were evaluated using alignment tool in UCSC genome browser (hg 19 assembly, released Feb. 2009) (33).

3.2.4 Transfection protocols

3.2.4.1 Transfection with plasmid constructs

To deliver plasmid vectors into cells, we used Lipofectamine 2000 transfection reagent for A549 and MCF-7 cells, and HiPerFect transfection reagent for HepG2 cells. For explanation on experimental plasmid constructs and the role of control vector, see 3.2.9 Luciferase reporter assay, which followed transfections with plasmid constructs. Each sample was done in triplicates.

3.2.4.1.1 Lipofectamine 2000 transfection protocol

Lipofectamine 2000 transfection reagent was used with A459 and MCF-7 cells. 10^4 cells/well were plated in 96-well luciferase plate in 200 µL of growth medium one day prior to transfection. The next day we removed the medium, added 50 µL cell growth medium (DMEM) supplemented with 10% FBS but with no antibiotics. We supplemented Opti-MEM® I Reduced Serum Media with 2% Lipofectamine 2000, gently mixed and incubated for 5 minutes. Then, we mixed it in 1:1 ratio with Opti-MEM® I Reduced Serum Media containing a single experimental plasmid construct and control vector, and incubated for 20 minutes. We added 50 µL of this mixture into each well, resulting in final amount of 200

ng experimental plasmid construct and 50 ng control vector per 100 µL well. 24-hour incubation was followed by luciferase assay.

3.2.4.1.2 HiPerFect transfection protocol

HiPerFect transfection was used with hard-to-transfect HepG2 cells. On the day of transfection 10^4 cells/well were plated in 96-well luciferase plate in 83 µL of cell growth media (DMEM) supplemented with 10% FBS but with no antibiotics. Experimental plasmid construct and control vector were added to Opti-MEM® I Reduced Serum Media supplemented with 2% HiPerFect transfection reagent and incubated for 10 minutes. We added 17 µL of this mixture into each well, resulting in final amount of 200 ng experimental plasmid construct and 50 ng control vector per 100 µL well. 24-hour incubation was followed by luciferase assay.

3.2.4.2 Transfection with siRNAs

One day prior to transfection 10^5 A549 cells/well and 1.5 10^5 MCF-7 cells/well were seeded in 6-well plates in 2 mL cell growth medium (DMEM supplemented with 10% fetal calf serum (FBS), 50 U/mL penicillin, and 50 U/mL streptomycin). The next day, transfection was carried out. Both cell lines were transfected in the same manner with Lipofectamine 2000. We removed growth medium and replaced it with 800 µL antibiotics free medium. We supplemented Opti-MEM® I Reduced Serum Media with 2% Lipofectamine 2000, gently mixed and incubated for 5 minutes. Then, we mixed it in 1:1 ratio with Opti-MEM® I Reduced Serum Media containing siRNA, and incubated for 20 minutes. We added 200 µL of this mixture into each well, resulting in final concentration of 40 nM siRNA. Either 24-hour or 48-hour incubation was followed by RNA extraction and RT-qPCR.

3.2.4.3 Co-transfection with plasmid constructs and siRNAs

10^4 cells/well were plated in 96-well luciferase plate in 200 µL of growth medium one day prior to transfection. The next day we removed the medium, added 35 µL cell growth media (DMEM) supplemented with 10% FBS but with no antibiotics. We supplemented Opti-MEM® I Reduced Serum Media with 2% Lipofectamine 2000, gently mixed and incubated for 5 minutes. Then, we mixed it in 1:1 ratio with Opti-MEM® I Reduced Serum Media containing a single experimental plasmid construct or control vector and a single siRNA, and incubated for 20 minutes. We added 35 µL of this mixture into each well, resulting in final amount of 200 ng experimental plasmid construct and 50 ng control vector per well (70 µL) and final concentration of 40 nM siRNA. 24-hour incubation was followed by luciferase assay.

3.2.5 RNA extraction

We used TRIzol® Reagent from Invitrogen (Karlsruhe, Germany) to extract RNA from cells according to manufacturer's instructions with some modifications. Detailed protocol is described in supplementary information (Supp. 5).

3.2.6 DNase I digest

To remove any residual DNA from RNA samples obtained with TRIzol® Reagent we did DNase I digest in a 10 µL reaction volume with 10 minute incubation at 37°C.

Reaction mixture	V [uL]	m [ug]
RNA		cca. 1
DNase I buffer (10x)	1	
DNase I	0.5	
DEPC treated water	ad 10	

Following this step we removed 1 µL from reaction mixture, added 1 µL 0.05 M EDTA and heat inactivated the enzyme at 75°C for 10 minutes.

3.2.7 Reverse transcription

To obtain complementary DNA from sample RNA we performed reverse transcription with High Capacity RNA-to-cDNA Kit in a 20 µL reaction volume using reaction mixture from DNase I digest as sample RNA source.

Reaction mixture	V [uL]
DNase I reaction mixture (contained cca. 1 ug RNA)	9
RT Buffer (2x)	10
Reverse Transcriptase	1

Reverse transcription was carried out at 37 °C for 60 minutes, following enzyme inactivation at 95 °C for 5 minutes. Obtained cDNA was then diluted to 100 uL by adding 80 uL ddH_2O and used for RT-qPCR.

3.2.8 Quantitative reverse transcription Polymerase chain reaction (RT-qPCR)

For RT-qPCR reactions, the Fast SYBR®Green Master Mix was used according to manufacturer's instructions. SYBR®Green dye emits fluorescence when bound to double-stranded DNA. It enables quantification of amplified DNA – more RT-qPCR product yields higher fluorescent signal. To achieve better RT-qPCR specificity, primers were designed in a way that their RT-qPCR product spanned the intron when detecting protein expression, or spanned Drosha cleavage site when detecting primary miRNA transcripts (Fig. 9).

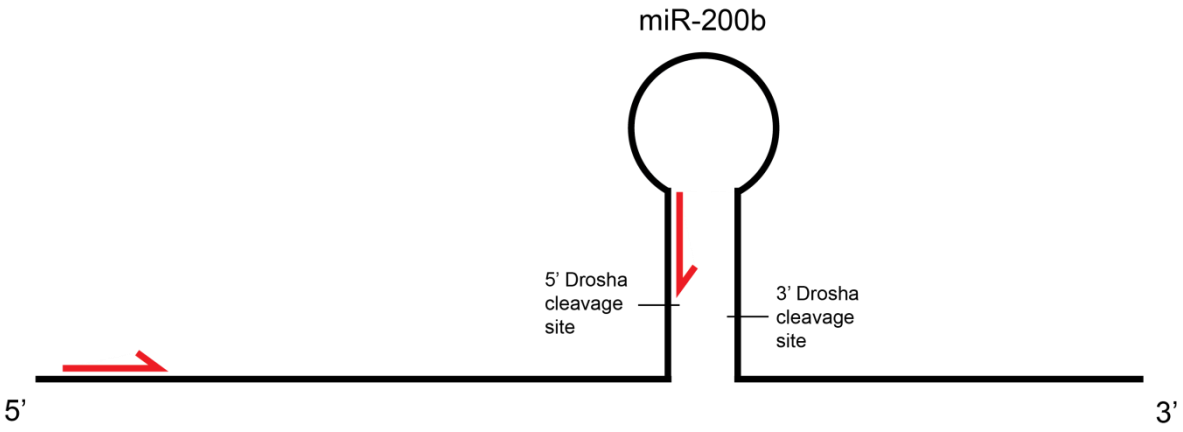

Figure 9: Primers (red) spanning Drosha cleavage site of miR-200b ensure specific detection of primary transcript.After the transcript is cut by Drosha amplicon does not form (34).

All RT-qPCR reactions were done in triplicates in 10 µL reaction volume.

Reaction mixture	**V [uL]**
Fast SYBR®Green Master Mix (2x)	5
Forward primer (10 µM)	0.25
Reverse primer (10 µM)	0.25
cDNA (from step 0)	2
ddH_2O	2.5

Prior to performing RT-qPCR reactions in Applied Biosciences 7900HT Fast Real-Time PCR System, reaction mixtures were spun down for 1 minute at 2000 rpm to expel air bubbles.

RT-qPCR protocol		T [°C]	t
Initial denaturation		94	10'
40 cycles	Denaturation	94	20''
	Annealing, extension, and read fluorescence	60	60''

Sample gene expression changes were analyzed relatively to endogenous control – constitutively expressed gene GAPDH.

3.2.9 Luciferase reporter assay

To evaluate basal promoter activity of putative promoter regions and later their activity upon knockdown of ZEB2, E-cadherin and Drosha, luciferase reporter assay was employed. The assay was performed with the Dual-Glo® Luciferase Assay System that enables consecutive quantification of luminescent signal from two reporter enzymes: *Photinus pyralis* (firefly) luciferase and *Renilla reniformis* (sea pansy) luciferase. Due to differential substrate requirements of the enzymes, their luminescent signal can be measured separately in a single sample. First, a substrate for firefly luciferase is added and generated luminescence is measured. Then, by adding the second reagent, firefly luciferase reaction is quenched while simultaneously *Renilla* luciferase reaction is activated. Cells were always co-transfected with two plasmids for luciferase reporter assay: pGL3-basic (described in 3.2.3), harboring the firefly luciferase gene under control of inserted putative promoter regions; and pRL-TK, an internal control vector harboring constitutively expressed *Renilla* luciferase gene, which thereby represents the baseline response of the system. For data analysis, firefly luciferase signal was normalized to *Renilla* luciferase signal to minimize effects of cell number and health, transfection efficiency and nonspecific cellular responses (35). The principle of this experiment is depicted in Fig. 10.

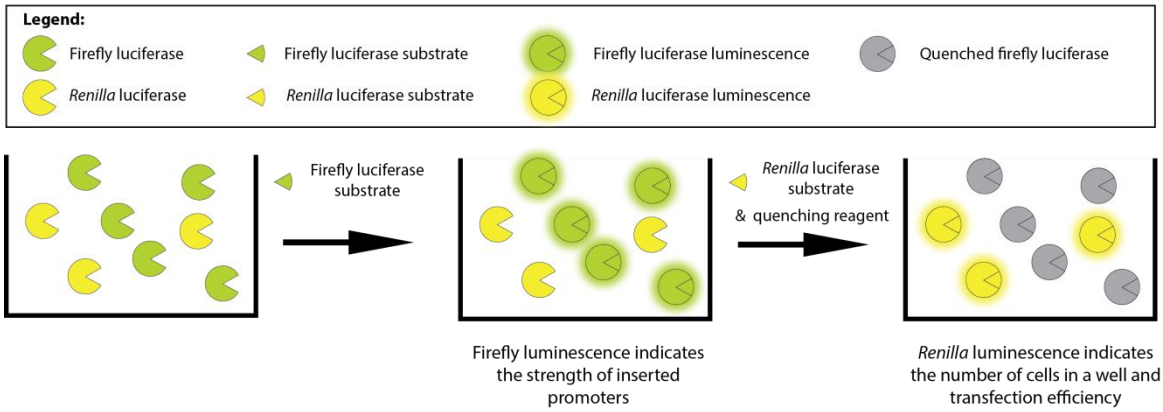

Figure 10: Principle of performed luciferase reporter assay. Expressed luciferase genes generated luminescence, which was measured twice: first firefly luminescence upon addition of firefly luciferase substrate, and then Renilla luciferase luminescence upon addition of reagent that quenched firefly luciferase reaction, and *Renilla* luciferase substrate, that initiated *Renilla* luminescence.

The protocol was performed following manufacturer's instructions with some adjustments of reagent volumes. Detailed protocol is described in supplementary information (Supp. 6).

IV. Results and discussion

First, we predicted putative promoter regions of miR-200b-200a-429 cluster with several annotations in UCSC genome browser and second, we compared expression levels of putative promoter regions with E-cadherin and ZEB2 – proteins that are involved in regulatory loop with miR-200 family – by RNA-seq data available for 11 cell lines. Then, we validated predicted promoter regions by means of luciferase reporter assay. To determine which promoter region is affected by ZEB2 and possibly E-cadherin, we coupled knockdown with luciferase reporter assay. Prior to merging of the two techniques, we assessed knockdown efficiency and subsequent alterations of miR-200b-200a-429 primary transcript levels by RT-qPCR.

4.1 Identification of alternative TSSs of miR-200b-200a-429 cluster

For prediction of promoter regions, we analyzed 10-kb region upstream of the first annotated miRNA in the examined cluster, miR-200b, as most miRNA TSSs are located up to 10 kb upstream of the gene (2). We visualized the following data in the UCSC genome browser (hg19 assembly, released Feb. 2009): 100 vertebrates basewise conservation by PhyloP, digital DNaseI hypersensitivity clusters in 125 cell types from ENCODE, H3K4Me3 mark on 7 cell lines from ENCODE, H3K27Ac mark on 7 cell lines from ENCODE, nucleus polyA+ RNA-seq plus signal from ENCODE/Cold Spring Harbor Laboratory for four cell lines (K562, A549, HepG2, MCF-7), Polymerase II transcription factor binding sites ChIP-seq density signal from ENCODE/OpenChrom-University of Texas at Arlington for four cell lines (K562, A549, HepG2, MCF-7), nucleus polyA+ CAGE plus start sites from ENCODE/RIKEN for four cell lines (K562, A549, HepG2, MCF-7).

These genomic characteristics were useful for our study, as they indicate high possibility of promoters or TSSs where they are present or where they initiate in the genome. Studied features for miR-200b-200a-429 TSS and promoter prediction are shown in Fig. 11.

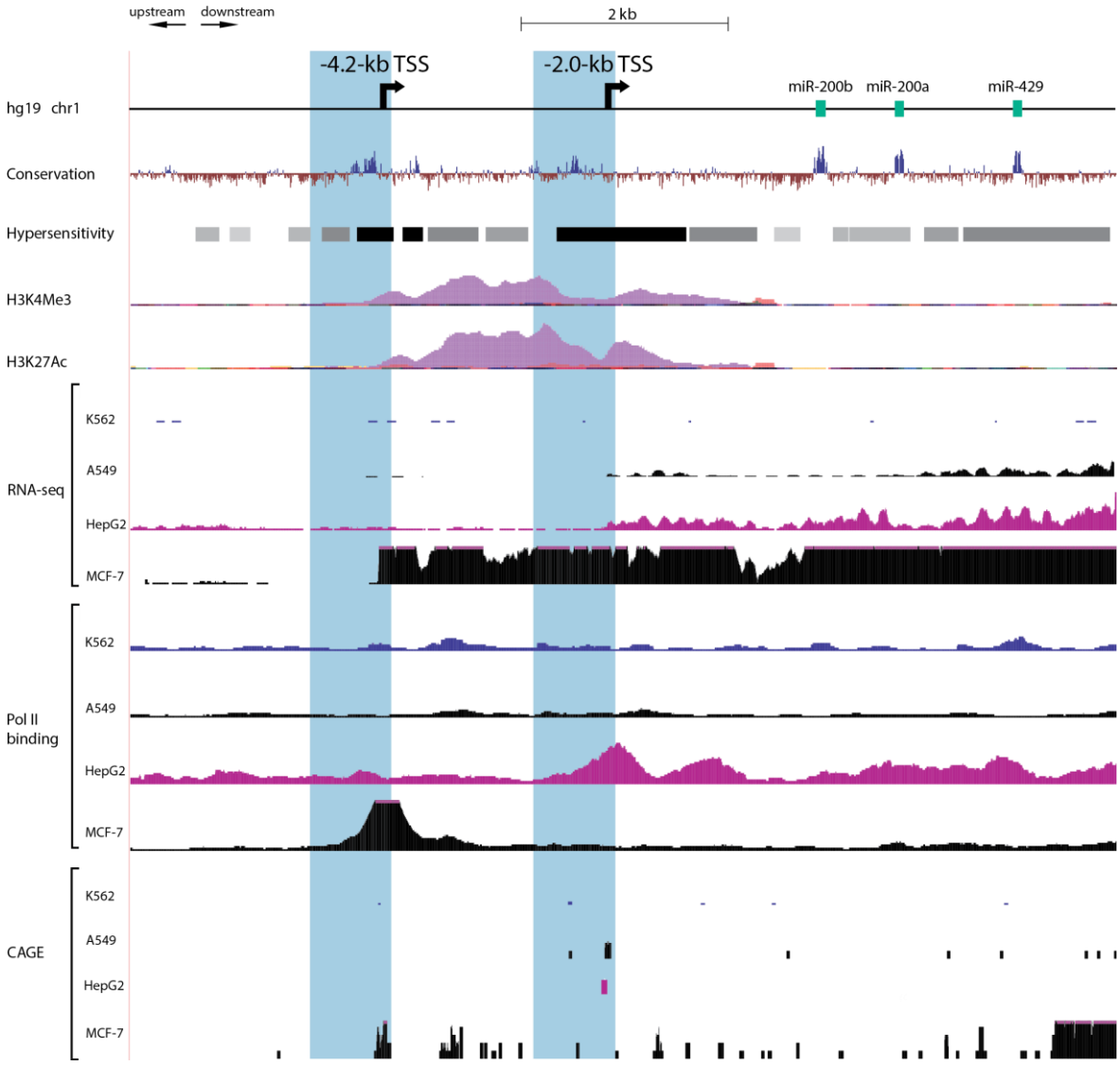

Figure 11: Alternative TSSs of miR-200b-200a-429 cluster. TSS and promoter prediction features of the 7-kb region upstream of the first annotated miRNA in miR-200b-200a-429 cluster (UCSC hg19 assembly, released Feb. 2009). RNA-seq reads and CAGE tags in four cell lines identify two differentially used TSSs, while conservation, hypersensitivity, histone marks H3K4Me3 and H3K27Ac, and Pol II binding support the existence of two distinct promoters of miR-200b-200a-429. Arrows represent identified TSSs 4.2 and 2.0 kb upstream of miR-200b. Blue shadings represent regions that contain corresponding promoters. Kb – kilo base pairs, TSS – transcription start site, chr1 – chromosome 1, RNA-seq – RNA sequencing, CAGE – cap analysis of gene expression.

Conservation is a measure of how well preserved a portion of the genome is through evolution. The more important the sequence, the more evolutionary conserved it is across species. MiRNA genes and their promoters are generally well conserved (2,12). We

observed high conservation rates (blue in conservation track, Fig. 11) in the regions 4.5 – 4.2 kb, 3.7 kb, and 2.3 kb upstream of miR-200b, in addition to regions where miRNA genes are encoded.

Hypersensitive sites denote regions of DNA where binding of proteins is feasible. This is due to less condensed chromatin structure; nucleosomes there are more sparsely packed (hypersensitive sites are also referred to as nucleosome depleted regions) and DNA is accessible to proteins. To sum up, hypersensitive sites are more prone to be influenced by transcription factors and enzymes and are therefore a characteristic of regulatory DNA sequences (36). We observed increased hypersensitivity (black - the darker the shade, the higher the score - in hypersensitivity track, Fig. 11) in the regions 4.4 – 3.7 kb, and 2.5 – 1.2 kb upstream of miR-200b.

Chromatin modifications are alterations of histone proteins that can influence transcription. They are generated by covalent binding of methyl, trimethyl, acetyl, etc., groups to aminoacids of histone proteins in nucleosomes. These changes can affect interactions with DNA and proteins. Histone marks, histone H3 lysine 4 trimethylation (H3K4Me3) and histone H3 lysine 27 acetylation (H3K27Ac), are thought to enhance transcription. H3K4Me3 is often found near promoters and H3K27Ac is often found near active regulatory elements (29,36). We observed the presence of H3K4Me3 and H3K27Ac histone marks in the region 4.2 – 1.0 kb upstream of miR-200b, increased signal initiating around 4.2, 3.7, and 2.0 kb upstream of miR-200b (H3K4Me3 and H3K27Ac tracks in Fig. 11).

Sequencing of full-length primary transcripts or RNA sequencing (RNA-seq) provides insight into cell's transcription profile, i.e., its transcriptome. It is a library of all RNA molecules present in a cell, sequenced and then aligned to the matching location in the genome. This way, RNA-seq data give information on what parts of the genome get transcribed from the DNA in the examined cell and at what amount, allowing for expression level assessment. We examined RNA-seq data in four cell lines. In K562 cells (blue in RNA-seq track, Fig. 11) full-length primary transcript of miR-200b-200a-429 cluster is not present, meaning that those miRNAs are not expressed in K562 cell line. In A549 cell line (upper black in RNA-seq track, Fig. 11) full-length primary transcript of miR-200b-200a-429 cluster in weakly expressed, initiating at 2.0 kb upstream of miR-200b. In HepG2 cell line (pink in RNA-seq track, Fig. 11) full-length primary transcript of miR-200b-200a-429 cluster is moderately expressed, initiating at 2.0 kb upstream of miR-200b. In

MCF-7 cell line (lower black in RNA-seq track, Fig. 11) full-length primary transcript of miR-200b-200a-429 cluster in highly expressed, initiating at 4.2 kb upstream of miR-200b.

RNA polymerase II is the enzyme most commonly catalyzing transcription of miRNAs (12). It binds to a promoter and initiates the transcription process, creating primary RNA transcript from the DNA. We examined Pol II binding data in four cell lines. In K562 (blue in Pol II binding track, Fig. 11) and A549 (upper black in Pol II binding track, Fig. 11) cells no significant peak of Pol II binding was observed. In HepG2 cells (pink in Pol II binding track, Fig. 11) Pol II binding peaks at 2.0 kb upstream of miR-200b. In MCF-7 cells (lower black in Pol II binding track, Fig. 11) Pol II binding peaks at 4.2 kb upstream of miR-200b.

CAGE – cap analiysis of gene expression tags – correspond to transcription initiation regions. They are obtained by sequestering short 5' ends of all primary transcripts in a cell, followed by sequencing and mapping to DNA (37). Similarly to RNA-seq data, CAGE data also provide information about the extent of expression but the signal is limited to TSSs and is ideally not present throughout the length of primary transcripts (38). We examined CAGE tags in four cell lines. There is no evident signal, which would not be considered noise, in K562 cells (blue in CAGE track, Fig. 11). In A459 (upper black in CAGE track, Fig. 11) and HepG2 (pink in CAGE track, Fig. 11) cells significant signal is present about 4.2 kb upstream of miR-200b. In MCF-7 cells (lower black in CAGE track, Fig. 11) CAGE tags are present through the length of primary transcript significantly peaking at 4.2 kb upstream of miR-200b (the peak is not displayed to its full size in Fig. 11), and less prominently at 3.3 kb upstream of miR-200b.

We found that miR-200b-200a-429 cluster is transcribed from two distinct promoters that are alternatively active in different cell types. We identified TSSs located 4.2 and 2.0 kb upstream of miR-200b. Corresponding promoters are expected to immediately precede TSSs and be a few hundred nts long. Additionally, we confirmed the assumption that this miRNA cluster is transcribed by Pol II and not Pol III.

TSSs were determined using the most trusted TSS predictive marks – CAGE tags and RNA-seq starting points. Other features (hypersensitivity, H3K4me3, H3K27ac, Pol II binding) were of a supportive nature, since at least one of them has commonly been found near TSS region – but none is necessarily present in all types of TSSs, whereas CAGE tags and RNA-seq starting points usually are (39). Nearly half of CAGE-identified TSSs and RNA-seq-identified TSSs genome-wide lie within 100 bp proximity indicating both methods detect the same TSSs (40). Therefore, in our study CAGE tags and RNA-seq starting points identify TSS 4.2 kb upstream of miR-200b (further designated -4.2-kb TSS)

in MCF-7 cell line, and TSS 2.0 kb upstream of miR-200b (further designated -2.0-kb TSS) in A549 and HepG2 cell lines.

Downsides of CAGE method are that it fails to detect rapidly degraded primary transcripts and of course promoters of those miRNAs that are not expressed in the analyzed cell line as seen in K562 cells, and sometimes yields signal where TSS is not present (2,39). Since CAGE method detects capped 5' ends of primary transcripts the signal should only be present at the TSS. The presence of CAGE tags throughout the primary transcript in MCF-7 cells, although decreased compared with -4.2 TSS, may be a consequence of primary transcript degradation and recapping of 5' ends of degradation products (39). To a lesser extent, the same phenomenon is visible also in A549 cells. Higher false CAGE signal in MCF-7 cell line (compared with A549 cell line) could arise due to higher expression level of primary transcript in MCF-7 cells, and thus higher concentration of degradation products. Advanced promoter prediction software can eliminate those false CAGE tags with an algorithm, but they are visible when viewing unfiltered data in the UCSC genome browser as seen in our case (39). Upon removal of CAGE signal from degradation products the remaining tags could indicate the existence of novel TSSs (38,39). It is unlikely that MCF-7 and A549 cell lines would have additional TSSs besides -4.2-kb and -2.0-kb TSS, since RNA-seq data initiate solely at those exact locations in cells where miR-200b-200a-429 are expressed. RNA-seq reads are consistent with CAGE tags indicating -4.2-kb TSS in MCF-7 cell line and -2.0-kb TSS in A549 and HepG2 cell lines. Additionally, supportive promoter indicative features are present around -4.2-kb and -2.0-kb TSS.

Pol II binding peaks at predicted TSSs in MCF-7 and HepG2 cells. Low expression of primary transcript in A549 cells (as deduced from RNA-seq data) might explain the missing Pol II binding signal at -2.0-kb TSS because the signal is dependent on the amount of bound Pol II, which is proportionate to gene's expression rate (41). Pol II binding signals at the predicted TSSs show that analyzed miRNA cluster is transcribed by Pol II and not Pol III. CAGE tags are confirmatory of this finding on the grounds that promoters of miRNAs generated by Pol III cannot be identified by means of CAGE tags, since this method is based on detecting 5' caps, which only occur on primary transcripts generated by Pol II (2,37).

Higher conservation rates prior to predicted TSSs fall within corresponding promoter regions. Previously annotated miRNA TSSs were found in close proximity (fewer than 500 bp) of increased hypersensitivity, the same appearing also around our predicted TSSs.

Histone marks H3K4Me3 and H3K27Ac were shown to arise immediately downstream from the TSS, which is consistent with our findings for predicted TSSs (36).

We cloned the blue shaded regions (Fig. 11) of 800 nts into a promoter reporter system for further experiments; 700 nts prior to each predicted TSS and 100 nts overlapping 5' end of each primary transcript, to assure that it encapsulated the promoter region and therefore transcription factor binding sites.

4.2 Differential promoter activity in several cell lines and connection with E-cadherin and ZEB2 expression

Since miR-200b-200a-429 cluster is a part of double negative feedback loop that controls epithelial to mesenchymal transition, alternative promoters could have a role in this process. To see whether there is a connection between the two predicted promoters and expression of related elements we compared their primary RNA transcript expression.

We analyzed RNA-sequencing data from ENCODE/Cold Spring Harbor Laboratory for 11 available cell lines for entities involved in the studied regulatory circuit: E-cadherin, ZEB2 protein, and both promoters of miR-200b-200a-429 cluster. Results are shown in Table I.

Table I: Primary transcript expression levels generated from -4.2-kb promoter and -2.0-kb promoter, respectively, expression of E-cadherin primary transcript, and expression of ZEB2 primary transcript.

Cell line	-4.2-kb promoter	-2.0-kb promoter	E-cadherin	ZEB2
GM12878	-	-	-	++
H1-hESC	++	+++	++	-
K562	-	-	-	+++
A549	-	+	+	+
HeLa-S3	-	-	-	++
HepG2	-	++	+	-
HUVEC	-	-	-	+++
IMR90	-	-	-	++
MCF-7	+++	+++	+++	-
SK-N-SH	-	-	-	++
NHEK	+	+++	++	-

Minus designates no expression, a plus low expression, double plus moderate expression, and triple plus high expression of primary transcript.

When -4.2-kb promoter is active E-cadherin is present and ZEB2 is not. This is seen in H1-hESC, MCF-7, and NHEK cell lines. When -2.0-kb promoter is active E-cadherin is present, though expressed in a lower degree, while ZEB2 is either weakly expressed (as in A549 cell line) or not at all (as in HepG2 cell line). When none of the promoters are active only ZEB2 is expressed (as seen in GM12878, K562, HeLa-S3, HUVEC, IMR90, SK-N-SH cell lines).

The fact that primary miR-200b-200a-429 transcript is expressed from first, second or none of the promoters in different cell lines indicates that alternative promoters are tissue specific.

We distinguished three expression profiles of miR-200b-200a-429 promoters, ZEB2, and E-cadherin in the analyzed cell lines, all are modeled in Fig. 12. In A, ZEB2 completely represses miR-200b-200a-429 cluster and E-cadherin, thus promoting mesenchymal cell state. In B, miR-200b-200a-429 cluster is highly expressed and ZEB2 is not, allowing for high expression of cell-cell adhesion protein E-cadherin. This expression profile promotes epithelial cell state (18). In C, none of the three elements, miR-200b-200a-429, ZEB2, or E-cadherin, is fully expressed or repressed. We dubbed it an interjacent expression profile. It is unclear which cell state it promotes. Here, miR-200b-200a-429 cluster is expressed from -2.0-kb promoter and not from -4.2-kb promoter.

One possible scenario for the interjacent expression profile would be that also E-cadherin has the ability to repress miR-200b-200a-429 by acting mainly on -4.2-kb promoter, while -2.0-kb promoter stays unaffected or at least not thoroughly deactivated – having the function of a secret switch-on leaking miR-200b-200a-429 expression. Poorly expressed miR-200b-200a-429 would not be able to entirely eliminate ZEB2 protein, which would in turn downregulate E-cadherin expression. When all events take action, stable state would result in interjacent expression profile as in A549 and HepG2 cells.

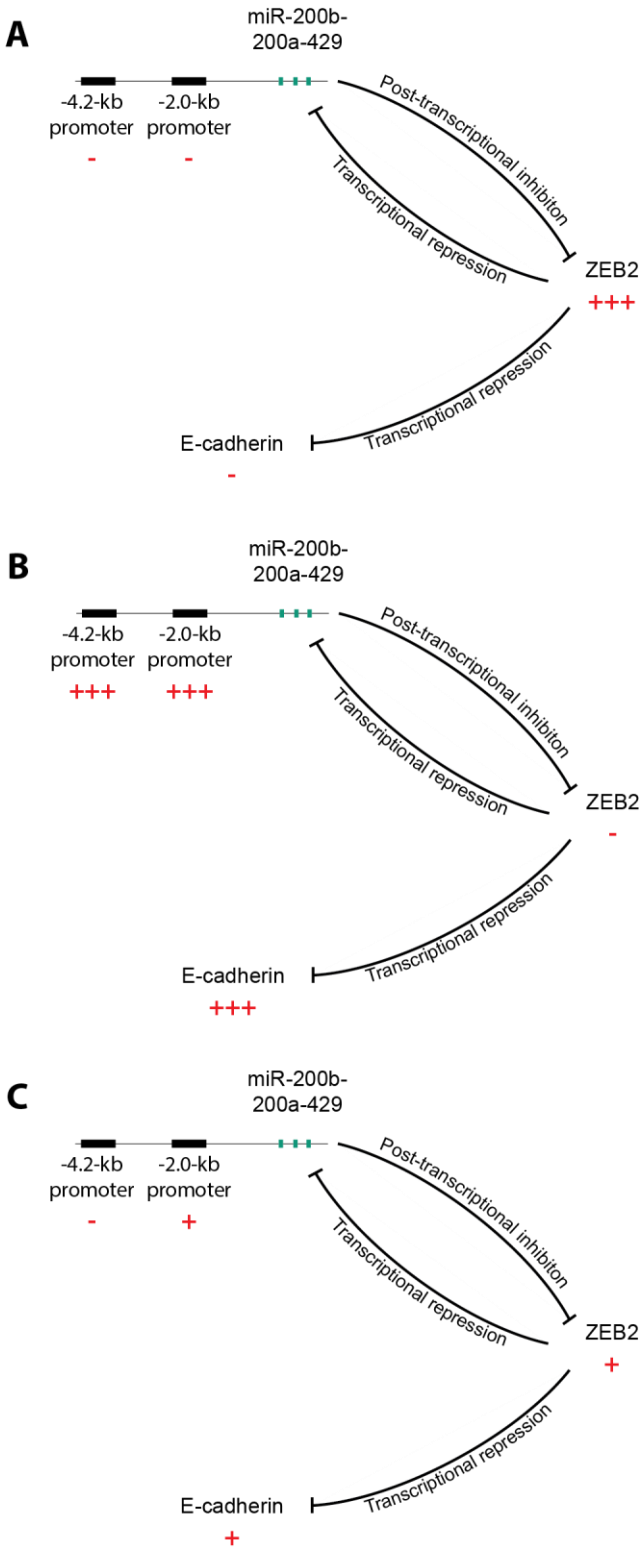

Figure 12: Three expression profiles of miR-200b-200a-429 promoters, ZEB2, and E-cadherin. **(A)** Expression profile that promotes mesenchymal state. **(B)** Expression profile that promotes epithelial state. **(C)** Interjacent expression profile.

Based on RNA-seq data analysis of predicted promoter regions of miR-200b-200a-429 in several cell lines, three were chosen for further experiments: MCF-7 cells, where both promoters are active, and A549 and HepG2 cells, where only -2.0-kb promoter is active.

4.3 Evaluation of predicted promoters in a luciferase assay

Predicted miR-200b-200a-429 promoters were experimentally evaluated by luciferase reporter assay. The experiment included assessing E-cadherin gene promoter strength as well.

Results are shown as an average of three independent biological repeats, where all samples were done in triplicates (Fig. 13). Due to inefficient transfection of HepG2 cells, we could not test promoter activity in that cell line. First indicatication of poor transfection was low *Renilla* luciferase signal which is expressed from control vector. To see if transfection would be more efficient with another transfection reagent, we transfected HepG2 cells with fluorescently labeled plasmid using three different transfection reagents and assessed transfection efficiency by fluorescence microscopy. None of the transfection reagents provided sufficient transfection efficiency, thus tests on HepG2 cells were abandoned.

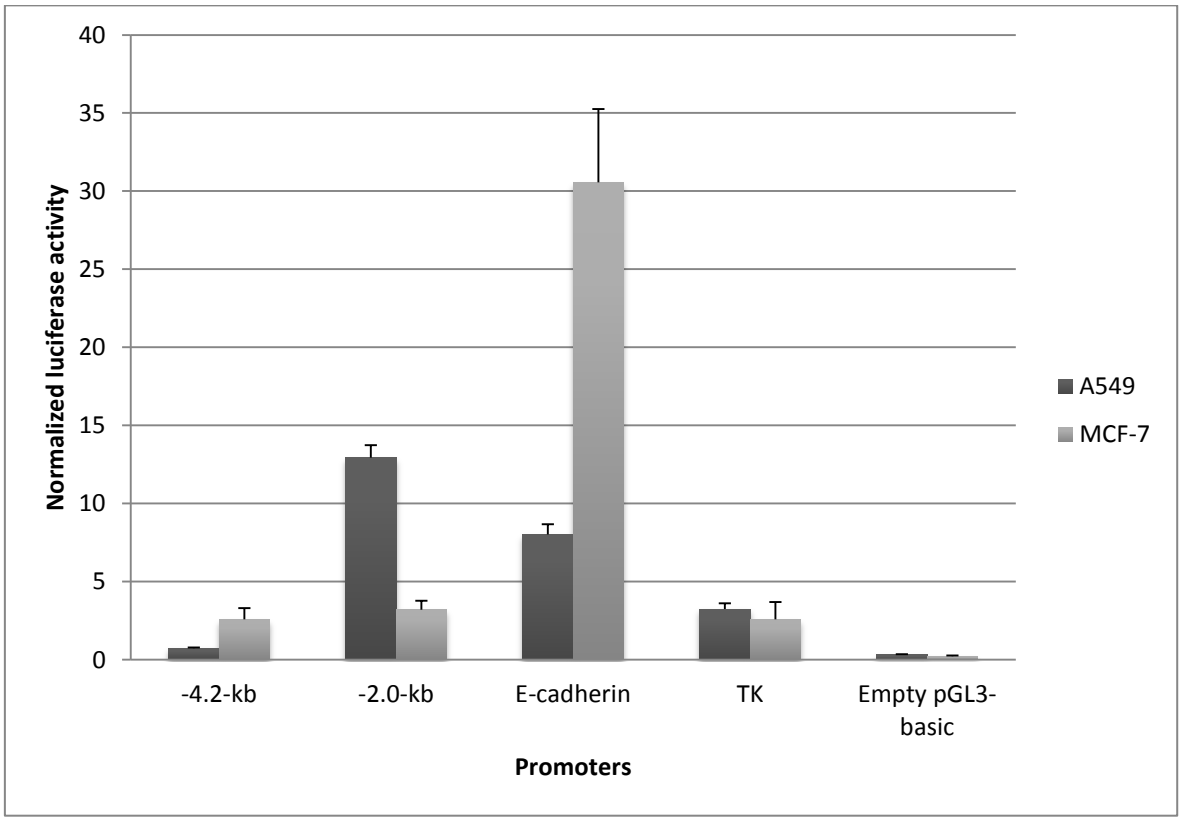

Figure 13: Captions on p. 47.

Figure 13: Promoter reporter assay results in A549 and MCF-7 cell lines. Strength of promoters (the two predicted miR-200b-200a-429 promoters and the E-cadherin promoter) was assessed by insertion into plasmid pGL3-basic, harboring firefly luciferase expression. Firefly luciferase signal was normalized to *Renilla* luciferase activity, expressed from co-transfected control plasmid. Thymidine kinase (TK) promoter with moderate promoter activity is a positive control, whereas pGL3-basic (no inserted promoter sequence) is a negative control of the system.

Only one of miR-200b-200a-429 promoters, -2.0-kb promoter, is active in A549 cells, while in MCF-7 cells both are active. This is consistent with predicted TSS for each cell line (see Fig. 11). E-cadherin promoter is moderately to highly active in A549 cells and highly active in MCF-7 cells compared with thymidine kinase promoter, which is typically of moderate activity. Compared with RNA-seq data analysis, higher activities of both miR-200b-200a-429 promoters were expected in MCF-7 cell line and lower activity of -2.0-kb promoter was expected in A549 cell line. High E-cadherin promoter activity in MCF-7 cells is consistent with RNA-seq data analysis, while its activity in A549 cells was expected to be lower. The discrepancy between the activities of promoter sequences tested in this assay with the extent of their expression by RNA-seq data analysis might be due to a myriad of transcription factors that can be cell-type and promoter sequence specific, and thus govern the activity of promoters differently for each promoter and differently in A549 and MCF-7 cell lines. Not only do transcription factors affect the rate of transcription but some also bind to promoter sequence of plasmid expression vector in the cytoplasm and mediate its import into the nucleus – upon it the transcription of reporter gene can occur and the signal in luciferase assay can subsequently be generated (42).

With this luciferase reporter assay we confirmed that both bioinformatically predicted promoter regions of miR-200b-200a-429 cluster do possess promoter activity in an *in vitro* system and that they are differentially active in tested cell lines.

4.4 Knockdown of Drosha, ZEB2 and E-cadherin

To evaluate knockdown efficiency of Drosha, ZEB2 and E-cadherin (important for the next experiment where knockdown is coupled with luciferase assay), we quantified changes of their mRNA levels by RT-qPCR. To evaluate the effect of Drosha, ZEB2 and E-cadherin on related elements in the analyzed double negative feedback loop, we also quantified changes in expression of the latter. Results are shown as an average of three independent biological repeats. All samples in a biological repeat were done in triplicates. The results were first normalized to GAPDH, a housekeeping gene (i.e., a gene that is

expressed constitutively regardless of tissue type or specific conditions in a cell). This is called relative quantification of gene expression. Then we compared average fold change of gene expression of knockdown samples relative to their respective negative controls. Standard deviation and statistical significance were calculated. Standard deviation of three independent repeats was generally high, especially when detecting primary miRNA transcript, which is expressed in a very low amount compared with Drosha, E-cadherin or ZEB2 mRNAs. The concentration of primary miR-200b-200a-429 transcript was up to 1000 times lower than the concentration of Drosha, E-cadherin and ZEB2 mRNAs (quantified relatively to GAPDH). Horwitz ratio predicts higher relative standard deviation for lower analyte concentration: expected RSD [%] = $2^{(1 - 0.5 * \log C)}$. Acceptable maximum relative standard deviation is two times of expected relative standard deviation. Because our technique only provides information about relative gene expression and not absolute concentration of the analyte, we cannot determine maximum allowed relative standard deviation, but can only expect it to be higher when detecting primary miR-200b-200a-429 transcript than when detecting Drosha, ZEB2 or E-cadherin. Although there is an apparent trend in fold change of gene expression throughout our results, they should be interpreted with caution where standard deviation is relatively high. Underlying reasons for high standard deviation could be different amounts of RNA (800 – 1300 µg) used for RT-qPCR among biological repeats or different quality of isolated RNA among samples. Normalization to a housekeeping gene should even out these differences to some extent, but different amount and quality of RNA used could still contribute to high standard deviation. Statistical significance was determined by one-tailed Student's t-test. P values of 0.01 – 0.05 were considered significant (designated with an asterisk, *). P values of 0.001 – 0.01 were considered very significant (designated with double asterisk, **).

4.4.1 Drosha knockdown increases levels of primary microRNA transcripts

Results of Drosha knockdown and its effect on primary miR-200b-200a-429 transcript level in A549 and MCF-7 cells are shown in Fig. 14.

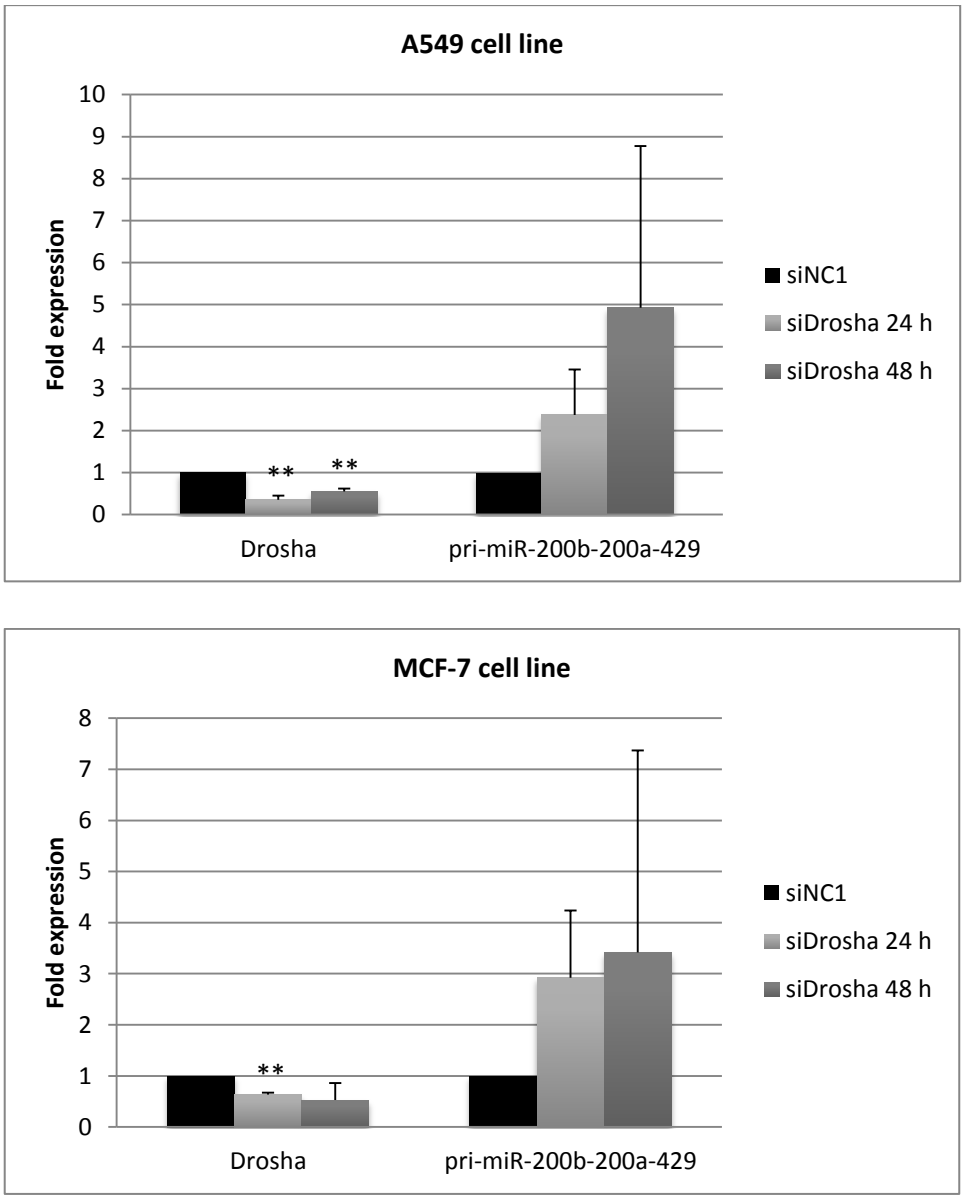

Figure 14: Changes in fold expression of Drosha mRNA and primary miR-200b-200a-429 transcript upon knockdown of Drosha in A549 cell line (upper graph) and MCF-7 cell line (lower graph). Black bars depict negative controls, light gray bars show expression 24 hours post transfection and dark gray bars show expression 48 hours post transfection. siNC1 – negative control siRNA, siDrosha – siRNA against Drosha.

In A549 cells we achieved 65% Drosha knockdown efficiency 24 hours post transfection ($P<0.01$), and 45% Drosha knockdown efficiency 48 hours post transfection ($P<0.01$). Primary miR-200b-200a-429 transcript increased on average 2.4 fold 24 hours post transfection, and 4.9 fold 48 hours post transfection.

In MCF-7 cells we achieved 36% Drosha knockdown efficiency 24 hours post transfection ($P<0.01$), and 48% Drosha knockdown efficiency 48 hours post transfection. Primary miR-

200b-200a-429 transcript increased on average 2.9 fold 24 hours post transfection, and 3.4 fold 48 hours post transfection.

We confirmed that pri-miR-200b-200a-429 is processed by Drosha. Blockage of miRNA maturation by Drosha knockdown should result in cell's inability to process primary miRNA transcripts, and thus their accumulation in the cell. Increased primary miRNA transcript levels upon Drosha knockdown have been reported in the past and are confirmative of the assumption that tested miRNA cluster is a substrate for the Microprocessor complex (43).

Drosha cleavage deteriorates the detection of full-length primary transcripts in sequencing assays, because it lowers analyte concentration (2). As shown here, Drosha knockdown increases the amount of primary miRNA transcripts, thus their concentration is more likely to be above the limit of detection of the sequencing method upon Drosha knockdown. Therefore, Drosha knockdown could be used prior to RNA sequencing when assessing expression of miRNAs (or determining their TSSs) in a cell.

We achieved sufficient siRNA-mediated Drosha knockdown in A549 and MCF-7 cell lines to proceed to the next experiment.

4.4.2 ZEB2 represses miR-200 family and E-cadherin

Results of ZEB2 knockdown and its effect on primary miR-200b-200a-429 transcript and E-cadherin mRNA levels are shown in Fig. 15. ZEB2 knockdown was performed only in A549 cells due to lack of its expression in MCF-7 cells.

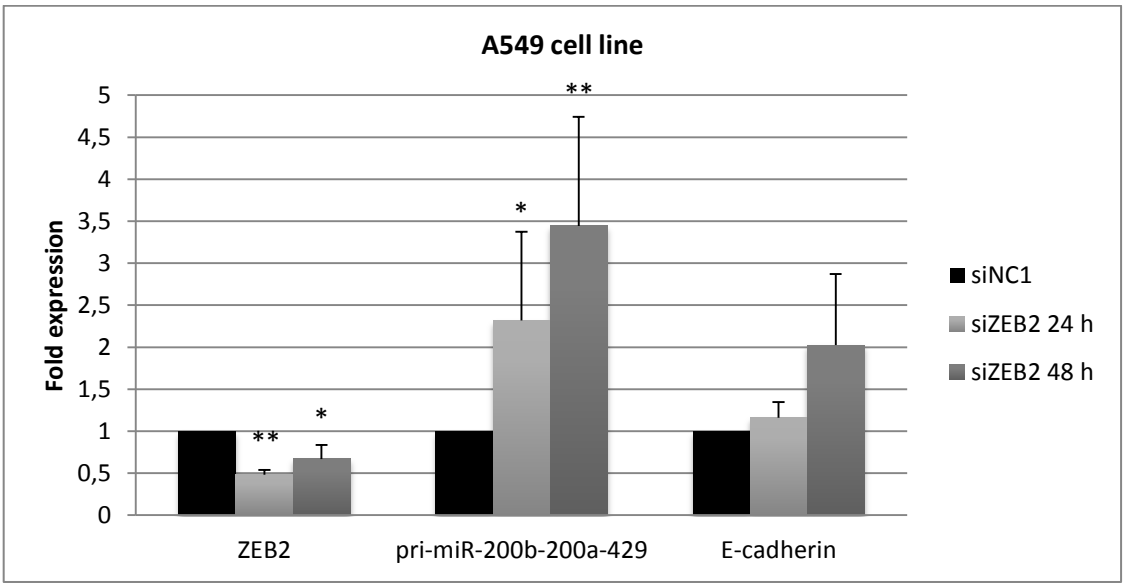

Figure 15: Captions on p. 51.

Figure 15: Changes in fold expression of ZEB2 mRNA, primary miR-200b-200a-429 transcript and E-cadherin mRNA upon knockdown of ZEB2 in A549 cell line. Black bars show negative controls, light gray bars show expression 24 hours post transfection and dark gray bars show expression 48 hours post transfection. siNC1 – negative control siRNA, siZEB2 – siRNA against ZEB2.

We achieved 52% ZEB2 knockdown efficiency 24 hours post transfection ($P<0.01$), and 33% ZEB2 knockdown efficiency 48 hours post transfection ($P<0.05$). Primary miR-200b-200a-429 transcript increased on average 2.3 fold 24 hours post transfection ($P<0.05$), and 3.4 fold 48 hours post transfection ($P<0.01$). E-cadherin mRNA expression increased on average 1.2 fold 24 hours post transfection, and 2.0 fold 24 hours post transfection.

Increased levels of both miR-200b-200a-429 and E-cadherin expression upon ZEB2 knockdown were expected, since ZEB2 is their transcriptional repressor (18). But why is the increase of pri-miR-200b-200a-429 and E-cadherin 48 hours post transfection more prominent than 24 hours post transfection, even though knockdown of ZEB2 is less efficient 48 hours post transfection? The reason is that with RT-qPCR we detected ZEB2 mRNA levels and not the concentration of the actual executer of transcriptional repression; the ZEB2 protein. While mRNA levels of ZEB2 were already rising 48 hours post transfection, due to transient effect of siRNAs, the protein might not have been synthesized yet, and thus could not suppress miR-200b-200a-429 and E-cadherin. Or, the opposite phenomenon could have occured 24 hours post transfection; there could still be some leftover ZEB2 protein, when mRNA levels were already decreased. Another analysis, Western blotting, would offer more confidence about the efficiency of ZEB2 protein knockdown, but it would not enable simultaneous measurement of miR-200b-200a-429 primary transcript (that cannot be detected by Western blotting) and ZEB2 protein level. Since we wanted to observe the dependency of those elements, RT-qPCR was the eligible method.

We achieved sufficient siRNA-mediated ZEB2 knockdown in A549 cell line to proceed to the next experiment.

4.4.3 E-cadherin may act as transcription factor that represses miR-200 family

Results of E-cadherin knockdown and its effect on primary miR-200b-200a-429 transcript level in A549 and MCF-7 cells are shown in Fig. 16.

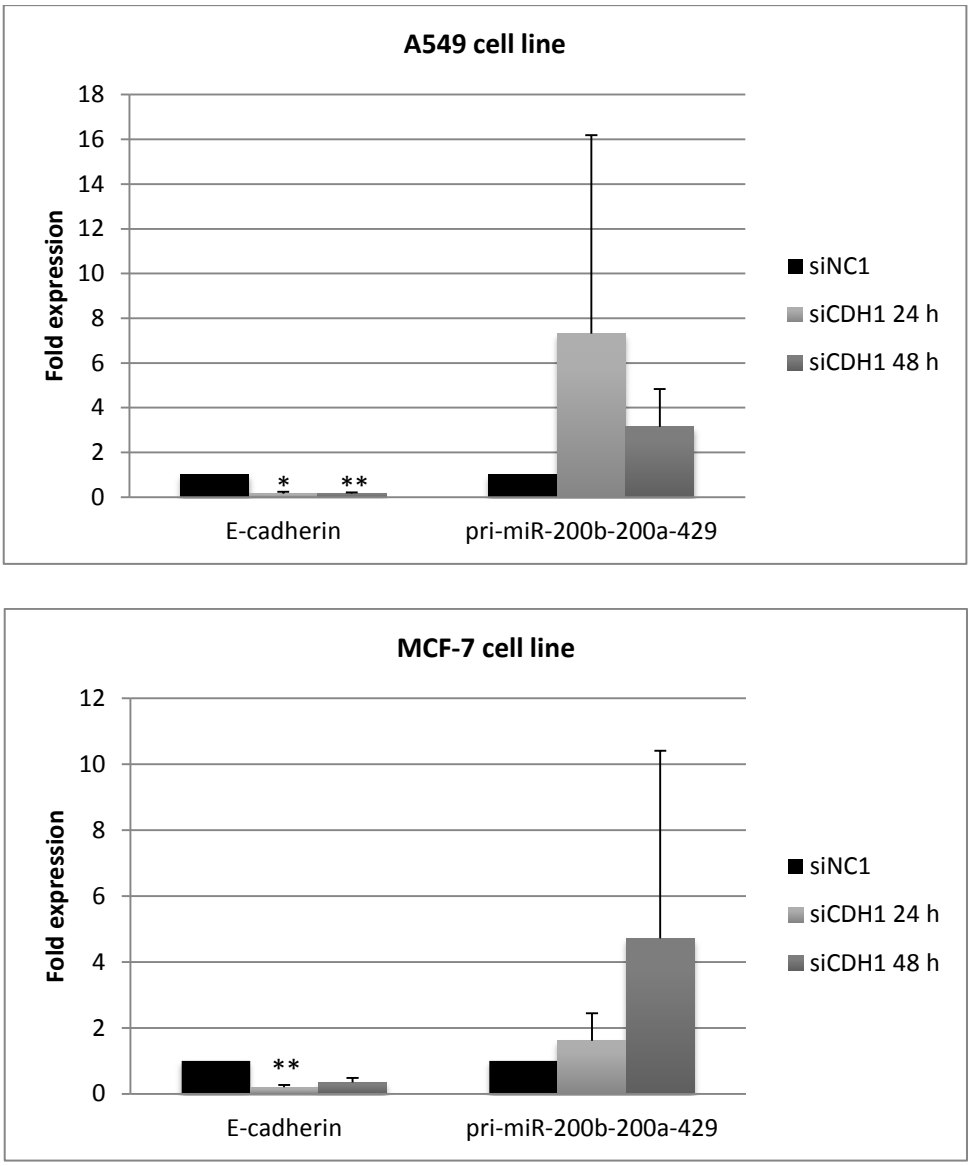

Figure 16: Changes in fold expression of E-cadherin mRNA and primary miR-200b-200a-429 transcript upon knockdown of E-cadherin in A549 cell line (upper graph) and MCF-7 cell line (lower graph). Black bars show negative controls, light gray bars show expression 24 hours post transfection and dark gray bars show expression 48 hours post transfection. siNC1 – negative control siRNA, siCDH – siRNA against E-cadherin.

In A549 cells we achieved 85% E-cadherin knockdown efficiency 24 hours post transfection ($P<0.05$), and 83% E-cadherin knockdown efficiency 48 hours post transfection ($P<0.01$). Primary miR-200b-200a-429 transcript increased on average 7.3 fold 24 hours post transfection, and 3.2 fold 48 hours post transfection.

In MCF-7 cells we achieved 80% E-cadherin knockdown efficiency 24 hours post transfection ($P<0.01$), and 65% E-cadherin knockdown efficiency 48 hours post

transfection. Primary miR-200b-200a-429 transcript increased on average 1.6 fold 24 hours post transfection, and 4.7 fold 48 hours post transfection.

In addition to previously discussed reasons for high standard deviation in all knockdown experiments, here, standard deviation of pri-miR-200b-200a-429 concentration is mainly a consequence of oddly high concentration of pri-miR-200b-200a-429 in third biological repeat – they were 10 times higher than in the first two biological repeats. This difference only occurred upon E-cadherin knockdown but not with the negative control. Due to high standard deviation apparent in pri-miR-200b-200a-429 quantification upon E-cadherin knockdown, more repeats of this experiment are required to confidently state that E-cadherin represses miR-200b-200a-429 expression. Nevertheless, the possibility of such regulation exists, since the interjacent expression profile could be explained by repressive effect of E-cadherin mainly on -4.2-kb promoter (Fig. 17). Supportive of this hypothesis are also studies where E-cadherin was found to have transcription factor activity; E-cadherin was proven to modulate transcription that is dependent on Wnt signaling pathway in colorectal cancer cells, and to increase promoter activity of activator protein 1 and cyclin D1 (19,44). Furthermore, nuclear localization (a prerequisite for a protein to function as a transcription factor) of E-cadherin, which is normally present on cell surface, has been noted in several tumor types (19).

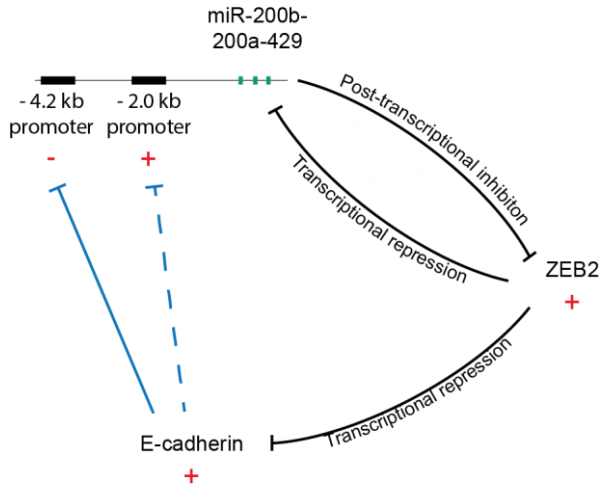

Figure 17: Possible role of E-cadherin as transcriptional repressor of -4.2-kb or both miR-200b-200a-429 promoters shown in interjacent expression profile, based on RNA-seq data (described before). Knockdown of E-cadherin increases the amount of pri-miR-200b-200a-429 transcript. Due to poor data quality, the assumption that E-cadherin acts as a transcriptional repressor merits limited trust (shown with blue lines). Based on RNA-seq

data E-cadherin could act as a repressor mainly on -4.2-kb promoter (full blue line), and possibly also on -2.0-kb promoter (dashed blue line).

We achieved sufficient siRNA-mediated E-cadherin knockdown in A549 and MCF-7 cell lines to proceed to the next experiment. With it, we will test whether the absence of E-cadherin truly increases promoter activity of -4.2-kb or -2.0-kb miR-200b-200a-429 promoters.

4.5 Activity of promoter regions when ZEB2, E-cadherin or Drosha are absent

To test whether ZEB2 or E-cadherin affect the activity of alternative miR-200b-200a-429 promoters differentially, we coupled knockdown of either ZEB2 or E-cadherin with a luciferase assay. ZEB2 knockdown was performed only in A549 cells due to the lack of its expression in MCF-7 cells. We included knockdown of Drosha, a representative of post-transcriptional miRNA regulation as opposed to transcriptional regulation by ZEB2 and E-cadherin.

This experiment included some foreseeable outcomes. Upon Drosha knockdown, activities of alternative promoters were not expected to change drastically, although blockage of miRNA maturation fails to decrease ZEB2 abundance and in turn represses E-cadherin expression – events that could affect the promoters. Activity of at least one miR-200b-200a-429 promoter should however increase in the absence of a known (ZEB2) and possible (E-cadherin) transcriptional repressor. Interjacent expression profile of the analyzed regulatory system in A549 cells indicates that E-cadherin probably acts mainly on -4.2-kb promoter, thus we expected an increase of its activity upon E-cadherin knockdown. We also predicted a rise in E-cadherin promoter activity upon ZEB2 knockdown, since ZEB2 is a well described transcriptional repressor of E-cadherin.

The question of how and which promoter is influenced by ZEB2 or E-cadherin remains unanswered due to inadequate responses of the controls in the experiment. Awry signals of controls were varying widely among repeats of the experiment. Experiment was repeated 9 times in A549 or MCF-7 cell lines, yet none of the repeats provided useful results. A representative experiment is shown in Fig. 18. Sometimes empty control vector pGL3-basic (without inserted promoter sequence) yielded a signal higher than acceptable (threshold set to 0.6). Other times thymidine kinase control vector with inserted sequence of moderate promoter activity had too varying signal among used siRNAs. When these random discrepancies in signal of controls were observed, we cleaned the detection

surface of the fluorescence spectrofluorometer, another member of our research group performed the assay to reduce the chance of human error, and we included another siRNA negative control (i.e., siRNA that does not produce a knockdown of any element included in the experiment). We achieved acceptable negative control (empty pGL3-basic) and positive control (thymidine kinase promoter inserted into pGL3-basic) scores in one repeat of the experiment, but the two siRNA negative controls did not yield comparable signals. We have observed that when siRNA against ZEB2 (siZEB2) was added more cells became detached but did not know how to prevent it (it could be an indication of siZEB2 cytotoxicity). Taken together, we could not generate valid results from this assay and could not identify reasons for its failure.

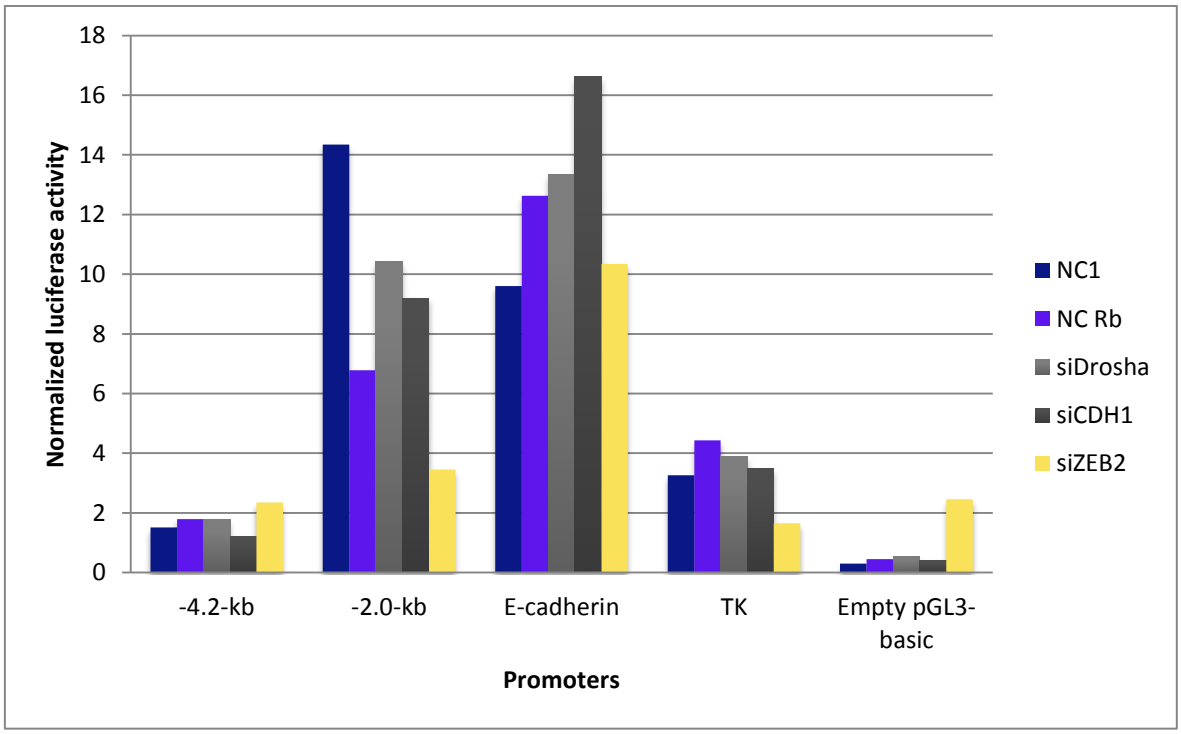

Figure 18: A representative repeat of luciferase experiment coupled with knockdown of Drosha, E-cadherin or ZEB2 in A549 cells. Here, all anomalies of controls we have encountered are present. Empty pGL3-basic should not have yielded signal higher than 0.6 no matter which siRNA cells were cotransfected with; here, cells cotransfected with siRNA against ZEB2 (siZEB2, yellow bars) and empty pGL3-basic produced signal higher than 2. Expected signal for promoter sequence of thymidine kinase is approximately 3 independent of cotransfected siRNA, and thus not as variable among samples with different siRNAs as here. The two negative controls should have yielded comparable signals. Here, the most evident deviation of negative control signals (blue bars) is seen when cells were cotransfected with -2.0-kb promoter.

Alternative method for this experiment could be DNA-protein interaction analysis by means of chromatin immunoprecipitation of ZEB2 and E-cadherin followed by sequencing (ChIP-seq). This method would give information about where do ZEB2 and E-cadherin bind to the genome; relevant here being -4.2-kb vs. -2.0-kb promoter regions. Chip-seq cannot indicate how the promoters are affected though (whether their activity increases or decreases upon binding of a protein). Thus, Chip-seq is not as informative replacement for predicting ZEB2 and E-cadherin effect on alternative promoters as knockdown coupled with a promoter reporter assay that we have suggested in this study.

V. Conclusion

We identified and experimentally confirmed the existence of two alternative promoters that govern the transcription of miR-200b-200a-429 cluster in a cell type-specific manner, located 4.2 and 2.0 kb upstream of the first annotated miRNA in the cluster. We observed three expression profiles of elements involved in a regulatory feedback: the two promoters of miR-200b-200a-429 cluster, ZEB2 and E-cadherin. Based on one of the expression profiles and increased amounts of pri-miRNA upon E-cadherin knockdown, we speculate that E-cadherin could act as a selective transcriptional repressor of primarily -4.2-kb promoter. To the best of our knowledge, no other study addressed the incorporation of alternative miRNA promoters in feedback circuits so far.

Alternative miRNA promoters are probably a genome-wide phenomenon, but their overall detection is currently beyond the scope of existent primary RNA sequencing methods; full-length primary miRNA transcript detection is dampened due to quick Drosha processing and thereby transient nature of those transcripts. Better sequencing depth and progress in bioinformatical TSS prediction algorithms should uncover more candidates for alternative miRNA promoters, but their looming existence is always to be verified experimentally. Alternative miRNA promoters could be selectively targeted in feedback circuits, allowing for yet another level of transcriptional regulation that could evade force feedback. The essence of alternative miRNA promoters needs to be further elucidated by more functional assays, such as the one proposed in this study.

It is assumed that aberrant miRNA expression in disease could eventually be restored to normal by gene therapeutics. Prior to any attempts to design pharmaceuticals that would manipulate the activity of miRNA promoters the exact principles of their regulation should be well described, including regulation by alternative promoters.

VI. References

1. Bratkovič T, Glavan G, Štrukelj B, Živin M, Rogelj B: **Exploiting microRNAs for cell engineering and therapy.** *Biotechnol Adv* 2012, 30 (3): 753-65.
2. Marsico A, Huska MR, Lasserre J, Hu H, Vucicevic D, Musahl A, Orom U, Vingron M: **PROmiRNA: a new miRNA promoter recognition method uncovers the complex regulation of intronic miRNAs.** *Genome Biol* 2013, 14 (8): R84.
3. Kim VN, Han J, Siomi MC: **Biogenesis of small RNAs in animals.** *Nat Rev Mol Cell Biol* 2009, 10: 126-139.
4. Winter J, Jung S, Keller S, Gregory RI, Diederichs S: **Many roads to maturity: microRNA biogenesis pathways and their regulation.** *Nat Cell Biol* 2009*,* 11 (3): 228-34.
5. Schanen BC, Li X: **Transcriptional regulation of mammalian miRNA genes.** *Genomics* 2011, 97: 1-6.
6. Kozomara A, Griffiths-Jones S: **miRBase: annotating high confidence microRNAs using deep sequencing data.** *Nucleic Acids Res* 2014, 42 (Database issue): D68-73.
7. Portnoy V, Huang V, Place **RF, Li LC: Small RNA and transcriptional upregulation.** *Wiley Interdiscip Rev RNA* 2011, 2 (5): 748-60.
8. Treiber T, Treiber N, Meister G: **Regulation of microRNA biogenesis and function.** *Thromb Haemost* 2012, 107 (4): 605-10.
9. Auyeung VC, Ulitsky I, McGeary SE, Bartel DP: **Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing.** *Cell* 2013. 152 (4): 844-58
10. Bartel DP: **MicroRNAs: Target Recognition and Regulatory Functions.** *Cell* 2009, 136 (2): 215-33.
11. Lai X, Vera J: **miRNA Clusters.** In: Dubitzky W. Wolkenhauer O. Cho K. Yokota H (Ed.) Encyclopedia of Systems Biology: SpringerReference (www.springerreference.com, accessed Jun. 2013). *Springer-Verlag Berlin Heidelberg* 2013.
12. Bartel DP: **MicroRNAs: genomics, biogenesis, mechanism, and function.** *Cell* 2004, 116 (2): 281-97.
13. Clancy S: **DNA transcription.** *Nature Education* 2008, 1 (1): 41.
14. Pennisi, Elizabeth: **ENCODE Project Writes Euology for Junk DNA.** *Science* 2012. 337: 1159-1161.

15. Han J, Lee Y, Yeom KH, Nam JW, Heo I, Rhee JK, Sohn SY, Cho Y, Zhang BT, Kim VN: **Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex.** *Cell* 2006, 125 (5): 887-901.
16. Ruby JG, Jan CH, Bartel DP: **Intronic microRNA precursors that bypass Drosha processing.** *Nature* 2007, 448 (7149): 83-6.
17. Hu S, Huang M, Li Z, Jia F, Ghosh Z, Lijkwan MA, Fasanaro P, Sun N, Wang X, Martelli F, Robbins RC, Wu JC: **MicroRNA-210 as a Novel Therapy for treatment of Ischemic Heart Disease.** *Circulation* 2010, 122 (11 Suppl): S124-31.
18. Bracken CP, Gregory PA, Kolesnikoff N, Bert AG, Wang J, Shannon MF, Goodall GJ: **A double-negative feedback loop between ZEB1-SIP1 and the microRNA-200 family regulates epithelial-mesenchymal transition.** *Cancer Res* 2008, 68 (19): 7846-54.
19. Rodriguez FJ, Lewis-Tuffin LJ, Anastasiadis PZ: **E-cadherin's dark side: possible role in tumor progression.** *Biochim Biophys Act.* 2012, 1826 (1): 23-31.
20. Christoffersen NR, Silahtaroglu A, Orom UA, Kauppinen S, Lund AH: **miR-200b mediates post-transcriptional repression of ZFHX1B.** *RNA* 2007, 13 (8): 1172-8.
21. Xu N, Papagiannakopoulos T, Pan G, Thomson JA, Kosik KS: **MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells.** *Cell* 2009, 137 (4): 647-58.
22. Deng Y, Wang CC, Choy KW, Du Q, Chen J, Wang Q, Li L, Chung TK, Tang T: **Therapeutic potentials of gene silencing by RNA interference: principles, challenges, and new strategies.** *Gene* 2014, 538 (2): 217-27.
23. Dykxhoorn DM, Novina CD, Sharp PA: **Killing the messenger: short RNAs that silence gene expression.** *Nat Rev Mol Cell Biol* 2003, 4 (6): 457-67.
24. Zhang Y, Wang Z, Gemeinhart RA: **Progress in microRNA delivery.** *J Control Release* 2013, 172 (3): 962-74.
25. Jurmeister S, Uhlmann S, Sahin ÌÐ: **MIR200C (microRNA 200c).** *Atlas Genet Cytogenet Oncol Haematol* 2012, 16 (2): 92-99.
26. **Promega Technical manual: pGL3 luciferase reporter vectors.** *Promega* 2008.
27. Harte RA, Diekhans M, Kent, WJ & Haussler D: **Guide to the UCSC Genome Browser.** *NPG Education* 2010.
28. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D: **The human genome browser at UCSC.** *Genome Res* 2002, 12 (6): 996-1006.

29. Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, Goldman M, Barber GP, Clawson H, Coelho A, Diekhans M, Dreszer TR, Giardine BM, Harte RA, Hillman-Jackson J, Hsu F, Kirkup V, Kuhn RM, Learned K, Li CH, Meyer LR, Pohl A, Raney BJ, Rosenbloom KR, Smith KE, Haussler D, Kent WJ: **The UCSC Genome Browser database: update 2011.** *Nucleic Acids Res* 2011, 39 (Database issue): D876-82.
30. Untergrasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG: **Primer3 - new capabilities and interfaces.** *Nucleic Acids Research* 2012, 40 (15): e115.
31. Koressaar T, Remm M: **Enhancements and modifications of primer design program Primer3.** *Bioinformatics* 2007, 23 (10): 1289-91.
32. Vincze T, Posfai J, Roberts RJ: **NEBcutter: a program to cleave DNA with restriction enzymes.** *Nucleic Acids Res* 2003, 31: 3688-3691.
33. Kent WJ: **BLAT - the BLAST-like alignment tool.** *Genome Res* 2002, 12 (4): 656-64.
34. Morlando M, Ballarino M, Gromak N, Pagano F, Bozzoni I, Proudfoot NJ: **Primary microRNA transcripts are processed co-transcriptionally.** *Nat Struct Mol Biol* 2008, 15 (9): 902-9.
35. **Dual-Glo® Luciferase Assay System.** (https://worldwide.promega.com/products/reporter-assays-and-transfection/reporter-assays/dual_glo-luciferase-assay-system/, accessed Jul. 2013)
36. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, Garg K, John S, Sandstrom R, Bates D, Boatman L, Canfield TK, Diegel M, Dunn D, Ebersol AK, Frum T, Giste E, Johnson AK, Johnson EM, Kutyavin T, Lajoie B, Lee BK, Lee K, London D, Lotakis D, Neph S, Neri F, Nguyen ED, Qu H, Reynolds AP, Roach V, Safi A, Sanchez ME, Sanyal A, Shafer A, Simon JM, Song L, Vong S, Weaver M, Yan Y, Zhang Z, Zhang Z, Lenhard B, Tewari M, Dorschner MO, Hansen RS, Navas PA, Stamatoyannopoulos G, Iyer VR, Lieb JD, Sunyaev SR, Akey JM, Sabo PJ, Kaul R, Furey TS, Dekker J, Crawford GE, Stamatoyannopoulos JA: **The accessible chromatin landscape of the human genome.** *Nature* 2012, 489 (7414): 75-82.
37. Shiraki T, Kondo S, Katayama S, Waki K, Kasukawa T, Kawaji H, Kodzius R, Watahiki A, Nakamura M, Arakawa T, Fukuda S, Sasaki D, Podhajska A, Harbers M, Kawai J, Carninci P, Hayashizaki Y: **Cap analysis gene expression for high-**

throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci U S A* 2003, 100 (26): 15776-81.

38. de Hoon M, Hayashizaki Y: **Deep cap analysis gene expression (CAGE): genome-wide identification of promoters, quantification of their expression, and network inference.** *Biotechniques* 2008, 44 (5): 627-8, 630, 632.
39. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakrabortty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Falconnet E, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena H, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Luo OJ, Park E, Persaud K, Preall JB, Ribeca P, Risk B, Robyr D, Sammeth M, Schaffer L, See LH, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Ruan X, Hayashizaki Y, Harrow J, Gerstein M, Hubbard T, Reymond A, Antonarakis SE, Hannon G, Giddings MC, Ruan Y, Wold B, Carninci P, Guigó R, Gingeras TR: **Landscape of transcription in human cells.** *Nature* 2012, 489 (7414): 101-8.
40. ENCODE Project Consortium: **An integrated encyclopedia of DNA elements in the human genome.** *Nature* 2012, 489 (7414): 57-74.
41. Cheng C, Alexander R, Min R, Leng J, Yip KY, Rozowsky J, Yan KK, Dong X, Djebali S, Ruan Y, Davis CA, Carninci P, Lassman T, Gingeras TR, Guigó R, Birney E, Weng Z, Snyder M, Gerstein M: **Understanding transcriptional regulation by integrative analysis of transcription factor binding data.** *Genome Res* 2012, 22 (9): 1658-67.
42. Lam AP, Dean DA: **Progress and prospects: nuclear import of nonviral vectors.** *Gene Ther* 2010, 17 (4): 439-47.
43. Ribas J, Ni X, Castanares M, Liu MM, Esopi D, Yegnasubramanian S, Rodriguez R, Mendell JT, Lupold SE: **A novel source for miR-21 expression through the alternative polyadenylation of VMP1 gene transcripts.** *Nucleic Acids Res* 2012, 40 (14): 6821-33.
44. Kuphal F, Behrens J: **E-cadherin modulates Wnt-dependent transcription in colorectal cancer cells but does not alter Wnt-independent gene expression in fibroblasts.** *Exp Cell Res* 2006, 312 (4): 457-67.
45. **AllPrep® DNA/RNA/Protein Mini Handbook.** *Qiagen* 2011.

46. **QIAquick Spin Handbook.** *Qiagen* 2008.
47. **QIAGEN® Plasmid Purification Handbook.** *Qiagen* 2012.
48. **TRIzol® Reagent Manual.** *Life technologies* 2012.
49. **Dual-Glo® Luciferase Assay System Technical Manual.** *Promega* 2011.

VII. Supplementary information

Supplement 1: Genomic DNA isolation.

Cells of one approximately 50% confluent 10 cm tissue culture dish were trypsinized and pelleted at 800 rpm for 3 minutes. To lyse the cells 350 µL of Buffer RTL were added and pipetted up and down to mix, then vortexed for 1 minute to homogenize the cells. The homogenized lysate was transferred to an AllPrep DNA spin column placed in a 2 mL collection tube and centrifuged for 30 seconds at 13,000 rpm. AllPrep DNA spin column was then placed into a new 2 mL collection tube. 500 µL of Buffer AW1 with ethanol were added and the tube was centrifuged for 15 seconds at 13,000 rpm to wash the spin column membrane. Flow-through was discarded and 500 µL of Buffer AW2 with ethanol were added and the tube was centrifuged 2 minutes at 13,000 rpm to wash the spin column membrane. The column was transferred into a new 1.5 mL collection tube, 100 µL of Buffer EB, preheated to 70 °C, were added. The column was incubated at room temperature for 2 minutes and then centrifuged for 1 minute at 13,000 rpm to elute the DNA. DNA yield was measured with UV-Vis Spectrophotometer NanoDrop 2000. Procedure was adapted from the protocol supplied with the kit (45).

Supplement 2: Gel extraction.

DNA in the gel was visualized using gel imager Bio-Rad Gel Doc 2000. Bands that contained amplified PCR product were excised with a scalpel using UV rays protective gear. The aim was not to exceed 400 mg weight limit per excised gel slice. 6 volumes of Buffer QG were added to 1 volume gel (100 mg ~ 100 µL), incubated at 50 °C for 10 min and occasionally vortexed to help dissolve gel. 1 volume of isopropanol was added to the sample and mixed. Sample was applied to a QIAquick spin column in a 2 mL collection tube and centrifuged for 1 minute at 13,000 rpm (the same rpm rate applies to all centrifugations in this protocol), flow- through was discarded. To wash the column, 0.75 mL buffer PE were added and the tube was centrifuged for 1 minute, flow-through was discarded. QIAquick column was centrifuged once more in a 2 mL collection tube for 1 minute to remove residual wash. QIAquick column was placed into a clean 1.5 mL tube. To elute DNA, 30 µL Buffer EB were added to the center of QIAquick membrane, the tube was subsequently centrifuged for 1 minute. Procedure was adapted from the protocol supplied with the kit (46).

Supplement 3: Purification of digested DNA fragment and plasmid.

5 volumes of Buffer PB were added to 1 volume of digestion reaction sample, mixed, applied to the QIAquick spin column placed in a 2 mL collection tube, and centrifuged for 30 - 60 seconds to bind DNA. All centrifugation steps were carried out at 13,000 rpm in a conventional tabletop microcentrifuge at room temperature. Flow-through was discarded, QIAquick column was centrifuged for additional 1 minute to remove residual ethanol from Buffer PE. QIAquick column was placed in a clean 1.5 mL microcentrifuge tube. To elute DNA, 30 µL of ddH_2O were added to the center of the QIAquick membrane, and the column was centrifuged for 1 minute. Procedure was adapted from the protocol supplied with the kit (46).

Supplement 4: Plasmid extraction.

Bacteria in overnight minicultures were harvested at 4,000 g for 15 minutes at 4°C. Supernatant was removed, bacterial pellet was resuspended in 0.3 mL of Buffer P1 with added RNase A and transferred in a microcentrifuge tube. 0.2 mL of Buffer P2 were added and mixed by vigorously inverting the tube 4-6 times and incubating at room temperature for 5 minutes. 0.3 mL of pre-chilled Buffer P3 were added and thoroughly mixed by vigorously inverting the tube 4-6 times, and incubated on ice for 5 minutes. The tube was centrifuged for 10 minutes at 13,000 rpm, supernatant containing plasmid DNA was then pippeted to the previously equilibrated QIAGEN-tip 20 column with 1 mL Buffer QBT, emptied by gravity flow. QIAGEN-tip 20 was washed with 2 x 2 mL Buffer QC, and eluted with 0.8 mL Buffer QF. Eluate was collected in a 1.5 mL microcentrifuge tube. DNA was precipitated by adding 0.7 volumes of room-temperature isopropanol to the eluate. Microcentrifuge tube was mixed and centrifuged immediately at 15,000 g for 30 minutes. Supernatant was carefully pippeted away and discarded. DNA pellet was washed with 1 mL of 70% ethanol and centrifuged at 15,000 g for 10 minutes. Supernatant was carefully pippeted away and discarded. DNA pellet was air-dried for 5-10 minutes, and redissolved in 30 µL of ddH_2O. DNA yield was measured with UV-Vis Spectrophotometer NanoDrop 2000. Procedure was adapted from the protocol supplied with the kit (47).

Supplement 5: RNA extraction.

Growth media was removed from culture dish. 1 mL TRIzol® Reagent was added directly to the cells in the 35 mm dish. Cells were lysed by pipetting up and down several times in the culture dish. Homogenized sample was transferred into a 1.5 mL microcentrifuge tube and incubated at room temperature to permit complete dissociation of the nucleoprotein complex for 5 minutes. 0.2 mL of chloroform were added, tube was capped securely and shaken vigorously for 15 seconds. Sample was incubated at room temperature for another

2-3 minutes, and then centrifuged at 12,000 g for 15 minutes at 4°C. The mixture separated into a lower red phenol-chloroform phase, an interphase, and a colorless upper aqueous phase, in which RNA remained. Aqueous phase was carefully pipetted into a new tube. Since we had small samples, 1 µL of RNase–free glycogen was added to the aqueous phase and briefly mixed. Glycogen was colored blue and provided better visibility of precipitated RNA in later steps. According to the manufacturer's instructions glycogen does not inhibit first-strand synthesis at concentrations ≤4 mg/mL, and does not inhibit PCR. 100% isopropanol was added to the aqueous phase, in quantities of at least 70% of the sample volume (sample being aqueous phase), to precipitate the RNA. Mixture was incubated for 10 minutes at room temperature and centrifuged at 12,000 g for 10 minutes at 4°C. RNA is invisible prior to centrifugation, but forms a gel-like pellet on the side and bottom of the tube after it. Supernatant was removed from the tube, leaving only the RNA pellet. The pellet was washed with 1 mL of 75% ethanol in RNase-free water, briefly vortexed, and centrifuged at 7,500 g for 5 minutes at 4°C. Supernatant was discarded, RNA pellet was air-dried for 5-10 minutes, and resuspended in 20-30 µL of RNase-free water (volume depended on the size of the pellet). Sample was incubated at 55-60°C for 10 minutes to achieve better RNA solubility and then kept on ice or stored at -18°C. RNA yield was measured with UV-Vis Spectrophotometer NanoDrop 2000. Procedure was adapted from the protocol supplied with the kit (48).

Supplement 6: Luciferase reporter assay.

24 hours post transfection medium was removed and replaced with 25 µL of PBS per well. 25 µL of Dual-Glo® Luciferase Substrate, diluted with supplied buffer, according to manufacturer's instructions, was added into each well and individually mixed by pipetting up and down several times. After horizontally shaking for 15 minutes at room temperature, firefly luciferase signal was measured. Addition of 25 µL of Dual-Glo® Stop & Glo® Substrate, diluted with supplied buffer, according to manufacturer's instructions, was added into each well and individually mixed by pipetting up and down several times. After incubating for 15 minutes at room temperature while horizontally shaking, firefly luciferase reaction was quenched and *Renilla* luciferase reaction was initiated. Measuring of *Renilla* luminescence followed. Procedure was adapted from the protocol supplied with the kit (49).