UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

PREDIÇÃO DE FUNÇÃO DE RNAs NÃO CODIFICANTES CARACTERÍSTICOS DO ESTADO TRONCO EMBRIONÁRIO HUMANO ATRAVÉS DE FERRAMENTAS DE BIOINFORMÁTICA

RAQUEL CALLONI

Tese submetida ao Programa de Pós-Graduação em Biologia Celular e Molecular da UFRGS como requisito parcial para a obtenção do grau de Doutora em Biologia Celular e Molecular.

Orientador: Prof. Dr. Diego Bonatto
Porto Alegre, junho de 2017

INSTITUIÇÕES E FONTES FINANCIADORAS

Agência financiadora

CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico

Instituição de origem:

Laboratório de Biologia Computacional e Molecular - LBCM

Centro de Biotecnologia da UFRGS

Universidade Federal do Rio Grande do Sul

Dedico este trabalho aqueles que abriram mão de si mesmos para que eu chegasse onde me encontro hoje, meu pai ***Adenir Antônio Calloni*** *e minha mãe* ***Carmem Bernardi Calloni****.*

AGRADECIMENTOS

Estava cá eu pensando em belas palavras para dispor aqui e eis que me ocorreu que, dentro em breve, esta folha será mais uma dentre outras milhares de folhas escondidas dentro de livros esquecidos sob uma película de poeira em um biblioteca qualquer.

Seria mesmo justo recompensar a quem pertence a minha gratidão com algumas palavras pintadas em uma folha esquecida? Penso que nascemos para marcar a vida de outras pessoas e isso faz-se frente a frente e não empilhando verbetes em uma folha.

Então, seguirei na contramão da tradição de dizer ao papel o que deveria dizer-se aos olhos e aos ouvidos da pessoa de carne e osso. Escolho encaixar-me dentro do abraço de cada ser que marcou meu caminho durante o desenrolar deste doutorado e dizer-lhe sobre a minha gratidão por tê-lo ao meu lado. Escolho a efemeridade de palavras que são ditas, posto que serão carregadas na memória de quem as ouve, à eternidade de palavras impressas esquecidas por debaixo de uma capa de livro.

Não cito, pois, aqui, os donos da minha gratidão porque dispensarei tempo meu para dizer-lhes pessoalmente o quão cheia de ventura sou por tê-los ao meu lado. Ademais, não necessito erigir listas dos destinatários do meu obrigado, posto que eles mesmos sabem exatamente quem são.

ESTRUTURAÇÃO DA TESE

Esta tese apresenta-se organizada em uma introdução geral, objetivos (gerais e específicos), três capítulos redigidos no formato de artigo, discussões, conclusões gerais e um item contendo as perspectivas do trabalho. Além disso, a tese apresenta, como anexo, um artigo de revisão redigido atrelado ao projeto inicial de doutorado, o qual foi posteriormente alterado.

A introdução geral apresenta os principais conceitos associados à RNAs não codificantes (ncRNAs), com foco em miRNAs, RNAs longos não codificantes, pseudogenes e RNAs circulares. Além disso, uma sessão é dedicada ao fenômeno de competição entre RNAs pela ligação a miRNAs, que é foco de dois dos artigos desta tese, e outra contém uma breve revisão bibliográfica sobre células-tronco embrionárias (CTEs).

O capítulo 1 aborda um estudo de predição de função de ncRNAs superexpressos em células-tronco embrionárias. O estudo utiliza a abordagem do tipo *guilty by association* e busca inferir, por meio de redes de interação entre ncRNAs e RNAs traduzidos a partir de genes codificantes para proteínas, o papel biológico de 15 ncRNAs obsevados em CTEs. O artigo encontra-se em fase de redação científica e o periódio ao qual será submetido ainda não foi decidido.

O capítulo 2 apresenta um revisão bibliográfica das principais características de RNAs competidores endógenos (ceRNAs), uma listagem de ceRNAs já documentados na literatura científica dentro do seu contexto biológico, bem como

fatores que influenciam a relação de competição entre RNAs. Por fim, uma sessão apresenta os conceitos e principais descobertas sobre redes biológicas de competição. O artigo encontra-se em fase de redação científica e será submetido ao periódico *Critical Reviews in Biochemistry and Molecular Biology*, cujo fator de impacto atual é 8,86.

O capítulo 3 aborda um estudo de predição de RNAs não codificantes com potencial de atuação como competidores endógenos em CTEs. O artigo enfoca RNAs não codificantes que, por meio de interações de competição, contribuam para a manutenção do estado tronco. O artigo encontra-se em fase de redação científica e será submetido ao periódico *Developmental Cell*, cujo fator de impacto atual é 9,36.

Os três capítulos estão seguidos por uma discussão geral englobando os conhecimentos discutidos em cada um deles, pelas conclusões geradas com o desenvolvimento deste trabalho de doutorado, bem como pelas perspectivas de continuação da pesquisa científica.

A tese é finalizada pelo anexo contendo o artigo entitulado *Scaffolds for artificial miRNA expression in animal cells.* O artigo versa sobre construções de vetores para a expressão de microRNAs artificiais e foi publicado no ano de 2015, no periódico *Human Gene Therapy Methods*, cujo fator de impacto atual é 4,08.

SUMÁRIO

LISTA DE ABREVIATURAS

ceRNA	*Competing endogenous RNA*
ceRNET	*Competing endogenous RNA network*
circRNA	*Circular RNA*
CT	Células-tronco
CTEs	Células-tronco embrionárias
DEG	*Differentially expressed gene*
eRNA	*Enhancer RNA*
ESCs	*Embryonic stem cells*
FDR	*False discovery rate*
GCP	Gene codificante de proteína
GO	*Gene ontology*
iPSCs	*Induced pluripotent stem cells*
lncRNA	*Long non-coding RNA*
miRLC	*miRNA-loading complex*
miRNA	microRNA
MRE	*microRNA response element*
ncRNA	*Noncoding RNA*
paRNA	*Promoter associated RNA*
PCT	*Probability of conserved targeting*
piRNA	piwiRNA
PWF	*Probability weighting function*
RISC	*RNA induced silencing complex*
RNAi	RNA de interferência
snoRNA	*Small nucleosomal RNA*
TF	*Transcription factor*
TPM	*Transcripts per million*
tRNA	*Transporter RNA*
UTR	*Untranslated region*

RESUMO

O genoma humano tem um tamano aproximado de 3,2 Gb e aproximadamente 74% das suas sequências são transcritas em RNAs. Destes transcritos, cerca de 33 mil não são traduzidos em proteínas, desempenhando seu papel funcional na célula na forma de RNAs. Dentre os trancritos que não codificam proteínas estão os lncRNAs, os pseudogenes e os circRNAs. Inicialmente considerados como ruídos transcricionais, esses RNAs têm chamado cada vez mais a atenção da comunidade científica e têm sido atribuídos a um série de funções celulares. Apesar do crescente número de trabalhos enfocando estes RNAs, pouco se sabe sobre essas moléculas em células-tronco embrionárias (CTEs) humanas, tanto no que se refere ao conjunto de moléculas expressas quanto sobre a função por elas desempenhada. Nesse sentido, este trabalho teve como objetivo identificar os lncRNAs, pseudogenes e circRNAs expressos em células-tronco embrionárias humanas a partir de dados de RNA-seq depositados no banco de dados GEO e inferir, através de ferramentas de bioinformática, as possíveis funções desempenhadas por estas moléculas no estado tronco. As análises revelaram, ao todo 1182 ncRNAs pertencentes as três classes em estudo (lncRNA, n=317; pseudogene, n=585; e circRNA, n=280) no conjunto de dados principal. Quando este foi comparado a outros conjuntos de dados, 87 lncRNAs, 51 pseudogenes e 10 circRNAs foram detectados como transcritos comuns. Dentre os lncRNAs e os pseudogenes, 15 foram selecionados para inferência de função pela abordagem de *guilty by association*. Foi possível associar RNA selecionados a

processos como splicing, organização de cromatina, mitose, ciclo celular, biogênese de ribossomos, reparo de DNA e resposta ao dano, apoptose, organização do citoesqueleto, desenvolvimento embrionário e regulação da diferenciação celular. Além disso, alguns dos RNAs em estudo mostraram-se fortes candidatos a atuarem como competidores endógenos de mRNAs expressos em CTEs. Análises de correlação, de força de ligação ao miRNA e do tipo de *seed* do sítio de ligação apontaram os pares de mRNAs e ncRNAs AHCY-RP11-253E3.3, F2RL1-EEF1A1P6, HSPD1-SNHG5, INO80-RP11-20D14.6, PARP1-RP11-20D14.6 e PRIM2-GAS5, juntamente com o RNA circular circ-HIPK3, como os RNAs com maior potencial para competirem pelos miRNAs expressos em CTEs. Estes resultados voltam a atenção da pesquisa básica para uma parte pouco conhecida da biologia das CTEs e direcionam futuros experimentos visando a elucidação da função de ncRNAs nesse contexto celular. Ensaios de modulação da expressão dos ncRNAs e a observação concomitante da respostas dos RNAs e proteínas a eles relacionados são o próximo passo para desvendar os mecanismos moleculares através dos quais estes RNAs desempenham seus papeis em CTEs.

ABSTRACT

The human genome has an average size of 3.2 Gb and approximately 74% of its sequences are transcribed into RNAs. From this transcripts, around 33 thousand are not translated into proteins, performing their cellular roles as RNAs molecules. Among the non-protein coding transcripts are the lncRNAs, the pseudogenes and the circRNAs. Inicially thought to be transcription noise, these molecules have been drawing attention of the scientific comunity and several cellular functions have been attributed to them. Although the rising number of studies focusing on these RNAs, little is known about them in human embryonic stem cells (ESCs), includind the set of expressed non-coding RNAs (ncRNA) and their functions. In this sense, this study aimed to identify lncRNAs, pseudogenes and circRNAs expressed in human ESCs from RNA-seq data deposited in GEO public database and infer, through bioinformatic tools, the possible functions played by these molecules in the stem state. The analyzis revealed 1182 ncRNAs belonging to the three classes (lncRNA, n=317; pseudogene, n=585; and circRNA, n=280) studied in the main dataset. When compared to other datasets, 87 lncRNAs, 51 pseudogenes and 10 circRNAs were detected as common to all datasets. Among the lncRNAs and the pseudogenes, 15 were selected for function inference through the *guilty by association* method. It was possible to associate the selected RNAs to cellular processes as RNA splicing, chromatin organization, mitosis, cell cycle, ribosome biogenesis, DNA repair and DNA damage response, apoptosis, cytoskeletton organization, embryonic

development and regulation of cell differentiation. Moreover, some of the RNAs showed to be strong candidates to act as endogenous competitors of the ESCs-expressed mRNAs. Pearson correlation analysis and investigation of the miRNA seed type and binding strenght indicate the following ncRNA-mRNA pairs AHCY-RP11-253E3.3, F2RL1-EEF1A1P6, HSPD1-SNHG5, INO80-RP11-20D14.6, PARP1-RP11-20D14.6 e PRIM2-GAS5, together with the circular RNA circ-HIPK3, as the RNAs with major potential to act as competing endogenous RNAs in ESCs. These results shift the attentions to a poorly known part of ESCs biology and give directions to future experiments aiming to elucidate the ncRNAs function in this cellular context. ncRNAs expression modulation assays and the concomitant observation of the positively correlated mRNAs response to these alterations are the next step in the unraveling of the specific molecular mechanisms through which ncRNAs exert their roles in ESCs.

1. INTRODUÇÃO GERAL

1.1. RNAs não codificantes

Considerando-se todo o DNA transcrito a partir do genoma humano, apenas 2% origina RNAs que são traduzidos em proteínas[1]. O restante dos transcritos compreende os RNAs não codificantes (ncRNAs). Os ncRNAs podem ser divididos em duas categorias: os estruturais e os regulatórios[2]. Dentro desta classificação, os ncRNAs estruturais compreendem os tRNAs, rRNAs, snoRNAs e snRNAs e os regulatórios incluem os ncRNAs curtos (miRNAs, piRNAs), médios (paRNAs) e os longos (lncRNAs, pseudogenes)[2]. A seguir, serão enfocados os ncRNAs alvos deste trabalho.

1.1.1. microRNAs

microRNAs são RNAs curtos com aproximadamente 22 nucleotídeos de tamanho envolvidos no controle da pós-transcricional da expressão gênica[3]. Atualmente há cerca de 1.881 sequências anotadas no banco de dados miRBase (*release* 21), mas estudos estimam que existam cerca de 25 mil precursores de miRNAs no genoma humano[4].

Os genes que codificam miRNAs podem aparecer de forma isolada ou agrupados em clusters[3,5.] Além disso, podem estar localizados em regiões intergênicas

ou dentro de sequências intrônicas pertencentes a genes codificantes de proteínas, além de éxons e íntrons de genes não codificantes[6,7.]

A biogênese e processamento dos miRNAs é dividida em duas etapas, iniciando-se núcleo e sendo concluída no citoplasma[3]. Inicialmente, os miRNAs são transcritos, majoritariamente pela RNA polimerase II, como uma molécula longa denominada pri-miRNA, a qual apresenta cap 5' e cauda poli(A)[3,8]. O pri-miRNA é então processado, ainda no núcleo, pela RNase Drosha, gerando uma molécula de aproximadamente 60-70 nucleotídeos denominada de pré-miRNA[9]. Após a sua geração, o pré-miRNA é exportado para o citoplasma por intermédio do receptor de transporte nuclear Exportina-5[10].

No citoplasma, o pré-miRNA associa-se ao complexo carregador de miRNA (miRLC, do inglês *miRNA-loading complex*), formado pelas proteínas Dicer, TRBP e Ago2[11]. Ocorre, então, o processamento do pré-miRNA pela enzima Dicer, gerando um RNA de fita dupla transiente com aproximadamente 22 nucleotídeos de extensão[11,12]. A fita de RNA que apresenta menor estabilidade de pareamento na extremidade 5' é mantida associada ao miRLC enquanto a outra é degradada[11,13]. Em seguida, as proteínas Dicer e TRBP são liberadas do complexo, permanecendo apenas Ago2 associada à fita simples de RNA[11]. Essa associação remanescente passa então a ser denominada de *RNA-induced silencing complex* (RISC) e é direcionada aos mRNAs alvo, promovendo a repressão da sua tradução ou a sua degradação[14].

A associação com o RNA alvo ocorre preferencialmente na extremidade 3'UTR, mas também há relatos da presença de sítos de ligação em 5'UTR e na porção

codificante de mRNAs[4]. Além dos RNAs codificantes para proteínas, também são alvos dos miRNAs os RNAs circulares, os lncRNAs e os pseudogenes[15–17.] Ademais, um único gene pode ser regulado por vários miRNAs e um mesmo miRNA pode apresentar mais de um alvo[18].

Quanto a sua função biológica, os miRNAs tem sido relacionados, desde a sua descoberta, a uma série de estados e funções celulares. Em humanos, miRNAs estão associados, por exemplo, ao desenvolvimento de câncer, podendo atuar tanto como oncogenes (oncomirs) quanto como supressores tumorais, além de outras patologias, como doenças cardiovasculares, inflamatórias, neurodegenarativas e autoimunes[19–24]. miRNAs também apresentam papel na regulação da autorrenovação e diferenciação de células-tronco, podendo, inclusive, ser utilizados na indução da reprogramação celular[25–27]. Além disso, foram hipotetizados como participantes de um novo nível regulação da expressão gênica proposto em 2011, onde RNAs competem entre si pela sua ligação[28].

1.1.2. RNAs longos não codificantes

A denominação de RNAs longos não codificantes (lncRNAs) é atribuída a RNAs não traduzidos em proteínas que apresentam tamanho igual ou superior a 200 nucleotídeos[29,30]. Essa categoria de RNAs é classificada de acordo com o local de origem no genoma, levando em consideração a sua relação com o gene codificante de proteína (GCP) mais próximo. Nesse sentido, os lncRNAs podem ser classificados em

(i) intergênicos, ou lincRNAs, os quais não se sobrepõem a nenhum GCP; e (ii) lncRNAs que se sobrepõem a um GCP. Este último grupo inclui lncRNAs que apresentam sobreposição com éxons, os que se originam de íntrons e os que apresentam genes codificantes para proteínas dentro de um de seus íntrons[31].

De modo geral, lncRNAs se caracterizam por apresentar níveis de expressão mais baixos quando comparados a GCPs, e o seu padrão de expressão é altamente tecido-específico[31.] Além disso, o processo de transcrição dos lncRNAs é similar aos GCP. Os transcritos são gerados pela RNA polimerase II e sofrem *splicing*, capeamento em 5' e apresentam cauda poli(A)[32]. A maioria dos lncRNAs é composta por apenas dois exons e o tamanho do transcrito geralmente é menor do que o de GCPs[31]. Assim como para mRNAs, também foi evidenciada a geração de transcritos alternativos em lcnRNAs[31.]

Quanto a função que desempenham nas células, os lncRNAs podem atuar como (i) sinalizadores, mediando, por exemplo, o silenciamento transcricional, a apoptose e o decaimento de RNAs; (ii) "iscas", associando-se a proteínas ou miRNAs e impedindo o desempenho de sua função; (iii) guias, ligando-se a proteínas e direcionando-as ao seu sítio de atuação; e (iv) *scaffolds*, servindo de plataforma para a montagem de complexos proteicos[29].

1.1.3. Pseudogenes

O termo pseudogene foi cunhado por Claude Jacq e colaboradores em 1977[33]. Estes elementos são caracterizados por variados graus de similaridade de sequência com genes funcionais, porém, perderam a sua capacidade de codificação original pela presença de códons de parada prematuros ou de alteração do quadro de leitura[34,35]. Pseudogenes podem ser classificados como processados ou não processados[36]. Especificamente, o grupo dos não processados é composto pelos pseudogenes unitários, que originam-se do acúmulo de mutações em um gene funcional de cópia única, e pelos duplicados, onde o pseudogene origina-se pela duplicação de um gene funcional somada ao acúmulo de mutações em uma das cópias. Nestes dois casos, a estrutura geral do gene e os elementos regulatórios são mantidas[34,36].

Os pseudogenes processados, também conhecidos como pseudogenes retrotranspostos, recebem esta denominação por apresentarem as mesmas características de um mRNA processado, como presença da cauda poli(A) e ausência de íntrons. Além disso, são flanqueados por regiões repetitivas, que estão associadas a sítios de inserção de elementos de transposição[36]. Isto ocorre, pois esta classe de pseudogenes origina-se da transposição de sequências originadas da transcrição reversa mRNAs[34,36]. Pseudogenes processados podem também apresentar retensão de sequências intrônicas, sendo denominados de “pseudogenes semiprocessados”, ou ainda, após o processo de geração do pseudogene, sofrer um processo de duplicação, gerando o que se nomeia de pseudogenes processados duplicados[37].

Originalmente, os pseudogenes foram considerados como elementos não funcionais do genoma. No entanto, estudos tem revelado um número cada vez maior destas entidades apresentando funções associadas, podendo ser transcritos e, inclusive, traduzidos[38,39]. Dentre as funções já reportadas para pseudogenes estão a (i) supressão da expressão gênica através da geração de transcritos antisenso ou RNAis contra o gene de origem do pseudogene, (ii) a atuação como competidores endógenos, funcionando como alvos alternativos para miRNAs (ver item 1.2 da Introdução), (iii) ou a codificação pequenos peptídeos ou proteínas funcionais[35,40].

Considerando que pseudogenes podem apresentar função no contexto celular, novas definição e nomenclatura para essas entidades foram propostas. Assim, pseudogenes seriam as sequências genômicas originadas de genes funcionais, mas que não codificariam o mesmo tipo de produto funcional que o gene original[41]. Estas sequências, por sua vez, receberiam a denominação de "pseudogenes adaptados", para diferenciá-las dos pseudogenes não funcionais[41].

1.1.4. RNAs circulares

Por muito tempo, o senso comum para moléculas de RNA não-virais foi de que estas seriam lineares. No entanto, este conceito foi revisto com a descoberta de RNAs com estrutura circular (circRNAs) em eucariotos[42]. Inicialmente vistos como acidentes genéticos ou artefatos[43], os circRNAs já tiveram sua expressão detectada em humanos, camundongos, peixes (*Danio rerio*), invertebrados (*Caenorhabditis elegans*,

Drosophila melanogaster), fungos (*Schizosaccharomyces pombe* e *Saccharomyces cerevisiae*), plantas (*Arabidopsis thaliana*), e protistas (*Plasmodium falciparum* e *Dictyostelium discoideum*)[44,45].

A gênese dos circRNAs dá-se durante o *splicing*, onde o spliceossoma aparentemente liga as extremidades 5' e 3' de éxons de um mesmo transcrito[46]. Especificamente, moléculas circulares são formadas quando o splicing ocorre entre um sítio aceptor de um éxon *upstream* e um sítio doador de um éxon *downstream*[46]. Nesse processo, sítios de *splicing* canônicos e crípticos podem ser usados[45].

A maioria dos circRNAs origina-se de eregiões exônicas, ou seja, é transcrita de genes previamente conhecidos por codificarem mRNAs[44,45.] Há, também, um pequeno número circRNAs derivados de UTRs, íntrons e regiões não anotadas do genoma[44]. Quanto a sua estrutura, os circRNAs podem ser constituídos de um ou mais éxons e não apresentam cauda poli(A)[45,46]. Além disso, pode existir mais de uma isoforma circular originada de uma mesma sequência genômica[43,45].

O perfil de expressão dos circRNAs caracteriza-se pela especificidade por tipo celular e estágio de desenvolvimento e, em muitos casos, os níveis de expressão do circRNA são comparáveis ao do transcrito linear canônico originado do mesmo gene[46]. Pelo seu formato circular, os circRNAs são mais estáveis dentro da célula que RNAs lineares e uma das estratégias para confirmar a sua identidade é a resistência à clivagem por RNase R, a qual é capaz de atuar apenas em RNAs lineares[44,45]. Além disso, circRNAs são predominantemente citoplasmáticos[43,46].

1.2. RNAs competidores endógenos

A hipótese sobre existência de RNAs atuando na modulação dos níveis de outros RNAs por competição foi inicialmente postulada em 2011[28]. Denominados de RNAs competidores endógenos ou ceRNAs, essas moléculas se caracterizam por apresentar, em comum com outros RNAs com os quais são co-expressos, sequências denominadas elementos de resposta a miRNAs (MREs). Desta forma, a atuação dos ceRNAs se daria pela competição com outros RNAs pela ligação de miRNAs. Em outras palavras, os ceRNAs funcionariam como "esponjas de miRNAs", monopolizando-os e deixando os outros RNAs alvo acessíveis à tradução ou ao desempenho de suas funções celulares. Esse mecanismo de regulação sempre apresentaria, portanto, três elementos de caráter ribonucleico, sendo que a relação entre dois deles é mediada por moléculas de miRNA.

Dentre as classes de RNAs que possuem relatos comprovando a atuação como competidores estão mRNAs, lncRNAs, pseudogenes e circRNAs[44,47,48]. O ceRNA mais conhecido e bem estudado é *PTENP1*, pseudogene processado homólogo ao gene PTEN. Ambas as sequências apresentam sítios comuns de ligação de miRNAs e o pseudogene regula positivamente PTEN[48]. Demonstrou-se que a superexpressão de *PTENP1* leva ao consequente aumento de PTEN, assim como a diminuição dos níveis de RNA de *PTENP1* leva à baixa produção do mRNA e da proteína PTEN[48]. A modulação dos níveis de *PTENP1* em células que não expressam Dicer não apresenta

os mesmos efeitos sobre PTEN, indicando que a relação entre os dois RNAs é efetivamente mediada por miRNAs[48].

Aproximadamente 60 casos de competição entre RNAs já foram documentados na literatura, tendo sido associados à vários contextos celulares diferentes. Dentre os papéis biológicos desempenhados pela competição estão a promoção ou a supressão do desenvolvimento de tumores, a manutenção do estado tronco, a diferenciação celular e associação com patologias como diabetes, osteoartrite, fibrose hepática e isquemia cerebral[48–54].

1.3. Células-tronco embrionárias

As células-tronco (CT) caracterizam-se, basicamente, pela capacidade de autorrenovação e de gerar células mais especializadas por meio da sua diferenciação[55.] Conforme o seu potencial de diferenciação e o período do desenvolvimento em que são obtidas, recebem diferentes denominações. As CT podem ser isoladas de embriões, fetos e indivíduos adultos[56–58].

As células originadas a partir do oócito fertilizado até o estágio de blastocisto são consideradas totipotentes, pois possuem o potencial de gerar um organismo inteiro[59]. A partir do blastocisto, as células começam a se especializar e são então consideradas pluripotentes, sendo mais conhecidas como células-tronco embrionárias (CTEs). As CTEs caracterizam-se pela capacidade de diferenciarem-se em células pertencentes aos três folhetos embrionários (endoderme, mesoderme e ectoderme),

mas já não são mais capazes de gerar um indivíduo[60]. Além disso, é possível mantê-las em cultura *in vitro* por longos períodos de tempo sem que as células percam suas características, devido à sua capacidade de autorrenovação indefinida[61].

Pela sua ampla capacidade de proliferação e de diferenciação, as CTEs tornaram-se interessantes do ponto de vista da terapia celular. Após o seu isolamento de blastocistos humanos, em 1998, uma série de estudos foram conduzidos com o objetivo de elucidar os mecanismos moleculares envolvidos na manutenção do estado tronco[58,62]. Além disso, inúmeras metodologias de indução de diferenciação de CTEs em células de interesse clínico foram desenvolvidas. Entre os tipos celulares diferenciados gerados a partir de CTEs estão cardiomiócitos, neurônios, hepatócitos, células beta pancreáticas, cones fotorreceptores, linhagens hematopoiéticas, entre outras[63–68]. Além disso, uma série de testes clínicos têm sido desenvolvidos para verificar o potencial destas células no tratamento de diversas patologias[69]

Por outro lado, o conhecimento dos mecanismos moleculares envolvidos na manutenção do estado tronco mostrou-se util no desenvolvimento de técnicas de indução do retorno de células diferenciadas ao estado tronco. A indução da expressão de genes superexpressos em CTEs, como KLF4, OCT4, NANOG e c-myc, ou mesmo miRNAs, como os pertencentes ao cluster mir-302-367, em células diferenciadas levou à geração das chamadas células-tronco pluripotentes induzidas (iPSCs, do inglês *induced Pluripotent Stem Cells*), as quais apresentam diversas características em comum com as CTEs[8,9].

1.3.1. RNAs não codificantes expressos em células-tronco embrionárias

Apesar do grande volume de conhecimento acumulado sobre as CTEs, pouco ainda se sabe sobre os ncRNAs nesse contexto celular, tanto no que diz respeito aos tipos de ncRNAs expressos quanto a função por eles desempenhada em CTEs. Genes codificantes para dois fatores de transcrição chaves na manutenção do estado tronco, OCT4 e NANOG, tem pseudogenes já documentados na literatura[71,72]. Além disso, um estudo voltado para a identificação de circRNAs em embriões em diferentes estágios prévios à implantação identificou cerca de 252 transcritos circulares no blastocisto, estágio em que são isoladas as CTEs[73]. Outro estudo, por sua vez, detectou entre 2 mil e mais de 9 mil candidatos a RNAs circulares em CTEs humanas[74]. Esses trabalhos indicam que há possibilidade de expressão de pseudogenes e de circRNAs em CTEs. No entanto, há poucos estudos enfocando na detecção e na determinção da função desses RNA em CTEs.

Os lncRNAs, por sua vez, têm sido alvo de um número maior de pesquisas em CTEs. A maior parte dos estudos em grande escala, no entanto, foi desenvolvida em CTEs oriundas de camundongos[75–78]. Há apenas dois estudos em larga escala investigando lncRNAs em CTEs humanas. Baseado em dados de RNA-seq, um dos trabalhos identificou cerca de 300 lincRNAs, sendo 78 deles expressos preferencialmente em CTEs[79]. No entanto, poucos lncRNAs tiveram a sua função estudada em detalhes nas CTEs humanas. Entre os já documentados estão lncRNA_ES1, lncRNA_ES2 e lncRNA_ES3, os quais são alvos dos fatores de

transcrição Oct4 e Nanog e apresentam papel na manutenção da pluripotência, e linc-RoR, reportado como competidor endógeno de Oct4, Sox2 e Nanog e apontado como um potencializador da reprogramação de células somáticas em iPSCs[54,80,81].

No que diz respeito a ceRNAs em CTEs, há apenas três RNAs descritos pela sua atuação como competidores em CTEs humanas. Como já mencionado acima, um dos ncRNAs descrito como competidor é linc-RoR. Este lncRNA encontra-se superexpresso no estado indiferenciado e compete com os mRNAs codificantes de Oct4, Sox2 e Nanog, pela ligação de hsa-miR-145[54]. Desta forma, a expressão de linc-ROR permite a expressão proteica de OCT-4, SOX2 e NANOG, tendo, portanto, papel na conservação do estado indiferenciado das ESCs[54]. Outro ncRNA que atua como ceRNA nesse contexto celular é FLJ11812, o qual compete com os mRNAs codificantes de CDC20B e ATG13 pelo miR-4459. Nesse caso, especula-se que o estado tronco seja favorecido por mecanismos relacionados ao ciclo celular e a autofagia[82]. O terceiro ncRNA associado a mecanismos de competição entre RNAs em CTEs é o lncRNA GAS5, o qual compete com o mRNA de NODAL pela ligaçao dos miRNAs miR-2467, -3200 e Let-7e[83].

Os miRNAs expressos em CTE humanas, por outro lado, já estão bem definidos na literatura[84–91]. Há algumas variações entre os miRNAs encontrados em diferentes linhagens de CTEs estudadas, mas algumas moléculas são consideradas consenso na litaretura, como os miRNAs expressos a partir dos clusters hsa-miR-302/367, -miR-520 e -miR-371/372/373[92]. Além disso, os miRNAs como hsa-miR-18b, -25, -93, -103a-2, -106b e -205 também são frequentemente detectados em CTEs[14–21].

2- JUSTIFICATIVA

Aproximadamente 74% do genoma humano é transcrito e mais de 33 mil RNAs são classificados como não codificantes, desempenhando seu papel sem serem traduzidos em proteínas. Estas moléculas, anteriormente consideradas produtos errôneos da transcrição, tem se mostrado atuantes em diversos processos biológicos. Consequentemente, RNAs produzidos a partir da transcrição de pseudogenes, lncRNAs e circRNAs estão entre as moléculas sobre as quais a ciência moderna tem se debruçado. Em adição a esse aumento do interesse por ncRNAs, um novo nível de regulação da expressão gênica, onde moléculas de RNA, incluindo lncRNAs, pseudogenes e circRNAs, competem com mRNAs pela ligação a miRNAs, foi proposto em 2011 e constitui um novo campo de investigação da atuação de ncRNAs.

Desde o isolamento das primeiras CTEs humanas, em 1998, muito conhecimento acumulou-se sobre a suas características moleculares, à exceção dos ncRNAs. A atual dimensão da escassez de informações sobre o papel desses RNAs em CTEs é dada pelo número de artigos disponíveis na literatura científica. Há apenas nove artigos que discutem ncRNAs no contexto de CTEs humanas, sendo que apenas três deles investigam processos de competição entre ncRNAs e mRNAs. Ademais, até o momento, nenhum estudo enfocando circRNAs em CTEs foi publicado.

A plena compreensão dos mecanismos moleculares característicos de CTEs, incluindo aqueles involvendo RNAs não codificantes, é importante devido ao seu alto potencial de diferenciação, o que as torna interessantes do ponto de vista da terapia

celular. Além disso, as descobertas feitas em CTEs podem ser usadas para o aprimoramento das técnicas de geração de iPCs. Dentro deste contexto, dados oriundos de análises em larga escala, como o RNA-seq, possibilitam a proposição de interações de competição envolvidas na manutenção do estado tronco, bem como a inferência de função de RNAs ou genes pela abordagem de *guilty by association*. As relações e funções hipotetizadas a partir destes dados servem para direcionar confirmações posteriores. Esse estudo, por conseguinte, possui potencial de ampliar o conhecimento atual sobre os papeis desempenhados por ncRNAs no estado tronco, além de contribuir para a elucidação da complexidade do novo nível de controle da expressão gênica envolvendo competição entre RNAs.

3- OBJETIVOS

3.1. Objetivo geral

Propor a função potencial de RNAs não codificantes superexpressos em células-tronco embrionárias humanas utilizando ferramentas de bioinformática, com enfoque especial em relações de competição entre RNAs.

3.2. Objetivos específicos

- Identificar, a partir de dados de RNA-seq depositados no banco de dados GEO, lncRNAs, pseudogenes, circRNAs e miRNAs diferencialmente expressos em células-tronco embrionárias humanas;

- Predizer, por meio da abordagem de *guilty by association*, empregando redes de co-expressão entre lncRNAs, pseudogenes e genes codificantes de proteínas, uma possível função de pseudogenes e lncRNAs superexpressos em células-tronco embrionárias humanas;

- Identificar potenciais sítios de ligação de miRNAs característicos do estado tronco embrionário nos lncRNAs, pseudogenes e circRNAs superexpressos nesse mesmo contexto celular;

- Identificar os mRNAs superexpressos em células-tronco embrionárias humanas que sejam alvo de miRNAs característicos deste mesmo contexto celular;

- Propor uma relação de competição envolvendo os lncRNAs, pseudogenes, circRNAS e mRNAs superexpressos em células-tronco embrionárias humanas e alvo dos miRNAs característicos deste tipo celular através da análise de correlação entre mRNAs e ncRNAs e análise dos sítios de ligação dos miRNAs nestas moléculas.

Capítulo I

Evaluating potential functions of ncRNAs upregulated in human embryonic stem cells

Artigo a ser submetido a um periódico a ser definido

Evaluating potential functions of ncRNAs upregulated in human embryonic stem cells

Raquel Calloni and Diego Bonatto

Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul, Departamento

de Biologia Molecular e Biotecnologia, Universidade Federal do Rio Grande do Sul

Abstract

Among the functional genomic segments are those transcribed in RNAs which are not translated in proteins, the non-coding RNAs (ncRNAs). These molecules have been implicated in many biological processes, but little is known about the ncRNAs expressed in embryonic stem cells (ESCs) and the functions they execute in these cells. In this study we identified 317 long non-coding RNAs and 585 pseudogenes upregulated in ESCs. Among the upregulated ncRNAs, 15 molecules had their function investigated through a guilty by associaton pipeline. Based on the observed gene ontologies (GOs), the ncRNAs were categorized in 10 major GOs: RNA splicing, chromatin organization, mitosis, cell cycle, ribosome biogenesis, DNA repair and response to DNA damage stimulus, apoptosis, cytoskeleton organization, embryonic development and regulation of cell differentiation. In addition, some of the investigated lncRNA and pseudogenes presented high potential to bind to splicing associated proteins.

Keywords: lncRNAs, pseudogenes, embryonic stem cells, guilty by association, gene fucntion prediction.

1. Introdução

The human genome has an approximated total length of 3.2 gigabases and, initially, most of its sequences were believed to have no function, being designated as "junk DNA"[1]. Nowadays, however, it is consensus that the concept of junk DNA is not correct as an average of 80.4% of our genome is known to be functional in at least one cell type[2]. In addition, about 74% of the sequences are transcribed in coding or non-coding RNAs[2]. Among all the transcripts produced, around of 33,000 do not code for proteins and most of them are transcribed from the 9,640 long non-coding RNA (lncRNAs) genes or from the 863 transcribed pseudogenes[2].

The designation of lncRNA is attributed to the non-translated RNAs larger than 200 nucleotides[3,4]. LncRNAs are expressed in low levels when compared to protein coding genes and their expression pattern is highly tissue specific[4]. In addition, lncRNAs act as signaling molecules, protein and miRNAs decoys, protein guides and, scaffolds complexes assembly[3]. Pseudogenes, on the other hand, are genomic sequences very similar to functional genes, but without the protein codification potential observed on its functional counterpart due to the presence of premature stop codons or frameshift alterations[5,6]. As the lncRNAs, pseudogenes were initially thought to be non-functional DNA elements. However, researches have shown that

many known pseudogenes are transcribed and also translated and have important functions in the cell[7,8]. They play roles in the gene expression silencing, through antisense transcripts, may act as competing endogenous RNAs and also code for functional small peptides and proteins[6,9].

The comprehension of the cellular roles of lncRNAs and pseudogenes are improving fast. Also, large amount of knowledge has been accumulated on embryonic stem cells (ESCs). However, little is known about the ncRNAs in this cellular context. Only few ESC-expressed lncRNAs had their function already documented in the literature. Among the molecules already described in detail are FLJ11812, lncRNA_ES1, lncRNA_ES2, lncRNA_ES3, LNCPRESS1 and linc-RoR, which demonstrated to have roles in the pluripotency maintainment[10–14.] In the case of pseudogenes, however, no study has been performed in order to elucidate their function in ESCs, even with the existence of OCT4 and NANOG pseudogenes reported in the literature[15,16]. In addition, solely one large scale study was performed on ncRNAs in ESCs, but it focused only in long intergenic RNAs (lincRNAs)[17].

In this context, the aim of this study was to identify lncRNAs and pseudogenes upregulated in ESCs using different RNA-seq datasets and infer their functions through a guilty by association approach, using co-expression networks and GO analysis.

2. Materials and Methods

2.1. RNA-seq data

Four RNA-seq data were used in this study: (i) RNA-seq data from embryonic stem cells (RUES2) and cardiomyocytes differentiated from them (GSE64417), (ii) RNA-seq data from WA-09 embryonic stem cells lineage and prefrontal cortex neurons differentiated from them (GSE56796), (iii) RNA-seq data from H9 embryonic stem cells lineage and retinal ganglion cells differentiated from them (GSE84639), and (iv) RNA-seq data from peripheral blood mononuclear cells-derived iPSCs and macrophages differentiated from them (GSE55536). All the libraries were generated from total RNA and are paired-end, except for those from the datasets GSE55536, which was originated from polyA RNA, and GSE56796, which is single-end.

2.2. Bioinformatics analysis

2.2.2. Reads alignment

Non-trimmed reads were aligned to the human genome (version GRCh38) using the software STAR (version 2.4.2a)[18]. The default parameters suggested by the developers were used in this step.

2.2.3. Differential expression analysis

Differential expression analysis was separately performed for each of the data sets included in this study. Read count files were generated from the alignment BAM files using the software HTSeq (version 0.6.1)[19]. These files were then used as input to the package DESeq2[20] to perform the identification of differentially expressed genes (DEG). It was considered differentially expressed those genes which presented $|\log_2 FC| \geqslant 1$ and adjusted p value < 0.05, estimated by the Benjamini-Hochberg method. Only genes upregulated in ESCs were considered for the analysis.

2.2.4. Network construction and gene ontology analysis

The correlation between ncRNAs and mRNAs was estimated using the web based tool Co-LncRNA (http://www.bio-bigdata.com/Co-LncRNA)[21]. The chosen co-expression method was Spearman Rank Correlation and p-value and correlation coefficient parameters were set to > 0.4 and < 0.05, respectively.

It was assumed that all the mRNAs detected in this study are potentially translated into protein. In this way, the proteins coded by the mRNAs detected as significantly positively correlated to ncRNAs were used as input in the web base tool STRING (version 10.0, http://string-db.org)[22] in order to create a protein interaction

network. For this step, the active interaction sources parameters chosen in STRING were experiments, databases, co-expression and co-occurrence. In addition to the inicial input, new protein nodes were added until the observation of the network saturation.

A network containing the interactions established by the correlation analysis and the interactions among the proteins detected with STRING was constructed for selected ncRNA using the software Cytoscape (version 3.4.0)[23]. All the networks were submited to a gene ontology (GO) analysis, performed using the Cytoscape's plugin BiNGO[24]. Significant biological processes were identified with hypergeometric test and Benjamini & Hochberg false discovery rate (FDR) as multitesting correction. A significance level of 0.05 was applied. A ncRNA was associated to a given biological process when at least 3 of its direct linked partners were associated to a common GO.

2.2.5. ncRNA protein binding prediction

The ncRNA potential to bind to proteins was analyzed with the web based tool catRAPID omics (http://service.tartaglialab.com/page/catrapid_group)[25]. The selected parameters for this analysis were set to study interactions to full-length proteins instead of nucleic acid domains and to include RNA- and DNA-binding proteins and disordered proteins. As the Star Rating Score ranges from 0 (lowest quality) to 3 (highest quality), the interactions with Star Rating Score ≥ 2 were considered as valid.

To perform the protein binding analysis it was necessary to determine the ncRNAs isoforms. This step was conducted using the software Salmon in the quasi-mapping-based mode, applying the default parameters suggested by the developers[24]. A fasta file containing the transcripts produced by human ncRNAs were used to build the indexes and perform the Salmon analysis. This file was downloaded from the Ensembl genome browser, release 79.

3. Results

3.1. Differentially expressed non-coding RNAs

From all differentially expressed genes detected, a total of 6,221 genes was observed as upregulated in ESCs of the dataset (i) when compared to differentiated cells. Among these DEGs, it was identified 317 long non-coding RNAS and 585 pseudogenes (Supplemental Material 1).

The ncRNAs observed as upregulated in the dataset (i) were compared to those observed in the three other datasets. ncRNAs common to all datasets were detected and included 87 lncRNAs and 51 pseudogenes (Supplemental Material 1). The 10 most upregulated ncRNAs common to all datasets were kept for function prediction analysis (Figure 1). Additionally, the already reported as ESC-expressed ncRNAs lncRNA_ES1 (LINC01108) and lncRNA_ES3 (LINC00458), together with the

pseudogenes of the transcritption factors Nanog and Oct4, NANOGP1, POU5F1P3 and POU5F1P4, were included in the analysis[11].

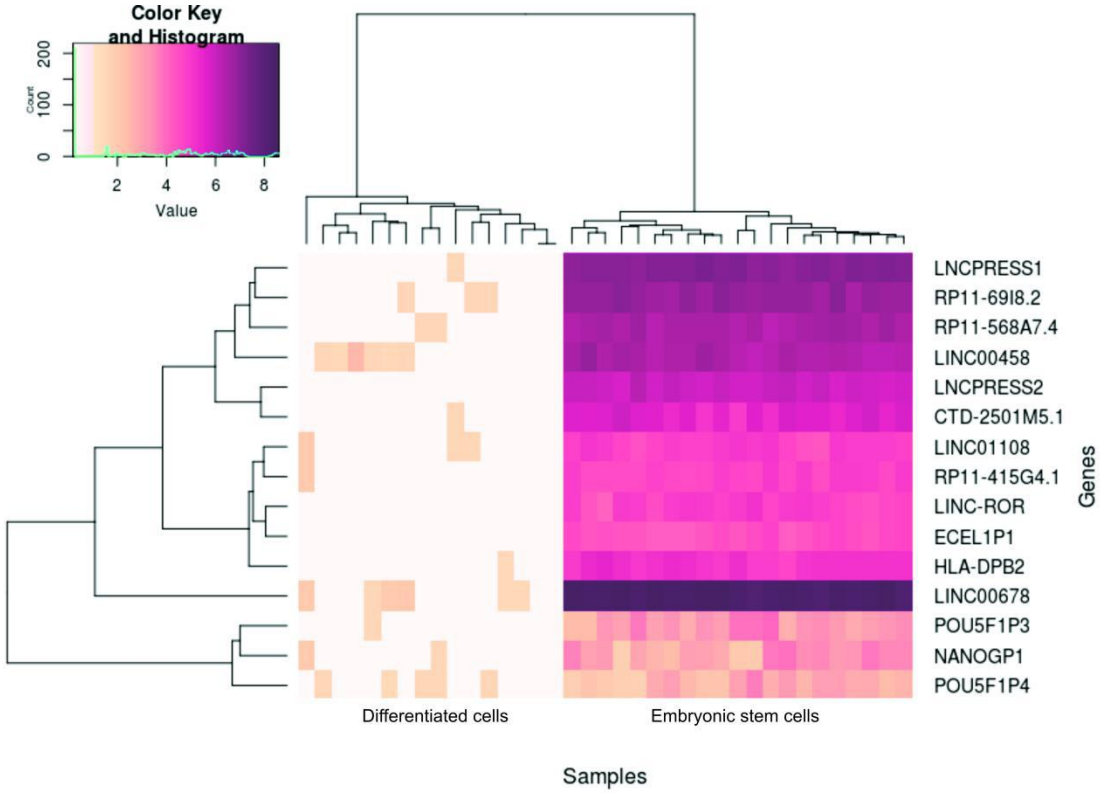


Figure 1: Heatmap showing the 15 ESCs upregulated ncRNAs chosen for function prediction analysis. Different color shades represent the gene associated log_2FC. Blue line in left above histogram represent the normalized read counts.

3.2. Network construction and ncRNAs function evaluation

A network was constructed for each ncRNA and the nodes were connected based in the correlations among ncRNAs and mRNAs and interactions established among the proteins coded by them (Figure 2). Interestingly, many mRNAs presented

positive correlation with more than one ncRNA and many ncRNA networks were observed to be connected (Figure 3).

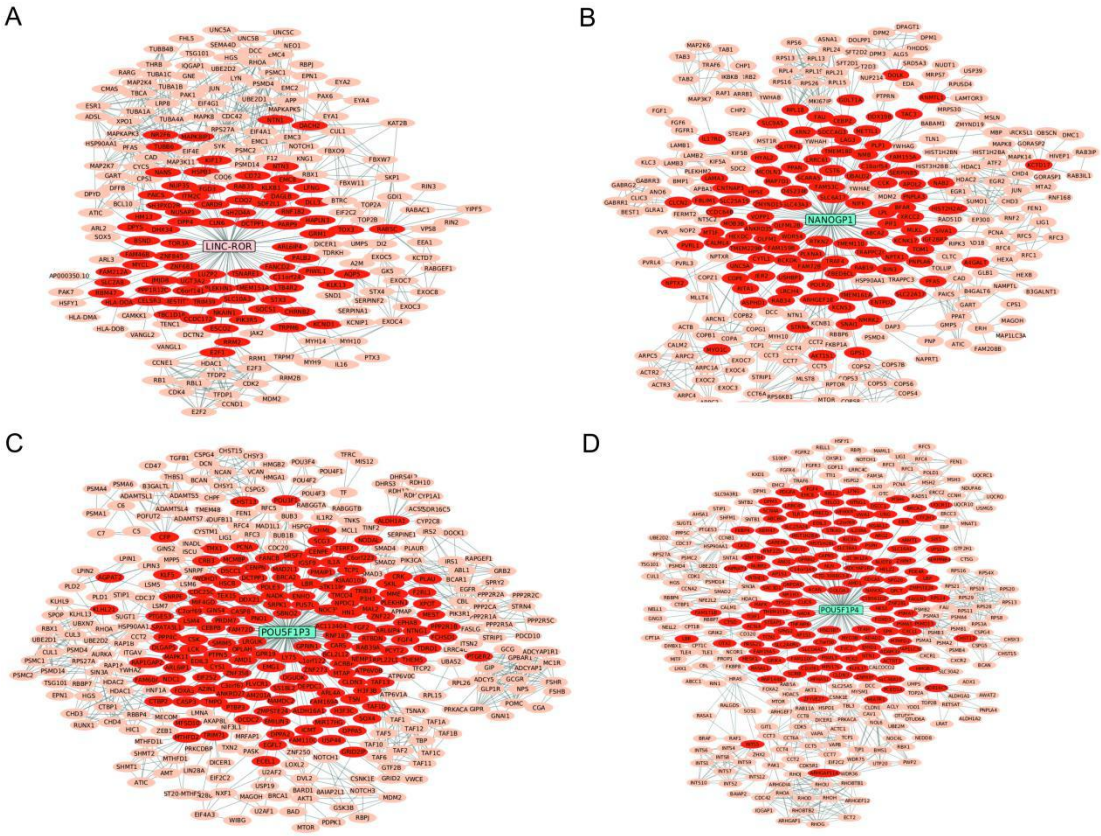


Figure 2: Some of the ncRNAs networks built from positively correlated genes and aditional interacting nodes added with STRING. Red ellipsis protein-coding mRNAs conected to ncRNAs by positive correlation. Orange ellipsis represent proteins added to the network using STRING. Blue ellipsis represent pseudogenes and pink rectangles represent lncRNAs.

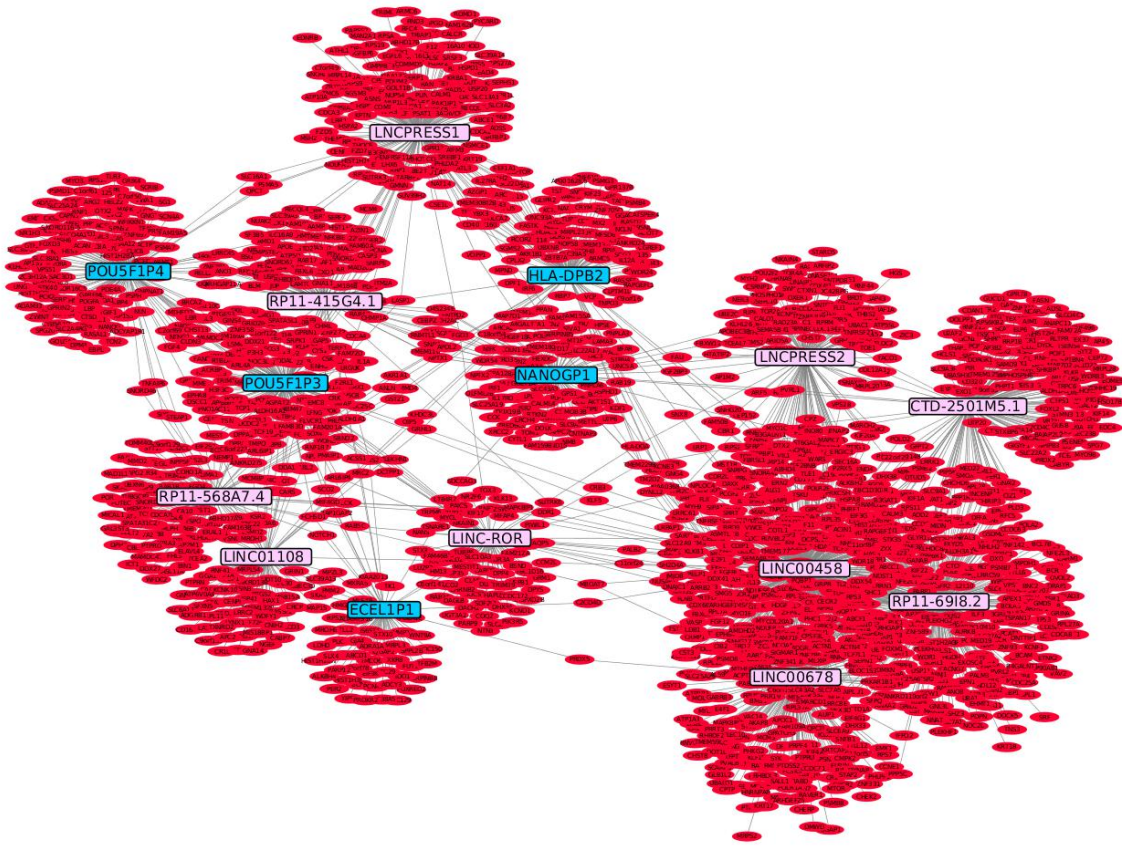


Figure 3: Many ncRNAs presented positive correlation with shared groups of mRNAs, establishing connections between the different ncRNAs networks. Red ellipsis represent protein-coding mRNAs, pink rectangles represent lncRNAs and blue rectangles represent pseudogenes.

A GO analysis was performed for each network and a ncRNA was associated to a given biological process when at least 3 of its direct linked partners were associated to a common GO (Table 1 and Supplemental Material 2). The GO analysis permitted to categorize the studied ncRNAs into 10 classes (Figure 4). Many ncRNAs could be associated with more than one biological process, which imply they may have function in more than one cellular mechanism.

Table 1: Main gene ontologies associated with the analyzed ncRNAs.

ncRNA	ncRNA-associated biological processes	p-value	Number of mRNAs*
CTD-2501M5.1	Cell cycle	$1,41.10^{-2}$	13
	M phase	$6.01.10^{-1}$	10
	Response to DNA damage stimulus	$3.78.10^{-2}$	7
	RNA splicing	$3.32.10^{-4}$	6
	Apoptosis	$4.39.10^{-2}$	4
	Chromosome segregation	$4.91.10^{-4}$	4
ECEL1P1	Cell cycle	$3.57.10^{-4}$	3
	Apoptosis	$2.33.10^{-3}$	3
	Embryonic development	$1.66.10^{-2}$	3
	Chromatin organization	$2.01.10^{-2}$	3
	Chromosome organization	$4.42.10^{-2}$	3
HLA-DPB2	Cell cycle	$1.79.10^{-14}$	13
	M phase	$3.74.10^{-10}$	7
	Chromatin organization	$3.70.10^{-3}$	5
	Chromosome organization	$6.06.10^{-4}$	5
	Cytoskeleton organization	$1.77.10^{-4}$	5
	RNA splicing	$5.15.10^{-3}$	3
	Spindle organization	$3.91.10^{-7}$	3
LINC00458	Cell cycle	$1.12.10^{-9}$	45
	RNA processing	$4.41.10^{-10}$	36
	Regulation of apoptosis	$3.02.10^{-2}$	25
	Cytoskeleton organization	$4.54.10^{-3}$	23

	Response to DNA damage stimulus	$7.67.10^{-3}$	19
	RNA splicing	$6.35.10^{-8}$	18
	Regulation of the cell differentiation	$1.26.10^{-2}$	17
	Chordate embryonic development	$1.26.10^{-2}$	15
	Ribosome biogenesis	$1.80.10^{-2}$	14
	Mitosis	$3.27.10^{-3}$	14
	Histone modification	$1.57.10^{-2}$	8
	Chromating remodeling	$7.90.10^{-4}$	7
	Ribosomal small subunit biogenesis	$2.50.10^{-2}$	4
	Chromatin silencing	$2.20.10^{-2}$	3
	Spliceosome assembly	$7.73.10^{-3}$	3
LINC00678	Cell cycle	$4.58.10^{-8}$	20
	RNA processing	$1.02.10^{-3}$	18
	Chromosome organization	$1.51.10^{-3}$	10
	Chromatin modification	$5.23.10^{-4}$	10
	M phase	$1.37.10^{-2}$	10
	Post-Golgi vesicle-mediated transport	$1.51.10^{-3}$	3
	Ribosomal small subunit biogenesis	$4.84.10^{-2}$	3
LINC01108	Cell cycle	$9.86.10^{-9}$	7
	M phase	$8.52.10^{-5}$	6
	Apoptosis	$8.49.10^{-3}$	5
	Cytoskeleton organization	$1.60.10^{-3}$	3
LINC-ROR	Cell cycle	$6.94.10^{-8}$	6
	Regulation of cell differentiation	$2.63.10^{-3}$	6
	Apoptosis	$1.52.10^{-3}$	4

LNCPRESS1	Cell cycle	$9.58.10^{-7}$	18
	Regulation of apoptosis	$1.04.10^{-4}$	15
	Apoptosis	$1.75.10^{-2}$	12
	Mitosis	$1.20.10^{-3}$	9
	Ribosome biogenesis	$4.46.10^{-4}$	7
	Chromosome organization	$1.95.10^{-3}$	6
	RNA splicing	$1.79.10^{-2}$	6
	Chromatin organization	$3.76.10^{-3}$	5
	Nucleotide-excision repair, DNA gap filling	$4.49.10^{-2}$	3
LNCPRESS2	Cell cycle	$9.55.10^{-3}$	9
	Regulation of apoptosis	$2.13.10^{-4}$	5
	Chromatin organization	$1.25.10^{-2}$	3
NANOGP1	Apoptosis	$2.16.10^{-2}$	5
	Translation	$2.43.10^{-3}$	3
POU5F1P3	Cell cycle	$2.94.10^{-9}$	12
	M phase	$1.68.10^{-5}$	12
	Embryonic development	$3.38.10^{-5}$	12
	Chromosome organization	$4.78.10^{-4}$	11
	Apoptosis	$1.63.10^{-2}$	10
	RNA processing	$2.84.10^{-3}$	9
	RNA splicing	$1.45.10^{-2}$	5
	Response to DNA damage stimulus	$2.26.10^{-2}$	4
	Chromatin organization	$3.13.10^{-2}$	4
	Chromosome segregation	$4.66.10^{-3}$	3
	DNA repair	$2.03.10^{-2}$	3

	DNA recombination	$1.59.10^{-2}$	3
	Nucleotide excision repair	$3.08.10^{-3}$	3
	Stem cell maintenance	$1.85.10^{-3}$	3
POU5F1P4	Cell cycle	$5.80.10^{-7}$	12
	Embryonic development	$3.25.10^{-2}$	8
	Regulation of cell differentiation	$3.80.10^{-3}$	8
	Chromosome organization	$1.54.10^{-2}$	6
	Negative regulation of cell differentiation	$9.04.10^{-3}$	5
	DNA recombination	$2.19.10^{-3}$	4
	DNA repair	$7.72.10^{-6}$	4
	Regulation of apoptosis	$2.41.10^{-3}$	4
	Regulation of cytoskeleton organization	$3.81.10^{-2}$	4
	RNA processing	$1.81.10^{-2}$	4
	Response to DNA damage stimulus	$7.18.10^{-5}$	4
RP11-415G4.1	Response to DNA damage stimulus	$1.34.10^{-3}$	4
	Regulation of apoptosis	$1.26.10^{-2}$	4
	DNA repair	$2.36.10^{-3}$	3
	Endocytosis	$1.77.10^{-2}$	3
	Mitotic cell cycle	$3.19.10^{-2}$	3
RP11-69I8.2	Cell cycle	$2.63.10^{-7}$	31
	Chromosome organization	$4.96.10^{-9}$	30
	RNA processing	$1.29.10^{-7}$	26
	Chromatin organization	$5.85.10^{-7}$	22
	Embryonic development	$5.05.10^{-9}$	22
	M phase	$9.94.10^{-4}$	19

	Cytoskeleton organization	$9.44.10^{-3}$	17
	Chromatin modification	$3.74.10^{-6}$	15
	Ribosome organization	$1.07.10^{-2}$	12
	RNA splicing	$1.72.10^{-6}$	12
	Chromosome segregation	$4.47.10^{-2}$	6
	Chromosome condensation	$2.87.10^{-2}$	4
	Histone modification	$4.81.10^{-5}$	4
	Ribosome small subunit biogenesis	$1.86.10^{-2}$	4
	Histone methylation	$2.53.10^{-2}$	3
	RNA export from nucleus	$1.51.10^{-2}$	3
RP11-568A7.4	Cytoskeleton organization	$2.47.10^{-4}$	7
	Cell cycle	$1.15.10^{-4}$	5
	Endocytosis	$8.53.10^{-3}$	4
	Mitosis	$4.18.10^{-4}$	4
	Apoptosis	$3.86.10^{-3}$	3
	RNA processing	$1.63.10^{-2}$	3

* The values represent the number of protein coding genes directly correlated to the ncRNA and associated to the cited GO.

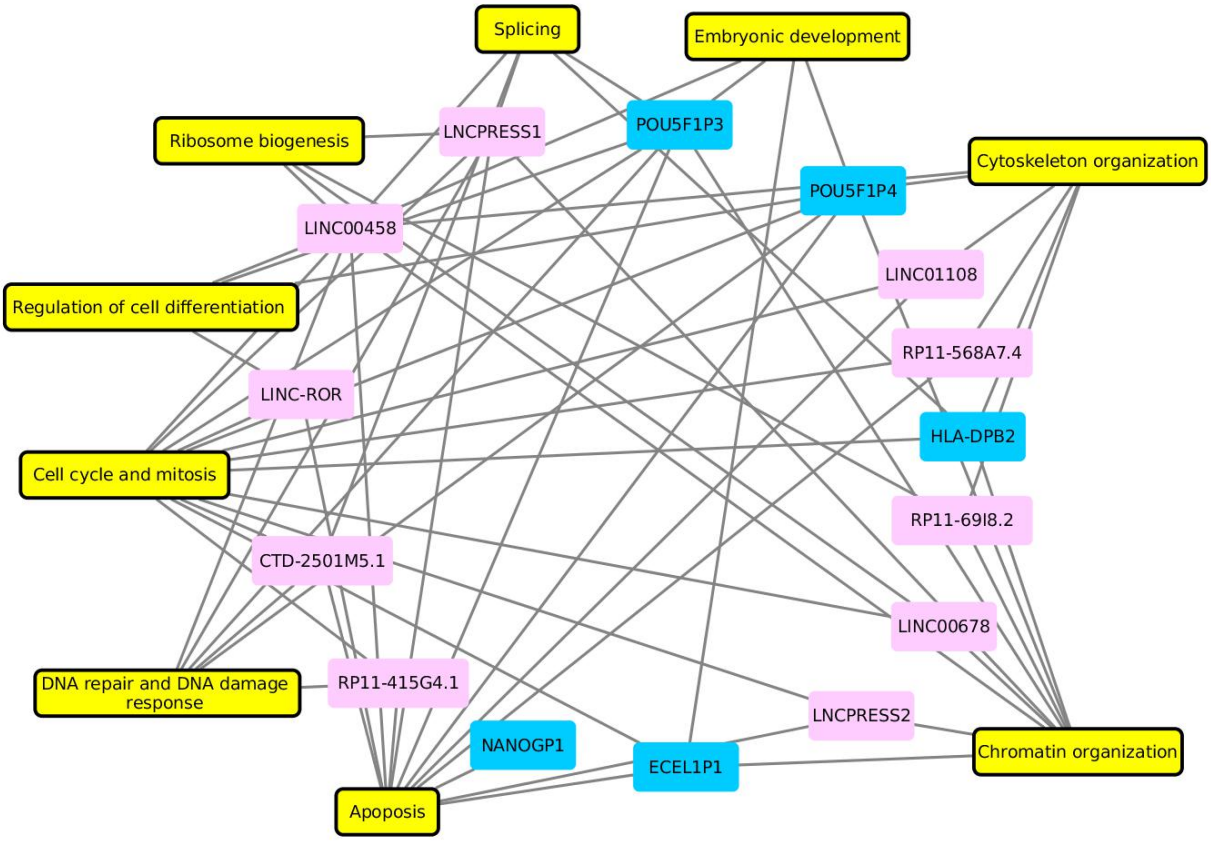


Figure 4: Biological processes categories for which the ncRNAs were classified. Yellow rectangles represent the biological processes categories, pink rectangles represent the lncRNAs and the blue rectangles represent the pseudogenes. Each RNA is connected with its categories by an edge. The categories “cell cycle” and “mitosis” were put together in a single node.

In addition, the chromosomal position of the genes coding for the mRNAs was investigated in order to verify the possibility of existence of cis regulatory sequences shared between the protein coding gene and the ncRNA gene. In the majority of the cases, however, the protein coding gene was placed in a chromosome different than the ncRNA gene (Supplemental Material 3). In the cases where both genes were located in the same chromosome, the distance between them were at least 5 Mb, except for seven cases, where the distance were of 1 Mb or less. These observations reinforces

that the occurrence of expression positive correlation is due to participation in a common cellular process not pervasive transcription.

Networks for each of the 10 ontology categories were constructed in order to best elucidate the role of the lncRNAs and pseudogenes in each biological process (Figure 5 and Supplemental material 4, Figure S1). In the ribosome biogenesis category, in addition to be correlated with mRNAs coding for proteins involved in the ribosome biogenesis pathway (Supplemental Material 4, Figure S2), the ncRNAs were positively correlated to RNAs that code for ribosome proteins, like RPL7A, RPL11, RPS6, RPS7, RPS16, RPS19 and RSL24D1, most of them part of the small subunit and upregulated in ESCs (Figure 5A and Figure 6). The ncRNAs associated with this category also showed positive correlation with some small nucleolar RNAs (Figura 5A), a RNA class associated with ribosome assembly.

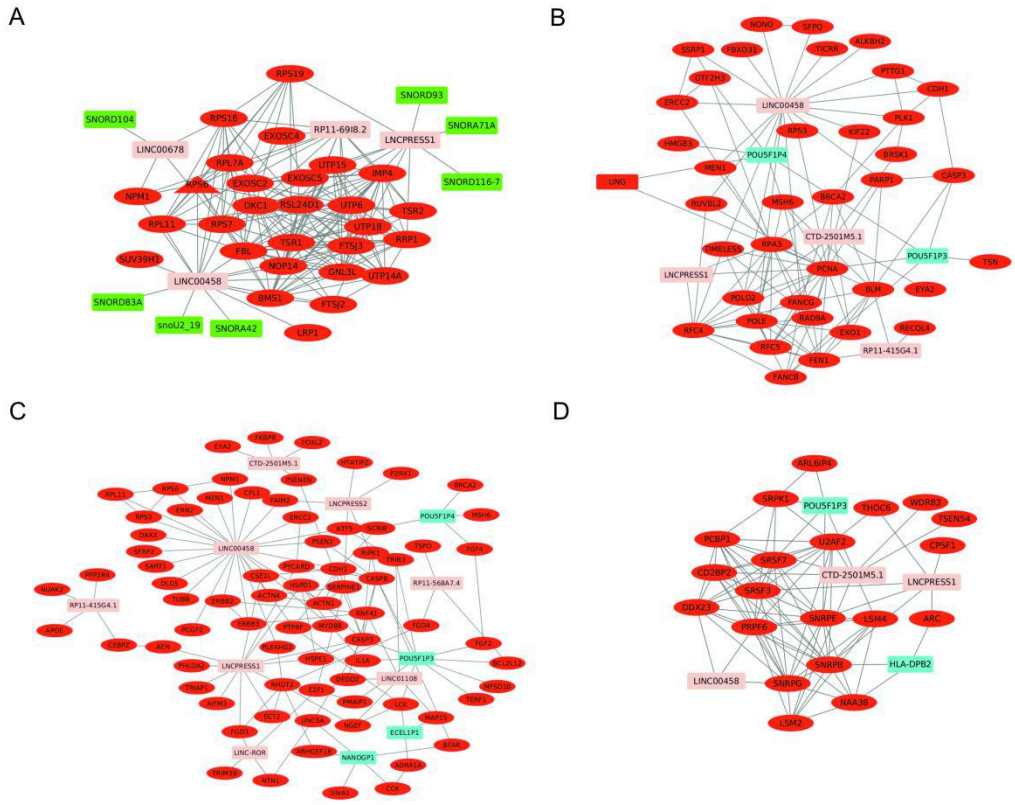


Figure 5: Networks representing some of the identified ontology categories. A, ribosome biogenesis. B, DNA repair and response to DNA damage stimulus. C, apoptosis. D, RNA splicing. Red elipses

represent protein-coding mRNAs, pink rectangles represent lncRNAs, blue rectangles represent pseudogenes and green rectangles represent snRNAs.

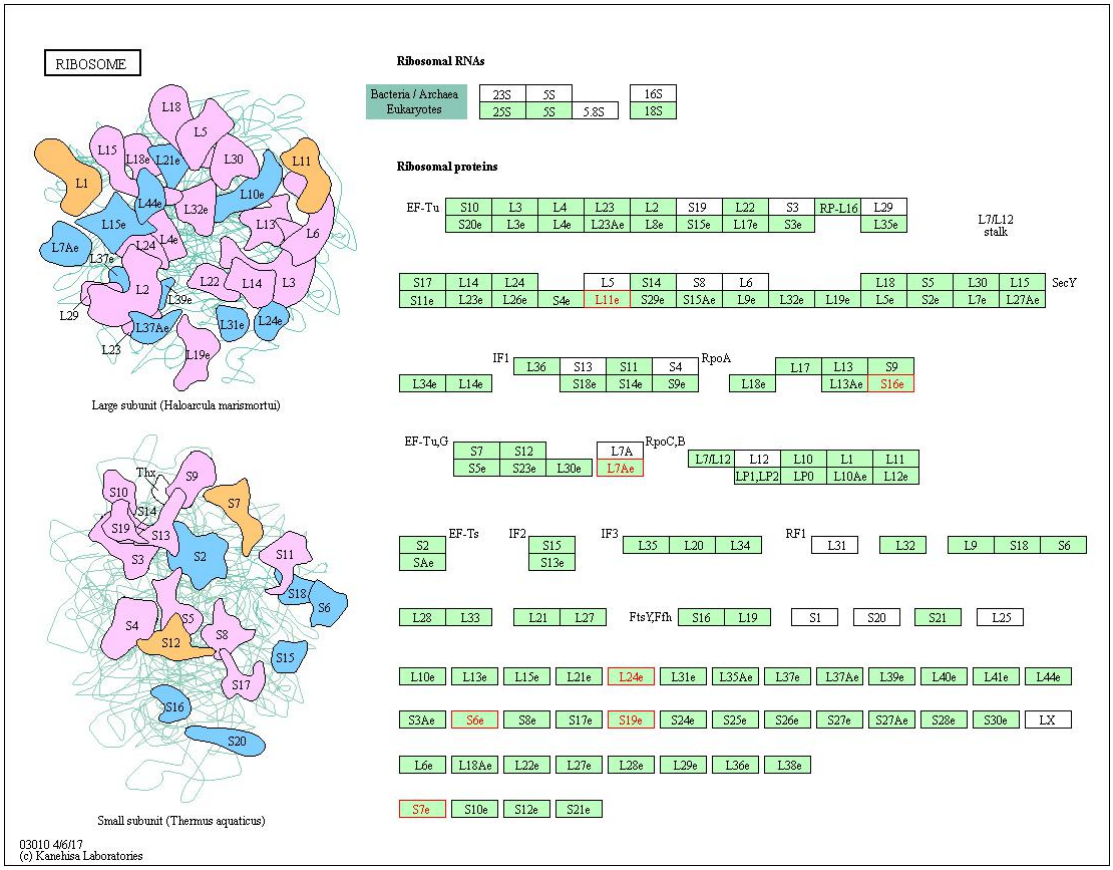


Figure 6: Proteins integrating the ribosome subunits. mRNAs coding for proteins highlighted in red are among the RNAs whose expression is positively correlated with the ribosome biogenesis-associated ncRNAs. Ilustration adapted from the KEGG PATHWAY database.

In the case of DNA repair and response to DNA damage stimulus category (Figure 5B), mRNAs coding for proteins involved in the base excision repair, nucleotide excision repair, mismatch repair and homologous recombination pathways were among the RNAs whose expression is positively correlated with the analyzed ncRNAs (Supplemental material 4, Figures S3, S4, S5 and S6).

The apoptosis network includes RNAs coding for proteins directly involved in the p53 signaling pathway, like PMAIP1, CASP3 and CASP8 (Figure 5C and Supplemental material, Figure S7). Besides, mRNAs coding for proteins associated with negative (n=18) and positive (n=28) regulation of apoptotic process are also present in the network.

In the RNA splicing category, most of the mRNAs positively correlated to the pseudogenes and lncRNAs code for spliceosome constituent proteins, like SRSF7, SRSF3, SNRPG, SNRPB, SNRPE, PRPF6 and DDX23 (Figure 5D and Supplemental Material, Figure S8).

The mitosis and cell cycle networks are very related to each other as many ncRNAs and correlated mRNAs are common to both categories. Among the mRNAs present in these two networks are those coding for cell cycle key proteins, as cyclins (CCNA1, CCNB2 and CCNE1), checkpoint related proteins (CHEK1, MAD1L1 and MAD2L1), transcription factors (E2F1 and FOXM1), and components of some complexes important for the mitosis progression, like chromosomal passenger complex (AURKC and INCENP) and the anaphase promoting complex/cyclosome, ANAPC11 (Supplemental material 4, Figures S9). At least three of the ncRNAs classified in the mitosis and cell cycle categories (CTD-2501M5.1, POU5F1P3, RP11-69I8.2) are positively correlated to chromosome segregation associated proteins.

Many ncRNAs associated with chromatin organization are related with nucleosome organization, since they were observed as positively correlated with mRNAs coding for histones (HIST1H2BL, HIST1H2BN, HIST1H2BA, HIST1H3J, HIST1H2AG, HIST2H2AB, H2AFZ, H3F3B), histone methyltransferases (SUV39H1, SUV39H2, EHMT1, CARM1, MEN1 and SETD1A), a histone deacetylase (HDAC7) and a histone demethylase (KDM4A), (Supplemental material 4, Figure S1).

In the cytoskeleton organization category, ncRNAs showed correlation to mRNAs coding for proteins related to actin (PFN1, MYH9 CCN2, DIAPH3, ACTN4, CAPZB and LIMA1) and microtubule (SASS6, TUBGCP2, TUBGCP6 and SPC25) cytoskeleton organization (Supplemental material 4, Figures S10).

The regulation of cell differenciation network comprises at least 18 nodes involved in the negative regulation of differentiation, including NANOG, FGF4, DMNT3D and NODAL. The network of embryonic development include mRNAs involved in the blastocist development (NODAL, NASP, FOXD3, BRCA2) and stem cell population maintenance (NODAL, NOTCH1, ZNF358, FOXD3, NANOG, FGF2, FGF4) (Supplemental material 4, Figure S1).

3.3. ncRNA protein binding prediction

The ncRNAs classified in the 10 categories were selected for a protein binding prediction analysis. As the used algorithm applies predictions of secondary structure, hydrogen bonding and van der Waals' contributions to estimate the binding propensity of RNA molecules, it was necessary to firstly investigate which isoforms of the selected ncRNAs were expressed in ESCs. It was observed that, for many lncRNAs and pseudogenes, more than one isoform is expressed in ESCs. Altogether, 16 of the studied ncRNAs presented more than one isoform and LINC00698 was the ncRNA with the largest number, with six detected transcripts. (Supplmental material 5).

In the analysis of potential binding of the ncRNAs to proteins, four ncRNAs were observed to have high protein bind potential: LINC00458, POU5F1P3, CTD-2501M5.1 and LNCPRESS1 Seven proteins were detected as possible targets ncRNAs: SRSF2, SRSF3, SRNPA, SSRP1 and PCBP1 (Figure 7).

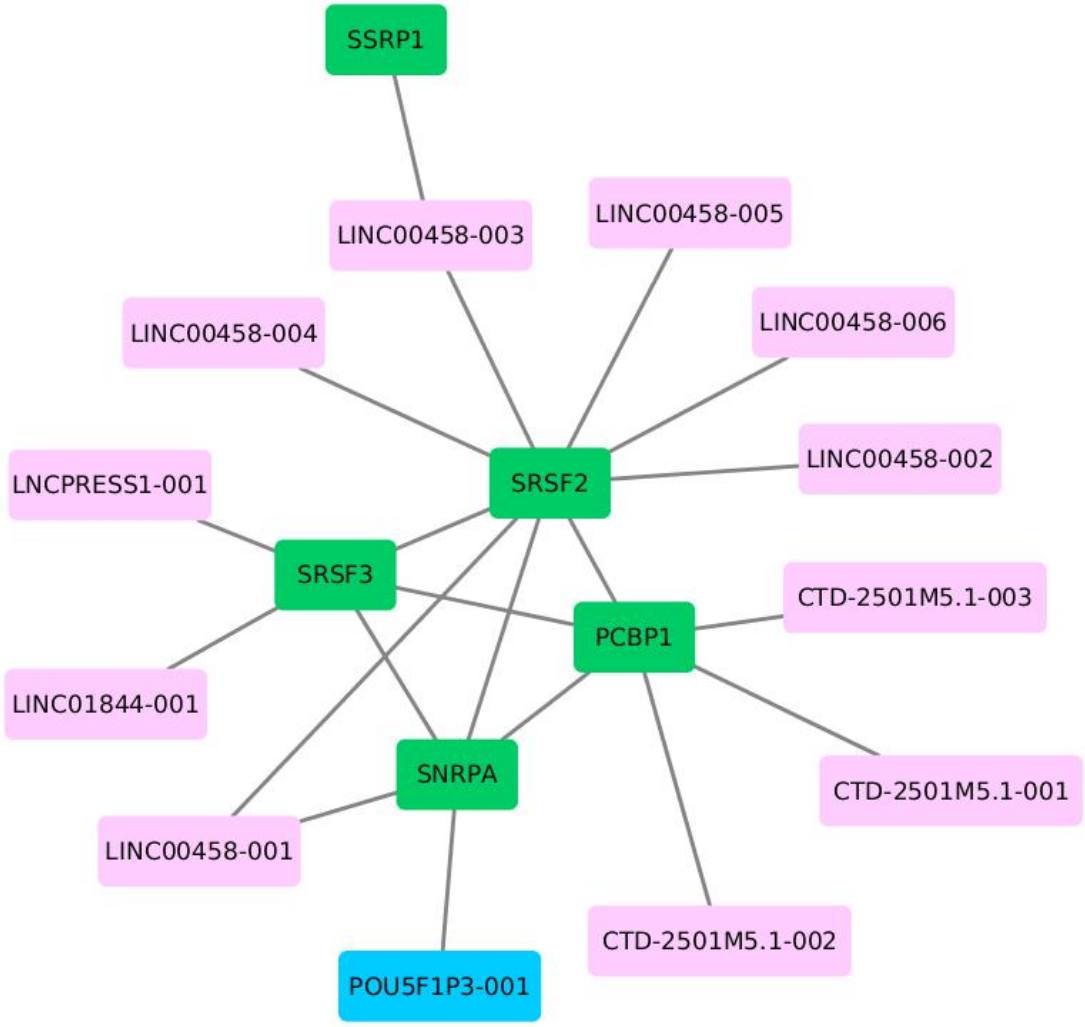


Figure 7: Proteins putatively bounding the ncRNAs in the study. Green rectangles are proteins, pink rectangles are lncRNAs and blue rectangles are pseudogenes.

4. Discussion

The human genome contains several segments transcribed into RNAs not translated into proteins and many of these RNAs have been shown to play important roles in the cells. However, a larger number of ncRNAs have their cellular functions barely known, even in the widely studied human ESCs. In this sense, the expression and potential function of lncRNAs and pseudogenes was

investigated in these cells and, to the best of our knowledge, this is the first large scale study on ncRNAs expressed in ESCs that includes pseudogenes.

Considering the ncRNAs expressed in ESCs, a total of 317 lncRNAS and 585 pseudogenes were identified as upregulated in these cells (main dataset). In addition, all lncRNAs previously detected in ESCs were observed in this study, except for FLJ11812 and lncRNA_ES2. A similar number of lncRNAs (n=300) was detected in a previous study, where 78 were reported as biasedly expressed in human ESCs[17]. These authors, however, investigated only intergenic lncRNAs (lincRNAs). On the other hand, we detected 87 lncRNAs that can be classified as ESC-associated lncRNAs, from which only 25 were among those observed on the previous study as ESC-specific. In addition, we also observed other classes of lncRNAs aside from lincRNAs. This differences may be due to the type of RNA-seq datasets applied by these authors, which have low replicates number and single end reads. Our data, on the other hand, have high number of replicates and paired end reads, which can confer higher confidence to the results. Regarding to pseudogenes, two RNAs previously reported as non-expressed in ESCs, NANOGP1 and NANOGP8, were detected in our analysis[26].

Many of the detected ncRNAs presented expression levels compared to those observed for important genes in the ESC context, like NANOG, which codes for a major ESC transcription factor. Considering this, is tempting to propose that these highly expressed ncRNAs may have an important function in these cells. In fact, the evaluation of potential function of selected ncRNAs revealed that some of these RNAs seem to be related to ESC-important processes whereas others may be associated to more basic and general cell functions.

At least four ncRNAs seems to be related to cell differentiation regulation. The pseudogenes POU5F1P3 and POU5F1P4 and the lncRNAs LINC00458 and LINC-ROR were observed to be

positively correlated to mRNAs coding for negative regulators of cell differentiation. The observation of LINC00458 and LINC-ROR as associated to this ontology is in agreement with previous studies that reported these two lncRNAs as acting in the embryonic stem state maintenance[11,12]. This result may indicate our method of function evaluation is effective in detecting the cell processes the ncRNAs are associated with. This presumption is also corroborated by the fact that many gene ontologies observed here, like mitosis, cell cycle, embryonic development and ribosome biogenesis, were also detected by Tang et al. (2013), in their study on ESCs lincRNAs[17].

Considering the biological characteristics of the ESCs, we supose some ncRNAs may be associated to ESCs-particular processes. ESCs proliferate intensively in culture and present particular cell cycle characteristics, like an abbreviated cycle due to a G1 phase shorter than the observed in differentiated cells[27,28]. Moreover, these cells present low frequency of mutations and elevated DNA repair capacity compared with their differentiated couterparts[29,30]. Levels of base excision and mismatch repair related mRNAs are higher in ESCs compared to differentiated cells, even when ESCs are not challenged with damage inducers[29,31]. In the case of nucleotide excision repair, there is not differential expression, but repair kinetics of ESCs is significantly faster[29]. In addition, homologous recombination is prevalent over non-homologous end joining in this cell type[32]. Considering the apoptotic process, ESCs seem to be primed to this mechanism. These cells present constitutively activated Bax in their cytoplasm, even when no programmed cell death stimulus is present[33]. However, activated Bax is not free, but stucked at the Golgi complex and the translocation to the mitochondria only occurs by the action of p53[33]. Moreover, the mitocondrias from these cells are also highly primed towards apoptosis when compared to differentiated cells[34].

In our guilty by association analysis, ncRNAs were observed associated to mRNAs coding for G1/S, S and M cyclins and proteins of complexes important for the mitosis progression. In the case of DNA repair-associated ncRNAs, there is positive correlation with proteins part of homologous recombination, mismatch repair, nucleotide and base excision repair, the exactly DNA repair pathways upregulated in ESCs. Regarding to ncRNAs related to apoptosis, they are correlated with mRNAs coding both pro- and anti-apoptitic proteins, in line with the apoptose primed condition of ESCs. Considering that, the ncRNAs associated to cell cycle, mitosis, DNA repair, DNA damage response and apoptosis are possibly associated to mechanisms particular to the ESCs.

In addition to these mechanisms, ncRNAs were also associated to basic cell processes, like ribosome biogenesis, RNA splicing, chromatin organization and cytoskeleton organization. However, as it is known that some genes have ESC-specific transcribed isoforms and the ESCs chromatin is enriched in bivalent promoters, i.e. silent, but poised to expression when the cells differentiate, it is not discarded that these ncRNAs may also play roles in ESC-especific mechanisms related to splicing and chromatin organization[35–38.]

Regarding to splicing associated ncRNAs, LINC00458 called our attention due to some characteristics similar to the lncRNA MALAT1, that was previously reported to act both in splicing and cell cycle. This lncRNA is enriched in cell nucleus, specifically in the nuclear speckles, subnuclear domains enriched in splicing factors, mainly due to its interaction with the proteins PRP6 and SON[39,40]. Moreover, MALAT1 physically interacts with splicing factors like SRSF1, SRSF2 and SRSF3 and modulate its localization to nuclear speckles[40]. In adition to its function in RNA splicing, MALAT1 is involved with cell cycle progression, specifically the G1/S transition and mitosis entering[41]. The role of MALAT1 in the cell cycle seems to be due to its regulation on

the transcription factor B-MYB alternative splicing, since depletion of MALAT1 resulted in alteration in the B-MYB mRNA and protein levels and aberrant splicing, together with reduction of FOXM1 transcript levels[41]. B-MYB, in turn, was previously reported to be important for the recruitment of FOXM1 to the promoter of mitotic genes[42].

In the case of LINC00458, this lncRNA is transcribed in four different isoforms and all of them potentially bind to at least one of the spliceosome associated proteins SRSF2, SSRP1 and SNRPA. LINC00458 was also associated to cell cycle and positively correlated to important transcription factors, like E2F1 and FOXM1, and the nuclear speckle-associated protein PRP6. In addition, previous studies in ESCs reported LINC00458 as predominantly located at the cell nucleus[11]. Therefore, considering the similarities of LINC00458 with MALAT1 and its previous reported role in stemness maintenance, we propose LINC00458 as an interesting target of further studies on ESC-specific splicing mechanisms and as starter point of studies to elucidate the roles played by ncRNAs em ESCs.

5. Conclusions

ESCs presents a large number of upregulated lncRNAs and pseudogenes in comparison to differentiated cells. The bioinfomatics analysis of a subgroup of these molecules revealed their putative association with biological processes like mitosis, cell cycle, embryo development, ribosome biogenesis, RNA splicing, chromatin organization, DNA repair and response to DNA damage stimulus, apoptosis, cytoskeleton organization and regulation of cell. In adition, some of the investigated RNAs presented high potential to bind to proteins associated with RNA splicing, apoptosis and DNA damage response.

The observations of this study shifts the attention to a poorly known part of the ESCs biology and provides directions to future experiments to elucidate the function of ncRNAs in this cellular context. ncRNAs expression modulation assays and the concomitant observation of the positively correlated mRNAs response to these alterations are the next step in the unraveling of the specific molecular mechanisms through which ncRNAs exert their roles in ESCs.

6. Acknowledgments

This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) under Grant 301149/2012-7.

7. Author Disclosure Statement

No competing financial interests exist.

8. References

1. Hattori M. Finishing the euchromatic sequence of the human genome. *Tanpakushitsu Kakusan Koso*. 2005;50(2):162-168.

2. Consortium EP, Dunham I, Kundaje A, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74.

3. Wang KC, Chang HY. Molecular Mechanisms of Long Noncoding RNAs. *Mol Cell*. 2011;43(6):904-914.

4. Consortium EP, Dunham I, Kundaje A, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74.

5. Podlaha O, Zhang J. Pseudogenes and Their Evolution. *Encycl Life Sci*. 2010;(November):1-8.

6. Li W, Yang W, Wang XJ. Pseudogenes: Pseudo or Real Functional Elements? *J Genet Genomics*. 2013;40(4):171-177.

7. Kandouz M, Bier A, Carystinos GD, Alaoui-Jamali M a, Batist G. Connexin43 pseudogene is expressed in tumor cells and inhibits growth. *Oncogene*. 2004;23(27):4763-4770.

8. Harrison PM, Zheng D, Zhang Z, Carriero N, Gerstein M. Transcribed processed pseudogenes in the human genome: An intermediate form of expressed retrosequence lacking protein-coding ability. *Nucleic Acids Res*. 2005;33(8):2374-2383.

9. Wen Y-Z, Zheng L-L, Qu L-H, Ayala FJ, Lun Z-R. Pseudogenes are not pseudo any more. *RNA Biol*. 2012;9(1):25-30.

10. Lu W, Han L, Su L, et al. A 3'UTR-associated RNA, FLJ11812 maintains stemness of human embryonic stem cells by targeting miR-4459. *Stem Cells Dev*. 2015;24(9):1133-1140.

11. Ng S-Y, Johnson R, Stanton LW. Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO J*. 2012;31(3):522-533.

12. Wang Y, Xu Z, Jiang J, et al. Endogenous miRNA Sponge lincRNA-RoR Regulates Oct4, Nanog, and Sox2 in Human Embryonic Stem Cell Self-Renewal. *Dev Cell*. 2013;25(1):69-80.

13. Loewer S, Cabili MN, Guttman M, et al. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat Genet*. 2010;42(12):1113-1117.

14. Jain AK, Xi Y, McCarthy R, et al. LncPRESS1 Is a p53-Regulated LncRNA that Safeguards Pluripotency by Disrupting SIRT6-Mediated De-acetylation of Histone H3K56. *Mol Cell*. 2016;64(5):967-981.

15. Hart AH, Hartley L, Ibrahim M, Robb L. Identification, Cloning and Expression Analysis of the Pluripotency Promoting Nanog Genes in Mouse and Human. *Dev Dyn*. 2004;230(1):187-198.

16. Poursani EM, Soltani BM, Mowla SJ. Differential expression of OCT4 pseudogenes in pluripotent and tumor cell lines. *Cell J*. 2016;18(1):28-36.

17. Tang X, Hou M, Ding Y, Li Z, Ren L, Gao G. Systematically profiling and annotating long intergenic non-coding RNAs in human embryonic stem cell. *From Asia Pacific Bioinforma Netw*. 2013:20-22.

18. Dobin A, Davis CA, Schlesinger F, et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21.

19. Anders S, Pyl PT, Huber W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31(2):166-169.

20. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):1-34.

21. Zhao Z, Bai J, Wu A, et al. Co-LncRNA: Investigating the lncRNA combinatorial effects in GO annotations and KEGG pathways based on human RNA-Seq data. *Database*. 2015;2015. doi:10.1093/database/bav082.

22. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015;43(D1):D447-D452.

23. Shannon P, Markiel A, Ozier O, et al. Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498-2504.

24. Maere S, Heymans K, Kuiper M. BiNGO: A Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. *Bioinformatics*. 2005;21(16):3448-3449.

25. Agostini F, Zanzoni A, Klus P, Marchese D, Cirillo D, Tartaglia GG. CatRAPID omics: A web server for large-scale prediction of protein-RNA interactions. *Bioinformatics*. 2013;29(22):2928-2930.

26. Ambady S, Malcuit C, Kashpur O, et al. Expression of NANOG and NANOGP8 in a variety of undifferentiated and differentiated human cells. *Int J Dev Biol*. 2011;54(11-12):1743-1754.

27. Amit M, Carpenter MK, Inokuma MS, et al. Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. *Dev Biol*. 2000;227(2):271-278.

28. Becker KA, Ghule PN, Therrien JA, et al. Self-renewal of human embryonic stem cells is supported by a shortened G1 cell cycle phase. *J Cell Physiol*. 2006;209(3):883-893.

29. Maynard S, Swistowska AM, Lee JW, et al. Human embryonic stem cells have enhanced repair of multiple forms of DNA damage. *Stem Cells*. 2008;26(1549-4918 (Electronic)):2266-2274.

30. Tichy ED. Mechanisms maintaining genomic integrity in embryonic stem cells and induced pluripotent stem cells. *Exp Biol Med (Maywood)*. 2011;236(9):987-996.

31. Seriola A, Spits C, Simard JP, et al. Huntington's and myotonic dystrophy hESCs: Down-regulated trinucleotide repeat instability and mismatch repair machinery expression bupon differentiation. *Hum Mol Genet*. 2011;20(1):176-185.

32. Adams BR, Golding SE, Rao RR, Valerie K. Dynamic dependence on ATR and ATM for double-strand break repair in human embryonic stem cells and neural descendants. *PLoS One*. 2010;5(4):e10001-e10001.

33. Dumitru R, Gama V, Fagan BM, et al. Human Embryonic Stem Cells Have Constitutively Active Bax at the Golgi and Are Primed to Undergo Rapid Apoptosis. *Mol Cell*. 2012;46(5):573-583.

34. Liu JC, Guan X, Ryan JA, et al. High mitochondrial priming sensitizes hESCs to DNA-damage-induced apoptosis. *Cell Stem Cell*. 2013;13(4):483-491.

35. Atlasi Y, Mowla SJ, Ziaee S a M, Gokhale PJ, Andrews PW. OCT4 spliced variants are differentially expressed in human pluripotent and nonpluripotent cells. *Stem Cells*. 2008;26:3068-3074.

36. Gabut M, Samavarchi-Tehrani P, Wang X, et al. An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell*. 2011;147(1):132-146.

37. Pan G, Tian S, Nie J, et al. Whole-Genome Analysis of Histone H3 Lysine 4 and Lysine 27 Methylation in Human Embryonic Stem Cells. *Cell Stem Cell*. 2007;1(3):299-312.

38. Zhao XD, Han X, Chew JL, et al. Whole-Genome Mapping of Histone H3 Lys4 and 27 Trimethylations Reveals Distinct Genomic Compartments in Human Embryonic Stem Cells. *Cell Stem Cell*. 2007;1(3):286-298.

39. Hutchinson JN, Ensminger AW, Clemson CM, Lynch CR, Lawrence JB, Chess A. A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics*. 2007;8:39.

40. Tripathi V, Ellis JD, Shen Z, et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell*. 2010;39(6):925-938.

41. Tripathi V, Shen Z, Chakraborty A, et al. Long Noncoding RNA MALAT1 Controls Cell Cycle Progression by Regulating the Expression of Oncogenic Transcription Factor B-MYB. *PLoS Genet*. 2013;9(3).

42. Down CF, Millour J, Lam EWF, Watson RJ. Binding of FoxM1 to G2/M gene promoters is dependent upon B-Myb. *Biochim Biophys Acta - Gene Regul Mech*. 2012;1819(8):855-862.

Supplemental Material 1 **(Adapted from the original Excel sheets)**

Table 1: 50 most upregulated lncRNAs in embryonic stem cells from the main dataset.

Ensembl Gene ID	Gene Name	Gene Type	log_2FC	p value	adjusted p value
ENSG00000232301	LNCPRESS1	antisense	9.66992288162666	3.66004666171958e-27	1.92723502013333e-26
ENSG00000225937	PCA3	antisense	3.26291243025736	0.000249657421655156	0.000506638541447119
ENSG00000234380	LINC01426	antisense	2.60935988414355	0.0166671376279974	0.0260514781153361
ENSG00000228962	HCG23	antisense	2.12838191311767	0.0237483732061736	0.035897962220637
ENSG00000248008	NRAV	antisense	1.30346435944361	2.59445565078132e-12	8.88454807648786e-12
ENSG00000233384	RP11-100E13.1	bidirectional promoter lncRNA	3.28428316511004	0.00317353017233328	0.00559242847358338
ENSG00000249152	LNCPRESS2	lincRNA	9.32757604765559	2.06868887786274e-19	8.99904466293286e-19
ENSG00000254934	LINC00678	lincRNA	8.98002898156571	1.6115487870555e-80	1.96637858725471e-79
ENSG00000280707	RP11-568A7.4	lincRNA	8.59488779574113	7.1249646207954e-28	3.81697038225121e-27
ENSG00000236673	RP11-69I8.2	lincRNA	8.52608710440864	1.43149417315855e-34	8.87433980291872e-34
ENSG00000253507	CTD-2501M5.1	lincRNA	8.23164610036708	4.48971394010088e-19	1.93171879789398e-18
ENSG00000258609	LINC-ROR	lincRNA	8.18003204809258	2.99924275887274e-15	1.14983297596234e-14
ENSG00000274090	RP11-415G4.1	lincRNA	7.38060314436838	3.64181663295141e-15	1.39176728346359e-14
ENSG00000234787	LINC00458	lincRNA	7.32870792592889	9.20594365453137e-42	6.47828603896031e-41
ENSG00000234261	RP11-146I2.1	lincRNA	7.30004497926441	8.3733339927582e-15	3.15857326168601e-14

ENSG00000253230	LINC00599	lincRNA	7.02403603187471	4.39431470757041e-23	2.10478456552133e-22
ENSG00000248131	LINC01194	lincRNA	6.96079775576223	1.46506648419699e-32	8.69928184069831e-32
ENSG00000237923	XXbac-BPG27H4.8	lincRNA	6.69601662532708	2.06834553533478e-12	7.10452539656545e-12
ENSG00000274652	RP11-457D13.4	lincRNA	6.64237075223624	3.56531904567749e-12	1.21455187771999e-11
ENSG00000227640	SOX21-AS1	lincRNA	6.56875484972313	3.00215882732538e-10	9.35010397079628e-10
ENSG00000226673	LINC01108	lincRNA	6.50466115793817	9.82550296443354e-20	4.32258586357699e-19
ENSG00000244342	LINC00698	lincRNA	6.29513391964392	1.19631519202551e-30	6.80287071578145e-30
ENSG00000231698	AP002856.5	lincRNA	6.2770930677899	7.65030799993716e-11	2.45284619809454e-10
ENSG00000241168	RP11-10O22.1	lincRNA	6.13316609191821	1.87682325107712e-17	7.68330401130519e-17
ENSG00000226506	AC007463.2	lincRNA	6.13264567655256	4.38526032816843e-09	1.28626949546672e-08
ENSG00000231566	RP5-1158E12.3	lincRNA	6.13094714751483	2.37686003444561e-10	7.43739663102234e-10
ENSG00000281383	CH507-513H4.5	lincRNA	6.1023509967848	1.28799740078705e-05	2.9905205952143e-05
ENSG00000269877	CTD-3022G6.1	lincRNA	5.82237494145826	2.67579651502381e-08	7.52541841589643e-08
ENSG00000229271	AC091493.2	lincRNA	5.77622312886619	3.45353193928336e-08	9.65534077245315e-08
ENSG00000233393	AP000688.29	lincRNA	5.70401155698609	5.5433639216066e-09	1.61811119943053e-08
ENSG00000249790	RP11-20D14.6	lincRNA	5.63505429702534	2.79354965377037e-138	6.13416162037055e-137
ENSG00000253252	RP11-10A14.6	lincRNA	5.62278089581703	8.30199879969647e-08	2.26255625413312e-07
ENSG00000225783	MIAT	lincRNA	5.60565675819002	3.06898006948017e-214	1.20587650094665e-212
ENSG00000227838	RP1-213J1P B.1	lincRNA	5.60332060805428	1.47938194744114e-11	4.90702756479561e-11
ENSG00000258649	CTD-2142D14.1	lincRNA	5.59256199075941	1.15620401748337e-08	3.3156836095966e-08
ENSG00000197462	AC005276.1	lincRNA	5.53386497139634	1.41803313860994e-07	3.80519543408052e-07
ENSG00000203635	AC144450.2	lincRNA	5.48881711374466	2.35705563451102e-08	6.65018409902825e-08
ENSG00000255666	RP11-867G2.6	lincRNA	5.4581080313201	2.82701207277417e-08	7.93975556513178e-08

ENSG00000248605	CTD-2306M5.1	lincRNA	5.44548027746923	7.8313774747783e-20	3.45367554438764e-19
ENSG00000228592	D21S2088E	lincRNA	5.43969255086136	5.3372596676314e-08	1.47376069909091e-07
ENSG00000262370	RP11-473M20.9	lincRNA	5.392479057566	2.72342432621498e-07	7.17602090861616e-07
ENSG00000267530	LINC01836	lincRNA	5.2793669456642	5.32587255093406e-07	1.37546777294233e-06
ENSG00000225077	LINC00337	lincRNA	5.24077072291107	8.62845621801421e-13	3.00839671575654e-12
ENSG00000261786	RP4-555D20.2	lincRNA	5.2291014211142	6.70190446748385e-07	1.71901480496687e-06
ENSG00000250993	RP11-404J23.1	lincRNA	5.20500735656894	7.42190254474316e-07	1.8969860030603e-06
ENSG00000273209	RP11-107N15.1	lincRNA	5.18811921805742	1.67003114050442e-07	4.46344084470413e-07
ENSG00000261757	AC005592.3	lincRNA	5.1509735342722	1.22796915536194e-17	5.05492295270779e-17
ENSG00000278041	RP5-984P4.6	lincRNA	5.1236985243595	1.24409515863112e-06	3.13080501885344e-06
ENSG00000235152	RP5-865N13.2	lincRNA	5.11432946671105	2.74801486089081e-07	7.23964670306653e-07
ENSG00000277200	RP11-74E22.8	lincRNA	4.98927148330316	3.19401527412296e-09	9.44579175171297e-09

Table 2: 50 most upregulated pseudogenes in embryonic stem cells from the main dataset

Ensembl Gene ID	Gene Name	Gene Type	log_2FC	p value	ajusted p value
ENSG00000204121	ECEL1P1	Unprocessed pseudogene	7.86408241161352	3.28814324845047e-14	1.21288356697249e-13
ENSG00000224557	HLA-DPB2	Transcribed unprocessed pseudogene	7.83112207679643	3.56824681183917e-17	1.44805362789178e-16
ENSG00000232293	PAICSP7	Processed pseudogene	7.19194770209906	4.31796123500801e-12	1.46513634256969e-11
ENSG00000179899	PHC1P1	Processed pseudogene	6.7883990197443	1.24094085731836e-35	7.86627856205766e-35
ENSG00000265735	RN7SL5P	miscRNA	6.23014201058318	2.1211433282628e-10	6.65957422638817e-10
ENSG00000237227	RP5-1155K23.4	Processed pseudogene	6.19438664233223	2.96022837369424e-09	8.76949526926731e-09
ENSG00000231001	CCNB1IP1P3	Processed pseudogene	5.99103861936496	6.72137533169675e-10	2.06129752353304e-09
ENSG00000213612	FAM220CP	Processed pseudogene	5.81113956220569	1.10834623862535e-15	4.30830016471369e-15
ENSG00000226539	AC012512.1	Processed pseudogene	5.80251426361997	1.49733402527702e-22	7.08872706877561e-22
ENSG00000235602	POU5F1P3	Processed pseudogene	5.78387120332948	3.00483226687036e-09	8.89840224722029e-09
ENSG00000233309	RPS6P12	Processed pseudogene	5.77035687108904	3.19381303813027e-09	9.44579175171297e-09
ENSG00000276170	AC124789.1	Processed transcript	5.76009977297457	3.7354961636718e-09	1.1004278305324e-08
ENSG00000233966	UBE2SP1	Processed pseudogene	5.65777300114222	6.35346628068112e-08	1.74433180019171e-07
ENSG00000224664	RPL36AP53	Processed pseudogene	5.63973478796982	7.70807900253316e-08	2.10596766455286e-07
ENSG00000207220	RNU1-57P	snRNA	5.60154526324832	8.67088836384439e-08	2.36052898829621e-07
ENSG00000225366	TDGF1P3	Transcribed processed pseudogene	5.44253335482373	1.84579723940816e-19	8.0465916503545e-19
ENSG00000270524	QTRT1P1	Processed pseudogene	5.43983160603056	9.11111589838937e-14	3.30588016795433e-13
ENSG00000251947	RN7SKP164	miscRNA	5.43254901619106	2.16516273613455e-07	5.74024794253534e-07

ENSG00000277103	RP11-520B13.8	Processed pseudogene	5.32969799067417	1.55494816817991e-46	1.17855542217759e-45
ENSG00000225728	RP6-191P20.3	Processed pseudogene	5.29239984394631	4.58201078819272e-07	1.1889995728846e-06
ENSG00000270354	RP11-547M24.1	Processed pseudogene	5.27695109982462	4.83539897058523e-07	1.25256205622808e-06
ENSG00000229150	CRYGEP	Transcribed unprocessed pseudogene	5.12602208042091	1.27716000606816e-06	3.21153858437718e-06
ENSG00000235001	EIF4A1P2	Processed pseudogene	5.12350174080167	1.23193423904502e-06	3.10187513674414e-06
ENSG00000176654	NANOGP1	Transcribed unprocessed pseudogene	5.1202326847204	1.36560349403265e-09	4.12409223891546e-09
ENSG00000227097	RPS28P7	Processed pseudogene	5.11345170423073	1.1166937549047e-09	3.38837324460071e-09
ENSG00000225573	RPL35P5	Processed pseudogene	5.08509579346198	1.50816771913459e-09	4.5453841946818e-09
ENSG00000255192	NANOGP8	Protein coding	5.08253565328283	2.56895469929503e-17	1.04668715799078e-16
ENSG00000244280	ECEL1P2	Transcribed unprocessed pseudogene	5.0674578204859	3.89554786502419e-07	1.01611622374796e-06
ENSG00000237158	RP11-472F14.4	Processed pseudogene	5.02569179264134	2.38534192636282e-06	5.87820260978103e-06
ENSG00000233762	AC007969.5	Processed pseudogene	5.01726701128388	2.32581260827961e-09	6.92778993664895e-09
ENSG00000188801	ZNF322P1	Processed pseudogene	4.98556686749698	1.13169339791208e-16	4.51791818420655e-16
ENSG00000213763	ACTBP2	Processed pseudogene	4.94393313639276	3.6998368723913e-06	8.9860757016461e-06
ENSG00000265260	RN7SL74P	miscRNA	4.89412914166562	9.49783846501121e-07	2.41022321877319e-06
ENSG00000159712	ANKRD18CP	Unprocessed pseudogene	4.89137605199483	1.21170582601285e-225	5.19903004349167e-224
ENSG00000231345	BEND3P1	Transcribed processed pseudogene	4.85782702585277	4.17196265009551e-06	1.00928298474416e-05
ENSG00000242477	CTD-2161E19.1	Processed pseudogene	4.80510577130715	5.36496949716478e-06	1.28645798649794e-05
ENSG00000216802	RP11-390P2.2	Processed pseudogene	4.78184906794015	6.24888172610063e-06	1.49118238778512e-05

ENSG00000223703	AC027612.4	Unprocessed pseudogene	4.75586676006984	6.91934925430349e-06	1.64444713480979e-05
ENSG00000204959	ARHGEF34P	Unprocessed pseudogene	4.75181723222866	2.02468024743675e-08	5.73322863925981e-08
ENSG00000249239	RP11-428L21.2	Processed pseudogene	4.6963718666409	3.18843561814103e-10	9.91890242030686e-10
ENSG00000268034	AC005795.1	Processed pseudogene	4.68286541367784	1.00208377508326e-05	2.34787883352602e-05
ENSG00000222375	RN7SKP127	miscRNA	4.65901838234247	3.88446600626061e-06	9.41895024776633e-06
ENSG00000227939	RPL3P2	Processed pseudogene	4.65646795018941	1.11968597358888e-05	2.6136678874481e-05
ENSG00000214182	PTMAP5	Transcribed processed pseudogene	4.64359520962292	2.76415282147476e-16	1.09228026052032e-15
ENSG00000248590	GLDCP1	Transcribed processed pseudogene	4.58845838424879	5.73702979857372e-06	1.37356360694167e-05
ENSG00000261963	CCDC92B	Unitary pseudogene	4.55536040151256	1.78067524436783e-05	4.08304832016141e-05
ENSG00000256663	RP11-424C20.2	Processed pseudogene	4.54592202359548	1.85576250114971e-05	4.24835989536146e-05
ENSG00000197847	SLC22A20	Transcribed unitary pseudogene	4.52550720476684	8.4524564669086e-06	1.99184186314593e-05
ENSG00000267102	RP11-686D22.7	Transcribed processed pseudogene	4.50863898782389	8.53786567591368e-06	2.01109789388208e-05
ENSG00000240327	RN7SL93P	miscRNA	4.48881166713518	2.61238084024887e-05	5.89858648750474e-05

Table 3: 50 most upregulated genes in embryonic stem cells from the main dataset. ncRNAs are highlightes in yellow

Gene ID	Associated Gene Name	log_2FC	pvalue	padj
ENSG00000240563	L1TD1	9,980616933	0	0
ENSG00000232301	LNCPRESS1	9,669922882	3,66005E-13	1,92724E-12
ENSG00000069482	GAL	9,614651401	1,5907E-06	7,14733E-06
ENSG00000129354	AP1M2	9,614477847	8,10147E-13	4,21964E-12
ENSG00000182866	LCK	9,415840678	9,57692E-06	4,21436E-05
ENSG00000162551	ALPL	9,403778439	1,60172E-11	8,11345E-11
ENSG00000164626	KCNK5	9,392900829	1,18948E-06	5,21398E-05
ENSG00000249152	LNCPRESS2	9,327576048	2,06869E-05	8,99904E-05
ENSG00000171246	NPTX1	9,290827847	2,88501E-05	1,24985E-05
ENSG00000039068	CDH1	9,264121231	6,41545E-81	9,20779E-80
ENSG00000277048	AL353644.10	9,133461799	0,001817052	0,007441411
ENSG00000121570	DPPA4	9,072972656	0	0
ENSG00000130182	ZSCAN10	9,034770028	2,26033E-09	1,09243E-08
ENSG00000164265	SCGB3A2	9,007041565	3,27553E-09	1,57492E-08
ENSG00000254934	LINC00678	8,980028982	1,61155E-67	1,96638E-65
ENSG00000133980	VRTN	8,966216587	1,55202E-49	1,53658E-48
ENSG00000226792	LINC00371	8,886388097	0,000981733	0,004053558
ENSG00000265992	ESRG	8,78345927	4,7018E-63	5,44215E-62
ENSG00000111319	SCNN1A	8,687168422	1,63552E-14	8,87533E-14
ENSG00000179059	ZFP42	8,681352072	9,9259E-132	2,3266E-130
ENSG00000280707	RP11-568A7.4	8,594887796	7,12496E-15	3,81697E-13
ENSG00000166105	GLB1L3	8,573354335	7,55554E-07	3,42121E-06

ENSG00000131203	IDO1	8,543103271	1,61724E-13	8,58635E-13
ENSG00000236673	RP11-69I8.2	8,526087104	1,43149E-20	8,87434E-20
ENSG00000167600	CYP2S1	8,525363448	2,50038E-20	1,5448E-19
ENSG00000111704	NANOG	8,514395971	2,75116E-73	3,62806E-72
ENSG00000163735	CXCL5	8,458685705	6,235E-13	3,26415E-12
ENSG00000198286	CARD11	8,333132509	1,29569E-05	5,67727E-05
ENSG00000043355	ZIC2	8,301176868	7,34987E-12	3,75568E-12
ENSG00000156959	LHFPL4	8,287475518	0,128202144	0,497451742
ENSG00000253507	CTD-2501M5.1	8,2316461	4,48971E-05	0,000193172
ENSG00000179046	TRIML2	8,187486872	1,49231E-38	1,25133E-37
ENSG00000241186	TDGF1	8,186408192	0	0
ENSG00000258609	LINC-ROR	8,180032048	0,299924276	1,149832976
ENSG00000067840	PDZD4	8,027746656	3,6201E-43	3,29349E-42
ENSG00000088836	SLC4A11	8,022412622	1,001998777	3,766666367
ENSG00000164076	CAMKV	8,013633384	5,22358E-10	2,56443E-09
ENSG00000230798	FOXD3-AS1	7,985679468	7,89799E-10	3,8594E-09
ENSG00000147676	MAL2	7,94376526	8,21207E-29	5,86612E-28
ENSG00000148357	HMCN2	7,8803099	3,02649094	11,18010885
ENSG00000205358	MT1H	7,871593301	0,002288143	0,00934373
ENSG00000204121	ECEL1P1	7,864082412	3,288143248	12,12883567
ENSG00000243004	AC005062.2	7,843155626	9,84467E-28	6,92478E-28
ENSG00000101194	SLC17A9	7,83777035	4,035257648	14,84443812
ENSG00000224557	HLA-DPB2	7,831122077	0,003568247	0,014480536
ENSG00000279560	AC007326.1	7,782175929	0,662044289	24,16124971
ENSG00000149418	ST14	7,768597673	0,006918156	0,002780547
ENSG00000243709	LEFTY1	7,731011253	9,13906E-64	1,07807E-62

ENSG00000141738	GRB7	7,722732728	0,011232262	0,004485761
ENSG00000160867	FGFR4	7,612559823	0,036378102	0,14314507

Table 4 : ncRNAs selected for the function evaluation analysis.

Gene ID	Gene Name	Gene type	log_2FC
ENSG00000253507	CTD-2501M5.1	lincRNA	8.23164610036708
ENSG00000204121	ECEL1P1	unprocessed pseudogene	7.86408241161352
ENSG00000224557	HLA-DPB2	transcribed unprocessed pseudogene	7.83112207679643
ENSG00000234787	LINC00458	lincRNA	7.32870792592889
ENSG00000254934	LINC00678	lincRNA	8.98002898156571
ENSG00000226673	LINC01108	lincRNA	6.50466115793817
ENSG00000258609	LINC-ROR	lincRNA	8.18003204809258
ENSG00000232301	LNCPRESS1	antisense	9.66992288162666
ENSG00000249152	LNCPRESS2	lincRNA	9.32757604765559
ENSG00000176654	NANOGP1	transcribed unprocessed pseudogene	5.1202326847204
ENSG00000235602	POU5F1P3	processed pseudogene	5.78387120332948
ENSG00000237872	POU5F1P4	processed pseudogene	3.33709498816713
ENSG00000274090	RP11-415G4.1	lincRNA	7.38060314436838
ENSG00000280707	RP11-568A7.4	lincRNA	8.59488779574113
ENSG00000236673	RP11-69I8.2	lincRNA	8.52608710440864

Table 5: lncRNAs commonly expressed in the four analyzed datasets. Information on log2FC, p value and adjusted p value are from the main dataset.

Ensembl Gene ID	Gene Name	log2FC	p value	Adjusted p value
ENSG00000232301	LNCPRESS1	9.66992288162666	3.66004666171958e-27	1.92723502013333e-26
ENSG00000249152	LNCPRESS2	9.32757604765559	2.06868887786274e-19	8.99904466293286e-19
ENSG00000254934	LINC00678	8.98002898156571	1.6115487870555e-80	1.96637858725471e-79
ENSG00000280707	RP11-568A7.4	8.59488779574113	7.1249646207954e-28	3.81697038225121e-27
ENSG00000236673	RP11-69I8.2	8.52608710440864	1.43149417315855e-34	8.87433980291872e-34
ENSG00000253507	CTD-2501M5.1	8.23164610036708	4.48971394010088e-19	1.93171879789398e-18
ENSG00000258609	LINC-ROR	8.18003204809258	2.99924275887274e-15	1.14983297596234e-14
ENSG00000274090	RP11-415G4.1	7.38060314436838	3.64181663295141e-15	1.39176728346359e-14
ENSG00000234787	LINC00458	7.32870792592889	9.20594365453137e-42	6.47828603896031e-41
ENSG00000234261	RP11-146I2.1	7.30004497926441	8.3733339927582e-15	3.15857326168601e-14
ENSG00000248131	LINC01194	6.96079775576223	1.46506648419699e-32	8.69928184069831e-32
ENSG00000237923	XXbac-BPG27H4.8	6.69601662532708	2.06834553533478e-12	7.10452539656545e-12
ENSG00000274652	RP11-457D13.4	6.64237075223624	3.56531904567749e-12	1.21455187771999e-11
ENSG00000226673	LINC01108	6.50466115793817	9.82550296443354e-20	4.32258586357699e-19
ENSG00000244342	LINC00698	6.29513391964392	1.19631519202551e-30	6.80287071578145e-30
ENSG00000231698	AP002856.5	6.2770930677899	7.65030799993716e-11	2.45284619809454e-10
ENSG00000241168	RP11-10O22.1	6.13316609191821	1.87682325107712e-17	7.68330401130519e-17
ENSG00000226506	AC007463.2	6.13264567655256	4.38526032816843e-09	1.28626949546672e-08
ENSG00000229271	AC091493.2	5.77622312886619	3.45353193928336e-08	9.65534077245315e-08
ENSG00000233393	AP000688.29	5.70401155698609	5.5433639216066e-09	1.61811119943053e-08

ENSG00000249790	RP11-20D14.6	5.63505429702534	2.79354965377037e-138	6.13416162037055e-137
ENSG00000227838	RP1-213J1P__B.1	5.60332060805428	1.47938194744114e-11	4.90702756479561e-11
ENSG00000258649	CTD-2142D14.1	5.59256199075941	1.15620401748337e-08	3.3156836095966e-08
ENSG00000197462	AC005276.1	5.53386497139634	1.41803313860994e-07	3.80519543408052e-07
ENSG00000203635	AC144450.2	5.48881711374466	2.35705563451102e-08	6.65018409902825e-08
ENSG00000255666	RP11-867G2.6	5.4581080313201	2.82701207277417e-08	7.93975556513178e-08
ENSG00000248605	CTD-2306M5.1	5.44548027746923	7.8313774747783e-20	3.45367554438764e-19
ENSG00000225077	LINC00337	5.24077072291107	8.62845621801421e-13	3.00839671575654e-12
ENSG00000250993	RP11-404J23.1	5.20500735656894	7.42190254474316e-07	1.8969860030603e-06
ENSG00000278041	RP5-984P4.6	5.1236985243595	1.24409515863112e-06	3.13080501885344e-06
ENSG00000235152	RP5-865N13.2	5.11432946671105	2.74801486089081e-07	7.23964670306653e-07
ENSG00000236714	LINC01844	4.93640782802276	1.75688211029079e-53	1.49545658668444e-52
ENSG00000250889	LINC01336	4.70396454565071	3.10057752942985e-06	7.57161450085585e-06
ENSG00000253686	LINC01484	4.60251278229943	5.62932559149013e-10	1.73290034140978e-09
ENSG00000236094	LINC00545	4.59517403225715	1.39771798758928e-05	3.23446404895573e-05
ENSG00000269842	CTC-339O9.1	4.55879164757735	1.6807221427673e-05	3.86145084475358e-05
ENSG00000256969	RP11-320N7.2	4.52585411445475	2.03195742047498e-05	4.63579329886452e-05
ENSG00000260834	RP11-256I9.2	4.50702516487852	4.63443460621265e-22	2.16485777914243e-21
ENSG00000260362	RP11-297M9.1	4.49566762639093	1.65892191837915e-09	4.98592769504014e-09
ENSG00000257086	RP11-783K16.13	4.31182688844693	2.18560905463858e-05	4.97105787489962e-05
ENSG00000251027	LINC01950	4.22679624594743	1.0092694003712e-06	2.55601213533859e-06
ENSG00000267055	RP11-486P11.1	4.17049052265183	4.45744157508285e-05	9.82147124027298e-05
ENSG00000250899	RP11-253E3.3	4.11841858513276	7.18495428689475e-106	1.15170902000828e-104

ENSG00000234273	AC073071.1	4.05077321445091	8.38073827718507e-05	0.000179206199231596
ENSG00000241743	XACT	3.98602130995229	4.18446246406558e-06	1.01223194826396e-05
ENSG00000230623	RP11-469A15.2	3.92993367075115	8.41944891217926e-23	4.00804557341416e-22
ENSG00000236481	LINC02195	3.9226466492599	0.000358467651572773	0.00071492094790827
ENSG00000250994	AC005355.1	3.86602362633577	0.00034005357907739	0.000679940567301588
ENSG00000236605	AC023115.4	3.86047278984303	0.000483345604339297	0.000949305303632285
ENSG00000250519	RP11-680H20.2	3.83608257384845	0.000214872422156536	0.000439049707650514
ENSG00000250046	RP11-148B6.2	3.78930756200017	1.0152894855842e-16	4.06265101510973e-16
ENSG00000266402	SNHG25	3.69713789953123	0.000674736596338995	0.00130232897934982
ENSG00000263571	CTC-304I17.5	3.68236936959442	0.000396285193707533	0.00078669164305316
ENSG00000229546	LINC00428	3.6395005744131	0.000470135958033571	0.000924584671548043
ENSG00000224506	RP1-293L8.2	3.63186059305498	2.42169212444725e-06	5.9641794574937e-06
ENSG00000229646	RP11-330A16.1	3.61062318947447	0.00113142258907898	0.00212184551741567
ENSG00000215866	LINC01356	3.57107236881295	4.49187937814619e-10	1.38877800828737e-09
ENSG00000231689	LINC01090	3.55892721526292	5.40668989987426e-15	2.05421501757848e-14
ENSG00000263082	LINC02187	3.54491219333261	0.000702965657341104	0.00135489254728184
ENSG00000258433	RP11-255M2.1	3.46947732075184	0.00181573961165814	0.0033147450408063
ENSG00000236372	RP5-865N13.1	3.31873802002789	0.00017226510335097	0.000355932427679293
ENSG00000238062	SPATA3-AS1	3.2669655506461	0.000235721724717378	0.00047934145578077
ENSG00000235532	LINC00402	3.16969677983312	0.00471947684499018	0.00811270836943727
ENSG00000232650	LINC01780	3.1681411904716	0.00285641317161219	0.00507026925139882
ENSG00000261175	LINC02188	3.13892507750009	1.03215058597306e-10	3.29054685932362e-10
ENSG00000250564	RP11-215P8.4	3.13527799801045	0.00337472976246327	0.00592810825284864

ENSG00000237087	AC068134.6	3.12949516809452	0.0052509774628577	0.00896074611253422
ENSG00000259203	RP11-209K10.2	3.08568793400865	0.00362001528940345	0.00633613855516183
ENSG00000230923	LINC00309	3.08422681919391	0.00366324388180823	0.00640768410718393
ENSG00000253170	RP11-431J17.1	2.87203072345707	0.00852526205858384	0.0140582221472087
ENSG00000280511	RP11-35J1.2	2.81388784430033	3.19998771384283e-08	8.96644630845242e-08
ENSG00000228648	RP11-568A7.2	2.7289551334825	0.0114260466680304	0.0184111367846688
ENSG00000267886	CTD-2291D10.4	2.7088826652323	7.58651511626577e-31	4.33143716692271e-30
ENSG00000260066	CTD-2587M23.1	2.62047496979668	0.0222326883047192	0.0337820158465118
ENSG00000231510	RP11-542C10.1	2.59142850401942	0.0268829135364915	0.040252253532972
ENSG00000266651	RP11-138I1.3	2.53554808129609	3.05419925511986e-15	1.17021490352234e-14
ENSG00000248927	CTD-2334D19.1	2.46609952821642	0.00016378219196658	0.000339328434777459
ENSG00000256288	RP11-277P12.10	2.41837691631284	0.00871112423130429	0.014345149707512
ENSG00000272524	RP11-254F7.4	2.38062632950798	0.0108130636906067	0.017498545652262
ENSG00000225218	AP001628.6	2.34449885496519	0.00121766096524519	0.00227545732485468
ENSG00000236989	AC142119.1	2.27284792210048	0.00516064588306705	0.00881950221783237
ENSG00000235215	RP11-145M4.2	2.04563089418597	0.0332363020971628	0.0488447086180826
ENSG00000281881	SPRY4-IT1	2.02167272231316	0.000390643757918068	0.000776247564723757
ENSG00000254277	RP11-1144P22.1	10.9181451865458	2.59297076307388e-26	1.33901125029082e-25
ENSG00000254339	RP11-267L5.1	10.3078417115964	5.0277242164638e-52	4.17723285282382e-51
ENSG00000233056	ERVH48-1	1.49519156520578	1.44590382795934e-30	8.21072205052667e-30
ENSG00000281566	RP11-485F13.1	1.1894309879669	7.45027219663493e-19	3.18494747007756e-18

Table 6: Pseudogenes commonly expressed in the four analized datasets. Information on log2FC, p value and adjusted p value are from the main dataset.

Ensembl Gene ID	Gene Name	log2FC	p value	Adjusted p value
ENSG00000204121	ECEL1P1	7.86408241161352	3.28814324845047e-14	1.21288356697249e-13
ENSG00000224557	HLA-DPB2	7.83112207679643	3.56824681183917e-17	1.44805362789178e-16
ENSG00000232293	PAICSP7	7.19194770209906	4.31796123500801e-12	1.46513634256969e-11
ENSG00000179899	PHC1P1	6.7883990197443	1.24094085731836e-35	7.86627856205766e-35
ENSG00000237227	RP5-1155K23.4	6.19438664233223	2.96022837369424e-09	8.76949526926731e-09
ENSG00000213612	FAM220CP	5.81113956220569	1.10834623862535e-15	4.30830016471369e-15
ENSG00000235602	POU5F1P3	5.78387120332948	3.00483226687036e-09	8.89840224722029e-09
ENSG00000224664	RPL36AP53	5.63973478796982	7.70807900253316e-08	2.10596766455286e-07
ENSG00000225366	TDGF1P3	5.44253335482373	1.84579723940816e-19	8.0465916503545e-19
ENSG00000277103	RP11-520B13.8	5.32969799067417	1.55494816817991e-46	1.17855542217759e-45
ENSG00000270354	RP11-547M24.1	5.27695109982462	4.83539897058523e-07	1.25256205622808e-06
ENSG00000229150	CRYGEP	5.12602208042091	1.27716000606816e-06	3.21153858437718e-06
ENSG00000176654	NANOGP1	5.1202326847204	1.36560349403265e-09	4.12409223891546e-09
ENSG00000255192	NANOGP8	5.08253565328283	2.56895469929503e-17	1.04668715799078e-16
ENSG00000244280	ECEL1P2	5.0674578204859	3.89554786502419e-07	1.01611622374796e-06
ENSG00000159712	ANKRD18CP	4.89137605199483	1.21170582601285e-225	5.19903004349167e-224
ENSG00000204959	ARHGEF34P	4.75181723222866	2.02468024743675e-08	5.73322863925981e-08
ENSG00000197847	SLC22A20	4.52550720476684	8.4524564669086e-06	1.99184186314593e-05
ENSG00000277883	NLRP3P1	4.37850136532689	1.0153598605451e-18	4.3202410954546e-18

ENSG00000275585	CH17-118O6.3	4.34843014623256	1.99487040006802e-05	4.55468207139508e-05
ENSG00000264164	RP11-640N20.6	4.33765105043286	2.04241188681723e-05	4.65866803551214e-05
ENSG00000226361	TERF1P5	4.30746894196695	2.99762074341093e-05	6.72888165226548e-05
ENSG00000228038	VN1R51P	4.24479093508731	7.60744620038334e-07	1.94349685732663e-06
ENSG00000185275	CD24P4	4.1399423576947	1.55668140354089e-06	3.88955931959116e-06
ENSG00000260459	FTLP14	3.97030177459403	0.000235016518659508	0.000478056277552347
ENSG00000224039	CDYLP1	3.96075671017873	5.3444820327136e-06	1.28182773124845e-05
ENSG00000229670	PKP4P1	3.89406001897807	0.000161132499155186	0.000334220452446648
ENSG00000237398	HLA-DPA3	3.83466676487445	0.000215466462372864	0.000440180875267394
ENSG00000228649	AC005682.5	3.79781096591682	1.73075142625901e-28	9.38745773591515e-28
ENSG00000197744	PTMAP2	3.54042496108463	9.44051056314008e-22	4.3729827054568e-21
ENSG00000229132	EIF4A1P10	3.53911333595371	2.93733360655624e-20	1.30952781159577e-19
ENSG00000257150	PGAM1P5	3.46476525559411	6.09364451331039e-06	1.45541423925544e-05
ENSG00000260860	RP11-615I2.3	3.46393325215828	0.00103346146951156	0.00194908934918797
ENSG00000233834	AC005083.1	3.35076085431944	1.08490067142029e-06	2.74265846416965e-06
ENSG00000237872	POU5F1P4	3.33709498816713	1.07913297413687e-07	2.91929162150099e-07
ENSG00000251271	ALG1L7P	3.27145405360878	0.00191103022682522	0.00347839673149547
ENSG00000262943	ALOX12P2	2.81899300931399	3.13126007360727e-50	2.53262085517524e-49
ENSG00000228519	RP11-298C3.2	2.74917043361241	0.000553967984116206	0.00108092282881334
ENSG00000234232	RP11-353N4.5	2.68265289174379	0.0128336343659133	0.0204961520819664
ENSG00000184844	CYCSP45	2.67579676322105	0.0140992721505561	0.0223599705794336
ENSG00000236253	SLC25A3P1	2.42942783324207	0.0289633955702361	0.0431202257960673
ENSG00000226668	RP11-526D8.7	2.37125178010505	0.030319867621196	0.0449531997260723

ENSG00000229020	AKR7A2P1	2.34724023773118	2.33287041730447e-09	6.94817886215685e-09
ENSG00000215086	NPM1P24	2.19978062668156	6.4024753951667e-06	1.52671895569257e-05
ENSG00000250071	CTD-2060C23.1	2.05210147552201	0.0127742504838565	0.020406302680963
ENSG00000239218	RPS20P22	1.91830158468737	1.84793941267911e-21	8.51162613948346e-21
ENSG00000215156	RP11-1023L17.2	1.87565645950208	3.32732128470582e-23	1.59934777842212e-22
ENSG00000229259	LRRC37A12P	1.86598083189539	0.0134636875180948	0.0214185785549735
ENSG00000235363	SNRPGP10	1.74430867004134	3.95152510634303e-14	1.45396887280046e-13
ENSG00000259151	CAP2P1	1.34971477731055	4.22476632969737e-06	1.02145183411535e-05
ENSG00000229298	TUBB8P1	1.31321980100883	0.00359005493372604	0.00628774734195262

Supplemental material 2

This supplemental material is too large and it was not included in this thesis. Please consult the original Excel sheets.

Supplemental material 3 **(Adapted from the original Excel sheets)**

Table 1: Chromosome location evaluation of the genes whose mRNAs are positively correlated with ncRNAs. Only the genes located in the same chromosome as the ncRNA they correlate with are shown. Each ncRNA and its respective correlated genes are separated by different color stripes. For further information, see the original Excel sheets.

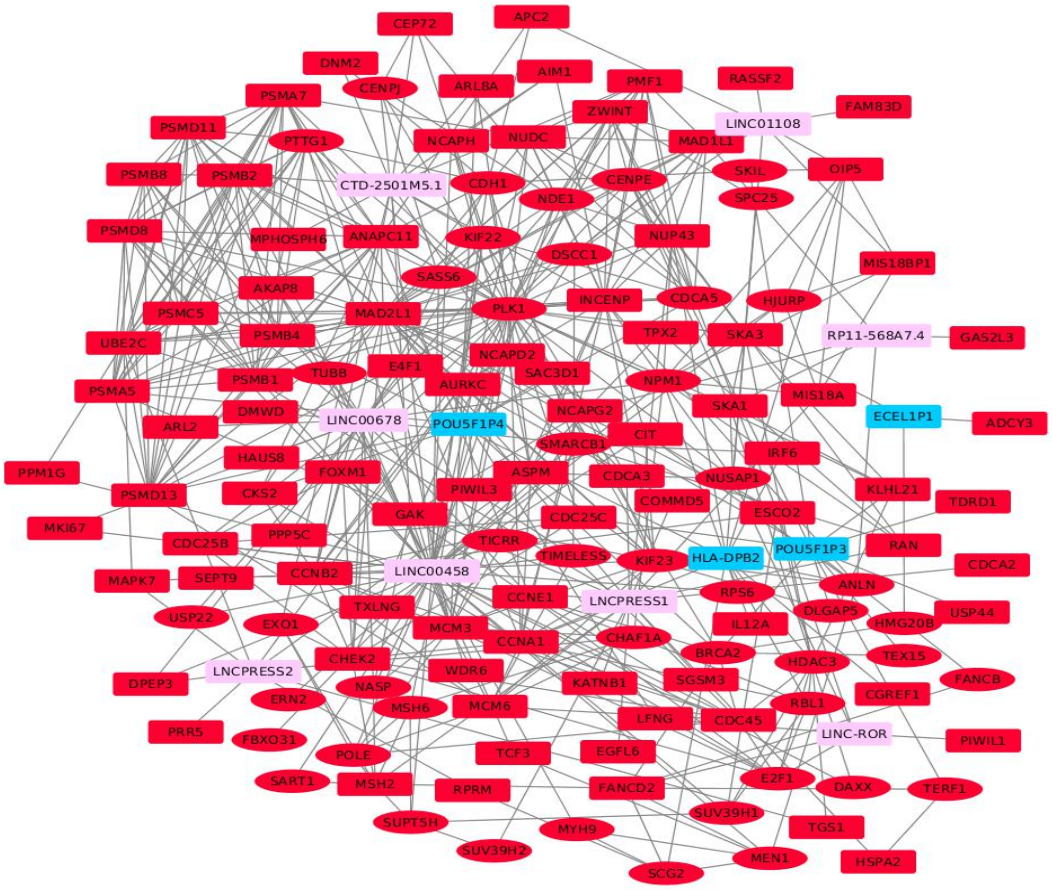
Gene ID	Gene name	Chromosome	Gene Start (bp)	Gene End (bp)
ENSG00000253507	**CTD-2501M5.1**	**8**	**131308545**	**131317632**
ENSG00000077782	FGFR1	8	38411138	38468834
ENSG00000168077	SCARA3	8	27633868	27676776
ENSG00000204121	ECEL1P1	2	232415818	232419816
ENSG00000138031	ADCY3	2	24819169	24919839
ENSG00000187833	C2orf78	2	73784189	73817147
ENSG00000144395	CCDC150	2	196639554	196763490
ENSG00000136720	HS6ST1	2	128236716	128318577
ENSG00000132326	PER2	2	238244038	238290102
ENSG00000206633	SNORA80B	2	10446714	10446849
ENSG00000224557	**HLA-DPB2**	**6**	**33112451**	**33129084**
ENSG00000156508	EEF1A1	6	73515750	73523797
ENSG00000145945	FAM50B	6	3849386	3851317
ENSG00000203908	KHDC3L	6	73362677	73364171
ENSG00000120253	NUP43	6	149724315	149749665
ENSG00000112494	UNC93A	6	167271169	167316019
ENSG00000234787	**LINC00458**	**13**	**54115783**	**54132866**
No correlated mRNA transcribed in the same chromosome.				
ENSG00000254934	**LINC00678**	**11**	**27617626**	**27634627**
ENSG00000213465	ARL2	11	65014113	65022184
ENSG00000075388	FGF4	11	69771016	69775403
ENSG00000149328	GLB1L2	11	134331874	134378341
ENSG00000182208	MOB2	11	1469457	1501247
ENSG00000182261	NLRP10	11	7959424	7965426
ENSG00000149782	PLCB3	11	64251523	64269150
ENSG00000126432	PRDX5	11	64318088	64321811

ENSG00000149136	SSRP1	11	57325985	57335877
ENSG00000168439	STIP1	11	64185272	64204543
ENSG00000226673	**LINC01108**	**6**	**14280127**	**14285454**
ENSG00000137225	CAPN11	6	44158811	44184402
ENSG00000124713	GNMT	6	42960758	42963880
ENSG00000203908	KHDC3L	6	73362677	73364171
ENSG00000137310	TCF19	6	31158542	31167159
ENSG00000232301	**LNCPRESS1**	**7**	**101299613**	**101301270**
ENSG00000070669	ASNS	7	97852118	97872542
ENSG00000160862	AZGP1	7	99966720	99976157
ENSG00000122783	C7orf49	7	135092363	135170795
ENSG00000004948	CALCR	7	93424487	93574730
ENSG00000133619	KRBA1	7	149714781	149734575
ENSG00000106399	RPA3	7	7636518	7718607
ENSG00000221740	SNORD93	7	22856613	22856686
ENSG00000196683	TOMM7	7	22812628	22822852
ENSG00000166925	TSC22D4	7	100463359	100479279
ENSG00000154978	VOPP1	7	55436056	55572988
ENSG00000249152	**LNCPRESS2**	**4**	**92268767**	**92277075**
ENSG00000145247	OCIAD2	4	48885019	48906937
ENSG00000258609	**LINC-ROR**	**18**	**57054559**	**57072119**
No correlated mRNA transcribed in the same chromosome.				
ENSG00000176654	**NANOGP1**	**12**	**7890801**	**7900140**
ENSG00000089692	LAG3	12	6772512	6778455
ENSG00000037897	METTL1	12	57768471	57772793
ENSG00000166886	NAB2	12	57088894	57095476
ENSG00000139405	RITA1	12	113185526	113192368
ENSG00000166863	TAC3	12	57010000	57028883
ENSG00000235602	**POU5F1P3**	**12**	**8133772**	**8134849**
ENSG00000166535	A2ML1	12	8822472	8887001
ENSG00000123064	DDX54	12	113157174	113185479
ENSG00000126749	EMG1	12	6970893	6979941
ENSG00000139146	FAM60A	12	31280584	31327058
ENSG00000111206	FOXM1	12	2857681	2877155
ENSG00000151948	GLT1D1	12	128853427	128984968
ENSG00000188375	H3F3C	12	31791185	31792241
ENSG00000135486	HNRNPA1	12	54280193	54287088
ENSG00000111602	TIMELESS	12	56416373	56449403
ENSG00000136014	USP44	12	95516560	95551490
ENSG00000237872	**POU5F1P4**	**1**	**155433178**	**155434262**

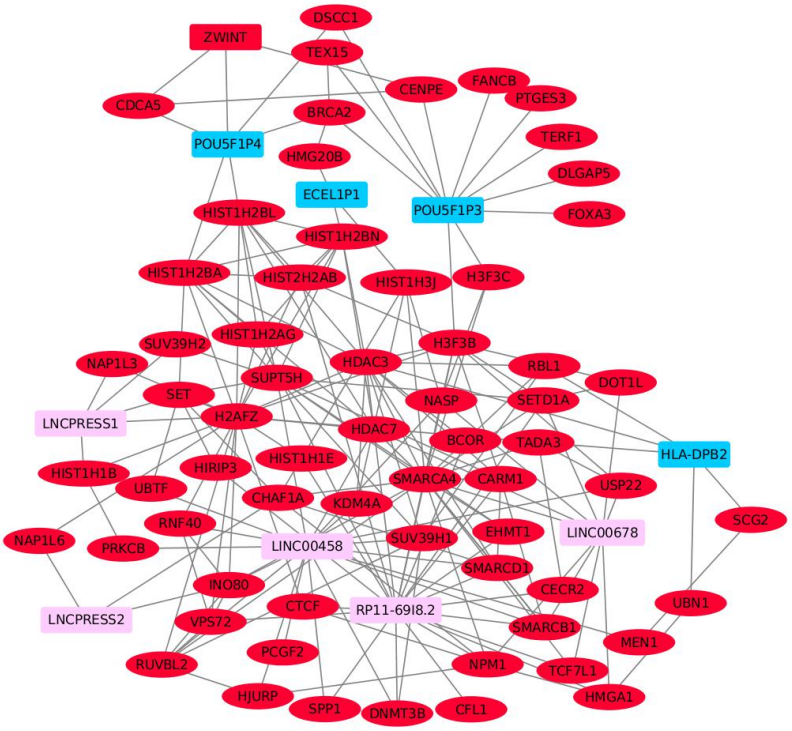
ENSG00000125462	C1orf61	1	156404250	156430701
ENSG00000179085	DPM3	1	155139891	155140595
ENSG00000187140	FOXD3	1	63323041	63325126
ENSG00000119285	HEATR1	1	236549005	236604504
ENSG00000173673	HES3	1	6244192	6245578
ENSG00000187961	KLHL17	1	960587	965715
ENSG00000143815	LBR	1	225401502	225428925
ENSG00000197769	MAP1LC3C	1	241995490	241999073
ENSG00000008130	NADK	1	1751232	1780457
ENSG00000162631	NTNG1	1	107140007	107483458
ENSG00000057757	PITHD1	1	23778405	23788232
ENSG00000143106	PSMA5	1	109399031	109426427
ENSG00000155380	SLC16A1	1	112911847	112957013
ENSG00000085491	SLC25A24	1	108134036	108200849
ENSG00000200913	SNORD46	1	44776490	44776593
ENSG00000196182	STK40	1	36339624	36385896
ENSG00000158710	TAGLN2	1	159918107	159925732
ENSG00000179403	VWA1	1	1434861	1442882
ENSG00000163874	ZC3H12A	1	37474552	37484379
ENSG00000084073	ZMPSTE24	1	40258107	40294184
ENSG00000162702	ZNF281	1	200404940	200410056
ENSG00000143067	ZNF697	1	119619422	119647773
ENSG00000274090	**RP11-415G4.1**	**13**	**55535697**	**55583524**
ENSG00000184564	SLITRK6	13	85792790	85799488
ENSG00000280707	**RP11-568A7.4**	**6**	**167228300**	**167237511**
No correlated mRNA transcribed in the same chromosome.				
ENSG00000236673	**RP11-69I8.2**	**6**	**131901963**	**131920565**
ENSG00000152661	GJA1	6	121435692	121449727
ENSG00000112312	GMNN	6	24774931	24786099
ENSG00000197153	HIST1H3J	6	27890382	27893106
ENSG00000112159	MDN1	6	89642499	89819723
ENSG00000008018	PSMB1	6	170535117	170553341

Supplemental material 4

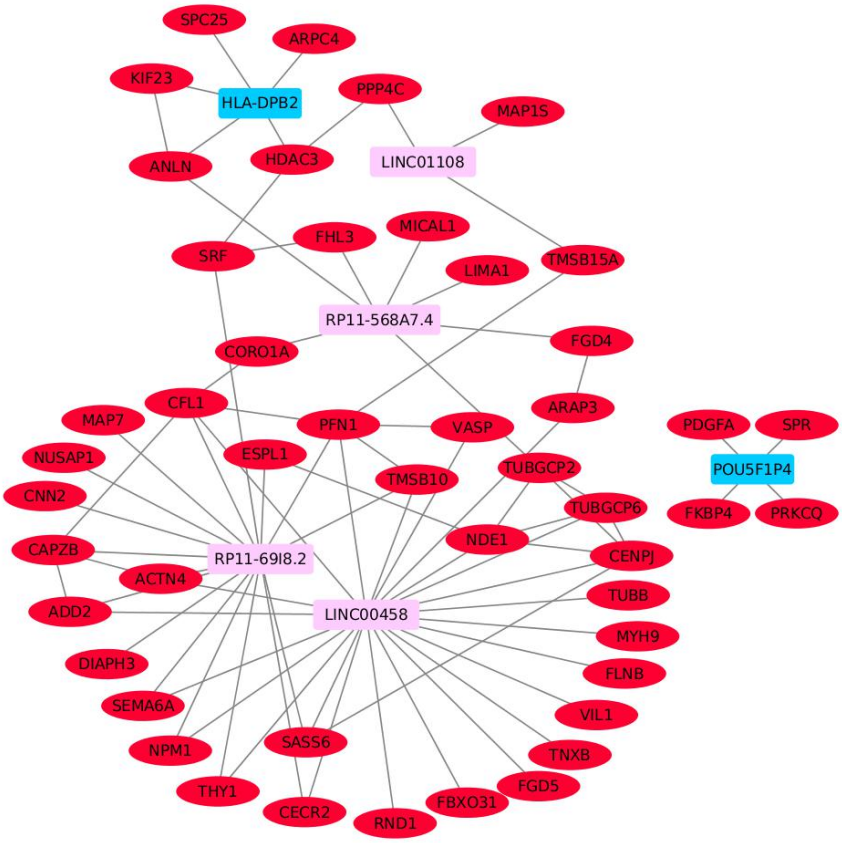
A



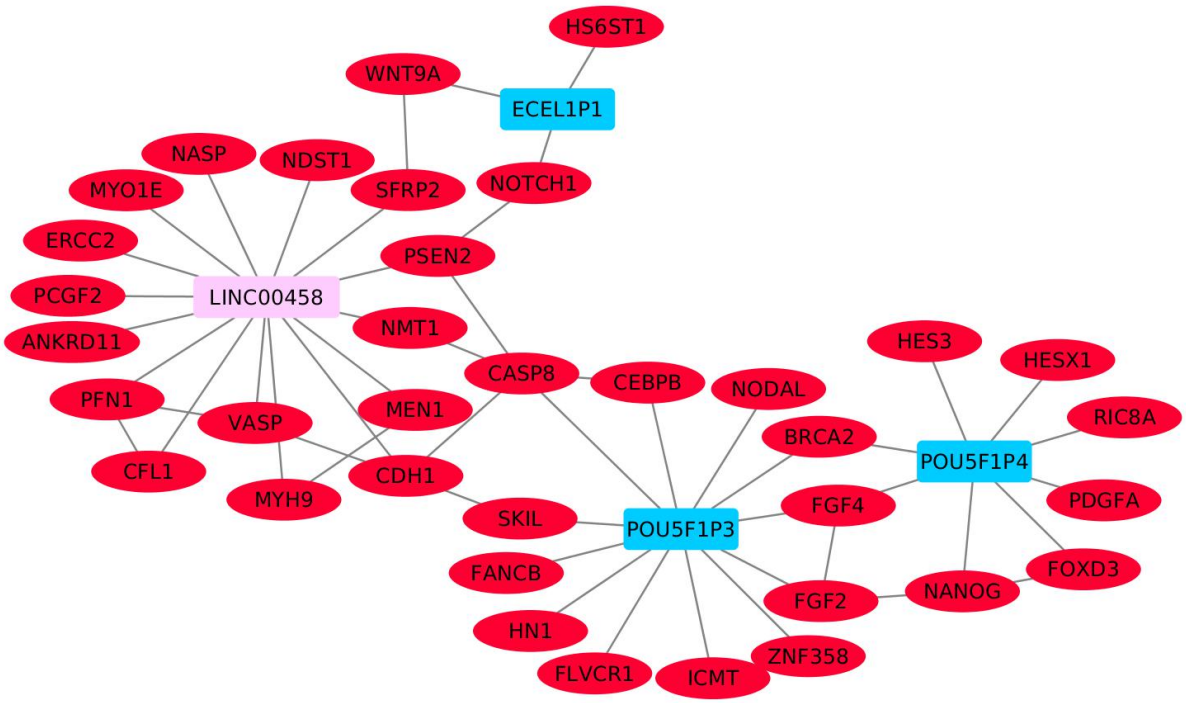
B



C


SPC25
ARPC4
KIF23
HLA-DPB2
PPP4C
MAP1S
ANLN
HDAC3
LINC01108
FHL3
MICAL1
SRF
LIMA1
TMSB15A
RP11-568A7.4
FGD4
CORO1A
CFL1
ARAP3
MAP7
PFN1
VASP
PDGFA
SPR
NUSAP1
ESPL1
TUBGCP2
POU5F1P4
TMSB10
CNN2
TUBGCP6
FKBP4
PRKCQ
NDE1
CAPZB
RP11-69I8.2
CENPJ
ACTN4
TUBB
ADD2
LINC00458
MYH9
DIAPH3
FLNB
SEMA6A
VIL1
NPM1
SASS6
TNXB
THY1
FGD5
CECR2
FBXO31
RND1

D


HS6ST1
WNT9A
ECEL1P1
NASP
NDST1
MYO1E
SFRP2
NOTCH1
ERCC2
PSEN2
PCGF2
LINC00458
ANKRD11
NMT1
HES3
HESX1
CASP8
CEBPB
NODAL
PFN1
MEN1
RIC8A
VASP
BRCA2
CFL1
CDH1
POU5F1P4
MYH9
FGF4
PDGFA
SKIL
POU5F1P3
FANCB
FOXD3
NANOG
FGF2
HN1
ZNF358
FLVCR1
ICMT

E

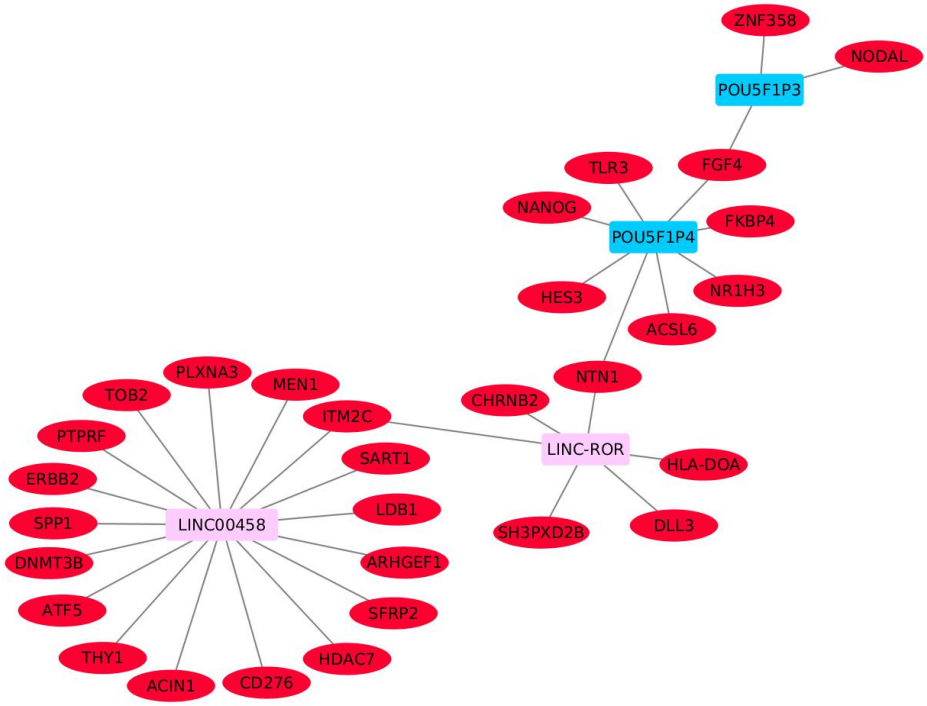


Figure S1: Networks of each ncRNA category. (A) Cell cycle and mitosis network. (B) Chromatin organization network. (C) Cytoskeleton organization network. (D) Embryonic development network. (E) Negative regulation of cell differentiation. Red elipses represent protein-coding mRNAs, pink rectangles represent lncRNAs and blue rectangles represent pseudogenes.

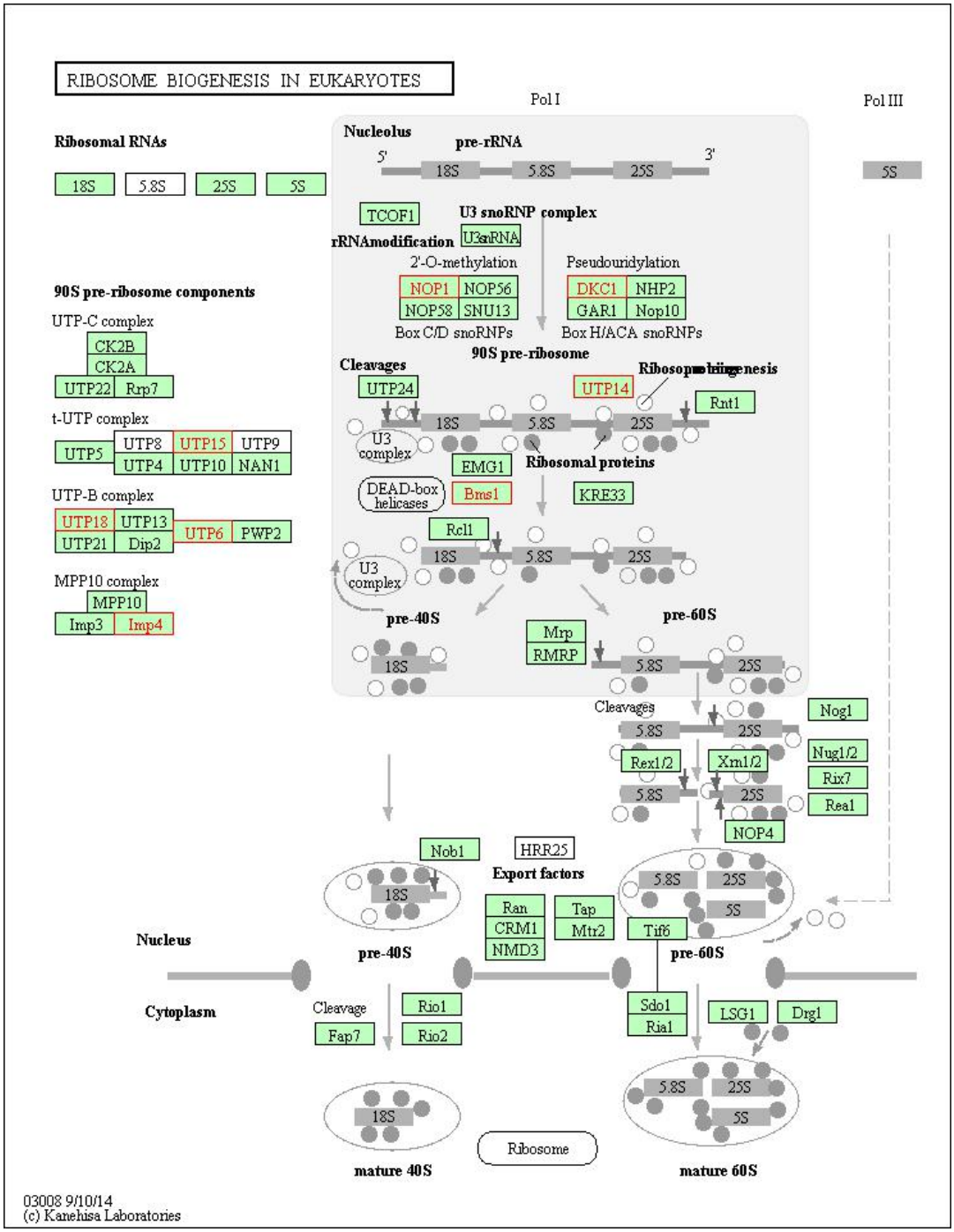


Figure S2: Ribosome biogenesis in eukaryotes. mRNAs coding for proteins highlighted in red are among the RNAs whose expression is positively correlated with the ribosome biogenesis-associated ncRNAs. Illustration adapted from the KEGG PATHWAY database.

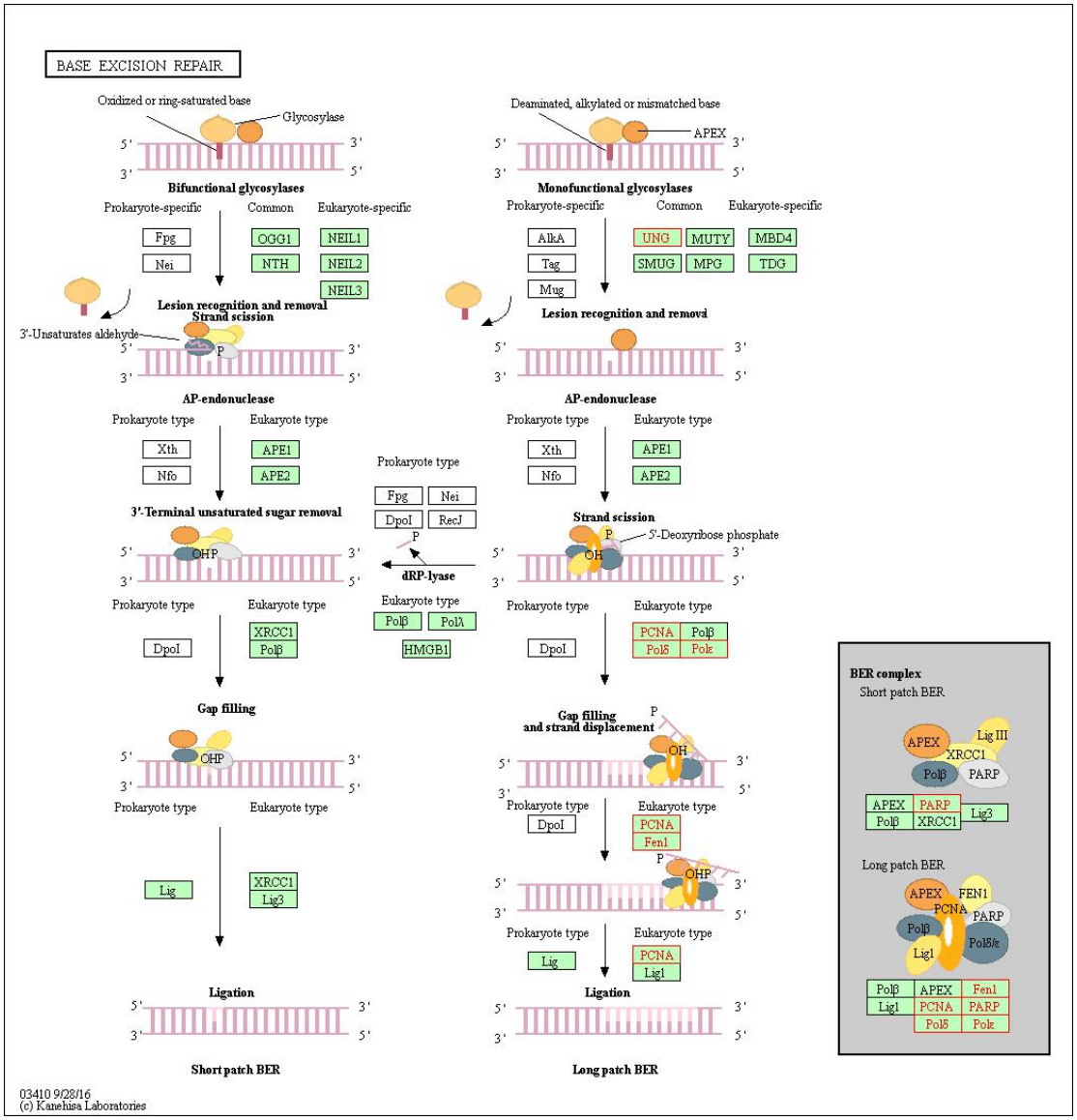


Figure S3: Base excision repair pathway. mRNAs coding for proteins highlighted in red are among the RNAs whose expression is positively correlated with the RNA repair-associated ncRNAs. Illustration adapted from the KEGG PATHWAY database.

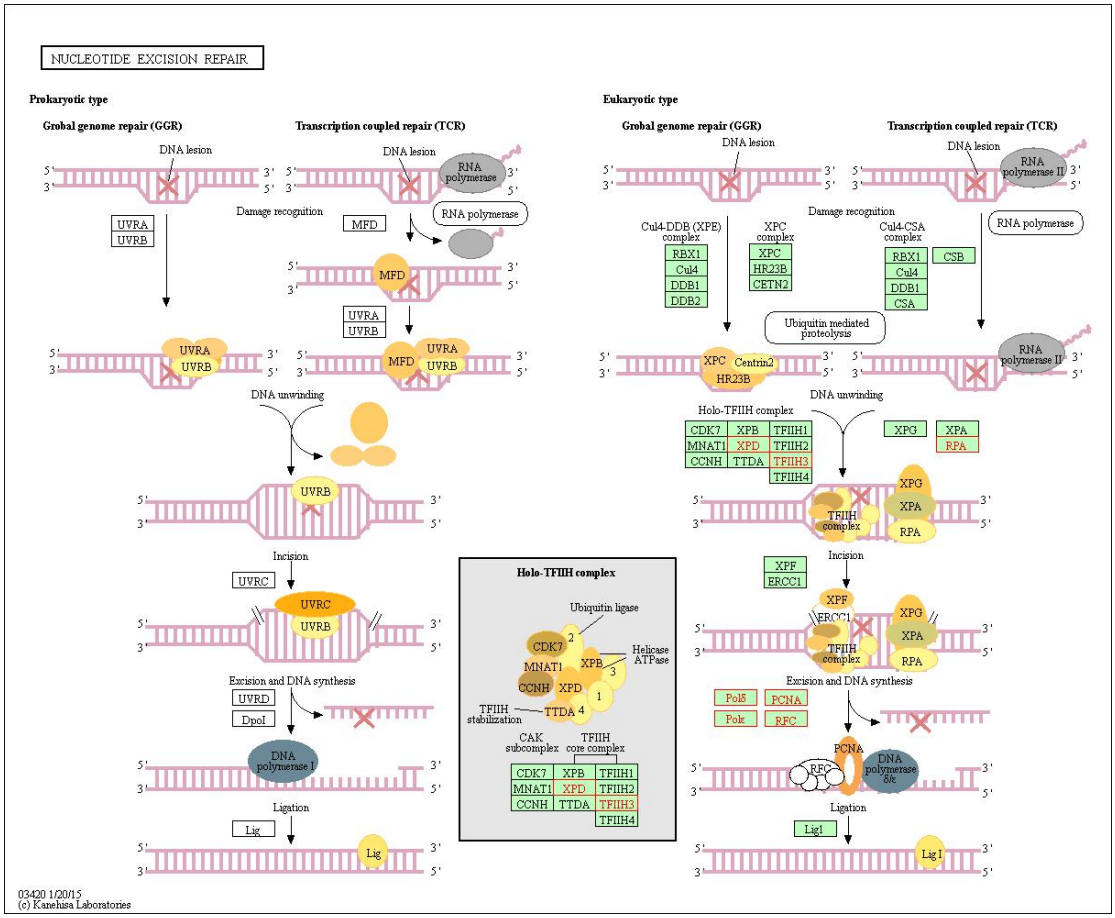


Figure S4: Nucleotide excision pathway. mRNAs coding for proteins highlighted in red are among the RNAs whose expression is positively correlated with the RNA repair-associated ncRNAs. Illustration adapted from the KEGG PATHWAY database.

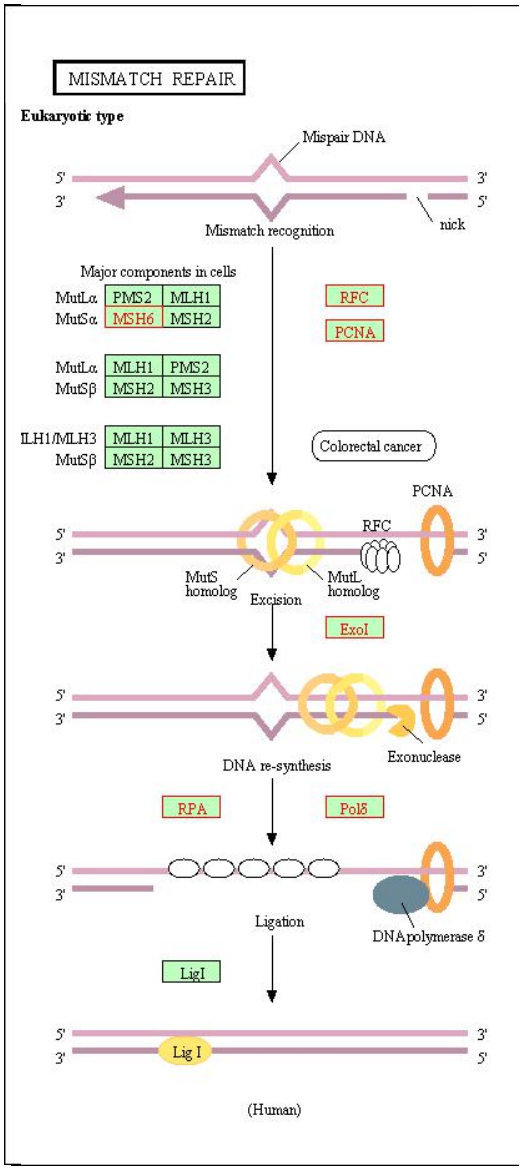


Figure S5: Mismatch repair pathway. mRNAs coding for proteins highlighted in red are among the RNAs whose expression is positively correlated with the RNA repair-associated ncRNAs. Illustration adapted from the KEGG PATHWAY database.

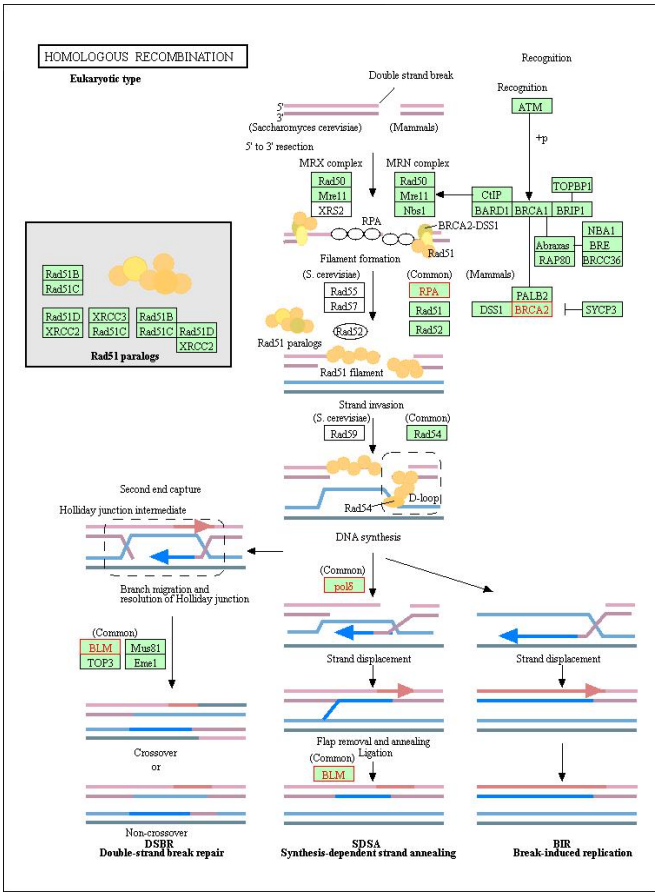


Figure S6: Homologous recombination pathway. mRNAs coding for proteins highlighted in red are among the RNAs whose expression is positively correlated with the apoptosis-associated ncRNAs. Illustration adapted from the KEGG PATHWAY database.

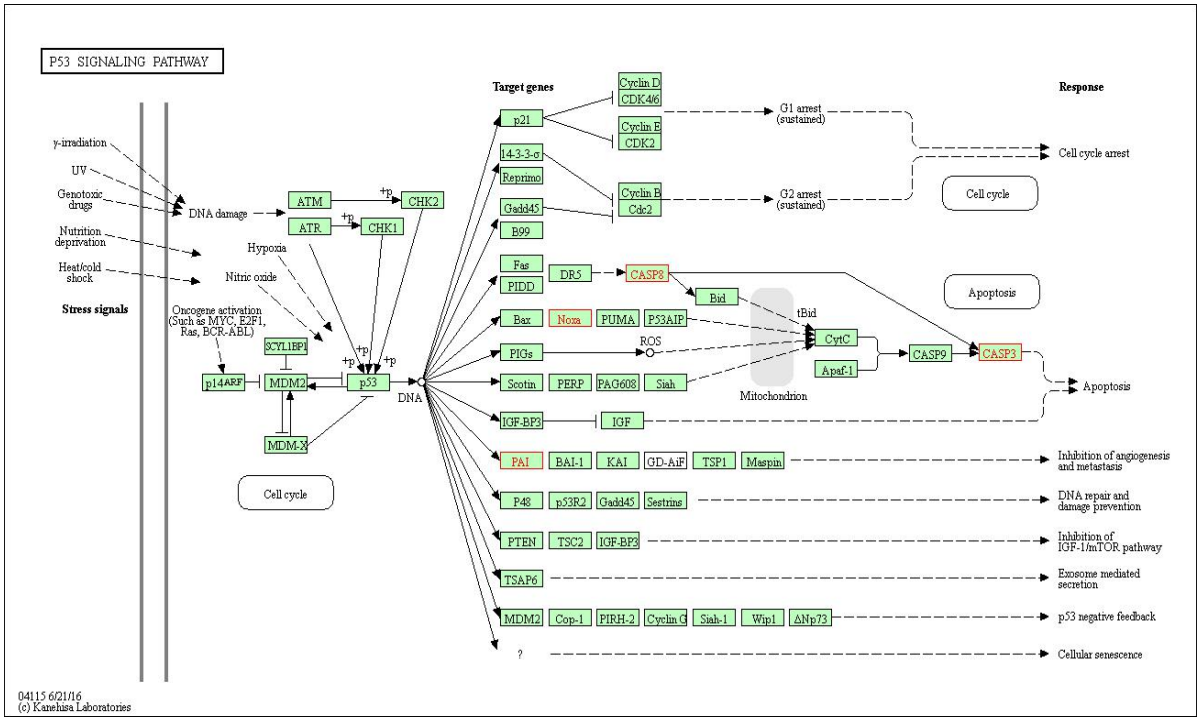


Figure S7: p53 signaling pathway. mRNAs coding for proteins highlighted in red are among the RNAs whose expression is positively correlated with the DNA repair and response to apoptosis-associated ncRNAs. Illustration adapted from the KEGG PATHWAY database.

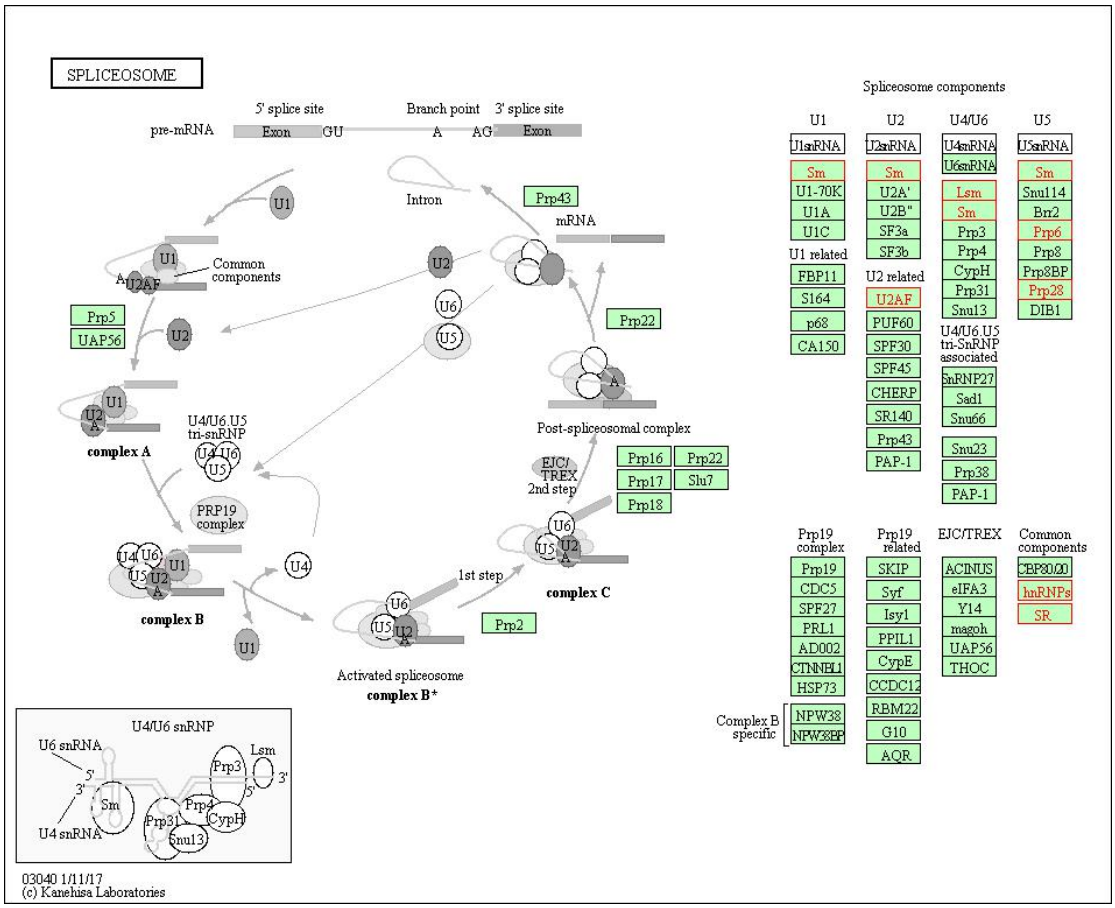


Figure S8: Spliceosome. mRNAs coding for proteins highlighted in red are among the RNAs whose expression is positively correlated with splicing-associated ncRNAs. Illustration adapted from the KEGG PATHWAY database.

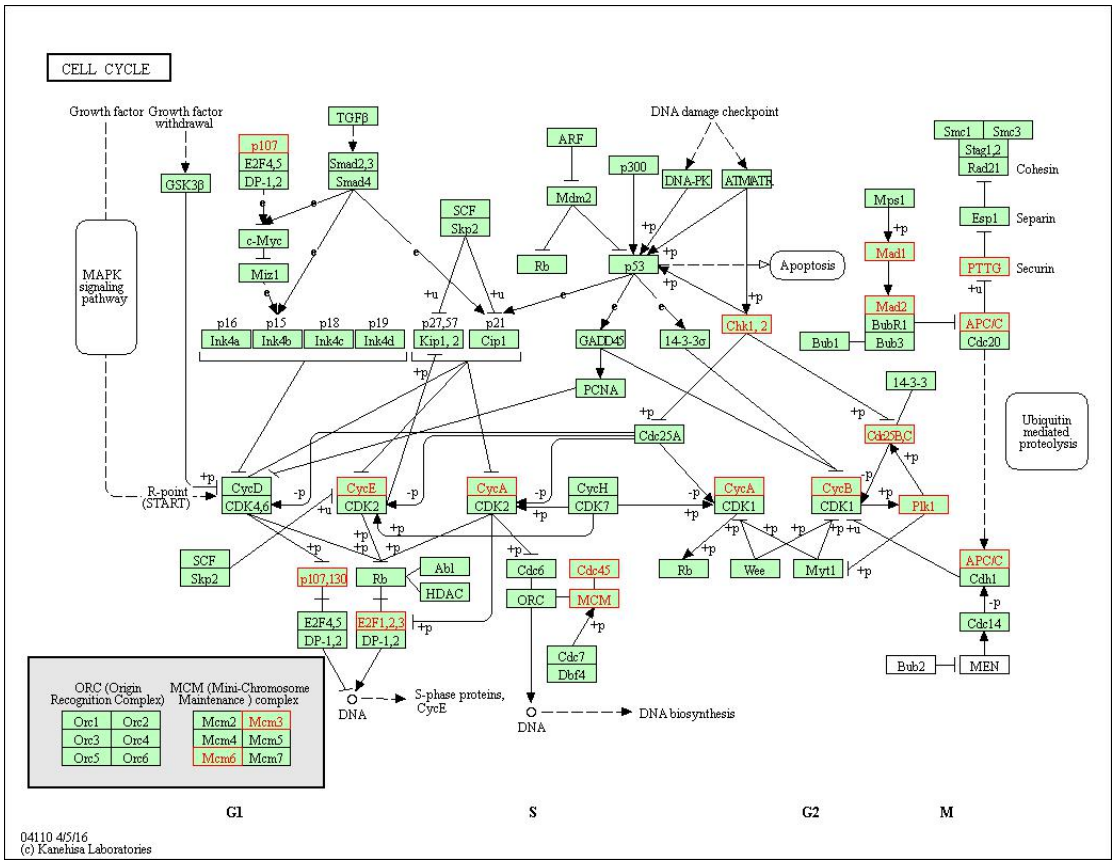
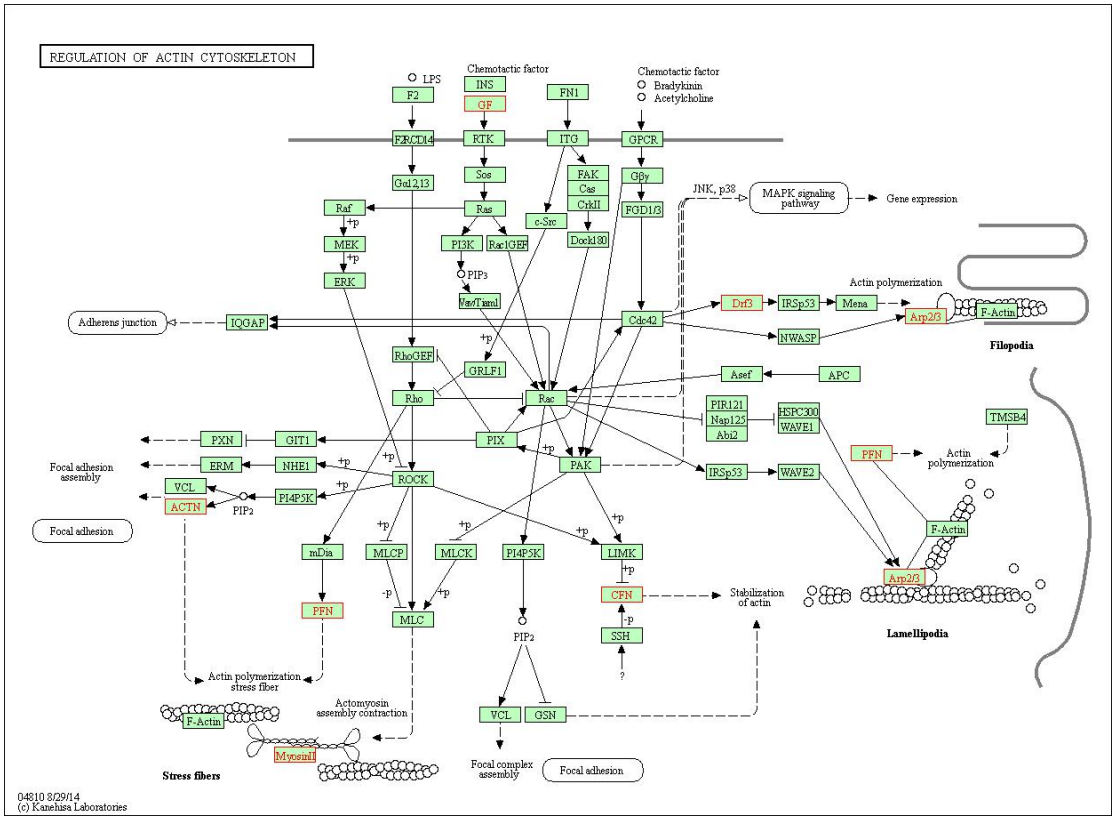


Figure S9: Cell cycle. mRNAs coding for proteins highlighted in red are among the RNAs whose expression is positively correlated with cell cycle-associated ncRNAs. Illustration adapted from the KEGG PATHWAY database.



Figures S10: Regulation of actin cytoskeleton. mRNAs coding for proteins highlighted in red are among the RNAs whose expression is positively correlated with cytoskeleton organization-associated ncRNAs. Illustration adapted from the KEGG PATHWAY database.

Supplemental material 5 **(Adapted from the original Excel sheets)**

Table 1: ncRNAs isoforms expressed in human embryonic stem cells

Gene ID	Gene name	Transcript ID	Transcript name	Transcript type	Transcript length	Strand
ENSG00000253507	CTD-2501M5.1	ENST00000519005	CTD-2501M5.1-003	lincRNA	731	1
	CTD-2501M5.1	ENST00000524275	CTD-2501M5.1-001	lincRNA	884	1
	CTD-2501M5.1	ENST00000519695	CTD-2501M5.1-002	lincRNA	854	1
ENSG00000204121	ECEL1P1	ENST00000373592	ECEL1P1-001	Unprocessed pseudogene	1742	-1
ENSG00000224557	HLA-DPB2	ENST00000435074	HLA-DPB2-002	Processed transcript	1188	1
	HLA-DPB2	ENST00000470997	HLA-DPB2-001	Transcribed unprocessed pseudogene	776	1
ENSG00000234787	LINC00458	ENST00000423442	LINC00458-003	lincRNA	1060	-1
	LINC00458	ENST00000606055	LINC00458-004	lincRNA	425	-1
	LINC00458	ENST00000427299	LINC00458-001	lincRNA	1030	-1
	LINC00458	ENST00000607494	LINC00458-006	lincRNA	849	-1
	LINC00458	ENST00000606706	LINC00458-005	lincRNA	833	-1
ENSG00000254934	LINC00678	ENST00000527083	LINC00678-001	lincRNA	786	-1
	LINC00678	ENST00000534757	LINC00678-002	lincRNA	798	-1
ENSG00000226673	LINC01108	ENST00000427276	LINC01108-001	lincRNA	3152	-1
ENSG00000258609	LINC-ROR	ENST00000553704	LINC-ROR-001	lincRNA	2603	-1
ENSG00000232301	LNCPRESS1	ENST00000429254	LNCPRESS1-001	antisense	756	1
ENSG00000249152	LNCPRESS2	ENST00000502518	LNCPRESS2-001	lincRNA	616	-1

	LNCPRESS2	ENST00000502518	LNCPRESS2-001	lincRNA	616	-1
ENSG00000176654	NANOGP1	ENST00000607111	NANOGP1-002	processedtranscript	1982	1
	NANOGP1	ENST00000525030	NANOGP1-001	Transcribed unprocessed pseudogene	915	1
ENSG00000235602	POU5F1P3	ENST00000428824	POU5F1P3-001	Processed pseudogene	1078	-1
ENSG00000237872	POU5F1P4	ENST00000413175	POU5F1P4-001	Processed pseudogene	1085	1
ENSG00000274090	RP11-415G4.1	ENST00000617598	RP11-415G4.1-001	lincRNA	809	-1
ENSG00000280707	RP11-568A7.4	ENST00000628362	RP11-568A7.4-001	lincRNA	582	1
	RP11-568A7.4	ENST00000625787	RP11-568A7.4-002	lincRNA	674	1
	RP11-568A7.4	ENST00000626424	RP11-568A7.4-003	lincRNA	403	1
ENSG00000236673	RP11-69I8.2	ENST00000454596	RP11-69I8.2-001	lincRNA	2305	1
	RP11-69I8.2	ENST00000421310	RP11-69I8.2-002	lincRNA	601	1
	RP11-69I8.2	ENST00000419695	RP11-69I8.2-003	lincRNA	712	1

Capítulo II

Characteristics of the competition among RNAs for the ligation of shared miRNAs

Artigo a ser submetido ao periódico Critical Reviews in Biochemistry and Molecular Biology

Characteristics of the Competition among RNAs for the Ligation of Shared miRNAs

Raquel Calloni and Diego Bonatto

Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul, Departamento de Biologia Molecular e Biotecnologia, Universidade Federal do Rio Grande do Sul

Abstract

Competing endogenous RNAs (ceRNAs) are RNAs that share common miRNA binding sites and compete with each other for miRNA ligation at these sites. The observation of this phenomenon in cells changed the status of miRNA target RNAs from molecules that are passively controlled by miRNAs to molecules that actively modulate miRNA activity. In this review, we build a general profile of ceRNA characteristics to facilitate ceRNA identification by researchers. The information summarized here contains a list of previously reported ceRNAs and classes of RNAs that can participate in this type of interaction, the expression behavior and characteristics of ceRNAs and miRNAs in the context of competition, the influence of shared MRE/miRNA numbers and miRNA binding strength on competition, reports on competition between RNAs in different subcellular localizations and the concept that ceRNAs may form a huge regulatory network in the cell.

Keywords: miRNAs, lncRNA, pseudogene, mRNA, competing endogenous RNAs, competition, ceRNETs

1. Introduction

Gene expression is strictly controlled at different levels and by different mechanisms in cells. One of those mechanisms has as the central molecules microRNAs (miRNAs), which are ~22-nucleotide-long RNAs that are endogenously expressed and involved in post-transcriptional gene silencing (Lee et al. 2002). miRNAs perform their biological role in association with the protein Ago2 as part of the RNA-induced silencing complex (RISC), where they act as guides, localizing the target RNAs through base pairing (Maniataki and Mourelatos 2005). Once paired with the target, RISC is able to impose two different fates to the bound mRNA: degradation or translation repression (Eulalio et al. 2008).

For years, the effects of the interaction between miRNAs and their targets were thought to be unidirectional (miRNA → target). The first evidence that challenged this idea came from plants, in a phenomenon termed target mimicry (Franco-Zorrilla et al. 2007). It was observed in Arabidopsis thaliana under phosphate starvation conditions, where both the non-coding RNA Induced by Phosphate Starvation 1 (IPS1) and miR-399 were expressed. IPS1 was shown to sequester miR-399, which caused the derepression of the mRNA coding for PHO2, a protein involved in phosphorous response and whose mRNA is also a target of miR-399 (Franco-Zorrilla et al. 2007).

In this sense, target RNAs are far from being passive substrates of miRNAs. In contrast, targets also exert an effect on miRNAs, influencing their availability to act on RNAs (Poliseno et al. 2010; Tay et al. 2011). Based on this concept, in 2011, a new layer of gene expression control was hypothesized (Salmena et al. 2011). The proposed regulatory mechanism relies on the idea that different targets may compete for ligation to their miRNAs, modulating the miRNA silencing

activity. Termed competing endogenous RNAs (ceRNAs), those molecules share miRNA response elements (MREs) with other co-expressed RNAs and act as miRNA sponges, leaving the mRNA targets free to be translated into proteins (Salmena et al. 2011).

Although many studies have been performed since the ceRNA hypothesis proposition, most of the information about the characteristics of this mechanism is still sparse in the literature. Also, many reviews about ceRNAs have been written, but they generally focus on describing types of molecules acting in the competition and the already-investigated and established ceRNAs (Ebert and Sharp 2010; Gupta 2014; Kartha and Subramanian 2014; Tay et al. 2014). This review covers the literature describing these interactions and summarizes some characteristics common to the ceRNAs already reported. In this way, the aim of this article was to build a general profile of ceRNA characteristics to facilitate ceRNA identification by researchers. The information summarized here contains (i) an updated list of ceRNAs already reported for humans and the classes of RNAs that can participate in this type of interaction, (ii) the expression behavior and characteristics of ceRNAs and miRNAs in the context of competition, (iii) the influence of the shared MRE/miRNA numbers and miRNA binding strength on the competition, (iv) reports on competition between RNAs in different subcellular localizations and (v) a discussion on the concept that ceRNAs may form a huge regulatory network in the cell.

2. RNAs acting as ceRNAs

Since the ceRNA hypothesis proposition in 2011, several RNAs were reported to act as competing endogenous RNAs (Poliseno et al. 2010; Tay et al. 2011; Karreth et al. 2015; Yan et al.

2015; Yu et al. 2015a; Wang et al. 2016a). Actually, the competition can occur among any RNAs that share MREs for the same miRNAs (Tay et al. 2011). Messenger RNAs (mRNAs), circular RNAs (circRNAs), pseudogenes, long non-coding RNAs (lncRNAs) and natural antisense RNAs (NATs) can act as ceRNAs (Figure 1) (Poliseno et al. 2010; Tay et al. 2011; Hansen et al. 2013; Kimura et al. 2015; Wang et al. 2016a). The competition can also occur between spliced isoforms, as reported for OCT4A and OCT4B, both products of the gene OCT4, or for CDKAL1-v1 and the full-length CDKAL1, both expressed by the a homonym gene (Zhou et al. 2014a; Li et al. 2015a). In this case, the involved RNAs have been named as alternative splicing-coupled competing endogenous RNAs (AS-ceRNAs) (Zhou et al. 2014a). Table 1 summarizes the human ceRNAs already validated and reported in the scientific literature, together with the biologicial context where they are important.

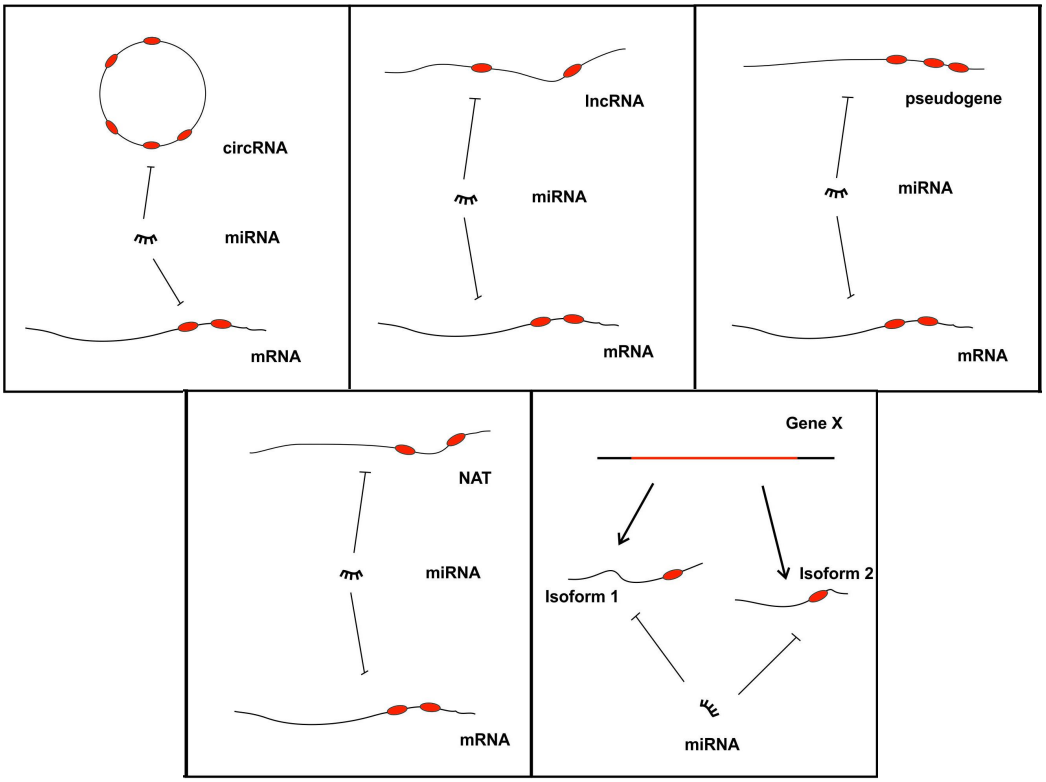


Figure 1: RNAs already reported to act as competing endogenous RNAs.

Table 1: List of curated human ceRNAs described in the literature.

Cancer promoters				
RNA	**mRNA**	**miRNA**	**Biological context**	**Reference**
BARD1 9' L	BARD1 isoforms	mirR-203 and miR-101	cancer cells	Pilyugin and Irminger-Finger 2014
BC032469	hTERT	miR-1207-5p	gastric cancer	Lü et al. 2015
BRAFP1	BRAF	miR-30a, miR -182, miR-876	lymphoma	Karreth et al. 2015
CCAT1	HMGA2 and c-Myc	let-7	hepatocellular carcinoma	Chen et al. 2016a
CYP4Z2P	CYP4Z1l	miR-211, miR-125a, miR-197, miR-1226, and miR-204	breast cancer	Zheng et al. 2015
H19	FOXM1	miR-342-3p	gallbladder cancer	Wang et al. 2016b
H19	ZEB1 and ZEB2	miR-138 and miR-200a	colorectal cancer	Liang et al. 2015
HMGA1P6 and HMGA1P7	HMGA1	miR-15, miR-16, miR-23b, miR-26a, miR-34b, miR-130b, miR-196a2, miR-326, miR-432, miR-548c-3p, miR-570, miR-603 and let-7a	pituitary tumors	Esposito et al. 2015
HOTAIR	CCND1	miR-1	esophageal cancer	Ren et al. 2016
HOTAIR	HER2	miR-331-3p	gastric cancer	Liu et al. 2014
HOTAIR	MAPK1	miR-1, miR-214-3p, and miR-330-5p	epithelial ovarian cancer	Yiwei et al. 2015
HOTAIR	PIK3R3	miR-214 or miR-217	ovarian cancer	Dong and Hui 2016
HULC	PRKACB	miR-372	liver cancer	Wang et al. 2010
KRAS1P	KRAS	miR-143 and let-7 family	prostate cancer cell line	Poliseno et al. 2010

lncARSR	AXL and c-MET	miR-34 and miR-449	renal carcinoma	Qu et al. 2016
lncRNA-ATB	ZEB1 and ZEB2	miR-200a, miR-200b, miR-200c	hepatocellular carcinoma	Yuan et al. 2014
Linc00152	ERBB4	miR-193a-3p	colon cancer	Yue et al. 2016
Malat1	ANXA2 and KRAS	miR-206	gallbladder cancer	Wang et al. 2016c
Malat1	MCL-1	miR-363-3p	gallbladder cancer	Wang et al. 2016a
Malat1	MMP14 and Snail	miR-22	malignant melanoma	Luan et al. 2016
Malat1	ZEB2	miR-200c	kidney carcinoma	Xiao et al. 2015
NEAT1	E2F3	miR-377-3p	non-small cell lung cancer	Sun et al. 2016
OCT4B	OCT4A	miR-335, miR-20a, miR-20b, miR-106a and miR-106b	invasive breast, prostate and colon cancer	Li et al. 2015a
OCT4-pg4	OCT4	miR-145	hepatocellular carcinoma	Wang et al. 2013
PIK3C2A	CD151	miR-124	hepatocellular carcinoma	Liu et al. 2016b
SNHG6-003	TAK1	miR-26a and miR-26b	hepatocellular carcinoma	Cao et al. 2016
SNHG6	ZEB1	miR-101-3p	hepatocellular carcinoma	Chang et al. 2016
TUG1	POU2F1	miR-9-5p	osteosarcoma	Xie et al. 2016
UCA1	MMP14.	miR-485-5p	ovarian cancer	Yang et al. 2016
UCA1	SOX4	miR-204	esophageal cancer	Jiao et al. 2016
Versican V0 and V1 isoforms	CD34, and fibronectin	miR-133a, miR-199a-3p, miR-144, and miR-431	hepatocellular carcinoma	Fang et al. 2013
Cancer inhibitors				
RNA	**mRNA**	**miRNA**	**Biological context**	**Reference**
CASC2	PIAS3	miR-18a	colorectal cancer	Huang et al. 2016
Cir-ITCH	ITCH	miR-7, miR-17 and miR-214	esophageal squamous cell carcinoma	Li et al. 2015b

CNOT6L	PTEN	miR-17, miR-19a, miR-19b, miR-20a, miR-20b, and miR-106b	prostate and colorectal cancer cell lines	Tay et al. 2011
CTNNAP1	CTNNA1	miR-141	colorectal cancer	Chen et al. 2016b
FER1L4	RB1	miR-106a-5p	gastric cancer	Xia et al. 2014
FER1L4	PTEN	miR-106a-5p	gastric cancer	Xia et al. 2015
FOXO3P and circ-FOXO3	FOXO3	miR-22, miR-136-3p, miR-138, miR-149-3p, miR-433, miR-762, miR-3614-5p and miR-3622b-5p	breast carcinoma	Yang et al. 2015
lncRNA-BGL3	PTEN	miR-17, miR-93, miR-20a, miR-20b, miR-106a and miR-106b	chronic myeloid leukemia	Guo et al. 2014
Linc-223	IRF4	miR-125-5p	acute myeloid leukemia	Mangiavacchi et al. 2016
MEG3	Bcl-2	miR-181a	gastric cancer	Peng et al. 2015
PTENP1	PTEN	miR-19b, miR-20a and miR-21	prostate cancer cell line and oral squamous cell carcinoma	Poliseno et al. 2010; Gao et al. 2016
RB1	PTEN	32 shared miRNAs	glioblastoma	Sumazin et al. 2011
SERINC1	PTEN	Not mentioned.	prostate and colorectal cancer cell lines	Tay et al. 2011
VAPA	PTEN	miR-17, miR-19a, miR-20a, miR-20b, miR-26b, miR-106a, and miR-106b	prostate and colorectal cancer cell lines	Tay et al. 2011
ZEB2	PTEN	miR-25, miR-92a, miR-181, and miR-200b	melanoma	Karreth et al. 2011
Other pathologies				
RNA	**mRNA**	**miRNA**	**Biological context**	**Reference**

CDKAL1-v1	CDKAL1	miR-494	type 2 diabetes	Zhou et al. 2014a
circRNA-CER	MMP13	miR-136	Osteoarthritis	Liu et al. 2016a
DKK1	PTEN	miR-217, miR-33a, miR-33b, miR-103a, miR-93, and miR-106a	diabetic cardiomyocytes	Ling et al. 2013
GAS5	p27	miR-222,	liver fibrosis	Yu et al. 2015a
H19	AQP3	miR-874	intestinal barrier function	Su et al. 2016
Malat1	Rac1	miR-101b	liver fibrosis	Yu et al. 2015b
MEG3	12/15-LOX	miR-181b	cerebral ischemic infarct and in hypoxia-induced neuron apoptosis.	Liu et al. 2016c
MIAT	VEGF	miR-150-5p	diabetic retinopathy	Yan et al. 2015
Organs				
RNA	**mRNA**	**miRNA**	**Biological context**	**Reference**
ciRS-7	Not mentioned	miR-7	brain	Hansen et al. 2013
Sry	Not mentioned	miR-138	testis	Hansen et al. 2013
Stem and differentiated cells				
RNA	**mRNA**	**miRNA**	**Biological context**	**Reference**
CARL	PHB2	miR-539	cardiomyocytes	Wang et al. 2014
FLJ11812	ATG13 and CDC20B	miR-4459	vascular endothelial and embryonic stem cells	Lu et al. 2015
IFN-α1 AS, IFN-α and IFN-α AS	IFN-α1	miR-1270	Human Namalwa B cells	Kimura et al. 2015
H19	β -catenin	miR-141 and miR-22	osteoblasts differentiation	Liang et al. 2016
H19	DICER	let-7	PA-1 cells	Kallen et al. 2013

linc-MD1	MAML1 and MEF2C	miR-133 and miR-135	myoblasts	Cesana et al. 2011
lincRNA-RoR	Oct4, Sox2, and Nanog	miR-145	embryonic stem cells	Wang et al. 2013
Malat1	SRF	mirR-133	myocyte differentiation	Han et al. 2015

3. Characteristics of the Competition Mechanism

3.1. Expression Characteristics of ceRNAs and miRNAs during Competition

The main characteristic that indicates whether two RNAs compete for the same miRNA pool is the behavior of the levels of one RNA regarding the experimental expression modulation of the other (Tay et al. 2011; Karreth et al. 2015). Superexpression of one RNA leads to the upregulation of the other, and its downregulation promotes the partner's downregulation (Karreth et al. 2015). This phenomenon is observed because when the number of molecules of one competitor increases, they titrate away the shared miRNAs, leaving the other RNA free to be translated or to execute its functions (Figure 2). This is not observed, however, when the cell is depleted for DICER or DROSHA, because of the decreased production of miRNAs shared by the competitors (Sumazin et al. 2011; Tay et al. 2011; Karreth et al. 2015). In this case, upregulation of one RNA does not affects the expression of the other (Sumazin et al. 2011; Tay et al. 2011; Karreth et al. 2014). On the other hand, when silencing the miRNA shared by the competing RNAs, a great improvement in the expression of both RNAs is observed (Kimura et al. 2015). However, in this case, the experimental modulation of levels of one RNA leads to very small alterations in the expression of its partner because the miRNA that mediates the crosstalk between the RNAs is no longer present (Sumazin et al. 2011; Tay et al. 2011). These last two observations indicate to the researcher that the relationship between the two RNAs is indeed mediated by miRNAs.

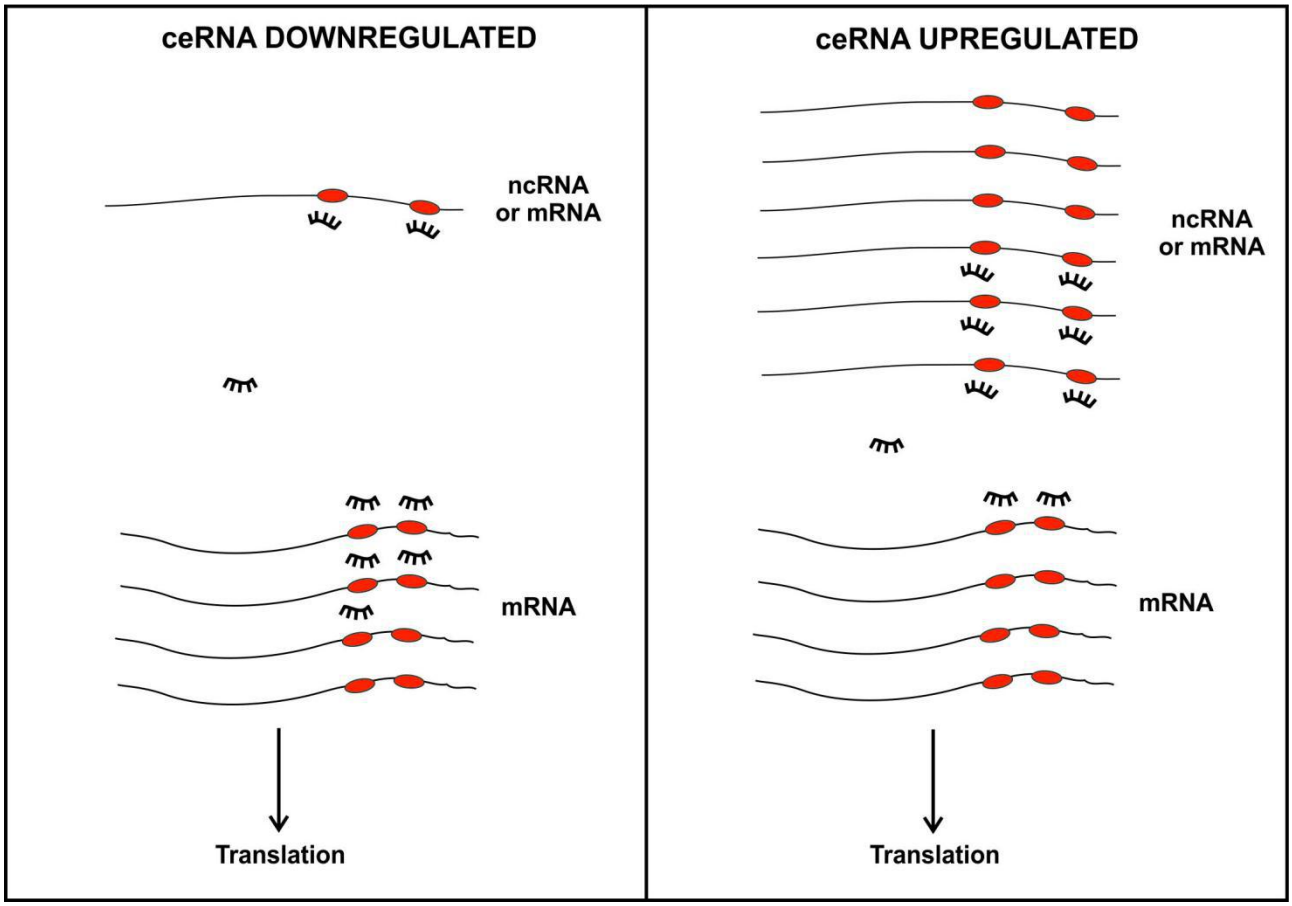


Figure 2: In the competition between two RNAs, the expression modulation of one RNA leads to the modification of the expression of the other.

In most cases, the crosstalk is reciprocal, as described above (i.e., alterations in the expression of one RNA leads to alterations in the expression of its partner competitor and vice-versa) (Sumazin et al. 2011; Tay et al. 2011). However, there are a few exceptions for to this rule, where the competition is not reciprocal (Karreth et al. 2011; Figliuzzi et al. 2013; ; Yuan et al. 2015; Zheng et al. 2015). This is the case of the ceRNA pair CYP4Z1-CYP4Z2P. Upregulation of the CYP4Z2P 3'UTR promotes the expression of CYP4Z1, but superexpression of the CYP4Z1-3' UTR, however, showed a weaker effect on CYP4Z2P levels (Zheng et al. 2015).

Modulation of the levels of the ceRNAs can possibly also affect the shared miRNAs levels. However, there is no consensus in the literature about this issue. Some authors report that alteration of the expression levels of the ceRNAs does not alter the miRNA levels in the cell and that the miRNA degradation is not induced after the binding to the ceRNA (Karreth et al. 2011; Deng et al. 2015; Zheng et al. 2015). However, some studies show the upregulation of the miRNA expression when one of the ceRNAs is downregulated or silenced and upregulation of the ceRNA promoting reduction in the expression and activity of the miRNA (Wang et al. 2010; Wang et al. 2014; Su et al. 2016).

Indeed, the expression levels of the shared miRNAs and the ceRNAs that compete for them are important for the competition to happen. It was hypothesized that the there is minimal competition when the total number of targets largely exceeds the number of miRNA molecules. The competition would be also be hampered when the miRNA molecules are much more abundant than ceRNAs (Ala et al. 2013). Moreover, it was shown that the optimal crosstalk between the two ceRNAs occurs in a near equimolar state of miRNA and ceRNAs (Ala et al. 2013). In agreement with this observation, it was revealed that miRNA:target ratios may influence the competition. Competition seems to occur only for ceRNAs with low total miRNA:target ratios. Targets of highly expressed miRNAs and with high miRNA:target ratios are likely not susceptible to derepression by ceRNA competition. The absence of competition is probably due to the buffering capacity provided by the high miRNA concentrations (Bosson et al. 2014).

Although the optimal crosstalk happens when the ceRNAs and miRNAs numbers of molecules are approximately equal, some reports show that the levels of each RNA of a ceRNA pair do not need to be the same. VAPA, one of the RNAs that competes with PTEN, is expressed at

higher levels than PTEN in several cell lines (Ala et al. 2013). Apparently, some ncRNAs can act as competitors even when they are less expressed than their RNA pair. That is the case of the pseudogene BRAFP1, which acts as the ceRNA of BRAF (Karreth et al. 2015). The same occurrs with CYP4Z1 and CYP4Z2P, where the pseudogene is less expressed than CYP4Z1 (Zheng et al. 2015). In addition, both RNAs also compete with CDK3 and are also less expressed than this gene (Zheng et al. 2016). FOXO3 shares MREs with FOXO3P and circ-FOXO3. Both ncRNAs effectively compete with FOXO3 for miRNAs binding. FOXO3, however, is much more highly expressed than the two ncRNAs (Yang et al. 2015). Although the competition happens in the situations above, the effect of loss of one partner is greater when both RNAs are present in similar quantities (Ala et al. 2013).

Moreover, the expression of both RNAs is generally observed to be positively correlated (Table 2), and this characteristic has been used as one of the criteria for predicting putative ceRNAs (Paci et al. 2014; Xia et al. 2014; Welch et al. 2015). The number of shared miRNAs seems to have influence on the correlation, as the co-expression increases with the number of shared miRNAs (Xu et al. 2015). Moreover, a significant negative correlation between miRNAs and ceRNAs was reported (Liu et al. 2014; Huang et al. 2016; Su et al. 2016; Sun et al. 2016). Some articles use the negative correlation of RNA pairs with their respective miRNAs as a filtering criterion in the ceRNA prediction methodology (Zhou et al. 2014b). However, the ceRNA expression is not always negatively correlated with the miRNA expression. In some cases, the negative correlation is low or even not significant (Sumazin et al. 2011; Wang et al. 2013; Wang et al. 2016a).

Table 2: RNAs that compete with each other for miRNA binding are generally positively correlated.

ceRNAs	Cell type	r value	p value	Reference
BRAF and BRAFP1[a]	diffuse large B cell lymphoma primary tumors	0.42	0.017	Karreth et al. 2015
BRAF and BRAFP1[a]	diffuse large B cell lymphoma cell lines	0.57	0.0029	Karreth et al. 2015
CTNNA1 and CTNNAP1[b]	colorectal cancer	0.631	<0.001	Chen et al. 2016b
HER2 and HOTAIR[c]	gastric cancer	0.861	Not supplied	Liu et al. 2014
HMGA1 and HMGA1P6[c]	growth hormone tumors	0.859	< 0.0001	Esposito et al. 2015
HMGA1 and HMGA1P7[c]	growth hormone tumors	0.677	< 0.0001	Esposito et al. 2015
KRAS and KRAS1P[c]	prostate cancer	0.486	0.0408	Poliseno et al. 2010
OCT4 and OCT4-pg4[a]	hepatocellular carcinoma	0.839	0.0001	Wang et al. 2013
PIAS3 and CASC2b	colorectal cancer	0.662	<0.01	Huang et al. 2016
PTEN and PTENP1[c]	prostate cancer	0.754	<0.0001	Poliseno et al. 2010
ZEB2 and MALAT1[c]	renal cancer	0.622	<0.0001	Xiao et al. 2015

[a] Pearson correlation. [b] Logistic regression analysis. [C] Bivariate correlation analysis through a method not specified.

3.2. miRNAs Shared by the ceRNAs

3.2.1. Influence of the Number of Shared miRNAS and MREs

The crosstalk between ceRNAs is mediated by miRNAs, specifically by the miRNA response elements (MREs). The number of shared miRNAs between ceRNAs is variable: BRAFP1 and BRAF bind at least three miRNAs, whereas the competition between PTEN and RB1 is mediated by 32 miRNAs (Sumazin et al. 2011; Karreth et al. 2015). However, the effect of a ceRNA expression modulation on its partner depends on how many common miRNAs the two RNAs share. In an in silico simulation, an increase of a ceRNA expression had major effects on RNAs whose

number of common binding miRNAs was high when compared with RNAs with low numbers of shared molecules. In this last case, the alteration of the ceRNA partner expression was mild, closer to the effect on RNAs with no miRNAs in common (Ala et al. 2013). Therefore, the competition is more pronounced among RNAs that share a large number of miRNAs (Ala et al. 2013).

Moreover, when the number of shared miRNAs is large, the effect of expression modulation of one miRNA on its target can be negligible. In other words, the targets regulation is unlikely to be significantly affected when only one miRNA has its expression altered (Sumazin et al. 2011).

In addition, the number of MREs of the shared miRNA can be variable in each RNA of the pair. This means that one ceRNA can have more MREs for a given shared miRNA than its partner, and this difference may influence the derepression levels by one ceRNA over another. The derepression occurs even when the competitor has less MREs than its partner. However, both the competition and the derepression are improved with the addition of MREs in the competitor sequence (Yuan et al. 2015).

3.2.2. Influence of miRNA Binding Characteristics

The ceRNA regulation efficiency is not only determined by the number of shared miRNAs or MREs; the binding strength between mRNA and miRNA also has an influence on the competition (Yuan et al. 2015). Evidence of this fact includes a study on ceRNAs applying synthetic gene circuits, which showed that the higher the MRE-miRNA binding affinity, the stronger was the ceRNA effect (Yuan et al. 2015). Moreover, the level of complementarity proved to also be important. When miRNAs incompletely bind to the MREs of the ceRNA pair, increasing the

expression of one of the RNAs caused improvement in the expression of its partner RNA. However, when one of the ceRNAs has MREs that bind perfectly to the miRNAs, modulations on its expression did not influence the expression of its pair (Yuan et al. 2015). In this case, a nonreciprocal competition was observed (Yuan et al. 2015).

In addition, there is an hierarchy of the miRNA binding to the MREs on the targets. miRNAs bind preferentially to the hight affinity 8- and 7-mer seed sites (Bosson et al. 2014). This miRNA behavior leads to the occurrence of two different target binding situations: (i) when the miRNA expression is lower than the target sites pool, these molecules will bind to 8- and 7-mer sites first; however, (ii) when the miRNA expression is increased, the interactions also spread to the 6-mer seed sites (Bosson et al. 2014). This characteristic influences the size of the competitors pool. In the case of low miRNA expression, the competitors pool is small and it is restricted to the high-affinity sites. However, when the miRNA levels are high, the pool is larger since the miRNAs will also bind to the low-affinity sites. In this last case, the target pool may become too large to be influenced by alterations in the expression of one ceRNA (Bosson et al. 2014).

Another miRNA binding characteristic is the interaction position on the RNA. In mRNAs, the majority of miRNAs binds to the 3' UTR sequence. Nevertheless, this is not necessarily true for ncRNAs, where the MREs can be located in the middle of the molecule (Wang et al. 2013). Moreover, different levels of miRNA interaction could occur among different sites for the same miRNA in a target. The lncRNA CCAT1 has three sites for miR-155, however, only one of the sites seems to have an strong interaction with the miRNA (Chen et al. 2016a). In addition, some authors observed that the predicted ΔG of the miRNA-lncRNA ligation was lower than the ΔG of the miRNA-mRNAs ligation (Cesana et al. 2011).

3.4. Subcellular Localization

The elements that participate in the competition (mRNAs, lncRNAs, pseudogenes, circRNAs and miRNAs) can naturally be found in different subcellular localizations (Liao et al. 2010; Cabili et al. 2015; Venø et al. 2015; Yan et al. 2015). However, for the competition to occur, both ceRNAs and miRNAs must share the same compartmental location in the cell. Many authors indeed support the opinion that there must be an overlap in the cellular localization of the miRNAs and ceRNAs involved in a given competition relationship (Salmena et al. 2011; de Giorgio et al. 2013; Tay et al. 2014; Qu et al. 2016). Thus, the knowledge of this cellular distribution information is very important when one proceeds with a ceRNA crosstalk prediction.

Nevertheless, evidences show that RNAs that do not share the same subcellular compartment but do share miRNAs target sites are not impeded to compete between each other. The competition may happen when the shared miRNAs are found in both ceRNAs' places. The lncRNA MALAT1 is found in the cytoplasm, but it is predominantly located in the nucleous. Its location, however, does not impairs its ceRNA activity on its partner, ZEB2, which is cytoplasmatic. In addition, miR-200c, a miRNA whose binding site is shared by both RNAs, is found in the cytoplasm and the nucleous of the cell (Xiao et al. 2015). The same is observed for MIAT and lncRNA-BGL3. MIAT is a nuclear lncRNA which that acts as a ceRNA of VEGF. The miRNA shared by both RNAs is found in the cytoplasm and cell nucleous, and it associates to MIAT in the nucleous (Yan et al. 2015). LncRNA-BGL3, in turn, acts as a ceRNA of PTEN, and it is found in both the nucleous and cytoplasm (Guo et al. 2014).

4. ceRNAs Networks: the ceRrNETs

Most of the articles describing ceRNAs report this type of interaction between only two molecules (Liu et al. 2014; Peng et al. 2015; Yu et al. 2015b; Wang et al. 2016a; Sun et al. 2016). However, in the real cell context, the competition is certainly not restricted to only two RNAs. In contrast, it probably occurs among several RNAs, forming a network, named the as ceRNET (Salmena et al. 2011; Sumazin et al. 2011; Kimura et al. 2015; Zheng et al. 2016). The complexity that a ceRrNET can achieve is evidenced by the work of Sumazin et al., (2011), which predicted that, in glioblastoma cells, approximately 248,000 pairwise RNA-RNA interactions are mediated by shared miRNAs, and with a highly conservative false discovery rate (Sumazin et al. 2011). Indeed, it has been speculated that the ceRNA layer of regulation may be as large as transcriptional regulation in terms of size (Sumazin et al. 2011).

Although most ceRNAs are investigated in pairs, some competition networks are starting to be unraveled. One of the ceRNETs that has continuously received new nodes is the PTEN network, indicating that this protein may be strongly regulated by ceRNAs. This fact is illustrated by a ceRNET constructed based on glioblastoma data, where PTEN showed a total of 534 interactions. Among the genes that may modulate PTEN levels are ABDHD13, CTBP2, HIAT1, HIF1A, KLF6, NRAS, RB1, TAF5 and TNKS2, aside from the known drivers of glioma tumorigenesis and glioblastoma subtypes PDGFRA, RUNX1, STAT3 and VEGFA (Sumazin et al. 2011). Other PTEN ceRNAs already described can be accessed in Table 1.

There is also a crosstalk among ceRNETs and transcription factors (TF). One of these competition networks includes CYP4Z1, its pseudogene CYP4Z2P and CDK3. CYP4Z1 and

CYP4Z2P compete with each other for the ligation of hsa-miR-125a-3p, regulating ERα activity (Zheng et al. 2015; Zheng et al. 2016). This regulation, however, is not direct. CYP4Z1 and CYP4Z2P also act as ceRNAs of CDK3, which phosphorylates ERα, promoting its transcriptional activity. As as result of this interaction network, Erα target genes, as c-Myc, Cyclin D1 and TFF1, are upregulated (Zheng et al. 2016). The lncRNA HULC also integrates a small ceRNET that involves a TF. HULK competes with the PRKACB mRNA for the ligation of miR-372. PRKACB protein, in its turn, promotes the phosphorylation of CREB, which binds the HULK promoter and induces its transcription (Wang et al. 2010).

Beyond the small networks reported above, large-scale ceRNETs have also been constructed and analyzed, together with the proposition of computational methods to build them (Nitzan et al. 2014; Paci et al. 2014; Zhou et al. 2014b; Xu et al. 2015). These ceRNETs may be useful to understand the crosstalk among ceRNAs in the whole cell context. In fact, experiments showed that the modulation of one ceRNET element can cause a propagated perturbation effect on the network (Nitzan et al. 2014). The introduction of a synthetic target for one of the network miRNAs (named here as X) promoted an elevation of the endogenous targets levels. This observation is an expected result, as the synthetic target probably acted as a ceRNA. Interestingly, levels of mRNAs which that are not targets of X, but share other miRNA target sites with X-targeted mRNAs, had their levels also increased (Nitzan et al. 2014). This is a clear evidence of crosstalk between distant ceRNAs in the network.

In addition, research studies aiming to find new drug therapy targets and predictive or prognostic cancer markers may also be benefitted from the construction of large-scale ceRNETs. Comparison of ceRNETs from normal and cancerous cells showed that the miRNA-mediated

interactions change completely from one condition to another (Paci et al. 2014). ceRNETs for 20 cancer types revealed that cancers originating from similar tissues share common ceRNAs. In addition, ceRNA interactions common to the 20 cancer types studied were also detected (Xu et al. 2015). Moreover, supporting the usefulness of ceRNETs in cancer research, the analysis of one breast-cancer-specific ceRNA network encountered 12 hub genes with strong metastasis prediction potential (Zhou et al. 2014b).

The network characteristics of mRNAs ceRNETs have also been studied. An experimentally determined network, including the miRNAs and the mRNAs that compete for them, showed a scale-free topology, which is a characteristic of most biological networks, and its nodes are highly interconnected (Nitzan et al. 2014).

The studies described above clearly show a movement of the ceRNA research towards to large-scale investigations. However, the large-scale networks already reported were constructed with only mRNAs or mRNAs and lncRNAs. A more realistic ceRNET should also include other RNAs that also act as competitors, such as pseudogenes , and circRNAs.

5. Conclusions

The study of competing endogenous RNAs is a young research area, and the knowledge about this mechanism is still sparse in the literature. However, analyzing all of the reported ceRNA cases permitted some conclusions about characteristics that may be universal for this type of RNA-mediated miRNA regulation. The general reciprocal crosstalk between the ceRNAs, their possible influence on the involved miRNA levels, the impact on competition occurrence and

strength of factors such as ceRNA and miRNA levels, their subcellular localizations, the number of shared miRNAs or MREs, and the characteristics of the miRNA target site, such as seed length, miRNA binding strength and level of complementarity in the miRNA-MRE base pairing, can be highlighted.

Although most of the ceRNA interactions described up to now involve only two competing RNAs, progress in this field indicates that the competition mechanism is organized into networks. The ceRNA field is still in its infancy, but it is clear that a movement towards large-scale investigations will contribute to our understanding of the real dimension and importance of this new level of gene expression regulation. In this sense, the advance in the research areas for ncRNAs, unraveling unknown nodes and connections, and on methods of competitors network construction and analysis will probably reveal important biological information on the role of ceRNAs in cells.

6. Acknowledgments

This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) under Grant 301149/2012-7.

7. Author Disclosure Statement

No competing financial interests exist.

8. References

Ala, U. et al., 2013. Integrated transcriptional and competitive endogenous RNA networks are cross-regulated in permissive molecular environments. Proceedings of the National Academy of Sciences of the United States of America, 110(18), 7154–7159.

Bosson, A.D., Zamudio, J.R. and Sharp, A., 2014. Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition. Molecular Cell, 56(3), 347–359.

Cabili, M.N. et al., 2015. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. Genome Biology, 16(1), 20.

Cao, C. et al., 2016. The long non-coding RNA, SNHG6-003, functions as a competing endogenous RNA to promote the progression of hepatocellular carcinoma. Oncogene, 36(8),1112-1122.

Cesana, M. et al., 2011. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell, 147(2), 358–369.

Chang, L. et al., 2016. Upregulation of SNHG6 regulates ZEB1 expression by competitively binding miR-101-3p and interacting with UPF1 in hepatocellular carcinoma. Cancer Letters, 383(2), 183–194.

Chen, L. et al., 2016a. Long Non-Coding RNA CCAT1 acts as a competing endogenous RNA to regulate cell growth and differentiation in acute myeloid leukemia. Molecules and Cells, 39(4), 330–336.

Chen, X. et al., 2016b. Downregulated pseudogene CTNNAP1 promote tumor growth in human cancer by downregulating its cognate gene CTNNA1 expression. Oncotarget, 7(34),55518-55528.

Deng, L. et al., 2015. Long noncoding RNA CCAT1 promotes hepatocellular carcinoma progression by functioning as let-7 sponge. Journal of experimental and clinical cancer research , 34, 18.

Dong, L. and Hui, L., 2016. HOTAIR promotes proliferation, migration, and invasion of ovarian cancer SKOV3 cells through regulating PIK3R3. Medical science monitor : international medical journal of experimental and clinical research, 22, 325–331.

Ebert, M.S. and Sharp, A., 2010. Emerging roles for natural microRNA sponges. Current Biology, 20(19),R858-R8561.

Esposito, F. et al., 2015. HMGA1-pseudogene expression is induced in human pituitary tumors. Cell cycle, 14(9), 1471–1475.

Eulalio, A., Huntzinger, E. and Izaurralde, E., 2008. Getting to the root of miRNA-mediated gene silencing. Cell, 132(1), 9–14.

Fang, L. et al., 2013. Versican 3'-untranslated region (3'-UTR) functions as a ceRNA in inducing the development of hepatocellular carcinoma by regulating miRNA activity. FASEB Journal, 27(3), 907–919.

Figliuzzi, M., Marinari, E. and De Martino, A., 2013. MicroRNAs as a selective channel of communication between competing RNAs: A steady-state theory. Biophysical Journal, 104(5), 1203–1213.

Franco-Zorrilla, J.M. et al., 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. Nature Genetics, 39(8), 1033–1037.

Gao, L. et al., 2016. PTENp1, a natural sponge of miR-21, mediates PTEN expression to inhibit the proliferation of oral squamous cell carcinoma. Molecular Carcinogenesis, doi: 10.1002/mc.22594. [Epub ahead of print].

de Giorgio, A. et al., 2013. Emerging roles of competing endogenous RNAs in cancer: insights from the regulation of PTEN. Molecular and cellular biology, 33(20), 3976–3982.

Guo, G. et al., 2014. A long noncoding RNA critically regulates Bcr-Abl-mediated cellular transformation by acting as a competitive endogenous RNA. Oncogene, 34, 1–12.

Gupta, K., 2014. Competing endogenous RNA (ceRNA): A new class of RNA working as miRNA sponges. Current Science, 106(6), 823–830.

Han, X. et al., 2015. Malat1 regulates serum response factor through miR-133 as a competing endogenous RNA in myogenesis. FASEB Journal, 29(7), 3054–3064.

Hansen, T.B. et al., 2013. Natural RNA circles function as efficient microRNA sponges. Nature, 495(7441), 384–388.

Huang, G. et al., 2016. The long noncoding RNA CASC2 functions as a competing endogenous RNA by sponging miR-18a in colorectal cancer. Scientific Reports, 6, 26524.

Jiao, C. et al., 2016. lncRNA-UCA1 enhances cell proliferation through functioning as a ceRNA of Sox4 in esophageal cancer. Oncology Reports, 2960–2966.

Kallen, A.N. et al., 2013. The imprinted H19 lncRNA antagonizes Let-7 microRNAs. Molecular Cell, 52(1), 101–112.

Karreth, F.A. et al., 2011. In vivo identification of tumor- suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. Cell, 147(2), 382–395.

Karreth, F.A. et al., 2015. The BRAF pseudogene functions as a competitive endogenous RNA and induces lymphoma in vivo. Cell, 161(2), 319–332.

Kartha, R. V. and Subramanian, S., 2014. Competing endogenous RNAs (ceRNAs): New entrants to the intricacies of gene regulation. Frontiers in Genetics, 5,8.

Kimura, T. et al., 2015. Interferon-alpha competing endogenous RNA network antagonizes microRNA-1270. Cellular and Molecular Life Sciences, 72(14), 2749–2761.

Lee, Y. et al., 2002. MicroRNA maturation: stepwise processing and subcellular localization. EMBO Journal, 21(17), 4663–4670.

Li, D. et al., 2015a. OCT4B modulates OCT4A expression as ceRNA in tumor cells. Oncology Reports, 33(5), 2622–2630.

Li, F. et al., 2015b. Circular RNA ITCH has inhibitory effect on ESCC by suppressing the Wnt/beta-catenin pathway. Oncotarget, 6(8), 6001–6013.

Liang, W.-C. et al., 2016. H19 activates Wnt signaling and promotes osteoblast differentiation by functioning as a competing endogenous RNA. Scientific Reports, 6, 20121.

Liang, W.-C. et al., 2015. The lncRNA H19 promotes epithelial to mesenchymal transition by functioning as miRNA sponges in colorectal cancer. Oncotarget, 6(26), 22513–25.

Liao, J.Y. et al., 2010. Deep sequencing of human nuclear and cytoplasmic small RNAS reveals an unexpectedly complex subcellular distribution of mirnas and tRNA 3’ trailers. PLoS ONE,5(5),e10563.

Ling, S. et al., 2013. Dickkopf-1 (DKK1) phosphatase and tensin homolog on chromosome 10 (PTEN) crosstalk via microRNA interference in the diabetic heart. Basic Research in Cardiology, 108(3),352.

Liu, X.H. et al., 2014. Lnc RNA HOTAIR functions as a competing endogenous RNA to regulate HER2 expression by sponging miR-331-3p in gastric cancer. Molecular Cancer, 13(1), 92.

Liu, Q. et al., 2016a. Circular RNA related to the chondrocyte ECM regulates MMP13 expression by functioning as a miR-136 “sponge” in human cartilage cegradation. Scientifc Reports, 6, 22572.

Liu, T. et al., 2016b. PIK3C2A mRNA functions as a miR-124 sponge to facilitate CD151 expression and enhance malignancy of hepatocellular carcinoma cells. Oncotarget, 7(28),43376-43389.

Liu, X. et al., 2016c. The Mechanism of long non-coding RNA MEG3 for neurons apoptosis caused by hypoxia: mediated by miR-181b-12/15-LOX signaling pathway. Frontiers in Cellular Neuroscience, 10, 201.

Lü, M. et al., 2015. Long noncoding RNA BC032469, a novel competing endogenous RNA, upregulates hTERT expression by sponging miR-1207-5p and promotes proliferation in gastric cancer. Oncogene, 35(27),3524-3534.

Lu, W. et al., 2015. A 3'UTR-associated RNA, FLJ11812 maintains stemness of human embryonic stem cells by targeting miR-4459. Stem cells and development, 24(9), 1133–1140.

Luan, W. et al., 2016. Long non-coding RNA MALAT1 acts as a competing endogenous RNA to promote malignant melanoma growth and metastasis by sponging miR-22. Oncotarget, 7(39), 63901–63912.

Mangiavacchi, A. et al., 2016. The miR-223 host non-coding transcript linc-223 induces IRF4 expression in acute myeloid leukemia by acting as a competing endogenous RNA. Oncotarget, 7(37),60155-60168.

Maniataki, E. and Mourelatos, Z., 2005. A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. Genes and Development, 19(24), 2979–2990.

Nitzan, M. et al., 2014. Interactions between distant ceRNAs in regulatory networks. Biophysical Journal, 106(10), 2254–2266.

Paci, P., Colombo, T. and Farina, L., 2014. Computational analysis identifies a sponge interaction network between long non-coding RNAs and messenger RNAs in human breast cancer. BMC systems biology, 8, 83.

Peng, W. et al., 2015. Long non-coding RNA MEG3 functions as a competing endogenous RNA to regulate gastric cancer progression. Journal of Experimental and Clinical Cancer Research, 34(79), 1–10.

Pilyugin, M. and Irminger-Finger, I., 2014. Long non-coding RNA and microRNAs might act in regulating the expression of BARD1 mRNAs. International Journal of Biochemistry and Cell Biology, 54, 356–367.

Poliseno, L. et al., 2010. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature, 465(7301), 1033–1038.

Qu, L. et al., 2016. Exosome-transmitted lncARSR promotes sunitinib resistance in renal cancer by acting as a competing endogenous RNA. Cancer Cell, 29(5), 653–668.

Ren, K. et al., 2016. Long noncoding RNA HOTAIR controls cell cycle by functioning as a competing endogenous RNA in esophageal squamous cell carcinoma. Translational Oncology, 9(6), 489–497.

Salmena, L. et al., 2011. A ceRNA hypothesis: the rosetta stone of a hidden RNA language? Cell, 146(3), 353–358.

Su, Z. et al., 2016. LncRNA H19 functions as a competing endogenous RNA to regulate AQP3 expression by sponging miR-874 in the intestinal barrier. FEBS Letters, 590(9), 1354–1364.

Sumazin, et al., 2011. An extensive MicroRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. Cell, 147(2), 370–381.

Sun, C. et al., 2016. Long non-coding RNA NEAT1 promotes non-small cell lung cancer progression through regulation of miR-377-3p-E2F3 pathway. Oncotarget,7(32),51784-51814.

Tay, Y. et al., 2011. Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. Cell, 147(2), 344–357.

Tay, Y., Rinn, J. and Pandolfi, P.P.,2014. The multilayered complexity of ceRNA crosstalk and competition. Nature, 505(7483), 344–52.

Venø, M.T. et al., 2015. Spatio-temporal regulation of circular RNA expression during porcine embryonic brain development. Genome biology, 16, 245.

Wang, J. et al., 2010. CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. Nucleic Acids Research, 38(16), 5366–5383.

Wang, K. et al., 2014. CARL lncRNA inhibits anoxia-induced mitochondrial fission and apoptosis in cardiomyocytes by impairing miR-539-dependent PHB2 downregulation. Nature communications, 5, 3596.

Wang, L. et al., 2013. Pseudogene OCT4-pg4 functions as a natural micro RNA sponge to regulate OCT4 expression by competing for miR-145 in hepatocellular carcinoma. Carcinogenesis, 34(8), 1773–1781.

Wang, S.H. et al., 2016a. The lncRNA MALAT1 functions as a competing endogenous RNA to regulate MCL-1 expression by sponging miR-363-3p in gallbladder cancer. Journal of Cellular and Molecular Medicine. 20(12),2299-2308.

Wang, S.H. et al., 2016b. Long non-coding RNA H19 regulates FOXM1 expression by competitively binding endogenous miR-342-3p in gallbladder cancer. Journal of Experimental and Clinical Cancer Research, 35(1), 160.

Wang, S.H. et al., 2016c. Long non-coding RNA Malat1 promotes gallbladder cancer development by acting as a molecular sponge to regulate miR-206. Oncotarget, 7(25),37857-37867.

Wang, Y. et al., 2013. Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. Developmental Cell, 25(1), 69–80.

Welch, J.D. et al., 2015. Pseudogenes transcribed in breast invasive carcinoma show subtype-specific expression and ceRNA potential. BMC genomics, 16(1), 113.

Xia, T. et al., 2014. Long noncoding RNA associated-competing endogenous RNAs in gastric cancer. Scientific reports, 4, 6088.

Xia, T. et al., 2015. Long noncoding RNA FER1L4 suppresses cancer cell growth by acting as a competing endogenous RNA and regulating PTEN expression. Scientific reports, 5, 13445.

Xiao, H. et al., 2015. LncRNA MALAT1 functions as a competing endogenous RNA to regulate ZEB2 expression by sponging miR-200s in clear cell kidney carcinoma. Oncotarget, 6(350,38005-38015.

Xie, C.-H. et al., 2016. Long non-coding RNA TUG1 contributes to tumorigenesis of human osteosarcoma by sponging miR-9-5p and regulating POU2F1 expression. Tumor Biology, 37(11), 15031–15041.

Xu, J. et al., 2015. The mRNA related ceRNA-ceRNA landscape and significance across 20 major cancer types. Nucleic Acids Research, 43(17), 8169–8182.

Yan, B. et al., 2015. LncRNA-MIAT regulates microvascular dysfunction by functioning as a competing endogenous RNA. Circulation Research, 116(7), 1143–1156.

Yang, W. et al., 2015. Foxo3 activity promoted by non-coding effects of circular RNA and Foxo3 pseudogene in the inhibition of tumor growth and angiogenesis. Oncogene, 35(30),3919-3931.

Yang, Y. et al., 2016. UCA1 functions as a competing endogenous RNA to suppress epithelial ovarian cancer metastasis. Tumor Biology, 37(8), 10633–10641.

Yiwei, T. et al., 2015. HOTAIR Interacting with MAPK1 Regulates Ovarian Cancer skov3 Cell Proliferation, Migration, and Invasion. Medical science monitor : international medical journal of experimental and clinical research, 21, 1856–1863.

Yu, F. et al., 2015a. Long non-coding RNA growth arrest-specific transcript 5 (GAS5) inhibits liver fibrogenesis through a mechanism of competing endogenous RNA. Journal of Biological Chemistry, 290(47), 28286–28298.

Yu, F. et al., 2015b. Malat1 functions as a competing endogenous RNA to mediate Rac1 expression by sequestering miR-101b in liver fibrosis. Cell Cycle, 14(24), 3885–3896.

Yuan, J. et al., 2014. A long noncoding RNA activated by TGF-β promotes the invasion-metastasis cascade in hepatocellular carcinoma. Cancer cell, 25(5), 666–681.

Yuan, Y. et al., 2015. Model-guided quantitative analysis of microRNA-mediated regulation on competing endogenous RNAs using a synthetic gene circuit. Proceedings of the National Academy of Sciences of the United States of America, 112(10), 3158–3163.

Yue, B. et al., 2016. Linc00152 functions as a competing endogenous RNA to confer oxaliplatin resistance and holds prognostic values in colon cancer. Molecular Therapy, 24(12), 2064–2077.

Zheng, L. et al., 2015. The 3′UTR of the pseudogene CYP4Z2P promotes tumor angiogenesis in breast cancer by acting as a ceRNA for CYP4Z1. Breast Cancer Research and Treatment, 150(1), 105–118.

Zheng, L. et al., 2016. Competing endogenous RNA networks of CYP4Z1 and pseudogene CYP4Z2P confer tamoxifen resistance in breast cancer. Molecular and Cellular Endocrinology, 427, 133–142.

Zhou, B. et al., 2014a. Identification of a splicing variant that regulates type 2 diabetes risk factor CDKAL1 level by a coding-independent mechanism in human. Human Molecular Genetics, 23(17), 4639–4650.

Zhou, X., Liu, J. and Wang, W., 2014b. Construction and investigation of breast-cancer-specific ceRNA network based on the mRNA and miRNA expression data. IET systems biology, 8(3), 96–103.

Capítulo III

Non-coding RNAs putatively acting as ceRNAs in human embryonic stem cells

Artigo a ser submetido ao periódico *Developmental Cell*

Non-coding RNAs putatively acting as ceRNAs in human embryonic stem cells

Raquel Calloni and Diego Bonatto

Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul, Departamento de Biologia Molecular e Biotecnologia, Universidade Federal do Rio Grande do Sul

Abstract

RNAs competing with each other for the binding of miRNAs have been proposed as a new level of gene expression control. Several examples concerning RNA competition have been reported in the literature, but few cases were identified in embryonic stem cells (ESCs). Here, we investigated available ESCs-associated RNA-seq datasets in order to propose RNAs that putatively act as competing endogenous RNAs (ceRNAs). Data analyses using different criteria for ceRNAs identification, including Pearson correlation and the investigation of the miRNA seed type and binding strength, showed that the mRNA-ncRNA pairs AHCY-RP11-253E3.3, F2RL1-EEF1A1P6, HSPD1-SNHG5, INO80-RP11-20D14.6, PARP1-RP11-20D14.6 and PRIM2-GAS5, together with the circular RNA circ-HIPK3, have the major potential to compete for miRNAs expressed in ESCs. In addition, these mRNA-ncRNA pairs are proposed as investigation starter point to elucidate the ESCs ceRNAs network.

Keyworks: ceRNAs, miRNA, lncRNA, pseudogene, circRNA, competition

1. Introduction

Approximately 80% of the human genome has an associated biochemical function, and 75% is transcribed into coding or non-coding RNAs[1]. The coding part of the genome is not, however, all transcribed at the same time. The expression pattern of these genome areas is specific for each cell type and is strictly controlled. This control is exerted at different levels and, in 2011, a new layer of gene expression regulation was proposed[2].

This new mechanism involves microRNAs (miRNAs) and coding and non-coding RNAs and is based on the premise that different RNAs can compete for the ligation of miRNAs, modulating the miRNA silencing activity[2]. Thus, some RNAs share miRNA response elements (MREs) with other co-expressed RNAs, acting as “sponges” of the shared miRNAs and allowing the other targets to escape from degradation and exert their cellular functions[2]. Due to their competition behavior, these ribonucleic acids were named as competing endogenous RNAs (ceRNAs)[2]. Among the RNA classes that can act as ceRNAs are mRNAs, long non-coding RNAs (lncRNAs), pseudogenes and circular RNAs (circRNAs)[3,4,5]. The most known and studied ceRNA is PTENP1, a processed pseudogene of PTEN[4]. Upregulation of PTENP1 results in the consequent rise in PTEN expression; conversely, PTENP1 downregulation leads to the diminishment of mRNA and protein levels of PTEN[4]. The expression modulation operated by PTENP1 is not observed, however, in cells lacking Dicer, implying that miRNAs must be involved in the process[4].

The competition mechanism seems to be present in several cell types, including cardiomyocytes, adipocytes, osteoblasts, neurons, cancer cells, and embryonic stem cells (ESCs)[6–11]. ESCs, in turn, are largely studied and known by their long-term self-renewing and their capacity to differentiate into cells from the three primary germ layers[12]. This characteristic makes ESCs

interesting for application in cell therapy and studies on human development[13,14]. Additionally, the knowledge of the molecular mechanisms involved in the stemness maintenance in ESCs may be useful to improve the technology of induced pluripotent stem cell (iPSC) generation.

The research field on ESCs has been improving since they were first isolated, and the same trend has occurred with the understanding of the competition between RNAs. It has been postulated that this competition occurs among several RNAs, forming a network named ceRNET. However, this mechanism has been poorly explored regarding its role in ESCs, and only a few nodes of the ESC ceRNET were elucidated: until now, only three lncRNAs were reported to act as competitors in ESCs, all of them being implicated in stemness maintenance[11,15,16].

Thus, the aim of this study was to search for, using in silico tools, ncRNAs that may act as ceRNAs in the context of ESCs. The proposed ceRNAs were investigated regarding their involvement in stemness maintenance and were proposed as start points for in vivo studies to elucidate the ESC ceRNET.

2. Materials and Methods

2.1. RNA-seq data

Four RNA-seq data were used in this study: (i) RNA-seq data from embryonic stem cells (RUES2) and cardiomyocytes differentiated from them (GSE64417), (ii) RNA-seq data from WA-09 embryonic stem cells lineage and prefrontal cortex neurons differentiated from them (GSE56796), (iii) RNA-seq data from H9 embryonic stem cells lineage and retinal ganglion cells

differentiated from them (GSE84639), and (iv) RNA-seq data from peripheral blood mononuclear cells-derived iPSCs and macrophages differentiated from them (GSE55536). All the libraries were generated from total RNA and are paired-end, except for those from the datasets GSE55536, which was originated from polyA RNA, and GSE56796, which is single-end.

2.2. Bioinformatics analysis

2.2.1. Reads alignment

Non-trimmed reads were aligned to the human genome (version GRCh38) using the software STAR, version 2.4.2a, following the default parameters suggested by the developers[17].

2.2.2. Differential expression analysis

Count files were generated from the alignment BAM files using the software HTSeq (version 0.6.1)[18]. These files were then used as input to the R package DESeq2[19] to perform the identification of differentially expressed genes (DEGs). It was considered differentially expressed those genes which presented $|\log_2 FC| \geqslant 1$ and adjusted p value < 0.05, estimated by the Benjamini-Hochberg method.

Among all of the DEGs detected, those upregulated in ESCs and also coding for miRNAs, as well as mRNAs, pseudogenes or lncRNAs predicted or validated as targets of those miRNAs, were considered for further analysis.

2.2.3. Circular RNA identification

The identification of circRNAs expressed in ESCs was performed through the alignment of the reads using the software MapSplice2[20]. The parameters *--min-fusion-distance*, set to 200, and *--gene-gtf*, required for the circRNA detection, were included in the default command line.

2.2.4. miRNA target identification

The pseudogenes and lncRNAs targeted by the miRNAs were identified using two approaches. Predicted targets were prospected using the miRcode database (version 11, released in 2012). As the list of upregulated ncRNAs was large, only the 50 most upregulated lncRNAs and pseudogenes were investigated for their binding to miRNAs. Validated targets were obtained from starBase (version 2.0, released in 2013). In this case, all ncRNAs targeted by a given miRNA were downloaded, and this list was compared to the pseudogenes and lncRNAs upregulated in ESCs.

In the case of mRNAs, the predicted and validated target mRNAs were identified using the databases TargetScanHuman (release 6.2, from 2012) and starBase (version 2,0, released in 2013), respectively. In the case of targets obtained from TargetScanHuman, only the targets with aggregate PCT (probability of conserved targeting) > 0.5 were considered.

Regarding circRNAs, the sequences of the detected molecules were screened to identify putative miRNA binding sites using the web-based tool miRmap (http://mirmap.ezlab.org/app/)[21].

2.2.5. Gene location investigation

The chromosomal gene location investigation was performed using the search mechanism available at the Ensembl genome browser, release 87, and human genome version GRCh38.

2.2.6. Transcription factor binding site identification

Transcription factor (TF) binding sites present in the sequence located at 1 kb from each side of the transcription start site (TSS) of the investigated genes were predicted using the Jaspar database[22], applying human TF matrix models. From the identified TFs, only those that were also observed as upregulated in ESCs [dataset (i)] were considered for further analysis.

2.2.7. Gene Ontology analysis

Gene ontology analysis was performed using the package Goseq, following the developer's instructions[23]. The probability weighting function (PWF) was estimated from the data distinguishing differentially and non-differentially expressed genes between ESCs and differentiated cells. The Wallenius approximation was calculated only for the differentially expressed genes of interest. All of the ontologies with false discovery rate (FDR) < 0.05 were considered significant.

2.2.8. Isoform identification

The detection of the ESC-expressed isoforms of the genes of interest was performed using the software Salmon in the quasi-mapping-based mode, applying the default parameters suggested by the developers[24].

Two fasta files, one containing the transcripts produced by coding genes and the other with ncRNAs transcripts, were used to build the indexes and perform the Salmon analysis. Both files were downloaded from the Ensembl genome browser, release 79.

2.2.9. Number of binding sites and miRNA binding strength detected isoforms

To investigate the presence and the number of relevant miRNA binding sites in the isoforms expressed in ESCs, the sequence of each isoform was screened using the web-based tool miRmap (http://mirmap.ezlab.org/app/)[21].

In addition, information about the seed type and the strength of the miRNA ligation to the putative site, measured by the parameters seed length and ΔG binding, was also recorded.

2.3. Statistical analysis

2.3.1 Pearson correlation

The coefficient and the associated p value of the correlation between mRNAs and pseudogenes or lncRNAs targeted by a common miRNA were estimated using the function *rcorr* of the R package *Hmisc*.

The numeric data used to estimate the correlation was the number of read counts for each gene that was normalized using the DESeq2 function *varianceStabilizingTransformation.* Only genes presenting transformed read counts with normal distribution were used in the analysis.

2.3.2 Partial correlation and estimation of the sensitivity correlation (S)

The partial correlation between mRNAs and pseudogenes or lncRNAs controlling for the TFs shared by them was estimated using the function *pcor* of the R package *ggm*. Because of the large number of shared TFs between mRNAs and ncRNAs, only the five most expressed ones were chosen for this analysis.

The sensitivity correlation (S) parameter was estimated by the difference between the Pearson and partial correlation coefficients, as defined by Paci et al., (2014)[25] **:**

$$S = \mathrm{corr}(mRNA,ncRNA) - \mathrm{corr}(mRNA,ncRNA|5TFs).$$

When the Pearson correlation is equal to the partial correlation (S=0), the correlation between the mRNA and the ncRNA is considered direct, and there is no influence of shared TFs on the correlation coefficient. When the partial correlation is lower than the Pearson correlation (S>0), it indicates that the correlation between the mRNA and ncRNA is not direct and could be mediated by the TFs. In the case of a null partial correlation, the correlation between the mRNA and the ncRNA is spurious and also possibly due to the shared TFs[25]. Considering that, putative ceRNA pairs with S $\geq$ 0.2 were discarded from the subsequent analysis.

3. Results

3.1. Differentially expressed genes

From all differentially expressed genes detected, a total of 6,221 genes were observed to be upregulated in ESCs compared with differentiated cells in the dataset (i) (Figure 1). Among these DEGs, 317 lncRNAs, 585 pseudogenes and 31 miRNAs were identified (Figure 2; Supplemental Material 1, Tables S1 to S3).

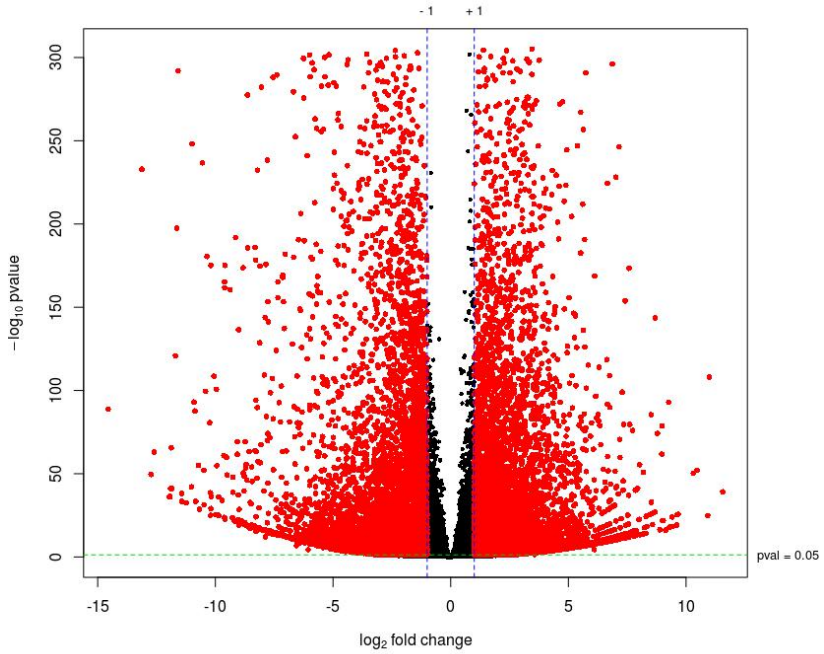


Figure 1: Volcano plot representing the differentially expressed genes detected by comparing ESCs with differentiated cells from the main dataset.

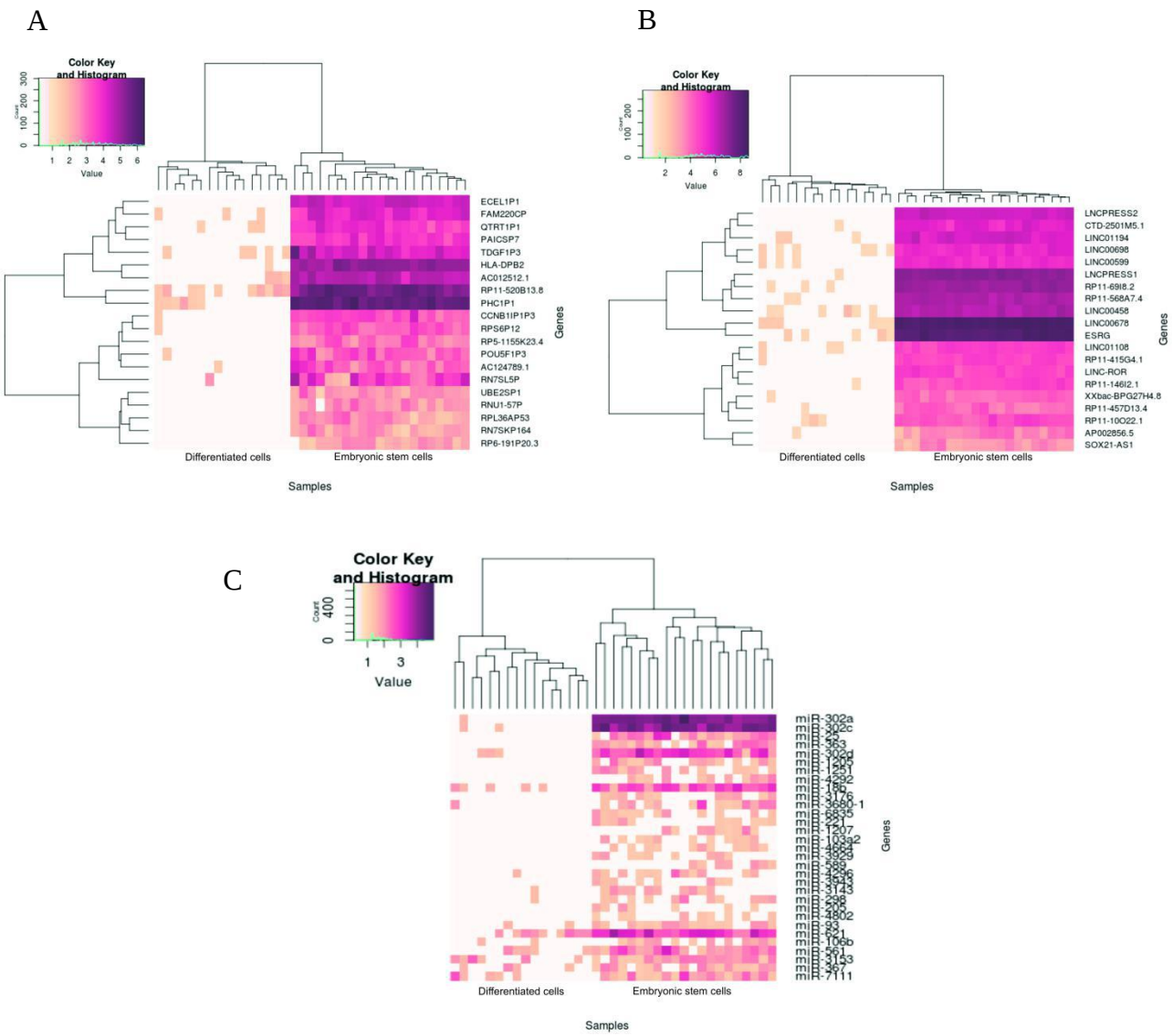


Figure 2: Heatmaps for the most upregulated pseudogenes (A), lncRNAs (B), and miRNAs (C) detected in ESCs.

miRNAs characteristically expressed by ESCs are well defined in the literature, and our analysis succeeded in detecting some of these miRNAs[26–34]. In this sense, only the miRNAs that were already reported in the literature as expressed in ESCs were chosen for further analysis. Therefore, the search for ESC ceRNAs was proceeded with the miRNAs hsa-miR-18b, -25, -93, -103a2, -106b, -205, -302acd, -363 and 367 (Supplemental Material 1, Table S4).

To identify the RNAs that are truly associated with the stem cell state, we compared the main dataset with three other RNA-seq datasets. The expression data was generated from two ESCs datasets (data from ESCs compared with prefrontal cortex neurons or retinal ganglion cells differentiated from them[35,36]) and IPSCs (data from IPSCs compared to macrophages differentiated from them[37]). Accordingly to the data comparison, only differentially expressed mRNAs, lncRNAs and pseudogenes detected concomitantly in the main dataset and at least one of the additional datasets were kept for further analyses.

3.2. ceRNAs prediction

3.2.1. Construction of the predicted ceRNA network

The first step toward predicting the potential ceRNAs in ESCs was to identify which of the ESC-upregulated RNAs are targeted by the miRNAs hsa-miR-18b, -25, -93, -103a2, -106b, -205, -302adc, -363 and/or -367.

The list of validated targets was composed of a total of 15 lncRNAs, 20 pseudogenes and 563 mRNAs. In the case of the predicted targets, 20 lncRNAs, 15 pseudogenes and 227 mRNAs were recovered (Supplemental Material 2). Using these lists, two networks of predicted and validated targets expressed in ESCs were constructed. The network resulting from the merge of these two (Figure 3) was used to perform a gene ontology (GO) analysis. Among the significant GOs detected are those related to the cell cycle, stem cell population maintenance, negative regulation of

differentiation, telomere maintenance and signaling pathways previously reported as being important to stem cell maintenance (Table 1 and Supplemental Material 3)[38–41].

Table 1: Significant ESC-related gene ontologies (FDR≤ 0.05) observed in the merged network of the chosen miRNAs targets.

Gene ontology ID	Biological process
GO:0007346	regulation of mitotic cell cycle
GO:0045787	positive regulation of cell cycle
GO:0051781	positive regulation of cell division
GO:0008284	positive regulation of cell proliferation
GO:0019827	stem cell population maintenance
GO:0017145	stem cell division
GO:0045595	regulation of cell differentiation
GO:0045596	negative regulation of cell differentiation
GO:0045599	negative regulation of fat cell differentiation
GO:0045665	negative regulation of neuron differentiation
GO:0010771	negative regulation of cell morphogenesis involved in differentiation
GO:0000723	telomere maintenance
GO:0010833	telomere maintenance via telomere lengthening
GO:0032201	telomere maintenance via semi-conservative replication
GO:0032200	telomere organization
GO:0038092	nodal signaling pathway
GO:0043410	positive regulation of MAPK cascade
GO:0070374	positive regulation of ERK1 and ERK2 cascade
GO:0043406	positive regulation of MAP kinase activity
GO:0043408	regulation of MAPK cascade
GO:0000186	activation of MAPKK activity
GO:0007179	transforming growth factor beta receptor signaling pathway
GO:0051092	positive regulation of NF-kappaB transcription factor activity
GO:0043123	positive regulation of I-kappaB kinase/NF-kappaB signaling
GO:0071456	cellular response to hypoxia

GO:0051896	regulation of AKT signaling cascade
GO:0038127	ERBB signaling pathway
GO:0043551	regulation of phosphatidylinositol 3-kinase activity
GO:0043552	positive regulation of phosphatidylinositol 3-kinase activity
GO:0008543	fibroblast growth factor receptor signaling pathway
GO:0044344	cellular response to fibroblast growth factor stimulus"
GO:0071774	response to fibroblast growth factor

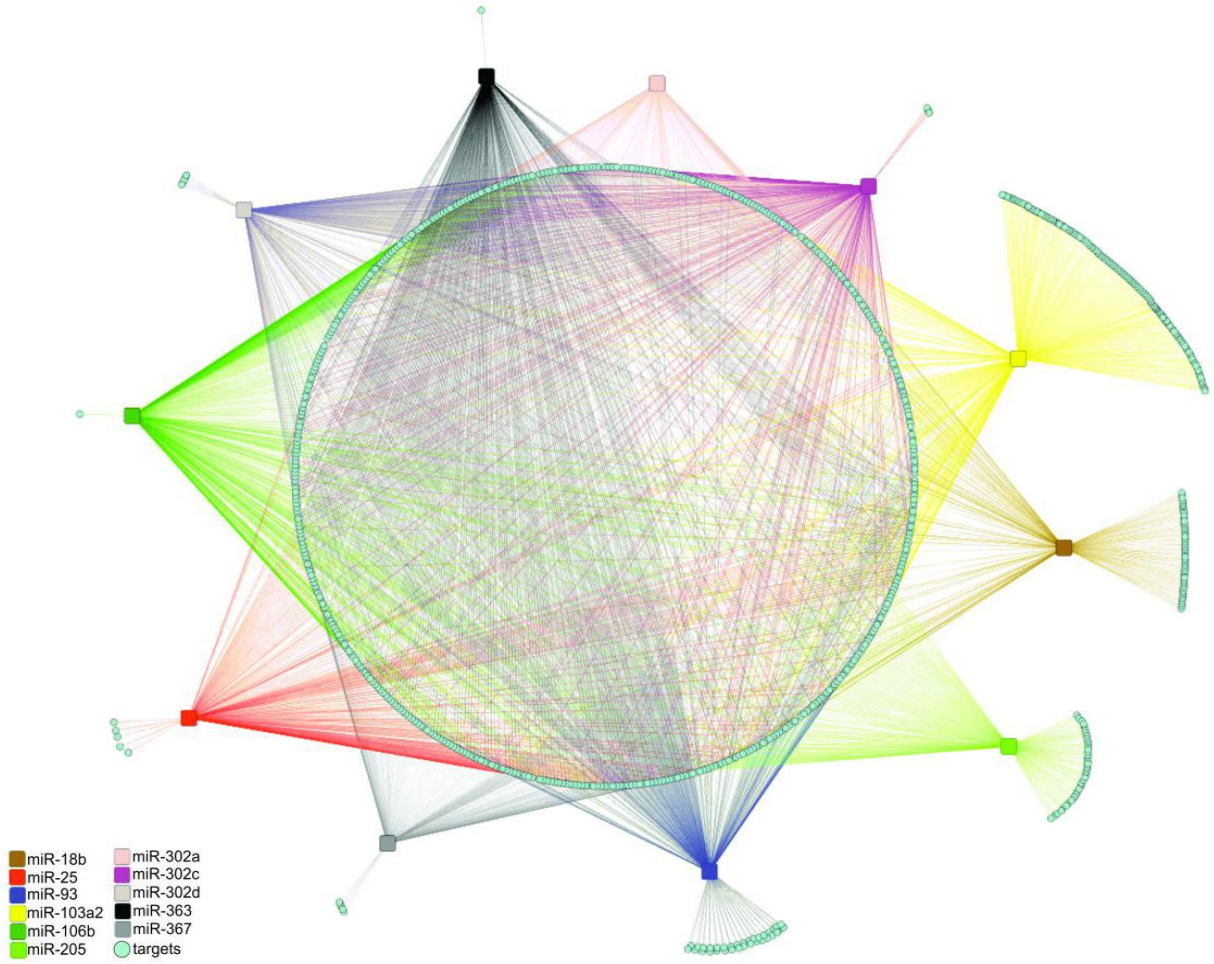


Figure 3: Network resulting from the merge of predicted and validated miRNAs targets networks.

3.2.2. Search for ncRNA-mRNA pairs with significant positive correlation

Using the data gathered in the previous analysis, all the pairs of mRNA-lncRNA or mRNA-pseudogene that share ESC-expressed miRNAs were evaluated using Pearson statistics in order to identify pairs with a significant positive correlation. A data frame containing the normalized read counts for each of the targets (mRNAs, pseudogenes and lncRNAs) was constructed, and, for each miRNA, the correlation among target mRNAs and target pseudogenes/lncRNAs was estimated. RNA pairs with a correlation coefficient equal to or greater than 0.6 ($p \leq 0.05$) were selected. (Supplemental Material 4).

The results indicated 50 lncRNA-mRNA pairs and 13 pseudogene-mRNA pairs from the validated miRNA targets and 9 lncRNA-mRNA pairs and 3 pseudogene-mRNA pairs from predicted miRNA targets. It should be noted that those pairs that were also significantly positively correlated in differentiated cells were not included in further analysis.

3.2.3. The positive correlation between mRNAs and ncRNAs is not due to divergent transcription

To investigate whether the positive correlation between ncRNAs and mRNAs observed in this work is due to shared regulatory sequencies, we examined the chromosomal location of the ncRNAs and of the correlated mRNAs. Among the lncRNAs and pseudogenes selected as putative ceRNAs, only two were located at the same chromosome as their mRNA pair (Figure 4). However, the distance between those ncRNAs and their respective mRNAs (~45 Mbp) was considered to be

too large to permit the sharing of regulatory elements. Therefore, the significant positive correlation observed in the lncRNA-mRNA pairs of this study is probably not due to common regulatory elements between the RNAs of each pair.

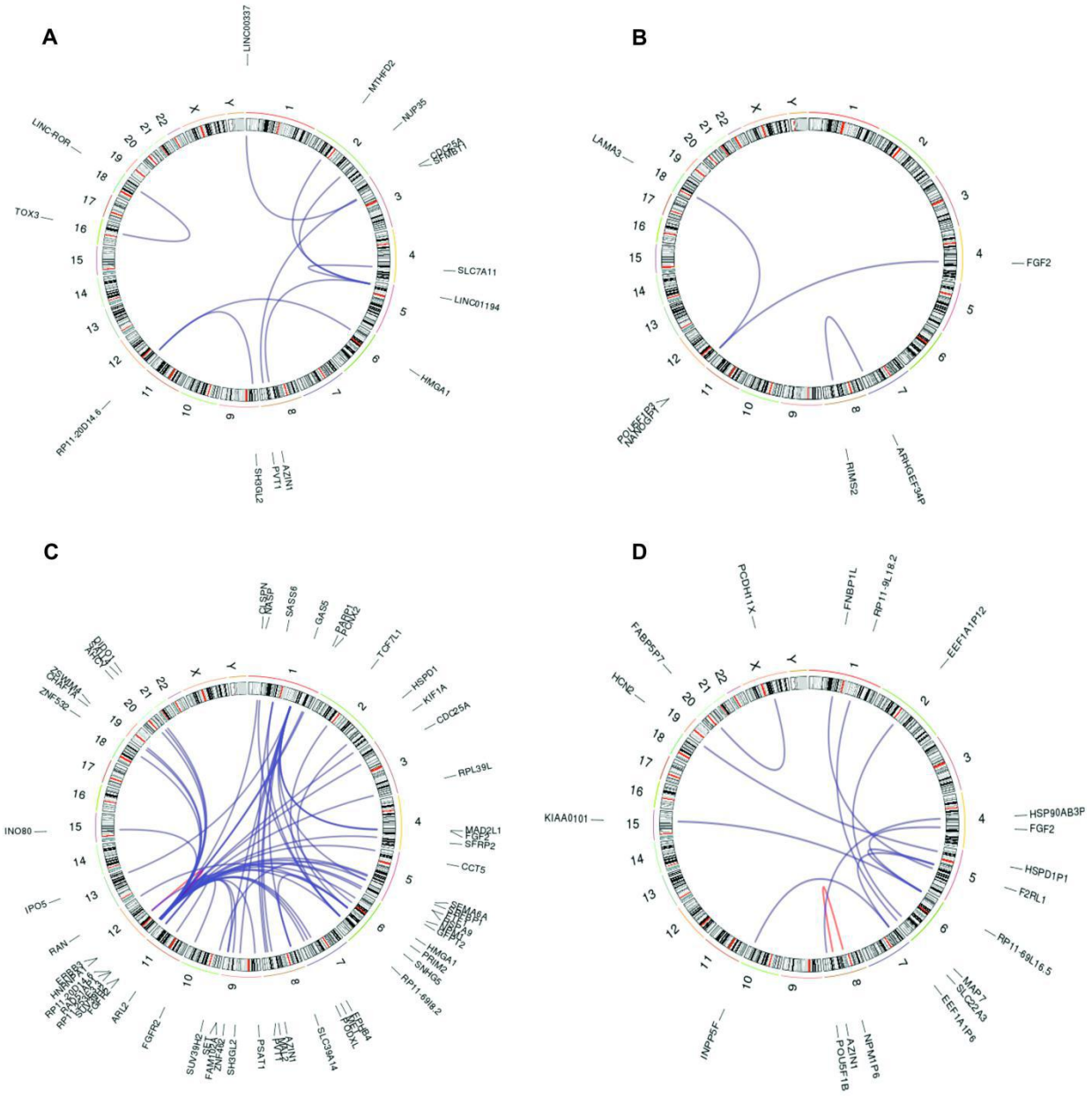


Figure 4: Chromosomal position of the RNAs from the putative ceRNA pairs. Blue lines indicate that each RNA of the pair is located in different chromosomes. Red lines indicate RNAs that are

located on the same chromosome. The data are divided into predicted (A and B) and validated miRNA targets (C and D). A and C: lncRNA-mRNA pairs; B and D: pseudogene-mRNA pairs.

3.2.4 The positive correlation between mRNAs and ncRNAs is not due to the regulation by the same transcription factors

The localization of the RNAs from the putative ceRNA pairs in different chromosomes is not impeditive of the regulation by the same TFs. Therefore, we analyzed the TFs that may putatively regulate the RNAs from each ceRNA pair. A sequence of 2 kb, 1 kb at each side of the TSS, was screened for TFs binding sites. The TFs not expressed in ESCs were excluded from the analysis. Surprisingly, all of the RNAs composing the ceRNA pairs shared an average of 55 common TFs binding sites in the region investigated (Supplemental Material 5). Therefore, we investigated the influence of the common TF binding sites on the correlation observed between mRNAs and ncRNAs estimating the factor S (see Material and Methods section). Among the most expressed TFs shared by the ncRNA-mRNA pairs and used in the estimation of S were VENTX, ETV4, SOX21, TFAP2C, INSM1, OTX2, FOXH1, SIPB and TEAD4.

As result of the analysis, only three ceRNA pairs presented S values superior to the established threshold of 0.2 (Figure 5) and were eliminated from subsequent analysis. These results indicate that the analyzed TFs may not be the responsible for the significant positive correlation observed between the RNAs that comprise the putative ceRNA pairs.

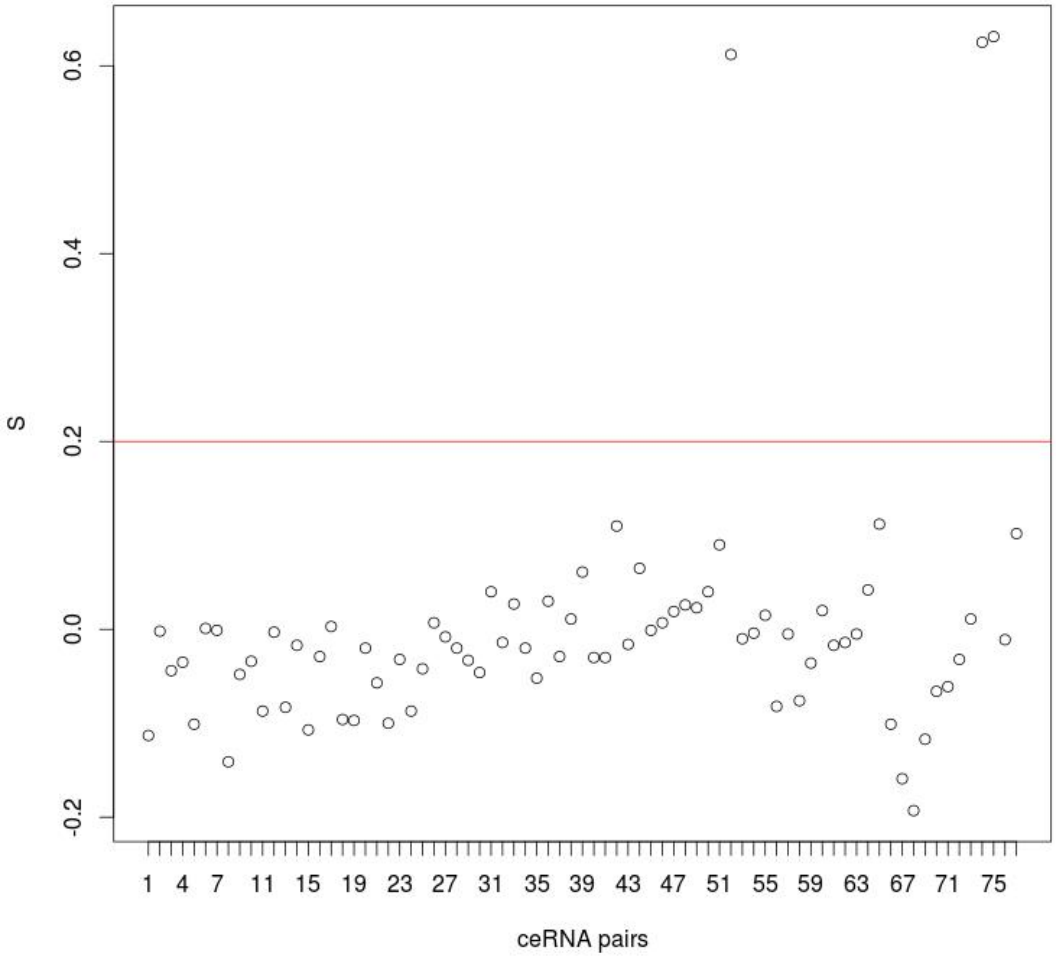


Figure 5: S values for the analyzed putative ceRNA pairs. Each number on the x axis corresponds to a ceRNA pair. 1-52: lncRNA-mRNA pairs from validated targets; 53-65: pseudogene-mRNA pairs from validated targets; 66-74: ncRNA-mRNA pairs from predicted targets; 75-77: pseudogene-mRNA pairs from predicted targets.

3.2.5. Selection of putative ceRNA pairs associated with ESCs maintenance

To select those putative ceRNA pairs that could have a role in the ESCs pluripotency maintenance and proliferation, we performed a GO analysis for those pairs that remained from the

filtering presented in the item above (Table 2 and Supplemental Material 6). ceRNA pairs whose mRNA had an associated ESCs -related ontology were chosen for further analysis (Table 3 and Supplemental Material 7).

In addition, all of the RNAs integrating the pairs are validated miRNA targets (Supplemental Material 2), except for POU5F1P3, which is a predicted target. Moreover, the chosen ceRNA pairs were only significantly positively correlated in ESCs. In other words, the correlation between the RNA pairs only exists in the cells that are expressing the miRNAs they compete for, the ESCs. In differentiated cells, where these miRNAs are not expressed or are downregulated, the correlation does not exist (Figure 6 and Supplemental Material 9, Figure S1). This observation could be considered a strong evidence that the chosen pairs may be real ceRNAs.

Table 2: Embryonic stem cell-related ontologies observed in the filtered list of putative ceRNAs.

Gene ontology ID	Biological process
GO:0007346	regulation of mitotic cell cycle
GO:0045787	positive regulation of cell cycle
GO:0090068	positive regulation of cell cycle process
GO:0051302	regulation of cell division
GO:0019827	stem cell population maintenance
GO:0017145	stem cell division
GO:0042659	regulation of cell fate specification
GO:0010453	regulation of cell fate commitment
GO:0000723	telomere maintenance
GO:0032200	telomere organization
GO:0043410	positive regulation of MAPK cascade
GO:0070374	positive regulation of ERK1 and ERK2 cascade
GO:0008543	fibroblast growth factor receptor signaling pathway
GO:0044344	cellular response to fibroblast growth factor stimulus"

GO:0071774	response to fibroblast growth factor
GO:0038127	ERBB signaling pathway
GO:0014068	positive regulation of phosphatidylinositol 3-kinase signaling
GO:0014066	regulation of phosphatidylinositol 3-kinase signaling
GO:0070371	ERK1 and ERK2 cascade
GO:0014065	phosphatidylinositol 3-kinase signaling
GO:0001666	response to hypoxia

Table 3: Putative ceRNA pairs possibly related to the ESCs proliferation and pluripotency maintenance.

mRNA	ncRNA	miRNAs	Related ontology
AHCY	RP11-253E3.3	hsa-miR-106b	response to hypoxia
ERBB3	RP11-69I8.2	hsa-miR-106b	positive regulation of MAPK cascade positive regulation of ERK1 and ERK2 cascade fibroblast growth factor receptor signaling pathway ERBB signaling pathway positive regulation of phosphatidylinositol 3-kinase signaling
F2RL1	EEF1A1P6 and RP11-69L16.5	hsa-miR-103a2	positive regulation of MAPK cascade positive regulation of ERK1 and ERK2 cascade positive regulation of phosphatidylinositol 3-kinase signaling
FGF2	RP11-69L16.5, POU5F1P3 and GAS5	hsa-miR-103a2 and -205	stem cell population maintenance regulation of cell fate specification regulation of cell fate commitment regulation of cell division
FGFR2	RP11-20D14.6	hsa-miR-103a2	positive regulation of cell cycle positive regulation of MAPK cascade positive regulation of ERK1 and ERK2 cascade fibroblast growth factor receptor signaling pathway ERBB signaling pathway
HSPD1	SNHG5	hsa-miR-25, -363, -367 and -302ad	response to hypoxia
INO80	RP11-253E3.3	hsa-miR-93 and -106b	regulation of mitotic cell cycle

NPM1	RP11-20D14.6	hsa-miR-25 and -302acd	positive regulation of cell cycle
PARP1	RP11-20D14.6	hsa-miR-103a2 and -302acd	telomere maintenance telomere organization
PRIM2	GAS5	has-miR-18b	telomere maintenance telomere organization
SALL4	RP11-20D14.6	hsa-miR-103a2	stem cell population maintenance
SFRP2	RP11-20D14.6	hsa-miR-205	regulation of cell fate specification regulation of cell fate commitment stem cell division regulation of cell fate specification regulation of cell division
TCF7L1	RP11-69I8.2	hsa-miR-93	stem cell population maintenance

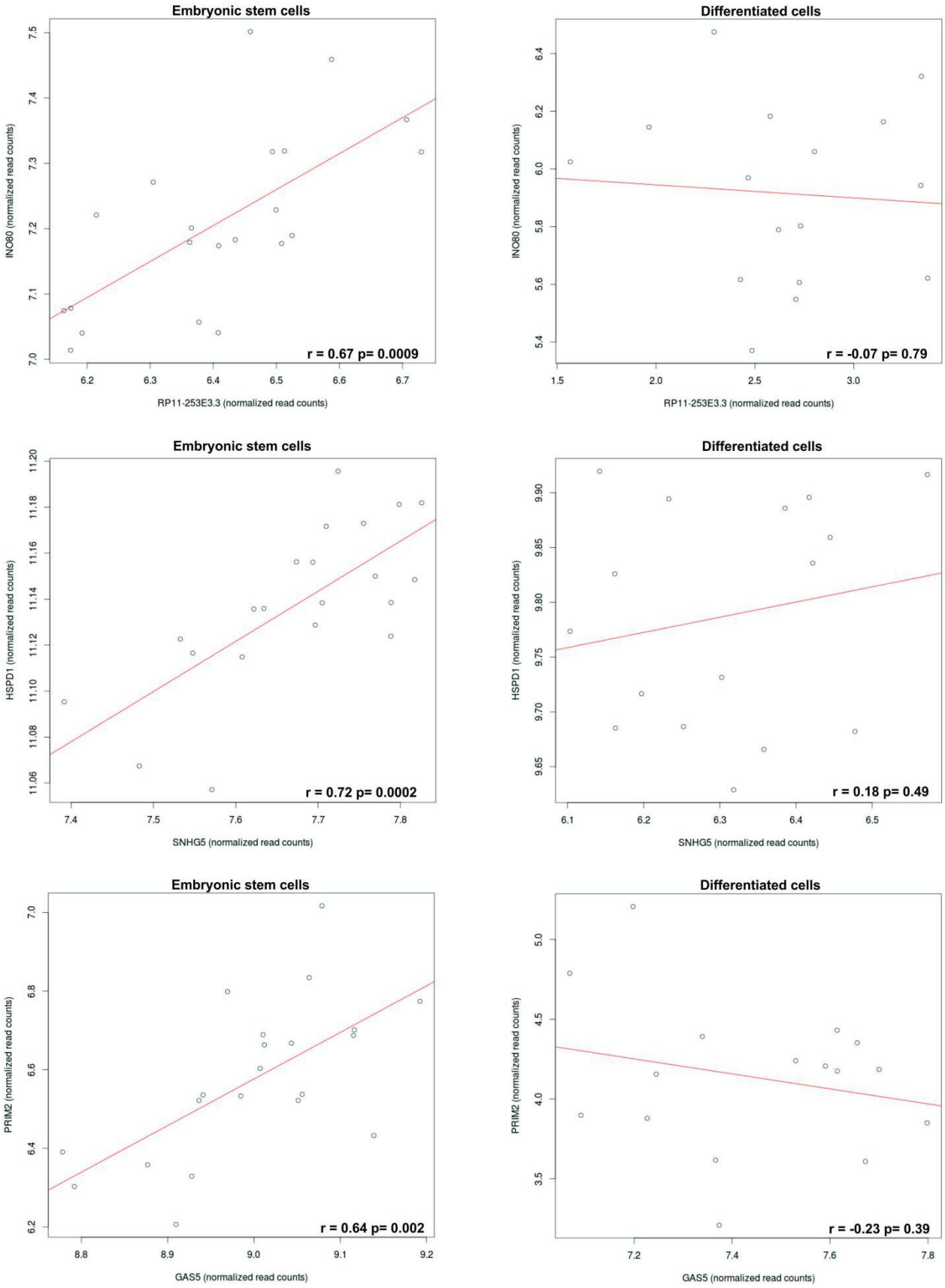


Figure 6: Correlation analysis between RNAs from the chosen ceRNA pairs in ESCs compared with differentiated cells.

3.2.6. Competition at the isoform level

The ceRNA pairs selected based on the GO analysis were further investigated at the isoform level. Both ncRNAs and mRNAs of each pair presented more than one isoform expressed in ESCs, with a few exceptions. In addition, some protein-coding genes were also transcribed into non-coding isoforms.

All isoforms had their sequences screened for binding sites of the investigated miRNAs. In addition, the strength of the miRNA binding, the seed type and the expression levels of each detected isoform were also estimated (Supplemental Material 8).

The ncRNAs of the pairs AHCY-RP11-253E3.3, FGFR2-RP11-20D14.6, F2RL1-EEF1A1P6-RP11-69L16.5, HSPD1-SNHG5, INO80-RP11-20D14.6, NPM1-RP11-20D14.6, PARP1-RP11-20D14.6, PRIM2-GAS5, SALL4-RP11-20D14.6 and SRFP2-RP11-20D14.6 presented shared miRNA sites with ΔG values similar or higher than those estimated for the sites on mRNAs (Figure 7 and Supplemental Material 9, Figure S2). In other words, the miRNA theoretically binds with approximately the same strength and stability to the MREs on the ncRNA and on the mRNA.

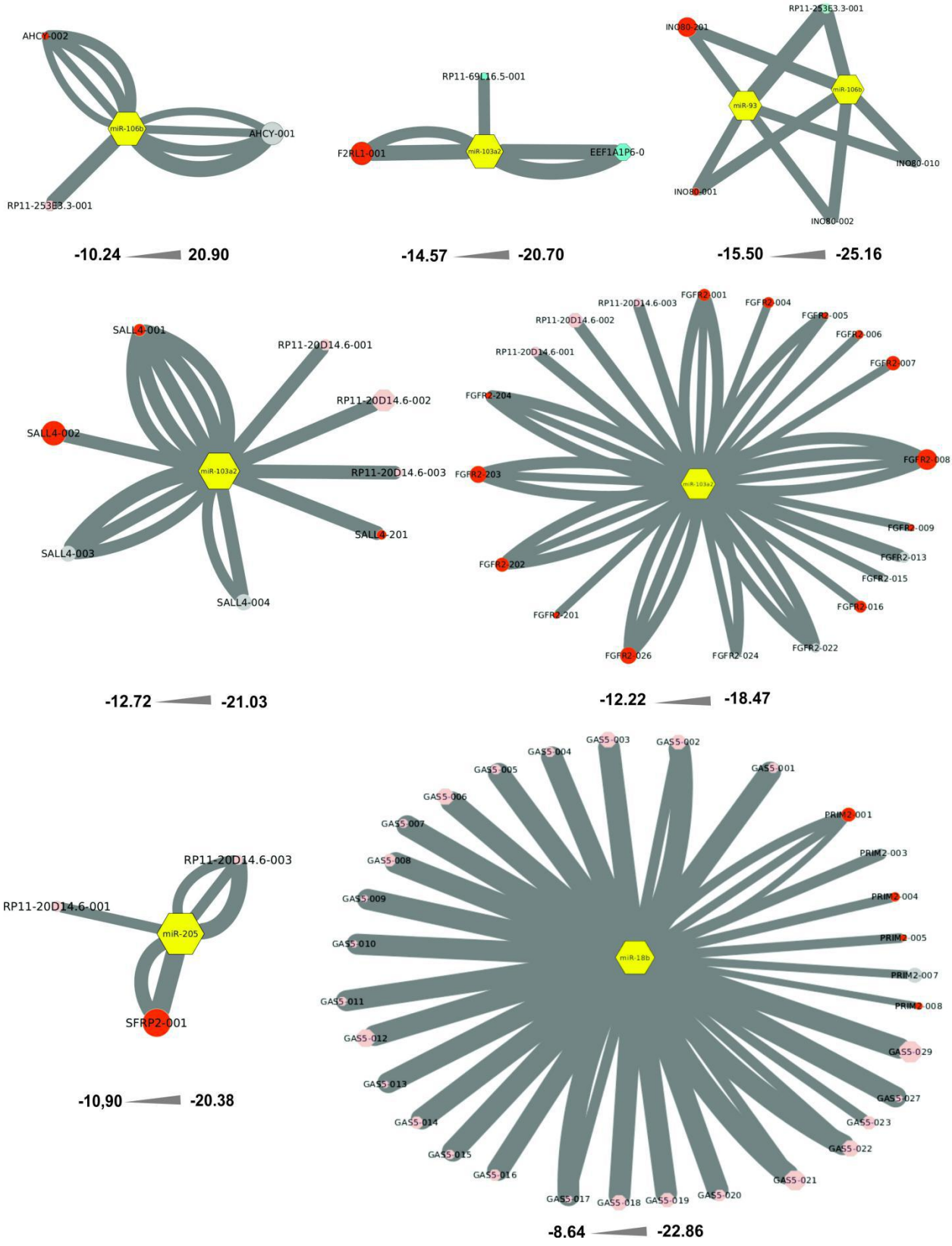


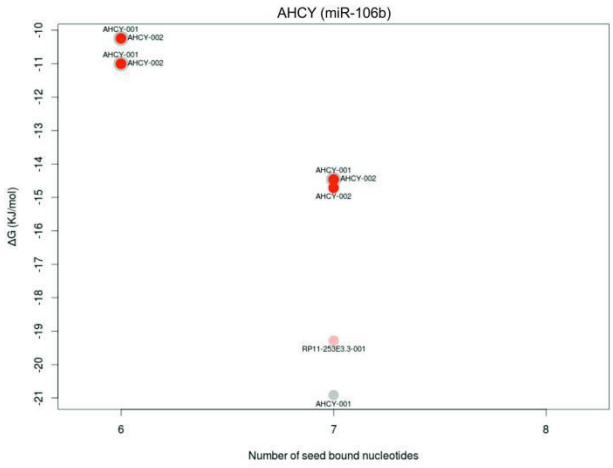
Figure 7: Schematic representation of the interaction between RNAs from the ceRNA pairs and miRNAs. The node sizes are proportional to the isoform's expression in transcripts per million

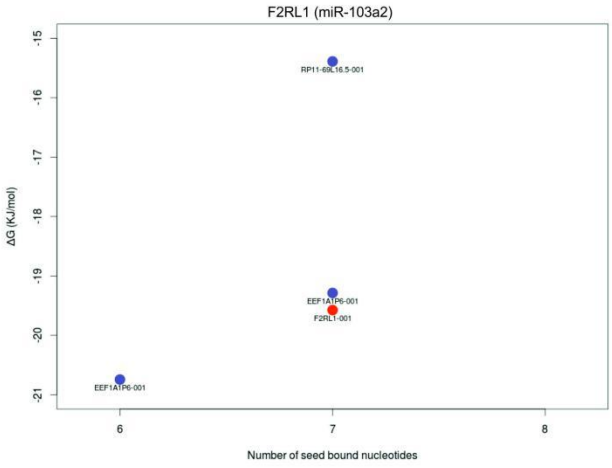
(TPM), except for miRNAs, whose expression is not exhibited. The connector width is proportional to the miRNA binding ΔG values ($kJ.mol^{-1}$). Red circles: mRNAs. Gray circles: non-coding RNAs isoforms expressed by a coding gene. Pink octagons: lncRNAs. Blue octagons: pseudogenes. Yellow hexagons: miRNAs.

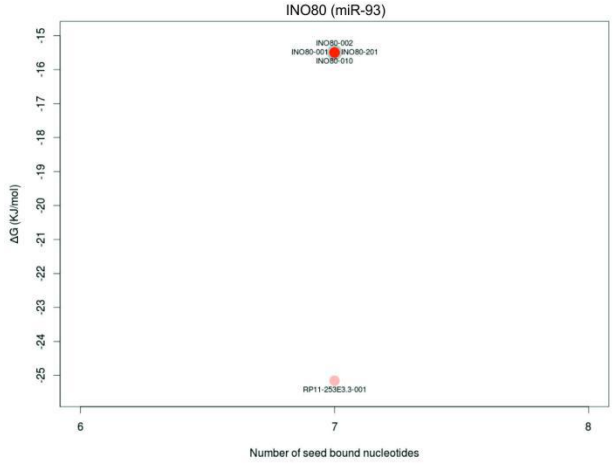
For these pairs, the type of miRNA seed binding was further investigated and, classified as 8-mer, 7-mer or 6-mer, and the miRNA binding strength values were plotted against the seed types (Figure 8 and Supplemental Material 9, Figure S3).

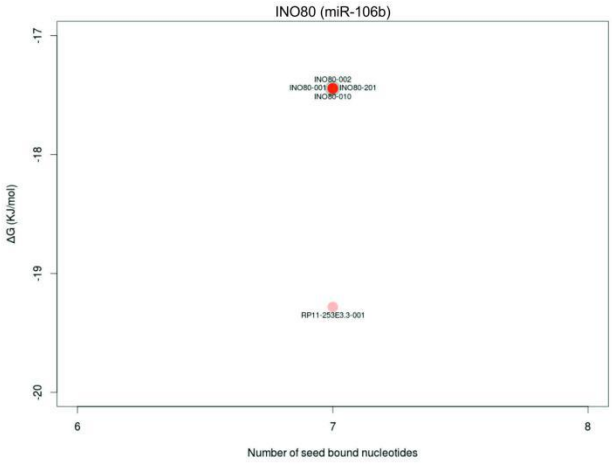
Both criteria, that is, seed type and miRNA-target binding strength, were used to further investigate the possibility that the RNAs of the pairs were indeed in a competition with each other. For the pair AHCY-RP11-253E3.3, RP11-253E3.3-001 binds to the miRNA more stably and presents a seed type with higher preference in the binding hierarchy. For the F2RL1-EEF1A1P6 pair, the seed types of both RNAs remain at the same preference level in the hierarchy, and the binding strength is very similar. The isoform RP11-69L16.5-001 also presents the same seed type as the mRNA, but with a lower binding strength; and therefore, it was discarded as a competitor. In the case of the HSPD1-SNHG5 pair, the isoforms of SNHG5 presented seed types with higher binding preference or a higher miRNA binding strength. In the pair INO80-RP11-20D14.6, the ncRNA presented highter miRNA binding strength, but the same seed type as the mRNA. For PARP1-RP11-20D14.6, the seeds types of the ncRNA were the same or had a higher binding preference than the mRNAs seeds. The same trend was observed for the binding ΔG. In the case of the PRIM2-GAS5 pair, GAS5 presents 8-mer seed types and a higher binding strength (Figure 8). In all of the other pairs, the ncRNAs presented seed types with a lower preference in the binding

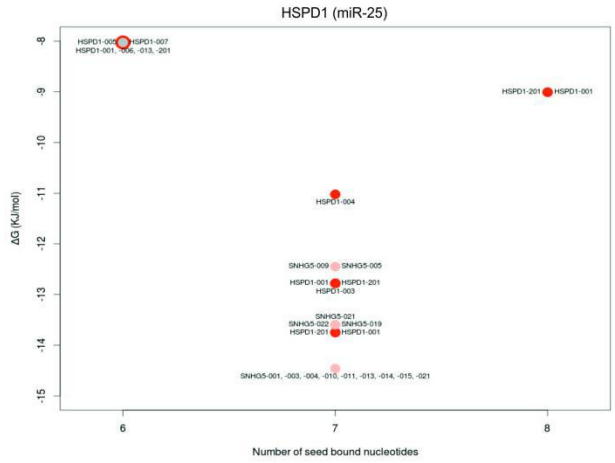
hierarchy or a miRNA binding strength that was lower than the that observed in for the mRNA. These pairs were disconsidered.

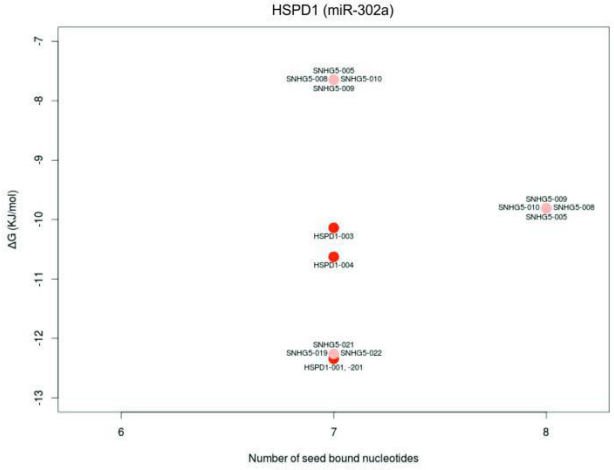

AHCY (miR-106b)
ΔG (KJ/mol)
Number of seed bound nucleotides

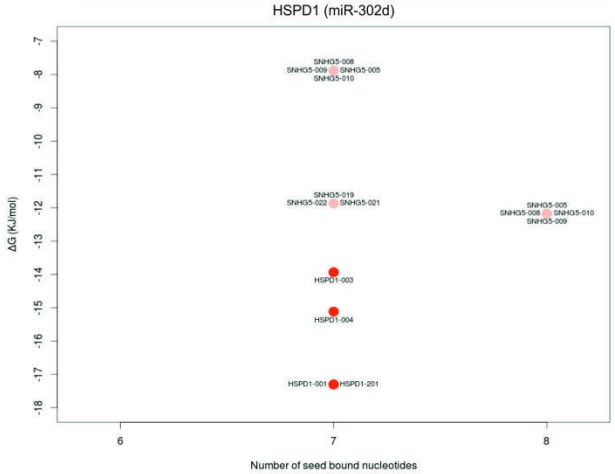

F2RL1 (miR-103a2)
ΔG (KJ/mol)
Number of seed bound nucleotides

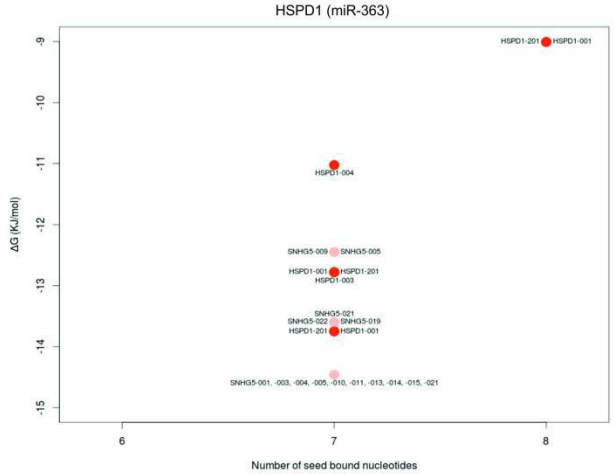

INO80 (miR-93)
ΔG (KJ/mol)
Number of seed bound nucleotides


INO80 (miR-106b)
ΔG (KJ/mol)
Number of seed bound nucleotides


HSPD1 (miR-25)
ΔG (KJ/mol)
Number of seed bound nucleotides


HSPD1 (miR-302a)
ΔG (KJ/mol)
Number of seed bound nucleotides


HSPD1 (miR-302d)
ΔG (KJ/mol)
Number of seed bound nucleotides


HSPD1 (miR-363)
ΔG (KJ/mol)
Number of seed bound nucleotides

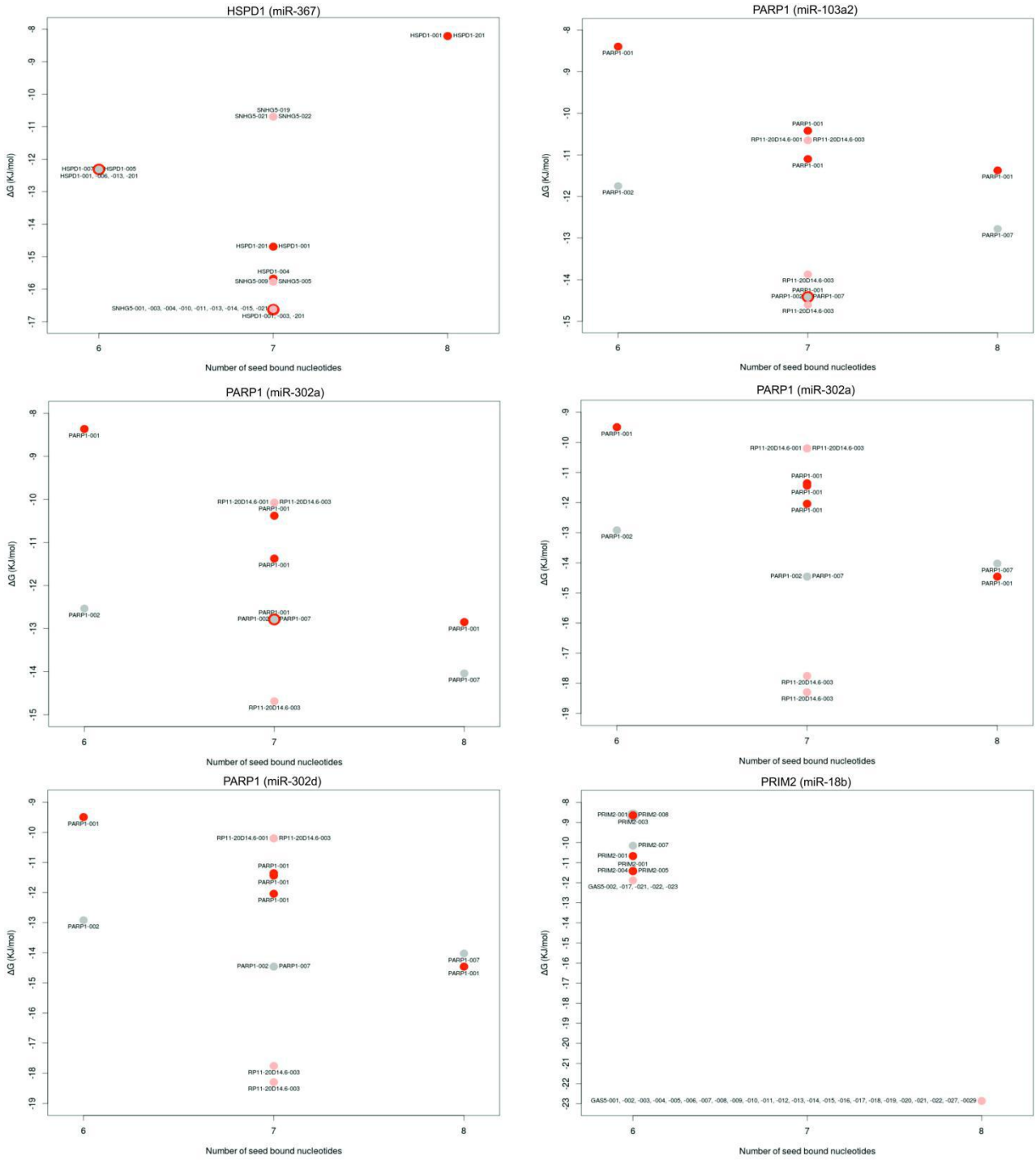


Figure 8: miRNA binding strength combined with seed type for each MRE detected on the isoforms of the genes composing the ceRNA pair. Red circles: mRNAs. Gray circles: non-coding isoforms expressed by a coding gene. Pink circles: lncRNAs. Blue circles: pseudogenes. In the case of overlapping circles, one of them is represented in a shorter diameter. For this case, identifications

above or bellow the circle corresponds to the larger one, and identifications on each side or both sides corresponds to the smaller circle.

Interestingly, some non-coding isoforms expressed by the coding genes were targeted by one or more miRNAs and were detected as upregulated in ESCs with ΔG values and seed types that were similar to or higher than those observed in their coding counterparts. This observation opens the possibility that these RNAs may also act as ceRNAs of their coding isoforms. This is the case for AHCY-001 from the pair AHCY-RP11-253E3.3, the isoforms INO80-002 and -201 in the pair INO80-RP11-20D14.6 and PARP1-002 and -007 in the case of the pair PARP1-RP11-20D14.6.

3.3. circRNAs that putatively act as ceRNAs in ESCs

A total of 280 different circRNAs were identified in the ESC replicates of the main dataset after the analysis with Mapsplice2. However, most of them were not detected in all of the samples. Therefore, we chose 10 circRNAs that were detected in most of the replicates. The circRNAs in case were derived from the backsplicing of some exons from the following genes: FAT1, FAT3, HIPK3, SPECC1, SETD3, SKIL, TNFRSF21, UGP2, XPO1 and ZNF124 (Table 4). Therefore we named these circRNAs by adding the prefix “circ” in front of the gene name. Besides, two further RNA-seq datasets, 2 and 3 (see Materials and Methods), were analyzed in order to confirm that those circRNAs were really expressed in ESCs. All of the circRNAs were detected, but only in the dataset 3.

In addition, we compared our circRNAs to those reported in a previous publication[42]. All circRNAs detected in our ESCs samples were also observed by those authors. The only exception was the circRNA derived from the gene coding for ZNF124, which presented three exons in our analysis and only two in the comparison report.

A comparison between the circRNAs detected in ESCs and those detected in differentiated cells were not conducted because there is no reliable tool to perform circRNA differential expression with RNA-seq data. In addition, most of them were also detected in differentiated cells. In this way, no inference about the role of detected circRNAs in ESC maintenance can be made. This fact, however, does not preclude a putative ceRNA behavior of these RNAs in stem cells.

Then, a screening of the circRNAs sequences was performed to identify putative miRNAs sites, their binding strength and seed type. Eight of the ten circRNAs presented binding sites for the investigated miRNAs (Table 4). Most of the ΔG values of the miRNA binding sites on circRNAs (~ -7 to -16 $kJ.mol^{-1}$) were lower than the lowest value tested by Yuan and coworkers[43], indicating that they possibly do not act as strong ceRNAs (Figure 8 and Supplemental Material 10). However, the miR-25 and miR-103a2 sites from circ-FAT1 (ΔG ~ -17 $kJ.mol^{-1}$), miR-103a2 from circ-HIPK3 (ΔG ~ -20 $kJ.mol^{-1}$) and miR18b from circ-TNFRSF21 (ΔG ~ -17 $kJ.mol^{-1}$) may contribute to a putative ceRNA role of these circRNAs (Figure 9 and Supplemental Material 9, Ffigure S4).

The seed types for these sites also corroborate the idea that they may act as putative ceRNAs. In the case of circ-FAT1, the seed types are 7-mer and 6-mer for the miR-103a2 and -25 binding sites, respectively. The seed for the miR-103a2 site on circ-HIPK3 is 8-mer, and that for the 18b site on circ-TNFRSF21 is 7-mer.

Table 4: circRNAs detected in embryonic stem cells

Chromosome	Number of exons	Exons	Gene name	Binding miRNAs
4	1	ENSE00001271254	FAT1	hsa-miR18b, -25, -103a2, -205, -302acd, -363 and 367
11	1	ENSE00001709176	FAT3	hsa-miR-18b, -93, -103a2, -106b and -205
11	1	ENSE00001479558	HIPK3	hsa-miR-25, -93, -103a2, -106b, -363 and -367
14	5	ENSE00003479105 ENSE00003524073 ENSE00001326746 ENSE00003574060 ENSE00003480744	SETD3	hsa-miR-103a2
3	1	ENSE00001594946	SKIL	hsa-miR-205 and -302acd
17	1	ENSE00001314252	SPECC1	hsa-miR-25, -103a2, -205, -363 and -367
6	2	ENSE00000973751 ENSE00001139134	TNFRSF21	hsa-miR-18b, -25, -93, -103a2,-205, -302acd, -363 and -367
2	2	ENSE00003665935 ENSE00003645549	UGP2	None
2	3	ENSE00003511604 ENSE00001273429 ENSE00003787921	XPO1	None
1	3	ENSE00003545914 ENSE00003609686 ENSE00002230988	ZNF124	hsa-miR-18b and -302acd.

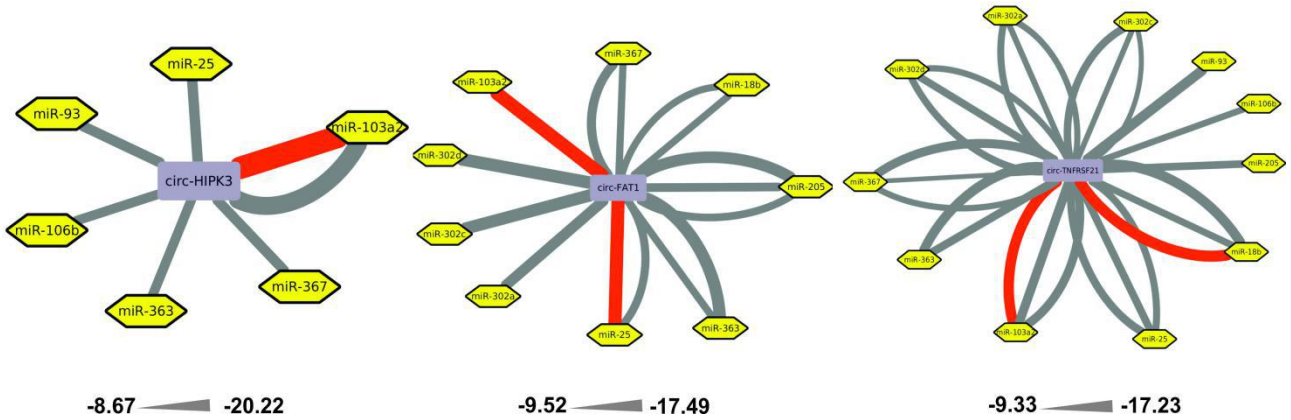


Figure 9: Schematic representation of the interaction among the detected circRNAs and the miRNAs upregulated in ESCs. The highestr and lowest miRNA binding ΔG values (kJ.mol^{-1}) are represented bellow each network, and the connector width is proportional to the ΔG values. The highest binding strengths observed are represented by red connectors.

4. Discussion

Competing endogenous RNAs have already been described in several cellular contexts, and there are more than 60 ceRNA pairs already described in the literature. However, only a few studies have been published concerning the role of ceRNAs in the stemness maintenance[11,15,16]. The first ncRNA reported to have a role in ESCs through competition was linc-ROR. This lncRNA helps to maintain the undifferentiated state of the stem cells through titrating away the miR-145, whose targets are Oct4, Sox2 and Nanog, the major transcription factors responsible for the characteristics of a ESCs. Another ncRNA that acts as a ceRNA in ESCs is FLJ11812, which establishes a competition crosstalk with CDC20B and ATG13 for the miRNA-4459. The stemness maintenance, in this case, was especulated to result from mechanisms related to the cell cycle and autophagy[15]. In addition, in 2016, a third lncRNA was reported to maintain ESCs pluripotency and self-renewal

through a competition mechanism. GAS5 competes with NODAL for the binding of miR-2467, -3200 and Let-7e[16].

None of the previously reported competition cases were, however, observed in this study. LINC-ROR was detected in our samples among the most expressed lncRNAs, but we could not identify the previously described competing crosstalk involving this lncRNA, mainly because the miRNA that intermediated the competition was not detected in our samples. Actually, miR-145 is not a miRNA characteristic of ESCs and it is expressed at low levels in these cells[43]. On the other hand, its expression is observed to rise when the stem cells initiate their differentiation[43]. In addition, no ceRNA pair containing this lncRNA was identified considering our ceRNA filtering criteria. We also failed to detect the crosstalk between FLJ11812 and the mRNAs CDC20B and ATG13 because neither FLJ11812 nor miRNA-4459 were detected as upregulated in any of the data sets analyzed. In the case of GAS5, we detected both Nodal and GAS5 RNAs, but we did not indentify any crosstalk between these two molecules because the miRNAs that mediate the competition were not detected in our datasets and are not reported as ESC-recognized miRNAs, the criterion adopted to select the miRNAs in our study. The fact that the miRNAs mediating the competition between GAS5 and NODAL improve their expression during ESCs differentiation may be an indicative that these small RNAs are not characteristic of ESCs and may explain why we did not observed them in our datasets[16].

Based on all the criteria applied in this work, we selected the pairs AHCY-RP11-253E3.3, F2RL1-EEF1A1P6, HSPD1-SNHG5, INO80-RP11-20D14.6, PARP1-RP11-20D14.6 and PRIM2-GAS5 as those with the major potential to truly compete for the shared miRNAs in the context of ESCs. However, none of the RNAs in our ceRNA pairs were previously reported to act

as miRNAs sponges, with the exception of GAS5. Beyond already reported role as a ceRNA in the ESCs, it was reported to act as a tumor suppressor by promoting the expression of PTEN through the sponging of miR-21 and miR-103 in breast and endometrial cancer cells, respectively[44,45].

Although the mRNAs of the chosen pairs were not reported in competition crosstalks previously, they are associated with processes that are important for ESCs, like the response to hypoxia (AHCY and HSPD1), regulation of the mitotic cell cycle (INO80), telomere maintenance (PARP1 and PRIM2) and positive regulation of ESCs -related signaling pathways (F2RL1). This fact makes their coded proteins possibly important for the ESCs biology.

The human embryo develops in an oxygen-poor environment at the beginning of the gestation, and it has been demonstrated that hypoxic ESCs culture conditions prevents the espontaneous differentiation[46,47]. In addition, hypoxia-inducible factors seems to regulate ESCs pluripotency and proliferation. OCT4, SOX2 and NANOG were downregulated and the ESCs number and colony size were significantly reduced when HIF2A and HIF3A, genes coding for factors related to the cellular hypoxia response, were silenced[48]. Moreover, the protein coded by HIF2A was observed to bind to the Oct-4 promoter and induce its expression[49]. AHCY and HSPD1 have never been investigated in ESCs previously. However, their association with hypoxia, even in the case of AHCY, whose hypoxia hypoxia-related ontology was inferred from electronic annotation, indicated that the proteins coded by these genes may be important in the stemness maintenance, which is reinforced by the possibility of their expression being maintained by competition. HSPD1 is an essential gene for early embryonic development because homozygous Hspd1 mutant mouse embryos die shortly after implantation[50]. In addition, HSPD1 is constitutively

expressed in murine embryonic carcinoma stem cells and had its protein expression downregulated when these cells were induced to differentiate[51].

INO80 codes for a chromatin remodeling ATPase part of the INO80 complex, which functions as a chromatin remodeler. Despite its associated ontology in mitotic cell cycle regulation, Ino80 was also reported to co-occupy pluripotency gene promoters with the master ESC transcription factors, helping to open the chromatin and recruit proteins to promote the gene expression[52]. The knockdown of INO80 leads to decreased expression of the transcription factors Oct4, Nanog, Sox2, Klf4, and Esrrb, which are important for the stemness maintenance, increased expression of differentiation markers and loss of ESC typical morphology. In addition, Ino80 interacts physically with Oct-4, that, together with WDR5, promote the recruiting of Ino80 to its target promoters[52].

One of the categories of transcripts enriched in ESCs is composed by of genes involved in signaling pathways[41]. Key components of the MAPK/ERK and PI3K/AKT pathways are downregulated during the ESCs differentiation[41]. In addition, the inhibition of these pathways with chemicals leads to ESCs differentiation together with a decrease in the expression of the ESCs markers, such as NANOG, OCT4, SOX2 and SSEA4[40,41]. F2RL1 encodes for a cell surface receptor member of the G-protein coupled receptor 1 family of proteins, known as PAR-2. PAR-2 has been shown to activate the PI3K/AKT pathway in cancer cells, as the introduction of a PAR-2 agonist in the cell culture leads to an increase in the PI3K and AKT phosphorylation[53]. Additionally, C-terminal PAR-2 mutants lead to a decrease in ERK1/2 phosphorylation[54]. No study has reported on the role of F2RL1/PAR-2 in ESCs, but its relationship with pathways that are important for these cells makes it an interesting target of ceRNA research.

One of the remarkable ESCs characteristics is the high telomerase activity together with the hight expression of hTERT, the catalytic reverse transcriptase subunit of the telomerase complex[55]. The homonymous protein enconded by the gene PARP-1 belongs to athe poly(ADP-ribose) polymerase (PARP) family that performs poly ADP-ribosylation on target proteins and has been associated with the telomere maintanance[56]. This association was confirmed by the cellular effects of PARP-1 silencing: reduced telomerase activity, inhibition of poly(ADP-ribosyl)ation of hTERT and reduction of TEP1/TP1, a telomerase- associated protein[57], together with telomere shortening[56]. Moreover, PARP-1 was observed to be downregulated during the murine ESC (mESCs) differentiation, and mESCs deficient in PARP-1 present a reduction in the expression of important ESC transcription factors such as Oct-4, Sox2 and Nanog[58]. Additionally, these cells differentiate more rapidly into cells from the three germ layers, indicating that PARP-1 may play a role in the ESC maintenance[58].

Also associated to with processes related to telomere maintenance, PRIM2 codes for one of the DNA primase subunits, and its protein was shown to interact with CTC1, which is part of a protein complex that is associated with telomere replication[59]. In addition, the mRNA for PRIM2 is important for the DNA replication[60], especially in cells that actively divide, such as ESCs. However, there are few publications on the protein coded by PRIM2 and no report on ESCs. Even so, considering the characteristics of the pair PRIM2-GAS5, the importance of the PRIM2-coded protein and the fact that GAS5 was previously reported to act as a ceRNA, an eventual competition between GAS5 and PRIM2 for miR-18b may be important.

Besides the visible importance of the mRNAs from chosen pairs in ESCs, there are several indications of the possible occurrence of a competition crosstalk between RNAs constituent of these

pairs. Firstly, they share binding sites for miRNAs that are highly expressed in ESCs. Also, they present positively correlated expression, that is not caused by shared regulatory elements. Moreover, although they share common TFs bindindg sites, this seems not to be the cause of the correlation either. Furthermore, it was proposed that stronger ceRNA crosstalk happens at higher miRNA-target binding ΔG and that available miRNA molecules distribute hierarchically on the target pool, where they bind preferentially to 8-mer sites, followed by 7-mer and 6-mer sites[61,62]. The ncRNAs that integrate the selected pairs are bound by the shared miRNAs with ΔG values equal to higher than the ΔG values of the mRNAs and present seed types with equal or higher preference in the binding hierarchy. These observations, together with the fact that the correlation between the RNAs is not observed in differentiated cells, where their shared miRNAs are not expressed, point to these pairs as strong candidates of to being ceRNAs.

In addition to the competition between ncRNAs and mRNAs, we also investigated the expression and putative action as ceRNAs of circRNAs in ESCs. circRNAs originate from a non-canonical splicing process called backsplicing[63]. In this process, a downstream splice donor is joined to an upstream splice acceptor, generating circular RNA molecules[63]. circRNAs seem to be regular RNA molecules expressed in cells and were also reported to act as ceRNAs in different cells types [9,64–67]. These molecules were detected in ESCs previously, but no study has reported them as ceRNAs in this cell type[42]. Here, we detected 280 different putative circRNAs in ESCs and, for the first time, proposed that three of them putatively act as ceRNAs in these cells. The circRNAs circ-FAT1, -HIPK3 and -TNFRSF21 were detected previously in ESCs[42] and were proposed here as putative ceRNAs because they have binding sites for miRNAs characteristic of these cells. An investigation of the ΔG binding values and the seed types revealed that their interaction with their

respective miRNA is indicative of a possible strong ceRNA effect, and their seed types favours their association with miRNAs in the binding hierarchy. All of these characteristics make these circRNAs interesting targets for an investigation about circRNAs acting as ceRNAs in the ESCs context, especially the circ-HIPK3, which presents the most stable binding to hsa-miR-103a2 and a seed type with a preference in the miRNA binding hierarchy (8-mer).

As the evidences gathered in this study indicates that the ncRNA-mRNA and circRNAs highlighted here may act as ceRNAs, they are proposed as preferential targets for in vivo testing aiming to broaden the knowledge about the RNAs competitions in ESCs.

5. Conclusions

Most of the ceRNAs crosstalks that have already been described have involved only pairs of RNAs. However, if the competition among RNAs for miRNA is a real new layer of gene expression control, it is more likely that the interactions are organized into networks and not in binary interactions. To the best of our knowledge, this is the first large-scale study of ceRNAs in ESCs, and the data presented here have the potential to support the existence and the importance of ceRNAs in ESCs. Additionally, it may help to start elucidating the complexity of this new level of expression control. In this sense, the pairs selected in this work, together with the detected circRNAs, are strong candidates to be the starting point of in vivo studies to disentangle the ceRNA network characteristic of ESCs.

6. Acknowledgments

This work was supported by the Conselho Nacional de Desenvolvimento Cientıfico e Tecnológico (CNPq) through Grant 301149/2012-7.

7. Author Disclosure Statement

No competing financial interests exist.

8. References

1. Consortium EP, Dunham I, Kundaje A, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74.
2. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: The rosetta stone of a hidden RNA language? *Cell*. 2011;146(3):353-358.
3. Cesana M, Cacchiarelli D, Legnini I, et al. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell*. 2011;147(2):358-369.
4. Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature*. 2010;465(7301):1033-1038.
5. Memczak S, Jens M, Elefsinioti A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. TL - 495. *Nature*. 2013;495 VN-(7441):333-338.
6. Wang K, Long B, Zhou L-Y, et al. CARL lncRNA inhibits anoxia-induced mitochondrial fission and apoptosis in cardiomyocytes by impairing miR-539-dependent PHB2 downregulation. *Nat Commun*. 2014;5:3596.

7. Li M, Sun X, Cai H, et al. Long non-coding RNA ADNCR suppresses adipogenic differentiation by targeting miR-204. *Biochim Biophys Acta - Gene Regul Mech*. 2016;1859(7):871-882.

8. Liang W-C, Fu W-M, Wang Y-B, et al. H19 activates Wnt signaling and promotes osteoblast differentiation by functioning as a competing endogenous RNA. *Sci Rep*. 2016;6(February):20121.

9. Hansen TB, Jensen TI, Clausen BH, et al. Natural RNA circles function as efficient microRNA sponges. *Nature*. 2013;495(7441):384-388.

10. Tay Y, Kats L, Salmena L, et al. Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell*. 2011;147(2):344-357.

11. Wang Y, Xu Z, Jiang J, et al. Endogenous miRNA Sponge lincRNA-RoR Regulates Oct4, Nanog, and Sox2 in Human Embryonic Stem Cell Self-Renewal. *Dev Cell*. 2013;25(1):69-80.

12. Amit M, Carpenter MK, Inokuma MS, et al. Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. *Dev Biol*. 2000;227(2):271-278.

13. Gerecht-Nir S, Itskovitz-Eldor J. Cell therapy using human embryonic stem cells. *Transpl Immunol*. 2004;12(3-4):203-209.

14. Dvash T, Ben-Yosef D, Eiges R. Human embryonic stem cells as a powerful tool for studying human embryogenesis. *Pediatr Res*. 2006;60(2):111-117.

15. Lu W, Han L, Su L, et al. A 3'UTR-associated RNA, FLJ11812 maintains stemness of human embryonic stem cells by targeting miR-4459. *Stem Cells Dev*. 2015;24(9):1133-1140.

16. Xu C, Zhang Y, Wang Q, et al. Long non-coding RNA GAS5 controls human embryonic stem cell self-renewal by maintaining NODAL signalling. *Nat Commun*. 2016;7:13287.

17. Dobin A, Davis CA, Schlesinger F, et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21.

18. Anders S, Pyl PT, Huber W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31(2):166-169.

19. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):1-34.

20. Wang K, Singh D, Zeng Z, et al. MapSplice: Accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res*. 2010;38(18).

21. Vejnar CE, Blum M, Zdobnov EM. miRmap web: Comprehensive microRNA target prediction online. *Nucleic Acids Res*. 2013;41.

22. Mathelier A, Fornes O, Arenillas DJ, et al. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*. 2015;44(November 2015):gkv1176.

23. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol*. 2010;11(2):R14.

24. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. *Salmon Provides Accurate, Fast, and Bias-Aware Transcript Expression Estimates Using Dual-Phase Inference*.; 2015. doi:10.1101/021592.

25. Paci P, Colombo T, Farina L. Computational analysis identifies a sponge interaction network between long non-coding RNAs and messenger RNAs in human breast cancer. *BMC Syst Biol*. 2014;8:83.

26. Suh M-R, Lee Y, Kim JY, et al. Human embryonic stem cells express a unique set of microRNAs. *Dev Biol*. 2004;270(2):488-498.

27. Bar M, Wyman SK, Fritz BR, et al. MicroRNA discovery and profiling in human embryonic stem cells by deep sequencing of small RNA libraries. *Stem Cells*. 2008;26(10):2496-2505.

28. Laurent LC, Chen J, Ulitsky I, et al. Comprehensive microRNA profiling reveals a unique human embryonic stem cell signature dominated by a single seed sequence. *Stem Cells*. 2008;26(6):1506-1516.

29. Morin RD, Connor MDO, Griffith M, et al. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res*. 2008:610-621.

30. Li SSL, Yu SL, Kao LP, et al. Target identification of micrornas expressed highly in human embryonic stem cells. *J Cell Biochem*. 2009;106(6):1020-1030.

31. Ren J, Jin P, Wang E, Marincola FM, Stroncek DF. MicroRNA and gene expression patterns in the differentiation of human embryonic stem cells. *J Transl Med*. 2009;7:20.

32. Parsons XH, Parsons JF, Moore D a. Genome-Scale Mapping of MicroRNA Signatures in Human Embryonic Stem Cell Neurogenesis. *Mol Med Ther*. 2012;1:2-3.

33. Zheng L, Li X, Gu Y, Lv X, Xi T. The 3′UTR of the pseudogene CYP4Z2P promotes tumor angiogenesis in breast cancer by acting as a ceRNA for CYP4Z1. *Breast Cancer Res Treat*. 2015;150(1):105-118.

34. Asikainen S, Heikkinen L, Juhila J, et al. Selective microRNA-Offset RNA expression in human embryonic stem cells. *PLoS One*. 2015;10(3):e0116668.

35. van de Leemput J, Boles NC, Kiehl TR, et al. CORTECON: A temporal transcriptome analysis of in vitro human cerebral cortex development from human embryonic stem cells. *Neuron*. 2014;83(1):51-68.

36. Gill KP, Hung SSC, Sharov A, et al. Enriched retinal ganglion cells derived from human embryonic stem cells. *Sci Rep*. 2016;6:30552.

37. Zhang H, Xue C, Shah R, et al. Functional Analysis and Transcriptomic Profiling of iPSC-Derived Macrophages and Their Application in Modeling Mendelian Disease. *Circ Res*. 2015;117(1):17-28.

38. KANG HB, KIM JS, KWON H-J, et al. Basic Fibroblast Growth Factor Activates ERK and Induces c-Fos in Human Embryonic Stem Cell Line MizhES1. *Stem Cells Dev*. 2005;586:576-586.

39. Vallier L, Reynolds D, Pedersen RA. Nodal inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway. *Dev Biol*. 2004;275(2):403-421.

40. Li J, Wang G, Wang C, et al. MEK/ERK signaling contributes to the maintenance of human embryonic stem cell self-renewal. *Differentiation*. 2007;75(4):299-307.

41. Armstrong L, Hughes O, Yung S, et al. The role of PI3K/AKT, MAPK/ERK and NFκβ signalling in the maintenance of human embryonic stem cell pluripotency and viability highlighted by transcriptional profiling and functional analysis. *Hum Mol Genet*. 2006;15(11):1894-1913.

42. Zhang XO, Wang H Bin, Zhang Y, Lu X, Chen LL, Yang L. Complementary sequence-mediated exon circularization. *Cell*. 2014;159(1):134-147.

43. 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):647-658.

44. Li W, Zhai L, Wang H, et al. Downregulation of LncRNA GAS5 causes trastuzumab resistance in breast cancer. *Oncotarget*. 2016;7(19).

45. Guo C, Song W-Q, Sun P, Jin L, Dai H-Y. LncRNA-GAS5 induces PTEN expression through inhibiting miR-103 in endometrial cancer cells. *J Biomed Sci*. 2015;22:100.

46. Rodesch F, Simon P, Donner C, Jauniaux E. Oxygen measurements in endometrial and trophoblastic tissues during early pregnancy. *Obstet Gynecol*. 1992;80(2):283-285.

47. Ezashi T, Das P, Roberts RM. Low O2 tensions and the prevention of differentiation of hES cells. *Proc Natl Acad Sci U S A*. 2005;102(13):4783-4788.

48. Forristal CE, Wright KL, Hanley NA, Oreffo ROC, Houghton FD. Hypoxia inducible factors regulate pluripotency and proliferation in human embryonic stem cells cultured at reduced oxygen tensions. *Reproduction*. 2010;139(1):85-97.

49. Covello KL, Kehler J, Yu H, et al. HIF-2{alpha} regulates Oct-4: effects of hypoxia on stem cell function, embryonic development, and tumor growth. *Genes Dev*. 2006;20(5):557-570.

50. Christensen JH, Nielsen MN, Hansen J, et al. Inactivation of the hereditary spastic paraplegia-associated Hspd1 gene encoding the Hsp60 chaperone results in early embryonic lethality in mice. *Cell Stress Chaperones*. 2010;15(6):851-863.

51. Afzal E, Ebrahimi M, Najafi SMA, Daryadel A, Baharvand H. Potential role of heat shock proteins in neural differentiation of murine embryonal carcinoma stem cells (P19). *Cell Biol Int*. 2011;35(7):713-720.

52. Wang L, Du Y, Ward JM, et al. INO80 facilitates pluripotency gene activation in embryonic stem cell self-renewal, reprogramming, and blastocyst development. *Cell Stem Cell*. 2014;14(5):575-591.

53. Sun Z, Cao B, Wu J. Protease-activated receptor 2 enhances renal cell carcinoma cell invasion and migration via PI3K/AKT signaling pathway. *Exp Mol Pathol*. 2015;98(3):382-389. doi:10.1016/j.yexmp.2015.03.018.

54. Zhu Z, Stricker R, yu Li R, Zündorf G, Reiser G. The intracellular carboxyl tail of the PAR-2 receptor controls intracellular signaling and cell death. *Cell Tissue Res*. 2015;359(3):817-827. .

55. Zeng S, Liu L, Sun Y, et al. Telomerase-mediated telomere elongation from human blastocysts to embryonic stem cells. *J Cell Sci*. 2014;127(Pt 4):752-762.

56. Beneke S, Cohausz O, Malanga M, Boukamp P, Althaus F, Bürkle A. Rapid regulation of telomere length is mediated by poly(ADP-ribose) polymerase-1. *Nucleic Acids Res*. 2008;36(19):6309-6317.

57. Ghosh U, Das N, Bhattacharyya NP. Inhibition of telomerase activity by reduction of poly(ADP-ribosyl)ation of TERT and TEP1/TP1 expression in HeLa cells with knocked down poly(ADP-ribose) polymerase-1 (PARP-1) gene. *Mutat Res - Fundam Mol Mech Mutagen*. 2007;615(1-2):66-74.

58. Roper SJ, Chrysanthou S, Senner CE, et al. ADP-ribosyltransferases Parp1 and Parp7 safeguard pluripotency of ES cells. *Nucleic Acids Res*. 2014;42(14):8914-8927.

59. Chen LY, Majerská J, Lingner J. Molecular basis of telomere syndrome caused by CTC1 mutations. *Genes Dev*. 2013;27(19):2099-2108.

60. Muzi-Falconi M, Giannattasio M, Foiani M, Plevani P. The DNA Polymerase α-Primase Complex: Multiple Functions and Interactions. *Mini-Review TheScientificWorldJOURNAL*. 2003;3:21-33.

61. Yuan Y, Liu B, Xie P, et al. Model-guided quantitative analysis of microRNA-mediated regulation on competing endogenous RNAs using a synthetic gene circuit. *Proc Natl Acad Sci U S A*. 2015;112(10):3158-3163.

62. Bosson AD, Zamudio JR, Sharp PA. Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition. *Mol Cell*. 2014;56(3):347-359.

63. Chen L. The biogenesis and emerging roles of circular RNAs. *Nat Rev Mol Cell Biol*. 2016;17(4):205-211.

64. Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One*. 2012;7(2).

65. Liu Q, Zhang X, Hu X, et al. Circular RNA Related to the Chondrocyte ECM Regulates MMP13 Expression by Functioning as a MiR-136 “Sponge” in Human Cartilage Degradation. *Sci Rep*. 2016;6:22572.

66. Li F, Zhang L, Li W, et al. Circular RNA ITCH has inhibitory effect on ESCC by suppressing the Wnt/beta-catenin pathway. *Oncotarget*. 2015;6(8):6001-6013.

67. Yang W, Du WW, Li X, Yee AJ, Yang BB. Foxo3 activity promoted by non-coding effects of circular RNA and Foxo3 pseudogene in the inhibition of tumor growth and angiogenesis. *Oncogene*. 2015;(October):1-13.

Supplemental Material 1

Table S1: The 50 most expressed lncRNAs in human embryonic stem cells detected in DESeq2 analysis of the main dataset.

Gene ID	Gene Name	log_2FC	p Value	Adjusted p value
ENSG00000232301	LNCPRESS1	9.66992288162666	3.66004666171958e-27	1.92723502013333e-26
ENSG00000249152	LNCPRESS2	9.32757604765559	2.06868887786274e-19	8.99904466293286e-19
ENSG00000254934	LINC00678	8.98002898156571	1.6115487870555e-80	1.96637858725471e-79
ENSG00000265992	ESRG	8.78345927029637	4.7017993192677e-76	5.4421535403307e-75
ENSG00000280707	RP11-568A7.4	8.59488779574113	7.1249646207954e-28	3.81697038225121e-27
ENSG00000236673	RP11-69I8.2	8.52608710440864	1.43149417315855e-34	8.87433980291872e-34
ENSG00000253507	CTD-2501M5.1	8.23164610036708	4.48971394010088e-19	1.93171879789398e-18
ENSG00000258609	LINC-ROR	8.18003204809258	2.99924275887274e-15	1.14983297596234e-14
ENSG00000274090	RP11-415G4.1	7.38060314436838	3.64181663295141e-15	1.39176728346359e-14
ENSG00000234787	LINC00458	7.32870792592889	9.20594365453137e-42	6.47828603896031e-41
ENSG00000234261	RP11-146I2.1	7.30004497926441	8.3733339927582e-15	3.15857326168601e-14
ENSG00000253230	LINC00599	7.02403603187471	4.39431470757041e-23	2.10478456552133e-22
ENSG00000248131	LINC01194	6.96079775576223	1.46506648419699e-32	8.69928184069831e-32
ENSG00000237923	XXbac-BPG27H4.8	6.69601662532708	2.06834553533478e-12	7.10452539656545e-12
ENSG00000274652	RP11-457D13.4	6.64237075223624	3.56531904567749e-12	1.21455187771999e-11

ENSG00000227640	SOX21-AS1	6.56875484972313	3.00215882732538e-10	9.35010397079628e-10
ENSG00000226673	LINC01108	6.50466115793817	9.82550296443354e-20	4.32258586357699e-19
ENSG00000244342	LINC00698	6.29513391964392	1.19631519202551e-30	6.80287071578145e-30
ENSG00000231698	AP002856.5	6.2770930677899	7.65030799993716e-11	2.45284619809454e-10
ENSG00000241168	RP11-10O22.1	6.13316609191821	1.87682325107712e-17	7.68330401130519e-17
ENSG00000226506	AC007463.2	6.13264567655256	4.38526032816843e-09	1.28626949546672e-08
ENSG00000231566	RP5-1158E12.3	6.13094714751483	2.37686003444561e-10	7.43739663102234e-10
ENSG00000281383	CH507-513H4.5	6.1023509967848	1.28799740078705e-05	2.9905205952143e-05
ENSG00000269877	CTD-3022G6.1	5.82237494145826	2.67579651502381e-08	7.52541841589643e-08
ENSG00000229271	AC091493.2	5.77622312886619	3.45353193928336e-08	9.65534077245315e-08
ENSG00000233393	AP000688.29	5.70401155698609	5.5433639216066e-09	1.61811119943053e-08
ENSG00000249790	RP11-20D14.6	5.63505429702534	2.79354965377037e-138	6.13416162037055e-137
ENSG00000253252	RP11-10A14.6	5.62278089581703	8.30199879969647e-08	2.26255625413312e-07
ENSG00000225783	MIAT	5.60565675819002	3.06898006948017e-214	1.20587650094665e-212
ENSG00000227838	RP1-213J1P B.1	5.60332060805428	1.47938194744114e-11	4.90702756479561e-11
ENSG00000258649	CTD-2142D14.1	5.59256199075941	1.15620401748337e-08	3.3156836095966e-08
ENSG00000197462	AC005276.1	5.53386497139634	1.41803313860994e-07	3.80519543408052e-07
ENSG00000203635	AC144450.2	5.48881711374466	2.35705563451102e-08	6.65018409902825e-08

ENSG00000255666	RP11-867G2.6	5.4581080313201	2.82701207277417e-08	7.93975556513178e-08
ENSG00000248605	CTD-2306M5.1	5.44548027746923	7.8313774747783e-20	3.45367554438764e-19
ENSG00000228592	D21S2088E	5.43969255086136	5.3372596676314e-08	1.47376069909091e-07
ENSG00000262370	RP11-473M20.9	5.392479057566	2.72342432621498e-07	7.17602090861616e-07
ENSG00000267530	LINC01836	5.2793669456642	5.32587255093406e-07	1.37546777294233e-06
ENSG00000225077	LINC00337	5.24077072291107	8.62845621801421e-13	3.00839671575654e-12
ENSG00000261786	RP4-555D20.2	5.2291014211142	6.70190446748385e-07	1.71901480496687e-06
ENSG00000250993	RP11-404J23.1	5.20500735656894	7.42190254474316e-07	1.8969860030603e-06
ENSG00000273209	RP11-107N15.1	5.18811921805742	1.67003114050442e-07	4.46344084470413e-07
ENSG00000261757	AC005592.3	5.1509735342722	1.22796915536194e-17	5.05492295270779e-17
ENSG00000278041	RP5-984P4.6	5.1236985243595	1.24409515863112e-06	3.13080501885344e-06
ENSG00000235152	RP5-865N13.2	5.11432946671105	2.74801486089081e-07	7.23964670306653e-07
ENSG00000277200	RP11-74E22.8	4.98927148330316	3.19401527412296e-09	9.44579175171297e-09
ENSG00000276529	AP001505.10	4.95631483136439	6.7447000369127e-07	1.72917670858412e-06
ENSG00000236714	LINC01844	4.93640782802276	1.75688211029079e-53	1.49545658668444e-52
ENSG00000197182	MIRLET7BHG	4.91913062057142	3.00044091027693e-06	7.33530971865547e-06
ENSG00000251598	RP11-789C2.1	4.82991039376195	5.05220181378315e-06	1.21377745290006e-05

Table S2: The 50 most expressed pseudogenes in human embryonic stem cells detected in DESeq2 analysis of the main dataset.

Gene ID	Gene Name	log_2FC	p Value	Adjusted p value
ENSG00000204121	ECEL1P1	7.86408241161352	3.28814324845047e-14	1.21288356697249e-13
ENSG00000224557	HLA-DPB2	7.83112207679643	3.56824681183917e-17	1.44805362789178e-16
ENSG00000232293	PAICSP7	7.19194770209906	4.31796123500801e-12	1.46513634256969e-11
ENSG00000179899	PHC1P1	6.7883990197443	1.24094085731836e-35	7.86627856205766e-35
ENSG00000265735	RN7SL5P	6.23014201058318	2.1211433282628e-10	6.65957422638817e-10
ENSG00000237227	RP5-1155K23.4	6.19438664233223	2.96022837369424e-09	8.76949526926731e-09
ENSG00000231001	CCNB1IP1P3	5.99103861936496	6.72137533169675e-10	2.06129752353304e-09
ENSG00000213612	FAM220CP	5.81113956220569	1.10834623862535e-15	4.30830016471369e-15
ENSG00000226539	AC012512.1	5.80251426361997	1.49733402527702e-22	7.08872706877561e-22
ENSG00000235602	POU5F1P3	5.78387120332948	3.00483226687036e-09	8.89840224722029e-09
ENSG00000233309	RPS6P12	5.77035687108904	3.19381303813027e-09	9.44579175171297e-09
ENSG00000276170	AC124789.1	5.76009977297457	3.7354961636718e-09	1.1004278305324e-08
ENSG00000233966	UBE2SP1	5.65777300114222	6.35346628068112e-08	1.74433180019171e-07
ENSG00000224664	RPL36AP53	5.63973478796982	7.70807900253316e-08	2.10596766455286e-07

ENSG00000207220	RNU1-57P	5.60154526324832	8.67088836384439e-08	2.36052898829621e-07
ENSG00000225366	TDGF1P3	5.44253335482373	1.84579723940816e-19	8.0465916503545e-19
ENSG00000270524	QTRT1P1	5.43983160603056	9.11111589838937e-14	3.30588016795433e-13
ENSG00000251947	RN7SKP164	5.43254901619106	2.16516273613455e-07	5.74024794253534e-07
ENSG00000277103	RP11-520B13.8	5.32969799067417	1.55494816817991e-46	1.17855542217759e-45
ENSG00000225728	RP6-191P20.3	5.29239984394631	4.58201078819272e-07	1.1889995728846e-06
ENSG00000270354	RP11-547M24.1	5.27695109982462	4.83539897058523e-07	1.25256205622808e-06
ENSG00000229150	CRYGEP	5.12602208042091	1.27716000606816e-06	3.21153858437718e-06
ENSG00000235001	EIF4A1P2	5.12350174080167	1.23193423904502e-06	3.10187513674414e-06
ENSG00000176654	NANOGP1	5.1202326847204	1.36560349403265e-09	4.12409223891546e-09
ENSG00000227097	RPS28P7	5.11345170423073	1.1166937549047e-09	3.38837324460071e-09
ENSG00000225573	RPL35P5	5.08509579346198	1.50816771913459e-09	4.5453841946818e-09
ENSG00000255192	NANOGP8	5.08253565328283	2.56895469929503e-17	1.04668715799078e-16
ENSG00000244280	ECEL1P2	5.0674578204859	3.89554786502419e-07	1.01611622374796e-06
ENSG00000237158	RP11-472F14.4	5.02569179264134	2.38534192636282e-06	5.87820260978103e-06
ENSG00000233762	AC007969.5	5.01726701128388	2.32581260827961e-09	6.92778993664895e-09
ENSG00000188801	ZNF322P1	4.98556686749698	1.13169339791208e-16	4.51791818420655e-16
ENSG00000213763	ACTBP2	4.94393313639276	3.6998368723913e-06	8.9860757016461e-06

ENSG00000265260	RN7SL74P	4.89412914166562	9.49783846501121e-07	2.41022321877319e-06
ENSG00000159712	ANKRD18CP	4.89137605199483	1.21170582601285e-225	5.19903004349167e-224
ENSG00000231345	BEND3P1	4.85782702585277	4.17196265009551e-06	1.00928298474416e-05
ENSG00000242477	CTD-2161E19.1	4.80510577130715	5.36496949716478e-06	1.28645798649794e-05
ENSG00000216802	RP11-390P2.2	4.78184906794015	6.24888172610063e-06	1.49118238778512e-05
ENSG00000223703	AC027612.4	4.75586676006984	6.91934925430349e-06	1.64444713480979e-05
ENSG00000204959	ARHGEF34P	4.75181723222866	2.02468024743675e-08	5.73322863925981e-08
ENSG00000249239	RP11-428L21.2	4.6963718666409	3.18843561814103e-10	9.91890242030686e-10
ENSG00000268034	AC005795.1	4.68286541367784	1.00208377508326e-05	2.34787883352602e-05
ENSG00000222375	RN7SKP127	4.65901838234247	3.88446600626061e-06	9.41895024776633e-06
ENSG00000227939	RPL3P2	4.65646795018941	1.11968597358888e-05	2.6136678874481e-05
ENSG00000214182	PTMAP5	4.64359520962292	2.76415282147476e-16	1.09228026052032e-15
ENSG00000248590	GLDCP1	4.58845838424879	5.73702979857372e-06	1.37356360694167e-05
ENSG00000261963	CCDC92B	4.55536040151256	1.78067524436783e-05	4.08304832016141e-05
ENSG00000256663	RP11-424C20.2	4.54592202359548	1.85576250114971e-05	4.24835989536146e-05
ENSG00000197847	SLC22A20	4.52550720476684	8.4524564669086e-06	1.99184186314593e-05
ENSG00000267102	RP11-686D22.7	4.50863898782389	8.53786567591368e-06	2.01109789388208e-05
ENSG00000240327	RN7SL93P	4.48881166713518	2.61238084024887e-05	5.89858648750474e-05

Table S3: miRNAs detected as differentially expressed in human embryonic stem cells by DESeq2 analysisof the main dataset.

Gene ID	Gene Name	log_2FC	p Value	Adjusted p value
ENSG00000207927	MIR302A	6.90048267115899	3.39279013151305e-13	1.20414916962411e-12
ENSG00000199102	MIR302C	6.33244213278602	9.77622949318177e-15	3.67630364403284e-14
ENSG00000207547	MIR25	4.98957157708772	2.76799252611148e-06	6.78634118956239e-06
ENSG00000207572	MIR363	4.48866264597365	2.53629758435981e-05	5.73432964440636e-05
ENSG00000199145	MIR302D	4.23551461457288	2.09195931666935e-08	5.91912093655873e-08
ENSG00000221771	MIR1205	3.97790483112655	0.000226535673323146	0.000461552652274265
ENSG00000221479	MIR1251	3.63106894342312	0.00100834558556303	0.00190370049492941
ENSG00000265806	MIR4292	3.45294949875045	0.00194953433645063	0.00354137386112856
ENSG00000239594	MIR18B	3.38981322507278	5.6038731620216e-08	1.54489755560899e-07
ENSG00000266235	MIR3176	3.31509865200913	0.00284789583932388	0.00505775520697067
ENSG00000265462	MIR3680-1	3.28997959680575	0.00227620331995399	0.00409987813344758
ENSG00000276404	MIR6835	3.16144489875363	0.00444638980330084	0.00767517021872379
ENSG00000207870	MIR221	3.13133952671024	0.00523088660544535	0.00893066193155564

ENSG00000221176	MIR1207	3.05535853760441	0.00823834023162377	0.0136132735446852
ENSG00000199024	MIR103A2	3.04879363849264	0.00629420788106062	0.0105824059400627
ENSG00000265660	MIR4664	2.9608594581941	0.00816234145738173	0.0134986210122786
ENSG00000265641	MIR3929	2.92976083226772	0.0101258783772607	0.0164968656209119
ENSG00000207973	MIR589	2.86043497913945	0.0166664580609437	0.0260514781153361
ENSG00000264572	MIR4296	2.70912137940815	0.0136876367464436	0.0217504727975898
ENSG00000264069	MIR3943	2.6936138595426	0.0210434169689983	0.0321492159717051
ENSG00000265565	MIR3143	2.58670275198245	0.0200340757577287	0.030859167923453
ENSG00000216031	MIR298	2.53116291844986	0.0215543776709191	0.0328445743751564
ENSG00000276335	MIR205	2.52676673667293	0.0340786927667634	0.0499570112758858
ENSG00000263642	MIR4802	2.4786927149464	0.031134410626263	0.0460396166735536
ENSG00000207757	MIR93	2.36639533539877	0.0106125587529758	0.0172090419349507
ENSG00000207652	MIR621	2.36245506018985	1.32090195650505e-09	3.99278751007248e-09
ENSG00000208036	MIR106B	2.28407707342218	0.0149970203135522	0.0236687470287643
ENSG00000207951	MIR561	2.27847662945195	0.000186861032352044	0.000384436168379998
ENSG00000276712	MIR7111	1.70471156934545	0.00688103274253322	0.0115119885796585
ENSG00000265112	MIR3153	1.58536562351453	0.00651472357584445	0.0109261979790967
ENSG00000199169	MIR367	1.48569131951868	0.03259883316195	0.0479834603256544

Table S4: miRNAs chosen for the ceRNAs investigation already reported as expressed in embryonic stem cells in other studies.

This study*	Suh(2004)	Bar(2008)	Laurent (2008)	Morin(2008)	Li (2009)	Ren(2009)	Parsons (2012)	Asikainen(2015)
hsa-miR-18b		hsa-miR-18b	hsa-miR-18b					hsa-miR-18b-5p
hsa-miR-25		hsa-miR-25		hsa-miR-25			hsa-miR-25	hsa-miR-25-5p
hsa-miR-93					hsa-miR-93		hsa-miR-93	hsa-miR-93-3p and 5p
hsa-miR-103	hsa-miR-103-3p	hsa-miR-103			hsa-miR-103			
hsa-miR-106b		hsa-miR-106b		hsa-miR-106b	hsa-miR-106b			hsa-miR-106b-5p
hsa-miR-205		hsa-miR-205	hsa-miR-205		hsa-miR-205			hsa-miR-205-3p and 5p
hsa-miR-221		hsa-miR-221		hsa-miR-221	hsa-miR-221		hsa-miR-221	
hsa-miR-302a	hsa-miR-302a-3p and 5p	hsa-miR-302a	hsa-miR-302a	hsa-miR-302a		hsa-miR-302a	hsa-miR-302a	hsa-miR-302a-3p and 5p
hsa-miR-302c	hsa-miR-302c-3p and 5p	hsa-miR-302c	hsa-miR-302c	hsa-miR-302c		hsa-miR-302c	hsa-miR-302c	hsa-miR-302c-3p and 5p
hsa-miR-302d	hsa-miR-302d-3p	hsa-miR-302d		hsa-miR-302d		hsa-miR-302d	hsa-miR-302d	hsa-miR-302d-3p and 5p
hsa-miR-363		hsa-miR-363	hsa-miR-363	hsa-miR-363			hsa-miR-363	hsa-miR-363-3p and 5p
hsa-miR-367	hsa-miR-367-3p		hsa-miR-367			hsa-miR-367		hsa-miR-367-3p and 5p
hsa-miR-1251							hsa-miR-1251-5p	

Supplemental material 2 **(Adapted from the original Excel sheets)**

Tabl1 1: lncRNAs targeted by the ESC-expressed miRNAs detected with miRcode.

hsa-miR-18b	hsa-miR-25/363/367	hsa-miR-93	hsa-miR-103a2	hsa-miR-106b	hsa-miR-205	hsa-miR-302acd
ENSG00000215866	ENSG00000226673	ENSG00000225077	ENSG00000215866	ENSG00000225077	ENSG00000225783	ENSG00000225077
ENSG00000225077	ENSG00000226792	ENSG00000225783	ENSG00000225783	ENSG00000225783	ENSG00000226673	ENSG00000225783
ENSG00000225783	ENSG00000238266	ENSG00000244342	ENSG00000225937	ENSG00000244342	ENSG00000226792	ENSG00000244342
ENSG00000231817	ENSG00000244342	ENSG00000248131	ENSG00000226673	ENSG00000248131	ENSG00000234787	ENSG00000248131
ENSG00000253230	ENSG00000248131	ENSG00000249859	ENSG00000226792	ENSG00000249859	ENSG00000235884	ENSG00000249859
ENSG00000253686	ENSG00000253230	ENSG00000253686	ENSG00000234787	ENSG00000253686	ENSG00000237767	ENSG00000253686
ENSG00000255794	ENSG00000254934	ENSG00000255794	ENSG00000235884	ENSG00000255794	ENSG00000238266	ENSG00000258609
ENSG00000258609	ENSG00000255794	ENSG00000258609	ENSG00000238266		ENSG00000244342	
			ENSG00000244342		ENSG00000248131	
			ENSG00000248131		ENSG00000249859	
			ENSG00000250889		ENSG00000255794	
			ENSG00000253230		ENSG00000258609	
			ENSG00000255794			

Table 2: Pseudogenes targeted by the ESC-expressed miRNAs detected with miRcode.

hsa-miR-18b	hsa-miR-25/363/367	hsa-miR-93	hsa-miR-103a2	hsa-miR-106b	hsa-miR-205	hsa-mir302acd
ENSG00000159712	ENSG00000204959	ENSG00000176654	ENSG00000188801	ENSG00000176654	ENSG00000159712	ENSG00000176654
ENSG00000204959	ENSG00000214077	ENSG00000204959	ENSG00000204121	ENSG00000232293	ENSG00000204959	ENSG00000204959
	ENSG00000233309	ENSG00000232293	ENSG00000213763		ENSG00000207220	
			ENSG00000224664		ENSG00000214077	
			ENSG00000235602		ENSG00000226361	
					ENSG00000237624	
					ENSG00000248590	

Table 3: mRNAs targeted by the ESC-expressed miRNAs detected with TargetScanHuman.

hsa-miR-18b	hsa-miR-25/363/367	hsa-miR-93/106b	hsa-miR-103a2	hsa-miR-205	hsa-miR-302a,c,d
ENSG00000178409	ENSG00000183186	ENSG00000146830	ENSG00000151208	ENSG00000113645	ENSG00000146830
ENSG00000160973	ENSG00000196557	ENSG00000175161	ENSG00000104967	ENSG00000065361	ENSG00000175161
ENSG00000146830	ENSG00000100346	ENSG00000147119	ENSG00000168824	ENSG00000153395	ENSG00000214575
ENSG00000131153	ENSG00000175161	ENSG00000100433	ENSG00000110400		ENSG00000198010
ENSG00000135116	ENSG00000183166	ENSG00000137642	ENSG00000176749		ENSG00000104413
ENSG00000160877	ENSG00000149798	ENSG00000124766	ENSG00000164134		ENSG00000184408
	ENSG00000164649	ENSG00000181026	ENSG00000196876		ENSG00000177551
	ENSG00000100604	ENSG00000154065	ENSG00000160460		ENSG00000155096
	ENSG00000147119	ENSG00000186716	ENSG00000143858		ENSG00000083067
	ENSG00000214575	ENSG00000165118	ENSG00000035403		ENSG00000165118
	ENSG00000176490	ENSG00000134508	ENSG00000143126		ENSG00000167670
	ENSG00000151208	ENSG00000122786	ENSG00000139146		ENSG00000158050
	ENSG00000198010	ENSG00000163888	ENSG00000137942		ENSG00000167106
	ENSG00000091073	ENSG00000103326	ENSG00000155980		ENSG00000104059
	ENSG00000163508	ENSG00000170160	ENSG00000177511		ENSG00000154783
	ENSG00000070886	ENSG00000143126	ENSG00000206557		ENSG00000183770
	ENSG00000104413	ENSG00000167670	ENSG00000198963		ENSG00000083307
	ENSG00000111432	ENSG00000116254	ENSG00000103449		ENSG00000162522
	ENSG00000172020	ENSG00000158050	ENSG00000029534		ENSG00000182481
	ENSG00000131459	ENSG00000196411	ENSG00000164076		ENSG00000106852
	ENSG00000152208	ENSG00000167106	ENSG00000105173		ENSG00000157193
	ENSG00000103044	ENSG00000135842	ENSG00000174950		ENSG00000099875

ENSG00000099822	ENSG00000159784	ENSG00000094804	ENSG00000134323
ENSG00000002746	ENSG00000104059	ENSG00000184731	ENSG00000177614
ENSG00000141401	ENSG00000139146	ENSG00000128610	ENSG00000113356
ENSG00000008083	ENSG00000133477	ENSG00000138685	ENSG00000135378
ENSG00000184408	ENSG00000154783	ENSG00000053108	ENSG00000198915
ENSG00000131914	ENSG00000137942	ENSG00000130222	ENSG00000182175
ENSG00000065911	ENSG00000171956	ENSG00000137309	ENSG00000121871
ENSG00000104722	ENSG00000183770	ENSG00000118513	ENSG00000177511
ENSG00000177551	ENSG00000136928	ENSG00000089250	ENSG00000078900
ENSG00000179846	ENSG00000152661	ENSG00000104147	ENSG00000206557
ENSG00000104967	ENSG00000083307	ENSG00000103335	ENSG00000121297
ENSG00000123572	ENSG00000029993	ENSG00000175175	ENSG00000157851
ENSG00000168824	ENSG00000169427	ENSG00000168411	ENSG00000102755
ENSG00000156453	ENSG00000162522	ENSG00000101695	ENSG00000053747
ENSG00000099715	ENSG00000155980	ENSG00000144285	ENSG00000108352
ENSG00000124225	ENSG00000182481	ENSG00000107295	ENSG00000198963
ENSG00000111110	ENSG00000106852	ENSG00000118160	ENSG00000165449
ENSG00000110400	ENSG00000157193	ENSG00000178235	ENSG00000143067
ENSG00000163694	ENSG00000135525	ENSG00000116649	ENSG00000144285
ENSG00000176406	ENSG00000099875	ENSG00000134207	ENSG00000198720
ENSG00000188322	ENSG00000125966	ENSG00000115112	ENSG00000164045
ENSG00000091490	ENSG00000134323	ENSG00000147202	ENSG00000114346
ENSG00000151012	ENSG00000181163	ENSG00000165521	ENSG00000139132
ENSG00000063176	ENSG00000177614	ENSG00000027869	ENSG00000162344
ENSG00000107742	ENSG00000051341	ENSG00000138771	ENSG00000132470

ENSG00000067715	ENSG00000113356	ENSG00000112246	ENSG00000158445
ENSG00000177426	ENSG00000135378	ENSG00000102172	ENSG00000243709
ENSG00000113645	ENSG00000059915	ENSG00000166860	ENSG00000143768
ENSG00000042286	ENSG00000165186		ENSG00000183742
ENSG00000155096	ENSG00000131242		ENSG00000184613
ENSG00000135083	ENSG00000198915		ENSG00000115884
ENSG00000176749	ENSG00000113319		ENSG00000082497
ENSG00000145920	ENSG00000182175		ENSG00000145087
ENSG00000147459	ENSG00000143365		ENSG00000143469
ENSG00000134317	ENSG00000117676		ENSG00000090447
ENSG00000145703	ENSG00000171848		ENSG00000103460
ENSG00000100433	ENSG00000256463		ENSG00000072657
ENSG00000164134	ENSG00000162105		ENSG00000143816
ENSG00000146083	ENSG00000155886		ENSG00000147874
ENSG00000196876	ENSG00000196517		ENSG00000182601
ENSG00000137642	ENSG00000121871		ENSG00000183873
ENSG00000124766	ENSG00000147481		ENSG00000198846
ENSG00000160460	ENSG00000164651		ENSG00000204366
ENSG00000143858	ENSG00000187678		
ENSG00000083067	ENSG00000177511		
ENSG00000035403	ENSG00000120659		
ENSG00000074657	ENSG00000149115		
	ENSG00000078900		
	ENSG00000206557		
	ENSG00000121297		

ENSG00000168234
ENSG00000143494
ENSG00000177076
ENSG00000105993
ENSG00000157851
ENSG00000065361
ENSG00000117525
ENSG00000121104
ENSG00000102302
ENSG00000102755
ENSG00000145907
ENSG00000112218
ENSG00000053747
ENSG00000144749
ENSG00000120539
ENSG00000065320
ENSG00000163002
ENSG00000108352
ENSG00000198963
ENSG00000103449
ENSG00000163935
ENSG00000144040
ENSG00000138606
ENSG00000165449
ENSG00000151718

ENSG00000206579
ENSG00000176105
ENSG00000166261
ENSG00000143067

Table 4: lncRNAs targeted by the ESC-expressed miRNAs detected with StarBase.

hsa-miR-93	hsa-miR-103a2	hsa-miR-106b	hsa-miR-205	hsa-miR-302acd	hsa-miR-363	hsa-miR367
ENSG00000175701	ENSG00000222041	ENSG00000175701	ENSG00000203875	ENSG00000263731	ENSG00000203875	ENSG00000203875
ENSG00000206337	ENSG00000238266	ENSG00000206337	ENSG00000234741		ENSG00000249790	ENSG00000249790
ENSG00000236673	ENSG00000248538	ENSG00000236673	ENSG00000238266			
ENSG00000249859	ENSG00000249790	ENSG00000249859	ENSG00000249790			
ENSG00000250899		ENSG00000250312				
ENSG00000261121		ENSG00000250899				
		ENSG00000261121				

Table 5 :Pseudogenes targeted by the ESC-expressed miRNAs detected with StarBase.

hsa-miR-25	hsa-miR-93	hsa-miR-103a2	hsa-miR-106b	hsa-miR-205	hsa-mir-302acd	hsa-miR-363	hsa-miR-367
ENSG00000215835	ENSG00000233426	ENSG00000228502	ENSG00000233426	ENSG00000234176	ENSG00000228502	ENSG00000215835	ENSG00000215835
ENSG00000157152	ENSG00000234964	ENSG00000249264	ENSG00000234964	ENSG00000256356	ENSG00000249264	ENSG00000213430	ENSG00000213430
ENSG00000213430	ENSG00000256356	ENSG00000241293	ENSG00000256356	ENSG00000224773	ENSG00000233426		
	ENSG00000242445	ENSG00000220472	ENSG00000242445	ENSG00000213442	ENSG00000234964		
	ENSG00000188933	ENSG00000236552	ENSG00000188933	ENSG00000181524	ENSG00000249353		
	ENSG00000183199	ENSG00000233476	ENSG00000183199	ENSG00000243680	ENSG00000225159		
		ENSG00000218426		ENSG00000231940	ENSG00000213881		
		ENSG00000214185		ENSG00000213740	ENSG00000234825		
		ENSG00000228205		ENSG00000127589	ENSG00000233476		
		ENSG00000214199		ENSG00000249565	ENSG00000214199		
		ENSG00000213430		ENSG00000239246	ENSG00000183199		
		ENSG00000215154		ENSG00000139239	ENSG00000250182		
		ENSG00000250182					

Table 6: mRNAs targeted by the ESC-expressed miRNAs detected with Starbase .

hsa-mir103a2	hsa-miR106b	hsa-mir-205	hsa-mir-302a	hsa-miR-302c	hsa-miR-302d	hsa-miR-363	hsa-miR-367
ENSG00000029993	ENSG00000035403	ENSG00000076604	ENSG00000051341	ENSG00000051341	ENSG00000051341	ENSG00000035403	ENSG00000035403
ENSG00000035403	ENSG00000053747	ENSG00000096063	ENSG00000053747	ENSG00000053747	ENSG00000053747	ENSG00000065150	ENSG00000065150
ENSG00000053108	ENSG00000065150	ENSG00000101003	ENSG00000065361	ENSG00000065361	ENSG00000065361	ENSG00000065911	ENSG00000065911
ENSG00000065150	ENSG00000065320	ENSG00000105974	ENSG00000074657	ENSG00000074657	ENSG00000076604	ENSG00000074657	ENSG00000074657
ENSG00000103449	ENSG00000065361	ENSG00000106105	ENSG00000076604	ENSG00000076604	ENSG00000077942	ENSG00000101191	ENSG00000101191
ENSG00000104147	ENSG00000065911	ENSG00000117724	ENSG00000077942	ENSG00000077942	ENSG00000096063	ENSG00000111962	ENSG00000111231
ENSG00000111231	ENSG00000073008	ENSG00000119335	ENSG00000097021	ENSG00000088035	ENSG00000099875	ENSG00000113368	ENSG00000111962
ENSG00000119335	ENSG00000074657	ENSG00000134508	ENSG00000099875	ENSG00000099875	ENSG00000100749	ENSG00000119866	ENSG00000113368
ENSG00000119866	ENSG00000075131	ENSG00000136830	ENSG00000099954	ENSG00000100749	ENSG00000101003	ENSG00000120756	ENSG00000119866
ENSG00000120254	ENSG00000081913	ENSG00000137942	ENSG00000100749	ENSG00000101003	ENSG00000102755	ENSG00000122786	ENSG00000120254
ENSG00000122786	ENSG00000088035	ENSG00000139146	ENSG00000101003	ENSG00000102755	ENSG00000104059	ENSG00000124532	ENSG00000122786
ENSG00000123609	ENSG00000096063	ENSG00000139800	ENSG00000102755	ENSG00000104059	ENSG00000105974	ENSG00000124766	ENSG00000123609
ENSG00000133119	ENSG00000100749	ENSG00000141682	ENSG00000104059	ENSG00000105974	ENSG00000105993	ENSG00000129351	ENSG00000124532
ENSG00000134057	ENSG00000101191	ENSG00000145555	ENSG00000105974	ENSG00000105993	ENSG00000106852	ENSG00000135842	ENSG00000124766
ENSG00000134323	ENSG00000102302	ENSG00000145703	ENSG00000105993	ENSG00000106852	ENSG00000108352	ENSG00000136231	ENSG00000129351
ENSG00000135525	ENSG00000102755	ENSG00000147676	ENSG00000106852	ENSG00000108352	ENSG00000111231	ENSG00000136603	ENSG00000135842
ENSG00000136244	ENSG00000103449	ENSG00000155096	ENSG00000108352	ENSG00000111231	ENSG00000112118	ENSG00000137642	ENSG00000136231
ENSG00000137265	ENSG00000105993	ENSG00000155561	ENSG00000111231	ENSG00000112118	ENSG00000113356	ENSG00000139132	ENSG00000136603
ENSG00000137713	ENSG00000108352	ENSG00000174469	ENSG00000111962	ENSG00000113356	ENSG00000113368	ENSG00000143476	ENSG00000137642
ENSG00000137942	ENSG00000111231	ENSG00000176105	ENSG00000112118	ENSG00000113368	ENSG00000115339	ENSG00000143494	ENSG00000139132
ENSG00000139146	ENSG00000111962	ENSG00000187514	ENSG00000113356	ENSG00000115339	ENSG00000119866	ENSG00000145703	ENSG00000143476
ENSG00000139800	ENSG00000112218	ENSG00000088538	ENSG00000113368	ENSG00000119866	ENSG00000120539	ENSG00000150594	ENSG00000143494

ENSG00000141682	ENSG00000112365	ENSG00000095203	ENSG00000115339	ENSG00000120539	ENSG00000121104	ENSG00000151718	ENSG00000144730
ENSG00000143126	ENSG00000113368	ENSG00000108064	ENSG00000119866	ENSG00000121104	ENSG00000122644	ENSG00000155096	ENSG00000144749
ENSG00000143799	ENSG00000117525	ENSG00000113013	ENSG00000120539	ENSG00000122644	ENSG00000122786	ENSG00000155561	ENSG00000145703
ENSG00000146477	ENSG00000119335	ENSG00000116133	ENSG00000121104	ENSG00000122786	ENSG00000124486	ENSG00000157193	ENSG00000151718
ENSG00000146757	ENSG00000120254	ENSG00000123737	ENSG00000121297	ENSG00000124532	ENSG00000124532	ENSG00000165449	ENSG00000155096
ENSG00000151304	ENSG00000120539	ENSG00000132341	ENSG00000122644	ENSG00000124766	ENSG00000124766	ENSG00000166986	ENSG00000157193
ENSG00000154065	ENSG00000120756	ENSG00000138685	ENSG00000122786	ENSG00000124875	ENSG00000124875	ENSG00000171246	ENSG00000163814
ENSG00000155096	ENSG00000121104	ENSG00000138771	ENSG00000124532	ENSG00000125966	ENSG00000125966	ENSG00000176749	ENSG00000165449
ENSG00000155760	ENSG00000122786	ENSG00000146670	ENSG00000124766	ENSG00000126953	ENSG00000126953	ENSG00000181026	ENSG00000166986
ENSG00000155980	ENSG00000124532	ENSG00000148484	ENSG00000124875	ENSG00000130119	ENSG00000130119	ENSG00000181163	ENSG00000167106
ENSG00000164045	ENSG00000124766	ENSG00000150051	ENSG00000125966	ENSG00000132274	ENSG00000132274	ENSG00000181544	ENSG00000171241
ENSG00000165449	ENSG00000125733	ENSG00000166526	ENSG00000126953	ENSG00000133101	ENSG00000133101	ENSG00000187772	ENSG00000171246
ENSG00000166900	ENSG00000126953	ENSG00000175063	ENSG00000130119	ENSG00000134323	ENSG00000133119	ENSG00000204516	ENSG00000176749
ENSG00000170779	ENSG00000128050	ENSG00000196781	ENSG00000132274	ENSG00000135378	ENSG00000134323	ENSG00000008083	ENSG00000187772
ENSG00000171241	ENSG00000129351	ENSG00000113407	ENSG00000133101	ENSG00000135842	ENSG00000135378	ENSG00000010278	ENSG00000008083
ENSG00000176749	ENSG00000130119	ENSG00000136014	ENSG00000134323	ENSG00000136425	ENSG00000135842	ENSG00000091490	ENSG00000010278
ENSG00000177156	ENSG00000130787	ENSG00000163694	ENSG00000135378	ENSG00000136603	ENSG00000136425	ENSG00000095203	ENSG00000091490
ENSG00000187772	ENSG00000132031	ENSG00000005100	ENSG00000136425	ENSG00000137713	ENSG00000136603	ENSG00000105750	ENSG00000095203
ENSG00000196411	ENSG00000135521	ENSG00000043355	ENSG00000136603	ENSG00000139132	ENSG00000137713	ENSG00000111432	ENSG00000105750
ENSG00000197971	ENSG00000135842	ENSG00000050344	ENSG00000137713	ENSG00000143067	ENSG00000137942	ENSG00000111799	ENSG00000109854
ENSG00000204516	ENSG00000136231	ENSG00000095002	ENSG00000137942	ENSG00000143799	ENSG00000139132	ENSG00000118513	ENSG00000111432
ENSG00000008083	ENSG00000136603	ENSG00000104611	ENSG00000139132	ENSG00000144730	ENSG00000143799	ENSG00000118515	ENSG00000111799
ENSG00000010165	ENSG00000137642	ENSG00000104635	ENSG00000143067	ENSG00000145103	ENSG00000145103	ENSG00000121057	ENSG00000118513
ENSG00000010278	ENSG00000137713	ENSG00000114346	ENSG00000143799	ENSG00000146477	ENSG00000146477	ENSG00000138685	ENSG00000118515
ENSG00000011009	ENSG00000138032	ENSG00000121211	ENSG00000145103	ENSG00000146757	ENSG00000146830	ENSG00000149798	ENSG00000121057

ENSG00000027869	ENSG00000139132	ENSG00000128059	ENSG00000146477	ENSG00000146830	ENSG00000151690	ENSG00000151208	ENSG00000138685
ENSG00000035499	ENSG00000139835	ENSG00000128567	ENSG00000146757	ENSG00000151690	ENSG00000152284	ENSG00000157214	ENSG00000149798
ENSG00000066468	ENSG00000140015	ENSG00000130713	ENSG00000146830	ENSG00000152284	ENSG00000152455	ENSG00000164134	ENSG00000151208
ENSG00000070371	ENSG00000143067	ENSG00000132003	ENSG00000151690	ENSG00000152455	ENSG00000154065	ENSG00000164136	ENSG00000157214
ENSG00000070669	ENSG00000143476	ENSG00000133706	ENSG00000152284	ENSG00000154065	ENSG00000155096	ENSG00000165591	ENSG00000164134
ENSG00000071564	ENSG00000143494	ENSG00000134317	ENSG00000152455	ENSG00000155096	ENSG00000155561	ENSG00000166526	ENSG00000164136
ENSG00000076248	ENSG00000143799	ENSG00000135248	ENSG00000154065	ENSG00000155561	ENSG00000156876	ENSG00000168280	ENSG00000165591
ENSG00000081181	ENSG00000144040	ENSG00000141480	ENSG00000155096	ENSG00000156876	ENSG00000157193	ENSG00000170340	ENSG00000166526
ENSG00000088205	ENSG00000144730	ENSG00000145423	ENSG00000155561	ENSG00000157193	ENSG00000158050	ENSG00000182010	ENSG00000168280
ENSG00000088538	ENSG00000144749	ENSG00000146038	ENSG00000155886	ENSG00000158050	ENSG00000159247	ENSG00000184731	ENSG00000182010
ENSG00000091490	ENSG00000145703	ENSG00000148459	ENSG00000156876	ENSG00000159247	ENSG00000162522	ENSG00000196247	ENSG00000184731
ENSG00000092421	ENSG00000146757	ENSG00000153395	ENSG00000157193	ENSG00000162522	ENSG00000163002	ENSG00000196781	ENSG00000196781
ENSG00000092621	ENSG00000150594	ENSG00000163755	ENSG00000158050	ENSG00000163002	ENSG00000164045	ENSG00000196876	ENSG00000196876
ENSG00000095203	ENSG00000151690	ENSG00000164109	ENSG00000159247	ENSG00000164045	ENSG00000165118	ENSG00000197147	ENSG00000197147
ENSG00000096384	ENSG00000151718	ENSG00000164985	ENSG00000162522	ENSG00000165118	ENSG00000165449	ENSG00000247626	ENSG00000015133
ENSG00000099399	ENSG00000152455	ENSG00000165175	ENSG00000163002	ENSG00000165449	ENSG00000166900	ENSG00000006128	ENSG00000063176
ENSG00000099800	ENSG00000155096	ENSG00000167085	ENSG00000164045	ENSG00000166986	ENSG00000167106	ENSG00000015133	ENSG00000067715
ENSG00000100167	ENSG00000155561	ENSG00000167721	ENSG00000165118	ENSG00000167106	ENSG00000167670	ENSG00000063176	ENSG00000092853
ENSG00000100304	ENSG00000156876	ENSG00000169992	ENSG00000165449	ENSG00000167670	ENSG00000168079	ENSG00000067715	ENSG00000099715
ENSG00000100410	ENSG00000157193	ENSG00000170786	ENSG00000166900	ENSG00000168079	ENSG00000171241	ENSG00000092853	ENSG00000099822
ENSG00000100522	ENSG00000161547	ENSG00000174951	ENSG00000167106	ENSG00000171241	ENSG00000174469	ENSG00000096717	ENSG00000100604
ENSG00000101115	ENSG00000163002	ENSG00000176490	ENSG00000167670	ENSG00000171246	ENSG00000174600	ENSG00000099715	ENSG00000102290
ENSG00000101220	ENSG00000163935	ENSG00000179750	ENSG00000168079	ENSG00000174469	ENSG00000177551	ENSG00000099822	ENSG00000103356
ENSG00000101695	ENSG00000165449	ENSG00000180198	ENSG00000168234	ENSG00000174600	ENSG00000181163	ENSG00000100604	ENSG00000104413
ENSG00000102172	ENSG00000166261	ENSG00000182004	ENSG00000171241	ENSG00000177551	ENSG00000182175	ENSG00000102290	ENSG00000107742

ENSG00000105173	ENSG00000166986	ENSG00000183943	ENSG00000174469	ENSG00000181163	ENSG00000182481	ENSG00000103356	ENSG00000110660
ENSG00000105202	ENSG00000167972	ENSG00000185917	ENSG00000174600	ENSG00000182175	ENSG00000184368	ENSG00000104413	ENSG00000111110
ENSG00000105255	ENSG00000171246	ENSG00000186197	ENSG00000177551	ENSG00000182481	ENSG00000184613	ENSG00000107742	ENSG00000113407
ENSG00000105750	ENSG00000176105	ENSG00000206337	ENSG00000181163	ENSG00000184368	ENSG00000187514	ENSG00000110090	ENSG00000123610
ENSG00000106070	ENSG00000179862	ENSG00000207001	ENSG00000182175	ENSG00000184613	ENSG00000196517	ENSG00000110660	ENSG00000124225
ENSG00000106785	ENSG00000181026	ENSG00000239900	ENSG00000182481	ENSG00000187514	ENSG00000196793	ENSG00000111110	ENSG00000125285
ENSG00000107295	ENSG00000181163		ENSG00000184368	ENSG00000187772	ENSG00000197971	ENSG00000113407	ENSG00000131459
ENSG00000108064	ENSG00000184368		ENSG00000184613	ENSG00000196517	ENSG00000198780	ENSG00000116761	ENSG00000132646
ENSG00000109854	ENSG00000185252		ENSG00000187514	ENSG00000196793	ENSG00000198825	ENSG00000123610	ENSG00000136014
ENSG00000110400	ENSG00000196172		ENSG00000187772	ENSG00000197971	ENSG00000198915	ENSG00000123975	ENSG00000141401
ENSG00000111247	ENSG00000196793		ENSG00000196517	ENSG00000198780	ENSG00000198963	ENSG00000124225	ENSG00000145536
ENSG00000111331	ENSG00000198042		ENSG00000196793	ENSG00000198825	ENSG00000204516	ENSG00000125285	ENSG00000147119
ENSG00000111432	ENSG00000198780		ENSG00000197971	ENSG00000198915	ENSG00000088205	ENSG00000131459	ENSG00000148840
ENSG00000111696	ENSG00000198825		ENSG00000198780	ENSG00000198963	ENSG00000111247	ENSG00000132646	ENSG00000151012
ENSG00000111799	ENSG00000198963		ENSG00000198825	ENSG00000204516	ENSG00000111432	ENSG00000136014	ENSG00000153898
ENSG00000112297	ENSG00000204516		ENSG00000198915	ENSG00000088205	ENSG00000115107	ENSG00000136159	ENSG00000162702
ENSG00000113013	ENSG00000212993		ENSG00000198963	ENSG00000111247	ENSG00000119969	ENSG00000141401	ENSG00000163694
ENSG00000113645	ENSG00000008083		ENSG00000204516	ENSG00000111432	ENSG00000121057	ENSG00000145536	ENSG00000163735
ENSG00000115107	ENSG00000010278		ENSG00000088205	ENSG00000115107	ENSG00000157851	ENSG00000147119	ENSG00000164649
ENSG00000115541	ENSG00000091490		ENSG00000095203	ENSG00000119969	ENSG00000166526	ENSG00000148840	ENSG00000164841
ENSG00000116133	ENSG00000095203		ENSG00000105750	ENSG00000121057	ENSG00000067715	ENSG00000151012	ENSG00000170571
ENSG00000116649	ENSG00000111432		ENSG00000111247	ENSG00000130294	ENSG00000111110	ENSG00000153898	ENSG00000172164
ENSG00000118513	ENSG00000118513		ENSG00000111432	ENSG00000157851	ENSG00000136159	ENSG00000163694	ENSG00000173404
ENSG00000118515	ENSG00000118515		ENSG00000115107	ENSG00000166526	ENSG00000154237	ENSG00000163735	ENSG00000177426
ENSG00000118640	ENSG00000138685		ENSG00000119969	ENSG00000067715	ENSG00000166845	ENSG00000164649	ENSG00000180767

ENSG00000119969	ENSG00000149798	ENSG00000121057	ENSG00000096717	ENSG00000170571	ENSG00000164841	ENSG00000183186
ENSG00000120885	ENSG00000151208	ENSG00000138709	ENSG00000110090	ENSG00000179299	ENSG00000167513	ENSG00000184254
ENSG00000121057	ENSG00000157851	ENSG00000150907	ENSG00000111110	ENSG00000114346	ENSG00000170571	ENSG00000188191
ENSG00000122140	ENSG00000164136	ENSG00000157851	ENSG00000136159	ENSG00000134317	ENSG00000172164	ENSG00000189184
ENSG00000123737	ENSG00000165591	ENSG00000166526	ENSG00000166845	ENSG00000065328	ENSG00000173404	ENSG00000197961
ENSG00000124194	ENSG00000168280	ENSG00000067715	ENSG00000170571	ENSG00000073536	ENSG00000177426	ENSG00000278259
ENSG00000124641	ENSG00000184731	ENSG00000096717	ENSG00000179299	ENSG00000076003	ENSG00000180767	ENSG00000128059
ENSG00000125901	ENSG00000196247	ENSG00000110090	ENSG00000114346	ENSG00000078269	ENSG00000183186	ENSG00000134317
ENSG00000125977	ENSG00000196781	ENSG00000111110	ENSG00000153395	ENSG00000078900	ENSG00000184254	ENSG00000148459
ENSG00000127399	ENSG00000006128	ENSG00000123610	ENSG00000180198	ENSG00000081277	ENSG00000188191	ENSG00000164985
ENSG00000130294	ENSG00000015133	ENSG00000166845	ENSG00000065328	ENSG00000090447	ENSG00000189184	ENSG00000176490
ENSG00000130540	ENSG00000063176	ENSG00000170571	ENSG00000073536	ENSG00000100068	ENSG00000197961	ENSG00000186197
ENSG00000130558	ENSG00000067715	ENSG00000179299	ENSG00000076003	ENSG00000103044	ENSG00000128059	ENSG00000147459
ENSG00000132182	ENSG00000092853	ENSG00000114346	ENSG00000078269	ENSG00000108604	ENSG00000134317	ENSG00000163811
ENSG00000132341	ENSG00000096717	ENSG00000153395	ENSG00000078900	ENSG00000109113	ENSG00000148459	ENSG00000174564
ENSG00000135632	ENSG00000099715	ENSG00000065328	ENSG00000081277	ENSG00000109606	ENSG00000109606	ENSG00000174574
ENSG00000135749	ENSG00000099822	ENSG00000073536	ENSG00000090447	ENSG00000111371	ENSG00000147459	ENSG00000144381
ENSG00000135916	ENSG00000100604	ENSG00000076003	ENSG00000100068	ENSG00000115884	ENSG00000163811	ENSG00000011376
ENSG00000137124	ENSG00000101444	ENSG00000078269	ENSG00000103044	ENSG00000117322	ENSG00000174564	ENSG00000030066
ENSG00000137309	ENSG00000102290	ENSG00000078900	ENSG00000108604	ENSG00000117394	ENSG00000174574	ENSG00000042286
ENSG00000138311	ENSG00000103356	ENSG00000081277	ENSG00000109113	ENSG00000118242	ENSG00000144381	ENSG00000078053
ENSG00000138685	ENSG00000104413	ENSG00000090447	ENSG00000109606	ENSG00000126749	ENSG00000011376	ENSG00000079337
ENSG00000138709	ENSG00000107742	ENSG00000100068	ENSG00000111371	ENSG00000136270	ENSG00000030066	ENSG00000091073
ENSG00000138771	ENSG00000110090	ENSG00000103044	ENSG00000115884	ENSG00000138030	ENSG00000042286	ENSG00000091127
ENSG00000139514	ENSG00000110660	ENSG00000108604	ENSG00000117322	ENSG00000140284	ENSG00000078053	ENSG00000104722

ENSG00000139998	ENSG00000111110	ENSG00000109113	ENSG00000117394	ENSG00000143768	ENSG00000079337	ENSG00000130935
ENSG00000141738	ENSG00000113407	ENSG00000109606	ENSG00000118242	ENSG00000143811	ENSG00000091073	ENSG00000135083
ENSG00000141750	ENSG00000116761	ENSG00000111371	ENSG00000126749	ENSG00000147459	ENSG00000091127	ENSG00000135486
ENSG00000143469	ENSG00000123610	ENSG00000115884	ENSG00000136270	ENSG00000147874	ENSG00000104722	ENSG00000136492
ENSG00000143858	ENSG00000123975	ENSG00000117322	ENSG00000138030	ENSG00000148143	ENSG00000130935	ENSG00000138496
ENSG00000146670	ENSG00000124225	ENSG00000117394	ENSG00000140284	ENSG00000162062	ENSG00000135083	ENSG00000139734
ENSG00000148484	ENSG00000125285	ENSG00000118242	ENSG00000143768	ENSG00000162458	ENSG00000135486	ENSG00000146083
ENSG00000148843	ENSG00000131459	ENSG00000126749	ENSG00000143811	ENSG00000163421	ENSG00000136492	ENSG00000147799
ENSG00000149554	ENSG00000132646	ENSG00000136270	ENSG00000147459	ENSG00000163811	ENSG00000138496	ENSG00000151503
ENSG00000149798	ENSG00000136014	ENSG00000138030	ENSG00000147874	ENSG00000163840	ENSG00000139734	ENSG00000157152
ENSG00000150051	ENSG00000136159	ENSG00000140284	ENSG00000148143	ENSG00000166123	ENSG00000146083	ENSG00000164934
ENSG00000150753	ENSG00000141401	ENSG00000143768	ENSG00000162062	ENSG00000167747	ENSG00000147799	ENSG00000173145
ENSG00000150907	ENSG00000145536	ENSG00000143811	ENSG00000162458	ENSG00000174564	ENSG00000151503	ENSG00000176974
ENSG00000151208	ENSG00000147119	ENSG00000147459	ENSG00000163421	ENSG00000180817	ENSG00000157152	ENSG00000182903
ENSG00000152939	ENSG00000148840	ENSG00000147874	ENSG00000163811	ENSG00000181274	ENSG00000164934	ENSG00000198131
ENSG00000154096	ENSG00000151012	ENSG00000148143	ENSG00000163840	ENSG00000185090	ENSG00000173145	ENSG00000050327
ENSG00000157214	ENSG00000153898	ENSG00000162062	ENSG00000166123	ENSG00000185274	ENSG00000176974	ENSG00000072041
ENSG00000157851	ENSG00000154237	ENSG00000162458	ENSG00000167747	ENSG00000187398	ENSG00000182903	ENSG00000106305
ENSG00000158813	ENSG00000162702	ENSG00000163421	ENSG00000174564	ENSG00000197385	ENSG00000198131	ENSG00000131914
ENSG00000159111	ENSG00000163694	ENSG00000163811	ENSG00000180817	ENSG00000198435	ENSG00000225697	ENSG00000151702
ENSG00000160193	ENSG00000163735	ENSG00000163840	ENSG00000181274	ENSG00000198846		ENSG00000166197
ENSG00000160208	ENSG00000164649	ENSG00000166123	ENSG00000185090	ENSG00000215417		ENSG00000175161
ENSG00000160460	ENSG00000164841	ENSG00000167747	ENSG00000185274	ENSG00000243709		
ENSG00000161243	ENSG00000166845	ENSG00000174564	ENSG00000187398	ENSG00000060140		
ENSG00000163170	ENSG00000167513	ENSG00000180817	ENSG00000197385	ENSG00000158373		

ENSG00000163923 ENSG00000170571	ENSG00000181274	ENSG00000198435	ENSG00000099849
ENSG00000164134 ENSG00000172164	ENSG00000185090	ENSG00000198846	ENSG00000126457
ENSG00000164136 ENSG00000173404	ENSG00000185274	ENSG00000215417	ENSG00000139921
ENSG00000164237 ENSG00000177426	ENSG00000187398	ENSG00000243709	ENSG00000144381
ENSG00000164251 ENSG00000179299	ENSG00000197385	ENSG00000060140	ENSG00000162493
ENSG00000164305 ENSG00000180767	ENSG00000198435	ENSG00000138750	ENSG00000164163
ENSG00000165072 ENSG00000183186	ENSG00000198846	ENSG00000158373	ENSG00000164181
ENSG00000165591 ENSG00000184254	ENSG00000214575	ENSG00000172379	
ENSG00000166165 ENSG00000188191	ENSG00000215417	ENSG00000174574	
ENSG00000166508 ENSG00000189184	ENSG00000243709	ENSG00000196497	
ENSG00000166526 ENSG00000197961			
ENSG00000166803 ENSG00000278259			
ENSG00000166860			
ENSG00000167600			
ENSG00000168280			
ENSG00000168411			
ENSG00000169710			
ENSG00000169715			
ENSG00000170340			
ENSG00000171126			
ENSG00000173153			
ENSG00000173207			
ENSG00000173457			
ENSG00000173898			
ENSG00000174371			

ENSG00000174607
ENSG00000175063
ENSG00000175931
ENSG00000176092
ENSG00000177076
ENSG00000177542
ENSG00000178235
ENSG00000178498
ENSG00000182010
ENSG00000184731
ENSG00000187017
ENSG00000188807
ENSG00000196247
ENSG00000196305
ENSG00000196781
ENSG00000196876
ENSG00000197147
ENSG00000197417
ENSG00000197594
ENSG00000197905
ENSG00000198039
ENSG00000198246
ENSG00000213186
ENSG00000213465
ENSG00000213988

ENSG00000247077
ENSG00000247626
ENSG00000253626

Supplemental Material 3 **(Adapted from the original Excel sheets)**

Table 1: Gene ontologies with FDR<= 0.05. Only the relevant biological processes.

GOID	Description
GO:0030154	cell differentiation
GO:0000904	cell morphogenesis involved in differentiation
GO:0045595	regulation of cell differentiation
GO:0007346	regulation of mitotic cell cycle
GO:0043408	regulation of MAPK cascade
GO:0000165	MAPK cascade
GO:0008284	positive regulation of cell proliferation
GO:0051093	negative regulation of developmental process
GO:0045596	negative regulation of cell differentiation
GO:0043410	positive regulation of MAPK cascade
GO:0045664	regulation of neuron differentiation
GO:0045597	positive regulation of cell differentiation
GO:0008543	fibroblast growth factor receptor signaling pathway
GO:0038127	ERBB signaling pathway
GO:0044344	cellular response to fibroblast growth factor stimulus
GO:0071774	response to fibroblast growth factor
GO:0045787	positive regulation of cell cycle
GO:0019827	stem cell population maintenance
GO:0071560	cellular response to transforming growth factor beta stimulus
GO:0000186	activation of MAPKK activity
GO:0071559	response to transforming growth factor beta
GO:0007179	transforming growth factor beta receptor signaling pathway
GO:0070372	regulation of ERK1 and ERK2 cascade
GO:0010833	telomere maintenance via telomere lengthening
GO:0070371	ERK1 and ERK2 cascade
GO:0000723	telomere maintenance
GO:0032200	telomere organization
GO:0032201	telomere maintenance via semi-conservative replication
GO:0035019	somatic stem cell population maintenance
GO:0043122	regulation of I-kappaB kinase/NF-kappaB signaling
GO:0051896	regulation of protein kinase B signaling
GO:0043406	positive regulation of MAP kinase activity
GO:0010771	negative regulation of cell morphogenesis involved in differentiation
GO:0007249	I-kappaB kinase/NF-kappaB signaling
GO:0048863	stem cell differentiation
GO:0045598	regulation of fat cell differentiation
GO:0030178	negative regulation of Wnt signaling pathway

GO:0045931	positive regulation of mitotic cell cycle
GO:0051092	positive regulation of NF-kappaB transcription factor activity
GO:0038092	nodal signaling pathway
GO:0043552	positive regulation of phosphatidylinositol 3-kinaseactivity
GO:0043123	positive regulation of I-kappaB kinase/NF-kappaB signaling
GO:0045665	negative regulation of neuron differentiation
GO:0010769	Term: regulation of cell morphogenesis involved in differentiation
GO:0000187	activation of MAPK activity
GO:0045599	negative regulation of fat cell differentiation
GO:0071456	cellular response to hypoxia
GO:1901992	positive regulation of mitotic cell cycle phase transition
GO:0017145	stem cell division
GO:0043551	regulation of phosphatidylinositol 3-kinase activity

Supplemental material 4 **(Adapted from the original Excel sheets)**

RNA pairs with correlation coefficient equal or above 0.6 ($p \leq 0,05$)

Table 1: mRNAs positively correlates with lncRNAs.

Predicted targets		
miRNA	**mRNA**	**lncRNA**
hsa-mir-25, -302a,c,d	AZIN1	LINC01194
hsa-miR-302a,c,d	CDC25A	PVT1
hsa-mir-103a2	HMGA1	RP11-20D14.6
hsa-mir-25	MTHFD2	LINC01194
hsa-mir-93 and -106b	NUP35	LINC01194
hsa-mir-93 and -106b	SFMBT1	LINC00337
hsa-mir-103a2	SH3GL2	RP11-20D14.6
hsa-mir-25	SLC7A11	LINC01194
hsa-miR-302a,c,d	TOX3	LINC-ROR
Validated targets		
miRNA	**mRNA**	**lncRNA**
hsa-mir-106b	AHCY	RP11-253E3.3
hsa-mir-103a2	ARL2	RP11-20D14.6
hsa-mir-18b	AZIN1	GAS5
hsa-mir-103a2	CCT5	RP11-20D14.6
hsa-mir-93	CDC25A	RP11-253E3.3
hsa-miR-302a,c,d	CHAF1A	RP11-20D14.6
hsa-mir-106b	CLSPN	RP11-253E3.3
hsa-mir-93 and -106b	DIDO1	RP11-253E3.3
hsa-mir-103a2	EPHB4	RP11-20D14.6
hsa-mir-106b	ERBB3	RP11-69I8.2
hsa-mir-205	ETF1	GAS5
hsa-miR-302a,c,d	FAM102A	RP11-20D14.6
hsa-mir-205	FGF2	GAS5
hsa-mir-103a2	FGFR2	RP11-20D14.6
hsa-mir-106b	GFPT2	RP11-253E3.3
hsa-mir-103a2	HMGA1	RP11-20D14.6
hsa-mir-25	HNRNPA1	RP11-20D14.6
hsa-mir-103a2 and -205	HSPA9	RP11-20D14.6
hsa-mir-25, -302a,d, -363 and -367	HSPD1	SNHG5
hsa-mir-93 and -106b	INO80	RP11-253E3.3

hsa-mir-18b	IPO5	GAS5
hsa-mir-103a2 and -302c	KIF1A	RP11-20D14.6
hsa-mir-205	MAD2L1	GAS5
hsa-mir-205	MAL2	GAS5
hsa-mir-25	MET	RP11-20D14.6
hsa-mir-93	NASP	RP11-69I8.2
hsa-mir-25 and -302a,c,d	NPM1	RP11-20D14.6
hsa-mir-103a2 and -302a,c,d	PARP1	RP11-20D14.6
hsa-mir-103a2	PCNX2	RP11-20D14.6
hsa-mir-205	PODXL	RP11-20D14.6
hsa-mir-18b	PRIM2	GAS5
hsa-mir-18b	PSAT1	GAS5
hsa-miR-302a,c,d	RAD51AP1	SNHG5
hsa-mir-205	RAN	SNHG5
hsa-mir-103a2	RPL39L	RP11-20D14.6
hsa-mir-103a2	SALL4	RP11-20D14.6
hsa-mir-93 and -106b	SASS6	RP11-69I8.2
hsa-mir-103a2	SEMA6A	RP11-20D14.6
hsa-mir-205	SET	GAS5
hsa-mir-205	SFRP2	RP11-20D14.6
hsa-mir-103a2	SH3GL2	RP11-20D14.6
hsa-mir-205	SLC39A14	GAS5
hsa-mir-205	SLC39A14	SNHG5
hsa-mir-93 and -106b	SORL1	RP11-253E3.3
hsa-mir-93	TCF7L1	RP11-69I8.2
hsa-mir-103a2	THY1	RP11-20D14.6
hsa-miR-302a,c,d	ZNF462	RP11-20D14.6
hsa-mir-25 and -302c	ZNF532	RP11-20D14.6
hsa-mir-205	ZSWIM4	RP11-20D14.6

Table 2: mRNAs positively correlates with pseudogenes.

Predicted targets		
miRNA	**mRNA**	**Pseudogene**
hsa-miR-103a2	FGF2	POU5F1P3
hsa-miR-106bb and -302a,c,d	LAMA3	NANOGP1
hsa-miR-25	RIMS2	ARHGEF34P
Validated targets		
miRNA	**mRNA**	**pseudogene**
hsa-miR-302a,c,d	AZIN1	NPM1P6
hsa-miR-103a2	F2RL1	RP11-69L16.5
hsa-miR-103a2	F2RL1	EEF1A1P6
hsa-miR-103a2	FGF2	RP11-69L16.5
hsa-miR-302a,d	FNBP1L	NPM1P27
hsa-miR-103a2	FNBP1L	RP11-69L16.5
hsa-miR-25, -363 and -367	HCN2	HSPD1P1
hsa-miR-302a,c,d	INPP5F	EEF1A1P6
hsa-miR-103a2	KIAA0101	RP11-69L16.5
hsa-miR-103a2	MAP7	EEF1A1P12
hsa-miR-106b	PCDH11X	FABP5P7
hsa-miR-93 and -106b	POU5F1B	HSP90AB3P
hsa-miR-25	SLC22A3	RP11-9L18.2

Supplemental material 5 **(Adapted from the original Excel sheets)**

Table 1: Transcription factors (TFs) with binding sites present at promoters of both mNRA and ncRNA. Highlighted in yellow are the 5 most expressed TFs used in the S parameter estimation. The pairs presented here are those selected as strong putative competition crosstalk at the end of the study. For further studied pairs, consult the original Supplemental material 5 file.

AHCY x RP11-253E3.3	**F2RL1 x EEF1A1P6**	**HSPD1 x SNHG5**	**INO80 x RP11-253E3.3**	**PARP1 x RP11-20D14.6**	**PRIM2 x GAS5**
AR	AR	AR	AR	AR	AR
ATF4	ATF4	BARX1	ATF4	BARX1	ATF4
BARX1	BARX1	CEBPA	BARX1	CEBPA	BARX1
CEBPA	CEBPA	CEBPB	CEBPA	CEBPB	CEBPA
CEBPB	CEBPB	CEBPG	CEBPB	E2F1	CEBPB
CTCF	E2F1	DBP	E2F1	E2F4	E2F1
E2F1	E2F4	E2F1	E2F4	ELK1	E2F4
E2F4	ELK1	E2F4	ELF3	EMX1	ELK1
EGR2	EMX1	EGR2	ELK1	ESRRA	EMX1
ELF3	EOMES	ELK1	EMX1	ETS1	EOMES
ELK1	ESRRA	EMX1	EOMES	ETV4	ERF
EMX1	ETS1	EOMES	ERF	FLI1	ETS1
EOMES	ETV4	ETS1	ETS1	FOXB1	ETV4
ETS1	FLI1	ETV4	ETV4	FOXH1	FLI1
ETV4	FOSL1	ETV5	FLI1	FOXO4	FOSL1
FLI1	FOXB1	FLI1	FOSL1	HESX1	FOXB1
FOSL1	FOXH1	FOSL1	FOXH1	HIC2	FOXH1
FOXH1	FOXO4	FOXB1	FOXO4	INSM1	FOXO4
FOXO4	GMEB2	FOXH1	GMEB2	IRF7	GMEB2
GMEB2	HESX1	FOXO3	HESX1	ISL2	GRHL1
HESX1	HIC2	FOXO4	HIC2	KLF16	HESX1
HIC2	HNF4G	HESX1	HNF4G	KLF5	HIC2
HNF4G	INSM1	HIC2	INSM1	LHX6	HNF4G
INSM1	ISL2	HNF4G	ISL2	LMX1B	INSM1
ISL2	KLF16	IRF7	JDP2	MAFF	ISL2
JDP2	KLF5	ISL2	KLF16	MAFK	JDP2
KLF16	LHX6	KLF16	KLF5	MEIS3	KLF16
KLF5	LMX1B	KLF5	LHX6	MNT	KLF5
LHX6	MAFF	LHX6	LMX1B	NFKB2	LHX6
LMX1B	MAFK	LMX1B	MAFK	NOTO	LMX1B

MAFK	MEIS3	MAFF	MEIS3	OTX2	MAFF
MEIS3	MLXIPL	MAFK	MNT	SP2	MAFK
MLX	MNT	MEIS3	MSC	SP8	MEIS3
MLXIPL	NFKB2	MGA	NFKB2	SPIB	MLXIPL
MNT	NOTO	MLXIPL	NOTO	SREBF1	MNT
MSC	OLIG3	MNT	OLIG3	TCF3	MSC
NFKB2	OTX2	MSC	OTX2	TEAD3	NOTO
NOTO	POU2F2	MYBL1	POU2F2	TEAD4	OLIG3
OLIG3	POU5F1B	NFYA	POU3F4	TFAP2C	OTX2
OTX2	SP2	NOTO	POU5F1B	TFAP4	POU2F2
POU2F2	SP8	OLIG3	SP2	TFE3	POU3F1
POU3F4	SPIB	ONECUT2	SP8	USF1	POU3F4
POU5F1B	SREBF1	OTX2	SPIB	VENTX	POU5F1B
SP2	TCF3	POU2F1	TCF3	ZBTB7A	SP8
SP8	TEAD3	POU2F2	TEAD3	ZBTB7B	SPIB
SPIB	TEAD4	POU3F1	TEAD4	ZNF263	SPIB
SRF	TFAP2C	POU3F4	TFAP2C		SREBF1
TCF3	TFAP4	POU2F2	TFAP4		SRF
TEAD3	TFE3	POU3F1	TFE3		TCF3
TEAD4	USF1	POU3F4	USF1		TEAD3
TFAP2C	VENTX	POU5F1B	VENTX		TEAD4
TFAP4	ZBTB7B	POU6F1	ZBTB7B		TFAP2C
TFE3	ZNF263	RFX5	ZNF263		TFAP4
USF1		SP2			TFE3
VENTX		SP8			TGIF1
ZNF263		SPIB			USF1

Supplemental material 6 **(Adapted from the original Excel sheets)**

Table 1: Relevant gene ontologies used to select putative ceRNA pairs that could play a role in the ESC pluripotency maintenance and proliferation

GOID	Description
GO:0010453	regulation of cell fate commitment
GO:0051302	regulation of cell division
GO:0043410	positive regulation of MAPK cascade
GO:0019827	stem cell population maintenance
GO:0042659	regulation of cell fate specification
GO:0017145	stem cell division
GO:0090068	positive regulation of cell cycle process
GO:0007346	regulation of mitotic cell cycle
GO:0000723	telomere maintenance
GO:0032200	telomere organization
GO:0014066	regulation of phosphatidylinositol 3-kinase signaling
GO:0014068	positive regulation of phosphatidylinositol 3-kinasesignaling
GO:0008543	fibroblast growth factor receptor signaling pathway
GO:0070371	ERK1 and ERK2 cascade
GO:0044344	cellular response to fibroblast growth factor stimulus
GO:0014065	phosphatidylinositol 3-kinase signaling
GO:0071774	response to fibroblast growth factor

Supplemental material 7 **(Adapted from the original Excel sheets)**

Table 1: mRNAs associated to stem cell related ontologies.

GOID: GO:0019827	stem cell population maintenance
ENSG00000101115	SALL4
ENSG00000138685	FGF2
ENSG00000152284	TCF7L1
GOID: GO:0042659	regulation of cell fate specification
ENSG00000138685	FGF2
ENSG00000145423	SFRP2
GOID: GO:0010453	regulation of cell fate commitment
ENSG00000066468	FGFR2
ENSG00000138685	FGF2
ENSG00000145423	SFRP2
GOID: GO:0017145	stem cell division
ENSG00000066468	FGFR2
ENSG00000145423	SFRP2
GOID: GO:0000723	telomere maintenance
ENSG00000143799	PARP1
ENSG00000146143	PRIM2
GOID: GO:0032200	telomere organization
ENSG00000143799	PARP1
ENSG00000146143	PRIM2
GO:0001666	response to hypoxia
ENSG00000101444	AHCY
ENSG00000144381	HSPD1
ENSG00000152455	SUV39H2

Table 1: mRNA associated to cell cycle related ontologies.

GOID: GO:0007346	regulation of mitotic cell cycle
ENSG00000066468	FGFR2
ENSG00000092853	CLSPN
ENSG00000128908	INO80
ENSG00000164109	MAD2L1
GO:0045787	positive regulation of cell cycle
ENSG00000066468	FGFR2
ENSG00000128908	INO80
ENSG00000164109	MAD2L1

ENSG00000181163	NPM1
GO:0090068	positive regulation of cell cycle process
ENSG00000128908	INO80
ENSG00000164109	MAD2L1
ENSG00000181163	NPM1
GO:0051302	regulation of cell division
ENSG00000138685	FGF2
ENSG00000066468	FGFR2
ENSG00000164109	MAD2L1
ENSG00000145423	SFRP2

Table 3: mRNA associated to signaling pathways related ontologies.

GO:0043410	positive regulation of MAPK cascade
ENSG00000065361	ERBB3
ENSG00000164251	F2RL1
ENSG00000138685	FGF2
ENSG00000066468	FGFR2
ENSG00000105976	MET
GO:0070374	positive regulation of ERK1 and ERK2 cascade
ENSG00000164251	F2RL1
ENSG00000138685	FGF2
ENSG00000066468	FGFR2
GO:0070371	ERK1 and ERK2 cascade
ENSG00000164251	F2RL1
ENSG00000138685	FGF2
ENSG00000066468	FGFR2
GO:0008543	fibroblast growth factor receptor signaling pathway
ENSG00000065361	ERBB3
ENSG00000138685	FGF2
ENSG00000066468	FGFR2
GO:0044344	cellular response to fibroblast growth factor stimulus
ENSG00000065361	ERBB3
ENSG00000138685	FGF2
ENSG00000066468	FGFR2
GO:0071774	response to fibroblast growth factor
ENSG00000065361	ERBB3
ENSG00000138685	FGF2

ENSG00000066468	FGFR2
GO:0038127	ERBB signaling pathway
ENSG00000065361	ERBB3
ENSG00000138685	FGF2
ENSG00000066468	FGFR2
ENSG00000107295	SH3GL2
GO:0014068	positive regulation of phosphatidylinositol 3-kinasesignaling
ENSG00000065361	ERBB3
ENSG00000164251	F2RL1
GO:0014066	regulation of phosphatidylinositol 3-kinase signaling
ENSG00000065361	ERBB3
ENSG00000164251	F2RL1
GO:0014065	phosphatidylinositol 3-kinase signaling
ENSG00000065361	ERBB3
ENSG00000164251	F2RL1

Supplemental material 8 **(Adapted from the original Excel sheets)**

Table 1: Binding energy of miRNAs in the mRNA and ncRNAs isoforms.

AHCY x RP11-253E3.3			
miRNA	**Gene**	**ΔG binding**	**TPM**
miR-106b	AHCY-001	-10.2441047434	291.259
miR-106b	AHCY-001	-11.0038044512	291.259
miR-106b	AHCY-001	-14.4638682616	291.259
miR-106b	AHCY-001	-20.9106337881	291.259
miR-106b	AHCY-002	-10.2441047434	218.545
miR-106b	AHCY-002	-11.0038044512	218.545
miR-106b	AHCY-002	-14.4638682616	218.545
miR-106b	AHCY-002	-14.7167000484	218.545
miR-106b	RP11-253E3.3-001	-19.2822340123	232.288
ERBB3 x RP11-69I8.2			
miRNA	**Gene**	**ΔG binding**	**TPM**
miR-106b	ERBB3-001	-5.72107018934	234.437
miR-106b	ERBB3-001	-6.49854477436	234.437
miR-106b	ERBB3-001	-14.5977836985	234.437
miR-106b	ERBB3-001	-14.9946355003	234.437
miR-106b	ERBB3-001	-18.3805313484	234.437
miR-106b	ERBB3-001	-21.1766022291	234.437
miR-106b	ERBB3-004	-5.72107018934	475.976
miR-106b	ERBB3-004	-6.49854477436	475.976
miR-106b	ERBB3-004	-14.5977836985	475.976
miR-106b	ERBB3-004	-14.9946355003	475.976
miR-106b	ERBB3-004	-21.1766022291	475.976
miR-106b	ERBB3-005	-6.49854477436	273.354
miR-106b	ERBB3-005	-9.93382611106	273.354
miR-106b	ERBB3-005	-13.4323350958	273.354
miR-106b	ERBB3-005	-14.9946355003	273.354
miR-106b	ERBB3-005	-21.1766022291	273.354
miR-106b	ERBB3-006	-16.8224603695	349.214
miR-106b	ERBB3-009	-6.49854477436	0.270993
miR-106b	ERBB3-009	-14.9946355003	0.270993
miR-106b	ERBB3-009	-18.3805313484	0.270993
miR-106b	ERBB3-009	-21.1766022291	0.270993
miR-106b	ERBB3-016	-18.3805313484	420.984
miR-106b	ERBB3-018	-6.49854477436	51.625

miR-106b	ERBB3-024	-6.49854477436	646.543
miR-106b	ERBB3-024	-14.9946355003	646.543
miR-106b	ERBB3-024	-18.3805313484	646.543
miR-106b	RP11-69I8.2-001	-11.7163325168	483.825
FGF2 x GAS5			
miRNA	**Gene**	**ΔG binding**	**TPM**
miR-103a2	FGF2-001	-12.3944196025	795.448
miR-103a2	FGF2-001	-12.8918865715	795.448
miR-205	FGF2-001	-14.030627294	795.448
miR-205	FGF2-001	-15.0980635566	795.448
miR-103a2	FGF2-001	-15.4705353869	795.448
miR-103a2	FGF2-001	-16.2583350131	795.448
miR-103a2	FGF2-001	-20.7629450665	795.448
miR-205	FGF2-002	-14.030627294	304.978
miR-103a2	FGF2-002	-15.4705353869	304.978
miR-103a2	FGF2-002	-20.7629450665	304.978
miR-103a2	FGF2-201	-12.3944196025	577.929
miR-103a2	FGF2-201	-12.8918865715	577.929
miR-205	FGF2-201	-14.030627294	577.929
miR-205	FGF2-201	-15.0980635566	577.929
miR-103a2	FGF2-201	-15.4705353869	577.929
miR-103a2	FGF2-201	-16.2583350131	577.929
miR-103a2	FGF2-201	-20.7629450665	577.929
miR-205	GAS5-002	-8.92381694167	448.894
miR-205	GAS5-006	-11.2166729071	537.467
miR-205	GAS5-007	-11.2166729071	101.823
miR-205	GAS5-011	-11.2166729071	101.925
miR-205	GAS5-017	-8.92381694167	0.199703
miR-205	GAS5-018	-8.92381694167	486.852
miR-205	GAS5-021	-8.92381694167	752.765
miR-205	GAS5-022	-8.92381694167	532.745
miR-205	GAS5-023	-8.92381694167	368.193
miR-103a2	POU5F1P3-001	-16.7331726531	132.804
miR-103a2	RP11-69L16.5-001	-15.3888396107	0.266265
FGFR2 x RP11-20D14.6			
miRNA	**Gene**	**ΔG binding**	**TPM**
miR-103a2	FGFR2-001	-12.2151831887	309.672
miR-103a2	FGFR2-001	-15.6214687065	309.672

miR-103a2	FGFR2-001	-16.694950574	309.672
miR-103a2	FGFR2-004	-12.2151831887	243.988
miR-103a2	FGFR2-005	-12.2151831887	0.185035
miR-103a2	FGFR2-005	-15.6214687065	0.185035
miR-103a2	FGFR2-006	-12.2151831887	128.249
miR-103a2	FGFR2-007	-12.2151831887	438.914
miR-103a2	FGFR2-008	-12.537390057	787.668
miR-103a2	FGFR2-008	-15.6214687065	787.668
miR-103a2	FGFR2-008	-16.694950574	787.668
miR-103a2	FGFR2-008	-18.474441035	787.668
miR-103a2	FGFR2-009	-12.2151831887	22.098
miR-103a2	FGFR2-013	-18.8242696625	227.564
miR-103a2	FGFR2-015	-12.2151831887	0.455911
miR-103a2	FGFR2-016	-12.2151831887	302.614
miR-103a2	FGFR2-022	-14.7946258086	149.613
miR-103a2	FGFR2-022	-16.119509286	149.613
miR-103a2	FGFR2-022	-16.4727101319	149.613
miR-103a2	FGFR2-024	-12.2151831887	0.0573969
miR-103a2	FGFR2-024	-15.6214687065	0.0573969
miR-103a2	FGFR2-026	-14.7946258086	571.539
miR-103a2	FGFR2-026	-16.119509286	571.539
miR-103a2	FGFR2-026	-16.4727101319	571.539
miR-103a2	FGFR2-201	-12.2151831887	0.12051
miR-103a2	FGFR2-202	-12.2151831887	417.877
miR-103a2	FGFR2-202	-15.6214687065	417.877
miR-103a2	FGFR2-202	-16.694950574	417.877
miR-103a2	FGFR2-203	-12.2151831887	581.006
miR-103a2	FGFR2-203	-15.6214687065	581.006
miR-103a2	FGFR2-203	-16.694950574	581.006
miR-103a2	FGFR2-204	-12.2151831887	3,08E+00
miR-103a2	FGFR2-204	-15.6214687065	3,08E+00
miR-103a2	FGFR2-204	-16.694950574	3,08E+00
miR-103a2	RP11-20D14.6-001	-16.5061016083	148.532
miR-103a2	RP11-20D14.6-002	-16.5061016083	417.286
miR-103a2	RP11-20D14.6-003	-16.5061016083	83.339

F2RL1 x RP11-69L16.5 and EEF1A1P6

miRNA	**Gene**	**ΔG binding**	**TPM**
miR-103a2	F2RL1-001	-14.5660569221	229.963

miR-103a2	F2RL1-001	-19.574608691	229.963
miR-103a2	RP11-69L16.5-001	-15.3888396107	0.266265
miR-103a2	EEF1A1P6-001	-19.2866512574	137.153
miR-103a2	EEF1A1P6-001	-20.7439292432	137.153
HSPD1 x SNHG5			
miRNA	**Gene**	**ΔG binding**	**TPM**
miR-363	HSPD1-001	-8.02735295901	962.607
miR-367	HSPD1-001	-8.20963476342	962.607
miR-363	HSPD1-001	-9.00790026962	962.607
miR-25	HSPD1-001	-9.69470838183	962.607
miR-25	HSPD1-001	-11.1619145211	962.607
miR-367	HSPD1-001	-12.3142764769	962.607
miR-302a	HSPD1-001	-12.3429701239	962.607
miR-363	HSPD1-001	-12.7791898108	962.607
miR-363	HSPD1-001	-13.7491373298	962.607
miR-367	HSPD1-001	-14.6915542846	962.607
miR-25	HSPD1-001	-15.0138912212	962.607
miR-367	HSPD1-001	-16.6302898312	962.607
miR-302d	HSPD1-001	-17.3038365027	962.607
miR-25	HSPD1-001	-22.3152772437	962.607
miR-302a	HSPD1-003	-10.139772588	453.373
miR-363	HSPD1-003	-12.7791898108	453.373
miR-302d	HSPD1-003	-13.9365769335	453.373
miR-25	HSPD1-003	-15.0138912212	453.373
miR-367	HSPD1-003	-16.6302898312	453.373
miR-302a	HSPD1-004	-10.6279198808	46.883
miR-302a	HSPD1-004	-10.6279198808	46.883
miR-363	HSPD1-004	-11.0234339487	46.883
miR-25	HSPD1-004	-11.3908606103	46.883
miR-302d	HSPD1-004	-15.1181857526	46.883
miR-302d	HSPD1-004	-15.1181857526	46.883
miR-367	HSPD1-004	-15.6719720146	46.883
miR-367	HSPD1-004	-15.6719720146	46.883
miR-363	HSPD1-005	-8.02735295901	103.173
miR-25	HSPD1-005	-9.69470838183	103.173
miR-367	HSPD1-005	-12.3142764769	103.173
miR-363	HSPD1-006	-8.02735295901	98.395
miR-25	HSPD1-006	-9.69470838183	98.395

miR-367	HSPD1-006	-12.3142764769	98.395
miR-363	HSPD1-007	-8.02735295901	128.978
miR-25	HSPD1-007	-9.69470838183	128.978
miR-367	HSPD1-007	-12.3142764769	128.978
miR-363	HSPD1-013	-8.02735295901	379.946
miR-25	HSPD1-013	-9.69470838183	379.946
miR-367	HSPD1-013	-12.3142764769	379.946
miR-363	HSPD1-201	-8.02735295901	883.958
miR-367	HSPD1-201	-8.20963476342	883.958
miR-363	HSPD1-201	-9.00790026962	883.958
miR-25	HSPD1-201	-9.69470838183	883.958
miR-25	HSPD1-201	-11.1619145211	883.958
miR-367	HSPD1-201	-12.3142764769	883.958
miR-302a	HSPD1-201	-12.3429701239	883.958
miR-363	HSPD1-201	-12.7791898108	883.958
miR-363	HSPD1-201	-13.7491373298	883.958
miR-367	HSPD1-201	-14.6915542846	883.958
miR-25	HSPD1-201	-15.0138912212	883.958
miR-367	HSPD1-201	-16.6302898312	883.958
miR-302d	HSPD1-201	-17.3038365027	883.958
miR-25	HSPD1-201	-22.3152772437	883.958
miR-363	SNHG5-001	-14.4607316613	376.568
miR-25	SNHG5-001	-14.8321091674	376.568
miR-367	SNHG5-001	-16.6153133685	376.568
miR-363	SNHG5-003	-14.4607316613	108.915
miR-25	SNHG5-003	-14.8321091674	108.915
miR-367	SNHG5-003	-16.6153133685	108.915
miR-363	SNHG5-004	-14.4607316613	122.067
miR-25	SNHG5-004	-14.8321091674	122.067
miR-367	SNHG5-004	-16.6153133685	122.067
miR-302a	SNHG5-005	-7.64671511184	121.642
miR-302d	SNHG5-005	-7.88575758935	121.642
miR-302a	SNHG5-005	-9.81007350898	121.642
miR-302d	SNHG5-005	-12.1755260663	121.642
miR-363	SNHG5-005	-12.448759984	121.642
miR-25	SNHG5-005	-15.7674060929	121.642
miR-367	SNHG5-005	-15.7726647407	121.642
miR-302a	SNHG5-008	-7.64671511184	207.388

miR-302d	SNHG5-008	-7.88575758935	207.388
miR-302a	SNHG5-008	-9.81007350898	207.388
miR-302d	SNHG5-008	-12.1755260663	207.388
miR-302a	SNHG5-009	-7.64671511184	189.434
miR-302d	SNHG5-009	-7.88575758935	189.434
miR-302a	SNHG5-009	-9.81007350898	189.434
miR-302d	SNHG5-009	-12.1755260663	189.434
miR-363	SNHG5-009	-12.448759984	189.434
miR-25	SNHG5-009	-15.7674060929	189.434
miR-367	SNHG5-009	-15.7726647407	189.434
miR-302a	SNHG5-010	-7.64671511184	254.896
miR-302d	SNHG5-010	-7.88575758935	254.896
miR-302a	SNHG5-010	-9.81007350898	254.896
miR-302d	SNHG5-010	-12.1755260663	254.896
miR-363	SNHG5-010	-14.4607316613	254.896
miR-25	SNHG5-010	-14.8321091674	254.896
miR-367	SNHG5-010	-16.6153133685	254.896
miR-363	SNHG5-011	-14.4607316613	0.38651
miR-25	SNHG5-011	-14.8321091674	0.38651
miR-367	SNHG5-011	-16.6153133685	0.38651
miR-363	SNHG5-013	-14.4607316613	457.202
miR-25	SNHG5-013	-14.8321091674	457.202
miR-367	SNHG5-013	-16.6153133685	457.202
miR-363	SNHG5-014	-14.4607316613	297.849
miR-25	SNHG5-014	-14.8321091674	297.849
miR-367	SNHG5-014	-16.6153133685	297.849
miR-363	SNHG5-015	-14.4607316613	137.701
miR-25	SNHG5-015	-14.8321091674	137.701
miR-367	SNHG5-015	-16.6153133685	137.701
miR-367	SNHG5-019	-10.6915250756	597.173
miR-302d	SNHG5-019	-11.8684933017	597.173
miR-302a	SNHG5-019	-12.2636004249	597.173
miR-25	SNHG5-019	-13.3169931975	597.173
miR-363	SNHG5-019	-13.5969527918	597.173
miR-367	SNHG5-021	-10.6915250756	269.631
miR-302d	SNHG5-021	-11.8684933017	269.631
miR-302a	SNHG5-021	-12.2636004249	269.631
miR-25	SNHG5-021	-13.3169931975	269.631

miR-363	SNHG5-021	-13.5969527918	269.631
miR-363	SNHG5-021	-14.4607316613	269.631
miR-25	SNHG5-021	-14.8321091674	269.631
miR-367	SNHG5-021	-16.6153133685	269.631
miR-367	SNHG5-022	-10.6915250756	0.481713
miR-302d	SNHG5-022	-11.8684933017	0.481713
miR-302a	SNHG5-022	-12.2636004249	0.481713
miR-25	SNHG5-022	-13.3169931975	0.481713
miR-363	SNHG5-022	-13.5969527918	0.481713
INO80 x RP11-253E3.3			
miRNA	**Gene**	**ΔG binding**	**TPM**
miR-93	INO80-001	-15.495178496	105.246
miR-106b	INO80-001	-17.443628019	105.246
miR-93	INO80-002	-15.495178496	0.000410983
miR-106b	INO80-002	-17.443628019	0.000410983
miR-93	INO80-010	-15.495178496	0.747679
miR-106b	INO80-010	-17.443628019	0.747679
miR-93	INO80-201	-15.495178496	916.993
miR-93	INO80-201	-17.443628019	916.993
miR-93	INO80-201	-25.1578926853	916.993
miR-106b	RP11-253E3.3-001	-19.2822340123	232.288
NPM1 x RP11-20D14.6			
miRNA	**Gene**	**ΔG binding**	**TPM**
mir-302a	NPM1-001	-10.986646934	296.943
mir-302c	NPM1-001	-11.3648948	296.943
mir-302d	NPM1-001	-12.1489791258	296.943
mir-302c	NPM1-001	-14.6168023029	296.943
mir-302a	NPM1-001	-15.574530946	296.943
mir-302d	NPM1-001	-17.4736645263	296.943
mir-302c	NPM1-002	-14.6168023029	473.459
mir-302a	NPM1-002	-15.574530946	473.459
mir-302d	NPM1-002	-17.4736645263	473.459
mir-302a	NPM1-003	-10.7926743754	260.887
mir-302c	NPM1-003	-10.8447741395	260.887
mir-302d	NPM1-003	-11.7603566549	260.887
miR-25	NPM1-003	-12.6067226502	260.887
mir-302c	NPM1-003	-14.6168023029	260.887
mir-302a	NPM1-003	-15.574530946	260.887

miR-25	NPM1-003	-16.8061286843	260.887
mir-302d	NPM1-003	-17.4736645263	260.887
mir-302c	NPM1-004	-14.6168023029	2429.43
mir-302a	NPM1-004	-15.574530946	2429.43
mir-302d	NPM1-004	-17.4736645263	2429.43
mir-302c	NPM1-005	-14.6168023029	363.852
mir-302a	NPM1-005	-15.574530946	363.852
mir-302d	NPM1-005	-17.4736645263	363.852
miR-302c	NPM1-008	-14.6168023029	16.984
miR-302a	NPM1-008	-15.574530946	16.984
miR-302d	NPM1-008	-17.4736645263	16.984
miR-302c	NPM1-009	-11.2835634772	206.671
miR-302a	NPM1-009	-14.0846806934	206.671
miR-302d	NPM1-009	-14.8560119101	206.671
miR-25	RP11-20D14.6-001	-6.45714394927	148.532
mir-302c	RP11-20D14.6-001	-10.0751148996	148.532
mir-302d	RP11-20D14.6-001	-10.2030306068	148.532
mir-302a	RP11-20D14.6-001	-10.6511439631	148.532
miR-25	RP11-20D14.6-002	-6.45714394927	417.286
miR-25	RP11-20D14.6-003	-6.45714394927	83.339
miR-25	RP11-20D14.6-003	-6.45714394927	83.339
mir-302c	RP11-20D14.6-003	-10.0751148996	83.339
mir-302d	RP11-20D14.6-003	-10.2030306068	83.339
mir-302a	RP11-20D14.6-003	-10.6511439631	83.339
mir-302a	RP11-20D14.6-003	-13.873880139	83.339
mir-302a	RP11-20D14.6-003	-14.6051267786	83.339
mir-302c	RP11-20D14.6-003	-14.6893145623	83.339
mir-302c	RP11-20D14.6-003	-15.4002129662	83.339
mir-302d	RP11-20D14.6-003	-17.7538593655	83.339
mir-302d	RP11-20D14.6-003	-18.2945132969	83.339
PARP1 x RP11-20D14.6			
miRNA	**Gene**	**ΔG binding**	**TPM**
mir-302c	PARP1-001	-8.36376892029	291.653
mir-302a	PARP1-001	-8.39772667846	291.653
mir-302d	PARP1-001	-9.49765759231	291.653
mir-302c	PARP1-001	-10.3786738304	291.653
mir-302a	PARP1-001	-10.4204422931	291.653
mir-302a	PARP1-001	-11.1003492748	291.653

mir-302d	PARP1-001	-11.360059412	291.653
mir-302a	PARP1-001	-11.3749689501	291.653
mir-302c	PARP1-001	-11.3751601311	291.653
mir-302d	PARP1-001	-11.4369603084	291.653
mir-302d	PARP1-001	-12.0414895119	291.653
mir-302c	PARP1-001	-12.7877430559	291.653
mir-302c	PARP1-001	-12.8492084309	291.653
mir-302a	PARP1-001	-14.414998462	291.653
mir-302d	PARP1-001	-14.4585104909	291.653
miR-103a2	PARP1-001	-16.7797536611	291.653
miR-302a	PARP1-002	-11.749153754	153.451
miR-302c	PARP1-002	-12.5338031181	153.451
miR-302c	PARP1-002	-12.7877430559	153.451
miR-302d	PARP1-002	-12.918326092	153.451
miR-302a	PARP1-002	-14.414998462	153.451
miR-302d	PARP1-002	-14.4585104909	153.451
miR-103a2	PARP1-002	-20.777661923	153.451
miR-103a2	PARP1-002	-22.5565711333	153.451
miR-302a	PARP1-007	-12.7800795932	363.799
miR-302c	PARP1-007	-12.7877430559	363.799
miR-302d	PARP1-007	-14.0231666253	363.799
miR-302c	PARP1-007	-14.0418952506	363.799
miR-302a	PARP1-007	-14.414998462	363.799
miR-302d	PARP1-007	-14.4585104909	363.799
miR-103a2	PARP1-008	-13.4957533693	492.294
mir-302c	RP11-20D14.6-001	-10.0751148996	148.532
mir-302d	RP11-20D14.6-001	-10.2030306068	148.532
mir-302a	RP11-20D14.6-001	-10.6511439631	148.532
miR-103a2	RP11-20D14.6-001	-16.5061016083	148.532
miR-103a2	RP11-20D14.6-002	-16.5061016083	417.286
mir-302c	RP11-20D14.6-003	-10.0751148996	83.339
mir-302d	RP11-20D14.6-003	-10.2030306068	83.339
mir-302a	RP11-20D14.6-003	-10.6511439631	83.339
mir-302a	RP11-20D14.6-003	-13.873880139	83.339
mir-302a	RP11-20D14.6-003	-14.6051267786	83.339
mir-302c	RP11-20D14.6-003	-14.6893145623	83.339
mir-302c	RP11-20D14.6-003	-15.4002129662	83.339
miR-103a2	RP11-20D14.6-003	-16.5061016083	83.339

mir-302d	RP11-20D14.6-003	-17.7538593655	83.339
mir-302d	RP11-20D14.6-003	-18.2945132969	83.339
PRIM2 x GAS5			
miRNA	**Gene**	**ΔG binding**	**TPM**
miR-18b	PRIM2-001	-8.64488355002	441.426
miR-18b	PRIM2-001	-10.6738630379	441.426
miR-18b	PRIM2-001	-11.4192804769	441.426
miR-18b	PRIM2-002	-13.0654368658	637.289
miR-18b	PRIM2-003	-8.64488355002	0.862382
miR-18b	PRIM2-004	-11.4192804769	164.296
miR-18b	PRIM2-005	-11.4192804769	0.241685
miR-18b	PRIM2-007	-10.1594345305	494.225
miR-18b	PRIM2-008	-8.64488355002	52.936
miR-18b	GAS5-001	-22.863007121	178.919
miR-18b	GAS5-002	-11.8815517925	448.894
miR-18b	GAS5-002	-22.863007121	448.894
miR-18b	GAS5-003	-22.863007121	514.209
miR-18b	GAS5-004	-22.863007121	162.967
miR-18b	GAS5-005	-22.863007121	160.841
miR-18b	GAS5-006	-22.863007121	537.467
miR-18b	GAS5-007	-22.863007121	101.823
miR-18b	GAS5-008	-22.863007121	309.777
miR-18b	GAS5-009	-22.863007121	0.276915
miR-18b	GAS5-010	-22.863007121	0.248645
miR-18b	GAS5-011	-22.863007121	101.925
miR-18b	GAS5-012	-22.863007121	624.254
miR-18b	GAS5-013	-22.863007121	0.372827
miR-18b	GAS5-014	-22.863007121	300.312
miR-18b	GAS5-015	-22.863007121	122.078
miR-18b	GAS5-016	-22.863007121	256.267
miR-18b	GAS5-017	-11.8815517925	0.199703
miR-18b	GAS5-017	-22.863007121	0.199703
miR-18b	GAS5-018	-22.863007121	486.852
miR-18b	GAS5-019	-22.863007121	419.861
miR-18b	GAS5-020	-22.863007121	265.986
miR-18b	GAS5-021	-11.8815517925	752.765
miR-18b	GAS5-021	-22.863007121	752.765
miR-18b	GAS5-022	-11.8815517925	532.745

miR-18b	GAS5-022	-22.863007121	532.745
miR-18b	GAS5-023	-11.8815517925	368.193
miR-18b	GAS5-027	-22.863007121	0.990705
miR-18b	GAS5-029	-22.863007121	848.227
SALL4 x RP11-20D14.6			
miRNA	**Gene**	**ΔG binding**	**TPM**
miR-103a2	SALL4-001	-15,38	189.469
miR-103a2	SALL4-001	-16,99	189.469
miR-103a2	SALL4-001	-17,25	189.469
miR-103a2	SALL4-001	-17,26	189.469
miR-103a2	SALL4-001	-17,45	189.469
miR-103a2	SALL4-002	-15,38	498.935
miR-103a2	SALL4-003	-14,43	272.329
miR-103a2	SALL4-003	-15,38	272.329
miR-103a2	SALL4-003	-21,03	272.329
miR-103a2	SALL4-004	-12,73	280.627
miR-103a2	SALL4-004	-15,38	280.627
miR-103a2	SALL4-201	-16,99	128.322
miR-103a2	RP11-20D14.6-001	-16,51	148.532
miR-103a2	RP11-20D14.6-002	-16,51	417.286
miR-103a2	RP11-20D14.6-003	-16,51	83.339
SFRP2 x RP11-20D14.6			
miRNA	**Gene**	**ΔG binding**	**TPM**
miR-205	SFRP2-001	-11.6914148891	741.881
miR-205	SFRP2-001	-14.5777814832	741.881
miR-205	SFRP2-001	-20.3761862454	741.881
miR-205	RP11-20D14.6-001	-10.8963454571	148.532
miR-205	RP11-20D14.6-003	-10.8963454571	83.339
miR-205	RP11-20D14.6-003	-11.4145610525	83.339
miR-205	RP11-20D14.6-003	-17.2280733842	83.339
TCF7L1 x RP11-69I8.2			
miRNA	**Gene**	**ΔG binding**	**TPM**
miR-93	TCF7L1-001	-6.38588688659	863.116
miR-93	TCF7L1-001	-13.1400308878	863.116
miR-93	TCF7L1-001	-20.1260231633	863.116
miR-93	TCF7L1-003	-10.6058932525	109.477
miR-93	RP11-69I8.2-001	-15.5044619143	483.825

Supplemental Material 9

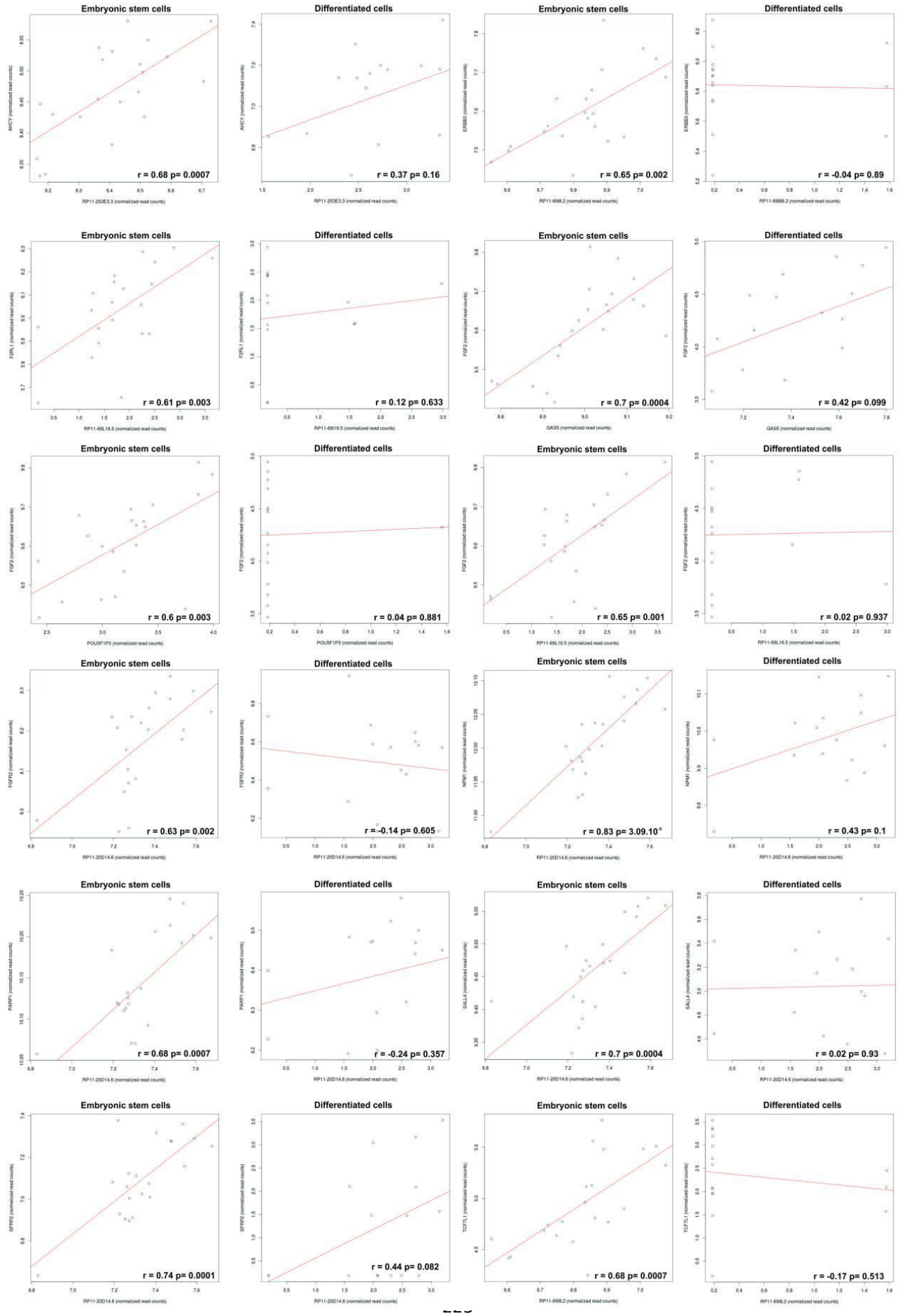


Figure S1: Correlation analysis between RNAs from the chosen ceRNA pairs in ESCs in comparison to differentiated cells.

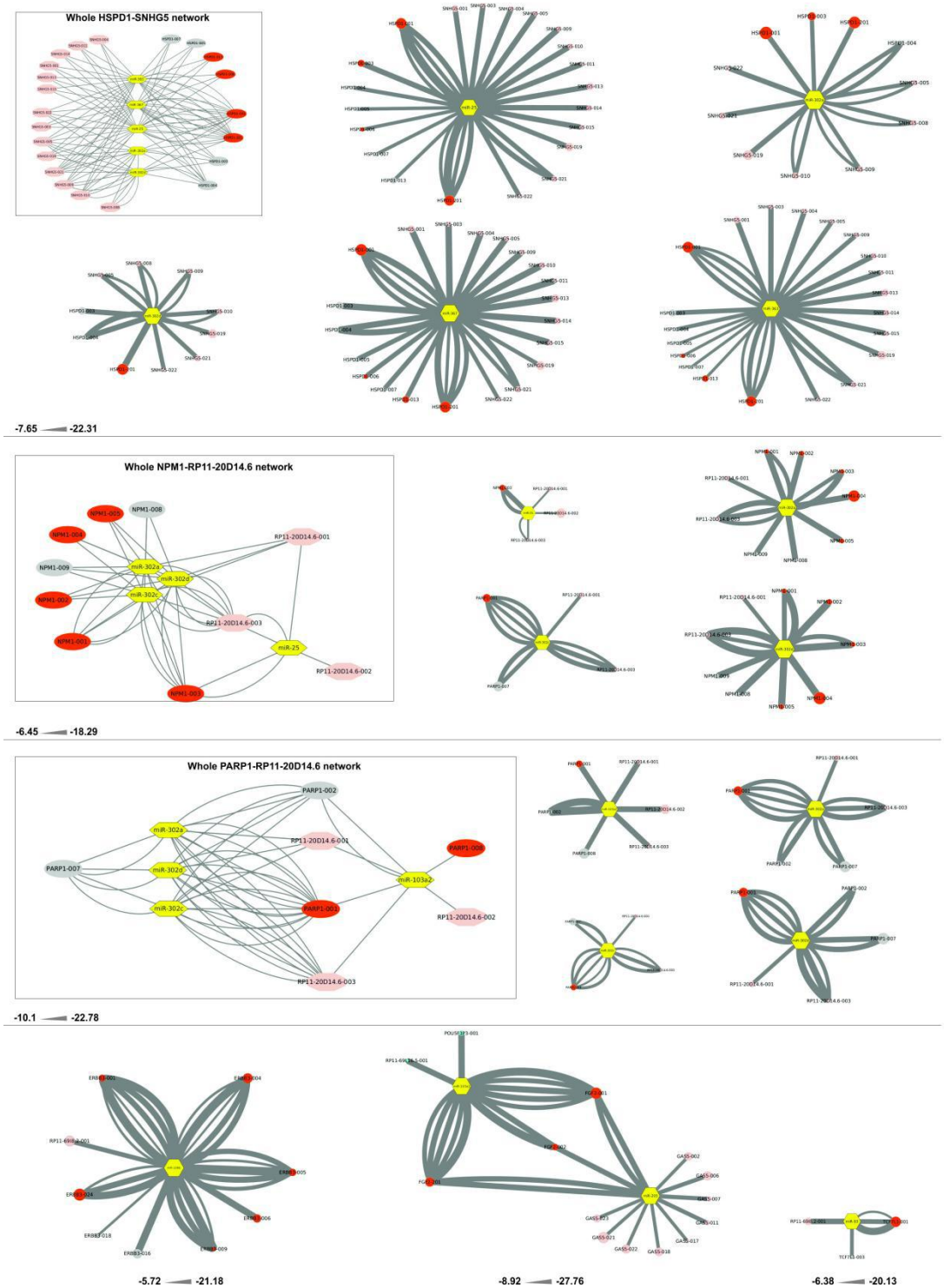


Figure S2: Schematic representation of the interaction between RNAs from the ceRNA pairs and miRNAs. Node sizes are proportional to the isoforms expression in transcripts per million (TPM), except for miRNAs, whose expression is not exibited. Conectors width is proportional to the miRNA binding ΔG values (kJ.mol^{-1}). Red circles: mRNAs. Grey circles: non-coding RNAs isoforms expressed by a coding gene. Pink octagons: lncRNAs. Blue octagons: pseudogenes. Yellow pentagons: miRNAs.

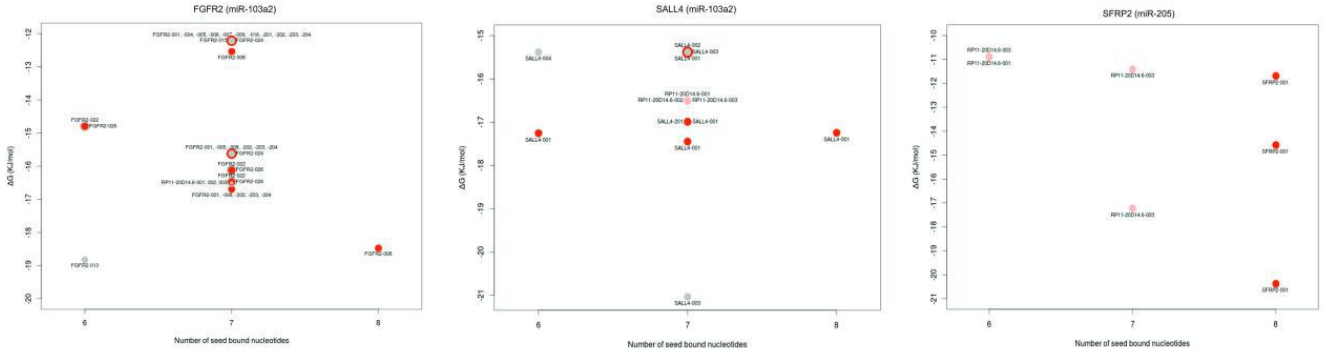


Figure S3: miRNA binding strength combined with seed type for each MRE detected on the isoforms of the genes composing the ceRNA pair. Red circles: mRNAs. Gray circles: non-coding isoforms expressed by a coding gene. Pink circles: lncRNAs. Blue circles: pseudogenes. In the case of overlapping circles, one of them is represented in a shorter diameter. For this case, identifications above or bellow the circle corresponds to the larger one and identifications in each or both sides corresponds to the small circle.

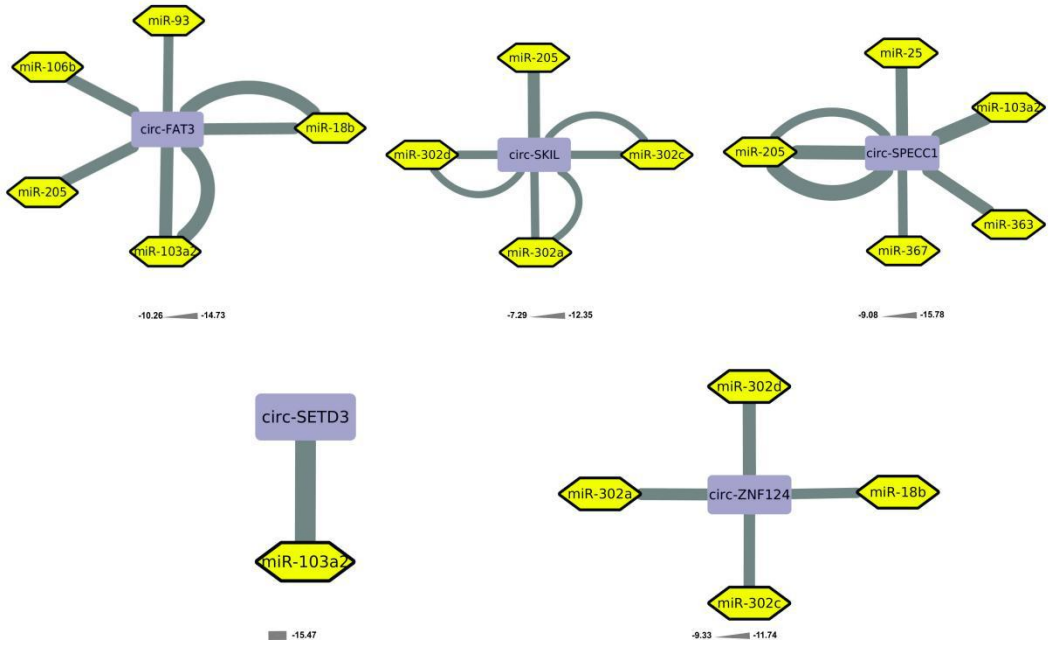


Figure S4: Schematic representation of the interaction among the detected circRNAs the miRNA upregulated in ESCs. The higher and lowest miRNA binding ΔG values ($kJ.mol^{-1}$) are represented bellow each network and the conector width is proportional to the ΔG values.

5. DISCUSSÃO GERAL

5.1. lncRNAs, pseudogenes e circRNAs expressos em CTEs humanas

As células tronco embrionárias humanas foram isoladas pela primeira vez em 1998 e, desde então, o conhecimento molecular sobre estas células evoluiu a ponto de possibilitar a criação artificial de células pluripotentes a partir de células já diferenciadas[58,70]. No que diz respeito especificamente aos RNAs sem potencial codificante, no entanto, somente a classe dos miRNAs apresenta um conjunto bem estabelecido e característico do estado tronco [84,86,88,90]. Outros ncRNAs, por sua vez, permanecem ainda bastante negligenciados nos estudos envolvendo CTEs. Considerando-se apenas os lncRNAs, há mais de três mil loci identificados como transcritos neste tipo celular, porém, somente sete destes RNAs já tiveram sua expressão e função investigadas em CTEs[93]. Além disso, pouco se conhece sobre quais pseudogenes estão expressos e a função que desempenham nessas células e nenhum estudo enfocando a função dos circRNAs no estado tronco foi realizado até o momento.

Desta forma, o presente trabalho é o primeiro a estudar ncRNAs em CTEs incluindo, além dos lncRNAs, os pseudogenes e os RNAs circulares. Ao todo, foram identificados 1182 ncRNAs pertencentes a estas três classes (lncRNA, n=317; pseudogene, n=585; e circRNA, n=280) no conjunto de dados principal. Quando este foi comparado a outros conjuntos de dados, 148 ncRNAS (87 lncRNAs, 51 pseudogenes e 10 circRNA) foram detectados como transcritos comuns. Além disso, muitos destes lncRNAs e pseudogenes encontram-se entre os genes com maior difrerença de expressão entre CTEs e células diferenciadas.

Com relação aos lncRNAs, Tang et al. (2013) detectaram um número similar ao observado no presente estudo (n=300) e apontaram 78 deles como exclusivamente expressos em CTEs[79]. No entanto, os lncRNAs reportados por estes autores incluiam apenas os intergenicamente codificados (lincRNAs), diferentemente do observado aqui, onde, além dos lincRNAs (n=303), foram detectados também lncRNAs classificados como antisenso, senso intrônico, senso sobreposto, transcrito processado e lncRNA de promotor bidirecional. Além disso, somente 36 dos lncRNAs apontados por Tang et al. (2013) como exclusivos de CTEs foram detectados como superexpressos em nosso estudo. Uma possível explicação para as diferenças observadas pode estar no tipo de dados empregado por Tang et al. (2013). Estes autores utilizaram dados de RNA-seq com poucas replicatas biológicas, sendo mais da metade deles com *reads* do tipo *single*. Os nossos dados, por outro lado, apresentam um número alto de replicatas e reads do tipo *paired*, o que confere mais confiabilidade aos resultados gerados.

Além disso, a verificação de que as análises empregadas neste trabalho foram efetivas na deteção de ncRNAs verdadeiramente expressos em CTEs dá-se pela deteção de todos os lncRNAs previamente observados neste tipo de célula, à exceção de FLJ11812 and lncRNA_ES2. Este último, no entanto, possivelmente não foi detectado por não estar oficialmente anotado na plataforma Ensembl, cujas anotações gênicas foram utilizadas para a identificação dos genes nos dados de RNA-seq.

No que diz respeito aos pseudogenes, pela primeira vez tem-se um panorama deste tipo de RNA em CTEs. Grande parte dos pseudogenes observados como superexpressos apresenta-se pelo menos oito vezes mais transcrito em CTEs do que em células diferenciadas, o que pode indicar uma possivel função associada ao estado tronco. Dentre os pseudogenes com maior diferença de

expressão estão ECEL1P1, HLA-DPB2 e PAICSP7, nenhum deles previamente estudado em detalhe. Além disso, foram detectados também pseudogenes dos genes codificantes dos fatores de transcrição Oct-4 (POU5F1), POU5F1P3 e POU5F1P4, e de Nanog, NANOGP1 e NANOGP8, estes dois últimos previamente reportados como não expressos em CTEs[94,95]. No entanto, a detecção destes dois pseudogenes em pelo menos quatro conjuntos diferentes de dados reforça a hipótese de que, ao contrário do observado nos estudos prévios, eles sejam transcritos nesse tipo celular.

Os circRNAs expressos em CTEs, por sua vez, já foram alvo de um estudo anterior que identificou pelo menos 1061 dessas moléculas na linhagem embrionária H9[74]. Este estudo, no entanto, não investiga a possível função desempenhada por estes RNAs no estado tronco, analisando apenas elementos de sequência que favorecem a circularização de RNAs durante o splicing[74]. Pelo menos 256 genes detectados como geradores de circRNAs por esses autores também tiveram transcritos circulares observados no presente trabalho[74]. No entanto, uma vez que nem todos os circRNAs detectados foram observados em todas as replicatas, apenas dez foram selecionados para avaliações posteriores. Destes, 9 foram identificados nas análises de Zhang et al. (2014)[74]. Essas informações indicam que o método de deteção de circRNAs foi eficiente, embora não tenha identificado o mesmo número de RNAs circulares observado no estudo prévio. Além disso, pode-se considerar que os circRNAs utilizados nas análises do potencial de competição são verdadeiramente expressos em CTEs, uma vez que as mesmas moléculas foram detectadas por dois métodos diferentes.

Considerando-se os resultados do desenvolvimento deste trabalho, pode-se propor os 138 lncRNAs e pseudogenes detectados como superexpressos em CTEs em quatro diferentes conjuntos de dados como característicos do estado tronco. Esta lista, no entanto, pode ser maior uma vez que

alguns ncRNAs podem ter deixado de ser detectados em mais de um conjunto de dados pois alguns deles apresentaram um número total de genes superexpressos bem menor do que o *dataset* principal. Por outro lado, os dez circRNAs selecionados para análises funcionais ainda não podem ser propostos como característicos de CTEs pois a sua expressão diferencial nessas células não pode ser confirmada devido a ausência de ferramentas de bioinformática confiáveis para tal análise. Pode-se concluir apenas que estes RNA estão expressos em CTEs.

O presente estudo amplia consideravelmente o conhecimento no que diz repeito a ncRNAs em CTEs. O passo seguinte consiste na confirmação da expressão destas moléculas por PCR em tempo real, o que permitirá tanto propor o conjunto de circRNAs com viés de expressão em CTEs quanto verificar se os 138 ncRNAs propostos como característicos do estado tronco embrionário realmente se enquadram nesta designação. Além disso, a expansão dos estudos no sentido de investigar outros tipos de RNAs não codificantes, como piRNAs, parRNAs e eRNAs também é um caminho interessante.

5.2. Potenciais papeis desempenhados por lncRNAs, pseudogenes e circRNAs expressos em CTEs humanas

RNAs circulares, lncRNAs ou aqueles oriundos da transcrição de pseudogenes foram considerados por muito tempo como ruído transcricional e inicialmente negligenciados nas pesquisas biológicas. Esse quadro tem se alterado com as contínuas descobertas atribuindo funções

celulares a essas moléculas, o que se reflete no número de publicações abordando esta temática. No caso dos lncRNAs, o número de artigos publicados subiu de 240 em 2011 para quase dois mil em 2016, enquanto que mais de 500 artigos foram publicados sobre circRNAs neste mesmo ano.

Dentro deste contexto, ncRNAs têm sido associados a uma série de processos celulares tais como apoptose, *splicing*, reparo de DNA, organização de cromatina, ciclo celular, manutenção do estado tronco, diferenciação e reprogramação celular bem como relacionados a processos tumorais[53,54,81,96–101]. Esses RNAs podem atuar, entre outros mecanismos, como competidores endógenos; repressores transcricionais; guias no direcionando proteínas ao seu sítio de atuação; "iscas" que retem proteínas de ligação ácidos nucleicos, impedindo sua função; plataformas para a montagem de complexos proteicos e como fontes de transcritos antisenso[29,102–106].

Apesar de haver estudos prévios em larga escala enfocando a detecção de lncRNAs e circRNAs em CTEs, a elucidação da função dessas moléculas nestas células, bem como o papel desempenhado pelos pseudogenes, tem recebido pouca atenção[74,79]. Nessse sentido, o presente estudo investigou, através da abordagem de *guilty by association*, a função de 15 dos ncRNAs propostos no item anterior desta discusssão como característicos do estado tronco. Além disso, devido a presença de relatos prévios de lncRNAs atuando como competidores endógenos em CTEs, também buscou-se predizer relações de competição nesse contexto celular.

A análise de *guilty by association* permitiu associar os RNAs estudados a pelo menos 11 processos celulares. A observação de correlação positiva entre pseudogenes ou lncRNAs e mRNAs traduzidos em proteínas envolvidas na regulação da diferenciação celular, no ciclo celular, no processo de mitose, no reparo de DNA e na apoptose aponta que alguns destes ncRNAs podem

estar envolvidos em processos-chave nas CTEs, uma vez que essas células proliferam ativamente e, em comparação a células diferenciadas, apresentam uma maior capacidade de reparo e, quando o reparo não é possível, são mais propensas a desencadear processos apoptóticos[61,107–109].

Por outro lado, outros ncRNAs parecem estar associados a processos celulares básicos importantes, mas que não estão diretamente ligados ao estado tronco, como biogênese de ribossomos, splicing, regulação do citoesqueleto e organização da cromatina.

Inferências específicas sobre como esses ncRNAs atuam nestes processos são dificultadas pois grande parte dos lncRNAs e pseudogenes analisados não possui qualquer outro estudo prévio de sua função em células humanas. No entanto, no que diz respeito aos ncRNAs LINC00458, POU5F1P3, CTD-2501M5.1 e LNCPRESS1, a atuação em CTEs pode se dar via associação física a proteínas, uma vez que estes RNAs apresentaram um alto potencial de ligação a fatores envolvidos no processo de *splicing*.

Por outro lado, recentemente propô-se que lncRNAs e pseudogenes podem regular a expressão de mRNAs competindo pelos miRNAs que ligam-se a essas moléculas. Desta forma as análises de predição de função do presente trabalho voltaram-se para RNAs que eventualmente poderiam competir pela ligação dos miRNAs superexpressos no estado tronco. Nesse sentido, enquanto as relações de competição propostas anteriormente estão centradas em miRNAs envolvidos com a diferenciação e que estão pouco expressos em CTEs, as relações de competição enfocadas neste trabalho atuariam protegendo moléculas importantes no estado tronco da ação de miRNAs expressos nesse mesmo contexto celular[54,82,110].

Considerando-se que a competição entre RNAs seja, como tem sido hipotetizado, um novo nível de regulação de expressão, todos RNAs que compartilham sítios para os mesmos miRNAs em um dado tipo celular potencialmente competem pela ligação desses pequenos RNAs. Nesse sentido, as relações de competição na célula formariam uma rede entre RNAs. De fato, trabalhos prévios tem demonstrado que há um grande número de RNAs que compartilham sítios de miRNAs dentro de uma mesma célula. Nitzan et al. (2014), analizaram os mRNAs alvos de 379 miRNAs na linhagem HEK293 e geraram uma rede com 6.299 nós, cujas conexões, que representavam miRNAs em comum, ultrapassaram o valor de um milhão[111]. Paci et al. (2014), por sua vez, observaram 1738 nós correspondendo a lncRNAs e mRNAs e apresentando mais de 32 mil conexões[112].

Nesse sentido, uma rede de competição possivelmente existe também em CTEs. De fato, o levantamento de alvos preditos e validados dos 11 miRNAs investigados neste estudo revelou 753 RNAs, entre mRNAs, lncRNAs e pseudogenes, que, a julgar pelos sítios compartilhados de miRNAs, estabeleceriam mais de 248.067 conexões. A análise de ontologia das proteínas codificadas pelos mRNAs revelou alvos associados a processos como ciclo celular, manutenção do estado tronco, regulação negativa da diferenciação, manutenção de telômeros e vias de sinalização características do estado tronco.

Sendo assim, pela primeira vez se reporta uma rede de potenciais competidores em CTEs, a qual aparentemente está associada a processos essenciais a este tipo celular. A rede, no entanto, pode possuir mais elementos do que os apresentados neste trabalho, uma vez que foram inclusos todos os RNAs-alvo validados e mRNAs-alvo preditos, mas a investigação de pseudogenes e lncRNAs-alvo preditos limitou-se ao 50 mais expressos.

Por outro lado, a proposição desta rede de competição é teórica e necessita de validação experimental. Desta forma, propõe-se os pares de RNAs AHCY-RP11-253E3.3, F2RL1-EEF1A1P6, HSPD1-SNHG5, INO80-RP11-20D14.6, PARP1-RP11-20D14.6 e PRIM2-GAS5 como alvos inicias de validação pelo fato de reunirem características que os tornam interessantes no contexto tronco. Primeiramente, os mRNAs das duplas propostas codificam proteínas associadas a processos importantes em CTEs, como resposta a hipóxia, regulação do ciclo celular e manutenção de telômeros. Além disso, os ncRNAs que integram as duplas ligam-se aos miRNAs compartilhados com valores de ΔG iguais ou maiores que os mRNAs e apresentam *seeds* com preferencia igual ou superior na hierarquia de ligação dos miRNAs. Isso leva a crer que os ncRNAs poderiam estar competindo com estes mRNAs de modo a manter os seus níveis ideais em CTEs e, por isso, contribuindo para a manutenção do estado tronco.

No entanto, é necessário considerar que outros RNAs possivelmente competem com os mRNAs das duplas propostas. Isso inclui outros mRNAs e alguns dos circRNAs detectados e que contem sítios para os miRNAs selecionados, como circ-HIPK3, circ-FAT1 e circ-TNFRSF21, que apresentaram maior potencial de atuação como competidores. Além disso, os lncRNAs e pseudogenes também podem competir com outros RNAs da rede apresentada. O lncRNA RP11-20D14.6, por exemplo, compartilha sítios de miRNAs com pelo menos outros 24 mRNAs. As duplas propostas, no entanto servem como um bom ponto de partida para iniciar a elucidação da rede de competição como um todo.

6. CONCLUSÕES GERAIS

Considerando-se os resultados obtidos e observações realizadas durante o desenvolvimento deste trabalho, conclui-se que:

- obteve-se sucesso em detectar miRNAs previamente reportados como característicos do estado tronco;

- um grande número de pseudogenes e lncRNAs apresenta-se superexpresso no estado tronco embrionário humano quando comparado a células diferenciadas;

- um grande número de circRNAs também foi detectado neste tipo celular. No entanto, conclusões sobre quais deles estão superexpressos em CTEs só poderão ser emitidas após análises por PCR em tempo real;

- pelo menos 137 ncRNAs detectados neste estudo são potencialmente característicos do estado tronco embrionário humano, uma vez que apresentam-se superexpressos em CTEs quando estas são comparadas a pelo menos quatro diferentes tipos de células diferenciadas;

- o subgrupo de ncRNAs cuja função foi predita utlizando-se a abordagem de *guilty by association* mostrou-se associado a processsos biológicos característicos do estado tronco embrionário, como os envolvidos na mitose, ciclo celular, reparo de DNA e resposta a dano, apoptose e regulação da diferenciação celular, ou processos celulares mais básicos, como biogênese de ribossomos, *splicing*, organização de cromatina e organização do citoesqueleto ;

- pelo menos quatros destes ncRNAs potencialmente se ligam a proteínas associadas ao mecanismo de *splicing*;

- ncRNAs e mRNAs superexpressos em CTEs humanas apresentam sítios de ligação para os miRNAs característicos desse tipo celular, podendo, por isso apresentar comportamento de competidores endógenos;

- a análise de ontologia da rede total de potenciais ceRNAs revelou a presença de processos relacionados ao ciclo celular, manutenção da população de células tronco, regulação negativa da diferenciação, manutenção de telômeros e vias de sinalização previamente apontadas como importantes na manutenção do estado tronco. Esses fatos indicam que os mecanismos de competição que potencialmente ocorrem em CTEs podem estar associados à manutenção do estado indifenciado destas células;

- os RNAs que potencialmente competem entre si pela ligação dos miRNAs identificados neste trabalho apresentam correlação positiva e descarta-se a ocorrência desta observação por compartilhamento de sequências regulatórias em *cis* ou devido à regulação pelos mesmos fatores de transcrição;

- análises em nível de isoformas da força de ligação do miRNA e do tipo *seed* presente no RNA alvo apontou as duplas de mRNAs e ncRNAs AHCY-RP11-253E3.3, F2RL1-EEF1A1P6, HSPD1-SNHG5, INO80-RP11-20D14.6, PARP1-RP11-20D14.6 como as de maior potencial de competição entre seus membros;

- os lncRNAs e pseudogenes identificados neste trabalho como potenciais ceRNAs podem possivelmente ter a sua atuação como competidores associada a manutenção do estado tronco embrionário em CTEs humanas;

- dentre os 10 circRNAs estudados, os circRNAs circ-FAT1, -HIPK3 e -TNFRSF21 apresentam grande potencial de atuação como competidores em CTEs.

7. PERSPECTIVAS

A conclusão deste trabalho deixa como perspectivas:

- confirmação da expressão dos pseudogenes, lncRNAs e circRNAs detectados neste trabalho por PCR em tempo real;

- identificação, também por PCR em tempo real, dos circRNAs superexpressos em CTEs quando comparadas a células diferenciadas;

- confirmação experimental de que a lista de 137 nRNAs é de fato característica do estado tronco embrionário humano;

- realização de ensaios de superexpressão e silenciamento para investigar as possíveis funções desempenhadas pelos circRNAs detectados neste estudo;

- testar experimentalmente a ocorrência de competição entre as duplas de mRNAs e ncRNAs, bem como circRNAs apontados como tendo grande potencial de atuação como ceRNAs;

- realização de ensaios de imunoprecipitação a fim de investigar e confirmar as possíveis interações de pseudogenes e lncRNAs com proteínas expressas em CTEs humanas;

- avaliar por meio da abordagem de *guilty by association*, os possíveis processos celulares associados aos demais ncRNAs detectados neste estudo;

- expandir os estudos sobre RNAs que atuem como competidores em CTEs de modo a gerar uma possível rede de competição presente neste tipo celular;

- avaliar o potencial de atuação como competidoras endógenas das isoformas não codificantes transcritas a partir do genes codificantes de proteínas em CTEs;

- identificar os elementos constituientes da rede de competição entre RNAs a nível de isoformas em CTES;

- estudar a expressão e a função de outros RNAs não codificantes, como eRNAs, paRNAs e piRNAs.

8. REFERÊNCIAS BIBLIOGRÁFICAS

1. Mattick JS. Non-coding RNAs: The architects of eukaryotic complexity. *EMBO Rep*. 2001;2(11):986-991.

2. Nie L, Wu HJ, Hsu JM, et al. Long non-coding RNAs: Versatile master regulators of gene expression and crucial players in cancer. *Am J Transl Res*. 2012;4(2):127-150.

3. Lee Y, Jeon K, Lee JT, Kim S, Kim VN. MicroRNA maturation: Stepwise processing and subcellular localization. *EMBO J*. 2002;21(17):4663-4670.

4. Miranda KC, Huynh T, Tay Y, et al. A Pattern-Based Method for the Identification of MicroRNA Binding Sites and Their Corresponding Heteroduplexes. *Cell*. 2006;126(6):1203-1217.

5. Mogilyansky E, Rigoutsos I. The miR-17/92 cluster: a comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease. *Cell Death Differ*. 2013;20(12):1603-1614.

6. Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A. Identification of mammalian microRNA host genes and transcription units. *Genome Res*. 2004;14(10 A):1902-1910.

7. Chien CH, Sun YM, Chang WC, et al. Identifying transcriptional start sites of human microRNAs based on high-throughput sequencing data. *Nucleic Acids Res*. 2011;39(21):9345-9356.

8. Lee Y, Kim M, Han J, et al. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J*. 2004;23(20):4051-4060.

9. Lee Y, Ahn C, Han J, et al. The nuclear RNase III Drosha initiates microRNA processing. *Nature*. 2003;425(6956):415-419

10. Lund E, Güttinger S, Calado A, Dahlberg JE, Kutay U. Nuclear export of microRNA precursors. *Science (80-)*. 2004;303(5654):95-98.

11. Maniataki E, Mourelatos Z. A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. *Genes Dev*. 2005;19(24):2979-2990.

12. Hutvagner G, McLachlan J, Pasquinelli AE, lint E, Tuschl T, Zamore PD. A Cellular Function for the RNA-Interference Temporal RNA Small let-7 Enzyme Dicer in the Maturation of the . *Science (80-)*. 2010;293(5531):1-6.

13. Khvorova A, Reynolds A, Jayasena SD. Functional siRNAs and miRNAs exhibit strand bias. *Cell*. 2003;115(2):209-216.

14. Eulalio A, Huntzinger E, Izaurralde E. Getting to the Root of miRNA-Mediated Gene Silencing. *Cell*. 2008;132(1):9-14.

15. Hansen TB, Jensen TI, Clausen BH, et al. Natural RNA circles function as efficient microRNA sponges. *Nature*. 2013;495(7441):384-388.

16. Leucci E, Patella F, Waage J, et al. microRNA-9 targets the long non-coding RNA MALAT1 for degradation in the nucleus. *Sci Rep*. 2013;3(October 2015):2535.

17. Chen X, Zhu H, Wu X, Xie X, Huang G. Downregulated pseudogene CTNNAP1 promote tumor growth in human cancer by downregulating its cognate gene CTNNA1 expression. *Oncotarget*. 2016;7(34).

18. Hashimoto Y, Akiyama Y, Yuasa Y. Multiple-to-Multiple Relationships between MicroRNAs and Target Genes in Gastric Cancer. *PLoS One*. 2013;8(5).

19. Li J, Qu W, Jiang Y, et al. miR-489 Suppresses Proliferation and Invasion of Human Bladder Cancer Cells. 2017;24(256):391-398.

20. Ma L, Teruya-Feldstein J, Weinberg R a. Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature*. 2007;449(7163):682-688.

21. Absalon S, Kochanek DM, Raghavan V, Krichevsky AM. MiR-26b, upregulated in Alzheimer's disease, activates cell cycle entry, tau-phosphorylation, and apoptosis in postmitotic neurons. *J Neurosci*. 2013;33(37):14645-14659.

22. Zhu X, Wang H, Liu F, et al. Identification of micro-RNA networks in end-stage heart failure because of dilated cardiomyopathy. *J Cell Mol Med*. 2013;17(9):1173-1187. doi:10.1111/jcmm.12096.

23. Dai Y, Huang Y-S, Tang M, et al. Microarray analysis of microRNA expression in peripheral blood cells of systemic lupus erythematosus patients. *Lupus*. 2007;16(12):939-946.

24. Stanczyk J, Leslie Pedrioli DM, Brentano F, et al. Altered expression of microRNA in synovial fibroblasts and synovial tissue in rheumatoid arthritis. *Arthritis Rheum*. 2008;58(4):1001-1009.

25. Ran X, Xiao C-H, Xiang G, Ran X-Z. Regulation of Embryonic Stem Cell Self-Renewal and Differentiation by MicroRNAs. *Cell Reprogram*. 2017;19(0):cell.2016.0048.

26. Kang H, Hata A. The role of microRNAs in cell fate determination of mesenchymal stem cells: Balancing adipogenesis and osteogenesis. *BMB Rep*. 2015;48(6):319-323.

27. Liao B, Bao X, Liu L, et al. MicroRNA cluster 302-367 enhances somatic cell reprogramming by accelerating a mesenchymal-to-epithelial transition. *J Biol Chem*. 2011;286(19):17359-17364.

28. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: The rosetta stone of a hidden RNA language? *Cell*. 2011;146(3):353-358.

29. Wang KC, Chang HY. Molecular Mechanisms of Long Noncoding RNAs. *Mol Cell*. 2011;43(6):904-914.

30. Consortium EP, Dunham I, Kundaje A, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74.

31. Derrien T, Johnson R, Bussotti G, et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res*. 2012;22(9):1775-1789. doi:10.1101/gr.132159.111.

32. Guttman M, Amit I, Garber M, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*. 2009;458(7235):223-227. doi:10.1038/nature07672.

33. Jacq C, Miller JR, Brownlee GG. A pseudogene structure in 5S DNA of Xenopus laevis. *Cell*. 1977;12(1):109-120. doi:10.1016/0092-8674(77)90189-1.

34. Podlaha O, Zhang J. Pseudogenes and Their Evolution. *Encycl Life Sci*. 2010;(November):1-8.

35. Li W, Yang W, Wang XJ. Pseudogenes: Pseudo or Real Functional Elements? *J Genet Genomics*. 2013;40(4):171-177.

36. Rouchka EC, Cha IE. Current Trends in Pseudogene Detection and Characterization. *Curr Bioinform*. 2009;4:112-119.

37. Zhang ZD, Cayting P, Weinstock G, Gerstein M. Analysis of nuclear receptor pseudogenes in vertebrates: How the silent tell their stories. *Mol Biol Evol*. 2008;25(1):131-143.

38. Kandouz M, Bier A, Carystinos GD, Alaoui-Jamali M a, Batist G. Connexin43 pseudogene is expressed in tumor cells and inhibits growth. *Oncogene*. 2004;23(27):4763-4770.

39. Harrison PM, Zheng D, Zhang Z, Carriero N, Gerstein M. Transcribed processed pseudogenes in the human genome: An intermediate form of expressed retrosequence lacking protein-coding ability. *Nucleic Acids Res*. 2005;33(8):2374-2383.

40. Wen Y-Z, Zheng L-L, Qu L-H, Ayala FJ, Lun Z-R. Pseudogenes are not pseudo any more. *RNA Biol*. 2012;9(1):25-30.

41. Zheng D, Gerstein MB. The ambiguous boundary between genes and pseudogenes: the dead rise up, or do they? *Trends Genet*. 2007;23(5):219-224.

42. MING-TA HSU & MIGUEL COCA-PRADOS. Electron microscopic evidence for the circular form of RNA in the cytoplasm of eukaryotic cells. 1979:339-340.

43. Cocquerelle C, Mascrez B, Hétuin D, Bailleul B. Mis-splicing yields circular RNA molecules. *FASEB J*. 1993;7(1):155-160.

44. Memczak S, Jens M, Elefsinioti A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. TL - 495. *Nature*. 2013;495 VN-(7441):333-338.

45. Salzman J. Circular RNA is expressed across the eukaryotic tree of life. *PLoS One*. 2014;9(6):e90859.

46. Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One*. 2012;7(2).

47. Cesana M, Cacchiarelli D, Legnini I, et al. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell*. 2011;147(2):358-369. .

48. Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature*. 2010;465(7301):1033-1038.

49. Karreth FA, Reschke M, Ruocco A, et al. The BRAF pseudogene functions as a competitive endogenous RNA and induces lymphoma in vivo. *Cell*. 2015;161(2):319-332.

50. Zhou B, Wei FY, Kanai N, Fujimura A, Kaitsuka T, Tomizawa zuhito. Identification of a splicing variant that regulates type 2 diabetes risk factor CDKAL1 level by a coding-independent mechanism in human. *Hum Mol Genet*. 2014;23(17):4639-4650.

51. Yu F, Zheng J, Mao Y, et al. Long non-coding RNA growth arrest-specific transcript 5 (GAS5) inhibits liver fibrogenesis through a mechanism of competing endogenous RNA. *J Biol Chem*. 2015;290(47):28286-28298.

52. Liu X, Hou L, Huang W, Gao Y, Lv X, Tang J. The Mechanism of Long Non-coding RNA MEG3 for Neurons Apoptosis Caused by Hypoxia: Mediated by miR-181b-12/15-LOX Signaling Pathway. *Front Cell Neurosci*. 2016;10(September):201.

53. Liang W-C, Fu W-M, Wang Y-B, et al. H19 activates Wnt signaling and promotes osteoblast differentiation by functioning as a competing endogenous RNA. *Sci Rep*. 2016;6(February):20121. doi:10.1038/srep20121.

54. Wang Y, Xu Z, Jiang J, et al. Endogenous miRNA Sponge lincRNA-RoR Regulates Oct4, Nanog, and Sox2 in Human Embryonic Stem Cell Self-Renewal. *Dev Cell*. 2013;25(1):69-80.

55. Watt FM, Hogan BL. Out of Eden: stem cells and their niches. *Science*. 2000;287(5457):1427-1430.

56. Gimble J, Guilak F. Adipose-derived adult stem cells: isolation, characterization, and differentiation potential. *Cytotherapy*. 2003;5(5):362-369.

57. Campagnoli C, Roberts IAG, Kumar S, Bennett PR, Bellantuono I, Fisk NM. Identification of mesenchymal stem/progenitor cells in human first-trimester fetal blood, liver, and bone marrow. *Blood*. 2001;98(8):2396-2402.

58. Thomson JA, Itskovitz-Eldor J, Shapiro SS, et al. Embryonic stem cell lines derived from human blastocysts. *Science*. 1998;282(5391):1145-1147.

59. Verfaillie CM, Pera MF, Lansdorp PM. Stem cells: hype and reality. *Hematology Am Soc Hematol Educ Program*. 2002:369-391.

60. Bongso A, Richards M. History and perspective of stem cell research. *Best Pract Res Clin Obstet Gynaecol*. 2004;18(6):827-842.

61. Amit M, Carpenter MK, Inokuma MS, et al. Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. *Dev Biol*. 2000;227(2):271-278.

62. International Stem Cell I, Adewumi O, Aflatoonian B, et al. Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. *Nat Biotechnol*. 2007;25(7):803-816.

63. Schuldiner M, Eiges R, Eden a, et al. Induced neuronal differentiation of human embryonic stem cells. *Brain Res*. 2001;913(2):201-205.

64. Mummery C. Differentiation of Human Embryonic Stem Cells to Cardiomyocytes: Role of Coculture With Visceral Endoderm-Like Cells. *Circulation*. 2003;107(>21):2733-2740. doi:10.1161/01.CIR.0000068356.38592.68.

65. Rambhatla L, Chiu CP, Kundu P, Peng Y, Carpenter MK. Generation of hepatocyte-like cells from human embryonic stem cells. *Cell Transplant*. 2003;12(1):1-11. doi:10.3727/000000003783985179.

66. Ledran MH, Krassowska A, Armstrong L, et al. Efficient hematopoietic differentiation of human embryonic stem cells on stromal cells derived from hematopoietic niches. *Cell Stem Cell*. 2008;3(1):85-98.

67. Zhang D, Jiang W, Liu M, et al. Highly efficient differentiation of human ES cells and iPS cells into mature pancreatic insulin-producing cells. *Cell Res*. 2009;19(4):429-438.

68. Zhou S, Flamier A, Abdouh M, et al. Differentiation of human embryonic stem cells into cone photoreceptors through simultaneous inhibition of BMP, TGFβ and Wnt signaling. *Development*. 2015;142(19):3294-3306.

69. Kimbrel E a., Lanza R. Current status of pluripotent stem cells: moving the first therapies to the clinic. *Nat Rev Drug Discov*. 2015;14(September):681-692.

70. Takahashi K, Tanabe K, Ohnuki M, et al. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell*. 2007;107(5):861-872. doi:10.1016/j.cell.2007.11.019.

71. Hart AH, Hartley L, Ibrahim M, Robb L. Identification, Cloning and Expression Analysis of the Pluripotency Promoting Nanog Genes in Mouse and Human. *Dev Dyn*. 2004;230(1):187-198.

72. Poursani EM, Soltani BM, Mowla SJ. Differential expression of OCT4 pseudogenes in pluripotent and tumor cell lines. *Cell J*. 2016;18(1):28-36.

73. Dang Y, Yan L, Hu B, et al. Tracing the expression of circular RNAs in human pre-implantation embryos. *Genome Biol*. 2016;17(1):130. doi:10.1186/s13059-016-0991-3.

74. Zhang XO, Wang H Bin, Zhang Y, Lu X, Chen LL, Yang L. Complementary sequence-mediated exon circularization. *Cell*. 2014;159(1):134-147. doi:10.1016/j.cell.2014.09.001.

75. Fort A, Yamada D, Hashimoto K, Koseki H, Carninci P. Nuclear transcriptome profiling of induced pluripotent stem cells and embryonic stem cells identify non-coding loci resistant to reprogramming. *Cell Cycle*. 2015;14(8):1148-1155. doi:10.4161/15384101.2014.988031.

76. Dinger ME, Amaral PP, Mercer TR, et al. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res*. 2008;18(9):1433-1445.

77. Tan JY, Sirey T, Honti F, et al. Extensive microRNA-mediated crosstalk between lncRNAs and mRNAs in mouse embryonic stem cells. *Genome Res*. 2015;125(5):655-666.

78. Guttman M, Donaghey J, Carey BW, et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*. 2011:1-11.

79. Tang X, Hou M, Ding Y, Li Z, Ren L, Gao G. Systematically profiling and annotating long intergenic non-coding RNAs in human embryonic stem cell. *From Asia Pacific Bioinforma Netw*. 2013:20-22.

80. Ng S-Y, Johnson R, Stanton LW. Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO J*. 2012;31(3):522-533.

81. Loewer S, Cabili MN, Guttman M, et al. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat Genet*. 2010;42(12):1113-1117. doi:10.1038/ng.710.

82. Lu W, Han L, Su L, et al. A 3'UTR-associated RNA, FLJ11812 maintains stemness of human embryonic stem cells by targeting miR-4459. *Stem Cells Dev*. 2015;24(9):1133-1140. doi:10.1089/scd.2014.0353.

83. Xu C, Zhang Y, Wang Q, et al. Long non-coding RNA GAS5 controls human embryonic stem cell self-renewal by maintaining NODAL signalling. *Nat Commun*. 2016;7:13287.

84. Suh M-R, Lee Y, Kim JY, et al. Human embryonic stem cells express a unique set of microRNAs. *Dev Biol*. 2004;270(2):488-498.

85. Bar M, Wyman SK, Fritz BR, et al. MicroRNA discovery and profiling in human embryonic stem cells by deep sequencing of small RNA libraries. *Stem Cells*. 2008;26(10):2496-2505.

86. Laurent LC, Chen J, Ulitsky I, et al. Comprehensive microRNA profiling reveals a unique human embryonic stem cell signature dominated by a single seed sequence. *Stem Cells*. 2008;26(6):1506-1516. doi:10.1634/stemcells.2007-1081.

87. Morin RD, Connor MDO, Griffith M, et al. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res*. 2008:610-621.

88. Li SSL, Yu SL, Kao LP, et al. Target identification of micrornas expressed highly in human embryonic stem cells. *J Cell Biochem*. 2009;106(6):1020-1030.

89. Ren J, Jin P, Wang E, Marincola FM, Stroncek DF. MicroRNA and gene expression patterns in the differentiation of human embryonic stem cells. *J Transl Med*. 2009;7:20.

90. Parsons XH, Parsons JF, Moore D a. Genome-Scale Mapping of MicroRNA Signatures in Human Embryonic Stem Cell Neurogenesis. *Mol Med Ther*. 2012;1:2-3.

91. Asikainen S, Heikkinen L, Juhila J, et al. Selective microRNA-Offset RNA expression in human embryonic stem cells. *PLoS One*. 2015;10(3):e0116668. doi:10.1371/journal.pone.0116668.

92. Calloni R, Cordero EAA, Henriques JAP, Bonatto D. Reviewing and updating the major molecular markers for stem cells. *Stem Cells Dev*. 2013;22(9):1455-1476.

93. Sigova AA, Mullen AC, Molinie B, et al. Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proc Natl Acad Sci U S A*. 2013;110(8):2876-2881.

94. Xu RH, Sampsell-Barron TL, Gu F, et al. NANOG Is a Direct Target of TGFβ/Activin-Mediated SMAD Signaling in Human ESCs. *Cell Stem Cell*. 2008;3(2):196-206.

95. Ambady S, Malcuit C, Kashpur O, et al. Expression of NANOG and NANOGP8 in a variety of undifferentiated and differentiated human cells. *Int J Dev Biol*. 2011;54(11-12):1743-1754.

96. Tripathi V, Shen Z, Chakraborty A, et al. Long Noncoding RNA MALAT1 Controls Cell Cycle Progression by Regulating the Expression of Oncogenic Transcription Factor B-MYB. *PLoS Genet*. 2013;9(3).

97. Mourtada-Maarabouni M, Pickard M, Hedge V, Farzaneh F, Williams G. GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. *Oncogene*. 2009;28373:195-208.

98. Tripathi V, Ellis JD, Shen Z, et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell*. 2010;39(6):925-938.

99. Gazy I, Zeevi DA, Renbaum P, et al. TODRA, a lncRNA at the RAD51 locus, is oppositely regulated to RAD51, and enhances RAD51-dependent DSB (double strand break) repair. *PLoS One*. 2015;10(7).

100. Kaneko S, Bonasio R, Saldaña-Meyer R, et al. Interactions between JARID2 and Noncoding RNAs Regulate PRC2 Recruitment to Chromatin. *Mol Cell*. 2014;53(2):290-300.

101. Sun C, Li S, Zhang F, et al. Long non-coding RNA NEAT1 promotes non-small cell lung cancer progression through regulation of miR-377-3p-E2F3 pathway. *Oncotarget*. 2016. doi:10.18632/oncotarget.10108.

102. Cao C, Zhang T, Zhang D, et al. The long non-coding RNA, SNHG6-003, functions as a competing endogenous RNA to promote the progression of hepatocellular carcinoma. *Oncogene*. 2016;(January):1-11.

103. Huarte M, Guttman M, Feldser D, et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell*. 2010;142(3):409-419. doi:10.1016/j.cell.2010.06.040.

104. Hung T, Wang Y, Lin MF, et al. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet*. 2011;43(7):621-629. doi:10.1038/ng.848.

105. Nagano T, Mitchell JA, Sanz LA, et al. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science*. 2008;322(5908):1717-1720. doi:10.1126/science.1163802.

106. Hawkins PG, Morris K V. Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5. *Transcription*. 2010;1(3):165-175.

107. Becker KA, Ghule PN, Therrien JA, et al. Self-renewal of human embryonic stem cells is supported by a shortened G1 cell cycle phase. *J Cell Physiol*. 2006;209(3):883-893.

108. Maynard S, Swistowska AM, Lee JW, et al. Human embryonic stem cells have enhanced repair of multiple forms of DNA damage. *Stem Cells*. 2008;26(1549-4918 (Electronic)):2266-2274.

109. Liu JC, Guan X, Ryan JA, et al. High mitochondrial priming sensitizes hESCs to DNA-damage-induced apoptosis. *Cell Stem Cell*. 2013;13(4):483-491.

110. Guo C, Song W-Q, Sun P, Jin L, Dai H-Y. LncRNA-GAS5 induces PTEN expression through inhibiting miR-103 in endometrial cancer cells. *J Biomed Sci*. 2015;22:100.

111. Nitzan M, Steiman-Shimony A, Altuvia Y, Biham O, Margalit H. Interactions between distant ceRNAs in regulatory networks. *Biophys J*. 2014;106(10):2254-2266.

112. Paci P, Colombo T, Farina L. Computational analysis identifies a sponge interaction network between long non-coding RNAs and messenger RNAs in human breast cancer. *BMC Syst Biol*. 2014;8:83.

REVIEW

Scaffolds for Artificial miRNA Expression in Animal Cells

Raquel Calloni and Diego Bonatto*

Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul, and Departamento de Biologia Molecular e Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

Artificial miRNAs (amiRNAs) are molecules that have been developed to promote gene silencing in a similar manner to naturally occurring miRNAs. amiRNAs are generally constructed by replacing the mature miRNA sequence in the pre-miRNA stem–loop with a sequence targeting a gene of interest. These molecules offer an interesting alternative to silencing approaches that are based on shRNAs and siRNAs because they present the same efficiency as these options and are less cytotoxic. amiRNAs have mostly been applied to gene knockdown in plants; they have been examined to a lesser extent in animal cells. Therefore, this article reviews the amiRNAs that have been developed for animal cells and focuses on the miRNA scaffolds that can already be applied to construct the artificial counterparts, as well as on the different approaches that have been described to promote amiRNA expression and silencing efficiency. Furthermore, the availability of amiRNA libraries and other tools that can be used to design and construct these molecules is briefly discussed, along with an overview of the therapeutic applications for which amiRNAs have already been evaluated.

INTRODUCTION

RNA INTERFERENCE (RNAi) is a biological process based on double-stranded RNA (dsRNA) that promotes posttranscriptional gene silencing.[1] RNAi has become useful for determining gene function, as an alternative to the knockout strategy, and has the potential to become a therapeutic application.[2,3] Small interfering RNA (siRNA) and short hairpin RNA (shRNA) are the most common RNAi approaches that are applied to mammalian cells.[1,4] However, as an alternative to using these synthetic molecules, gene silencing can also be achieved using molecules that are designed based on microRNAs.[5,6]

miRNAs are ~22-nucleotide (nt)-long endogenous RNA molecules that are involved in the posttranscriptional control of gene expression.[7] miRNA biogenesis begins in the cell nucleus, where the pri-miRNA is processed by Drosha to generate the stem–loop named pre-miRNA.[7,8] pre-miRNA is then exported to the cytoplasm, where it is processed by Dicer to generate a 22-nt-long dsRNA molecule.[9,10] One of the strands becomes associated with the Ago2 protein to form the RNA-induced silencing complex (RISC); the other is degraded.[9,11] As part of RISC, miRNA acts as a guide, localizing to the target mRNAs and loading them into the complex via base pairing.[12] Once paired, RISC is able to promote either the degradation of the mRNA target or the repression of its translation.[13]

Gene silencing using synthetic molecules that mimic miRNAs was first described in animal cells[14,15] and has recently been applied to vegetal cells.[16] The basic principle of this approach is to replace, in the pre-miRNA stem–loop, the sequence of the mature miRNA with a sequence targeting the gene of interest[5,6,17,18] (Fig. 1). Such a design permits the artificial hairpin to be processed by the same pathway as natural miRNAs.[19] Therefore, the structural features of the hairpin precursor are critical to its correct and efficient processing. For this reason, the original structural characteristics of the miRNA, such as the hairpin loop and the bulges and mispairing in the stem, must be maintained.[20,21]

This technique has received several denominations in the literature, such as "second generation hairpin,"[5,22] "shRNA-miR," "synthetic miRNA,"[5]

*Correspondence: Prof. Diego Bonatto, Departamento de Biologia Molecular e Biotecnologia, Universidade Federal do Rio Grande do Sul, Avenida Bento Gonçalves 9500-Prédio 43421, Caixa Postal 15005, Porto Alegre–Rio Grande do Sul, Brazil. E-mail: diegobonatto@gmail.com

DOI: 10.1089/hgtb.2015.043

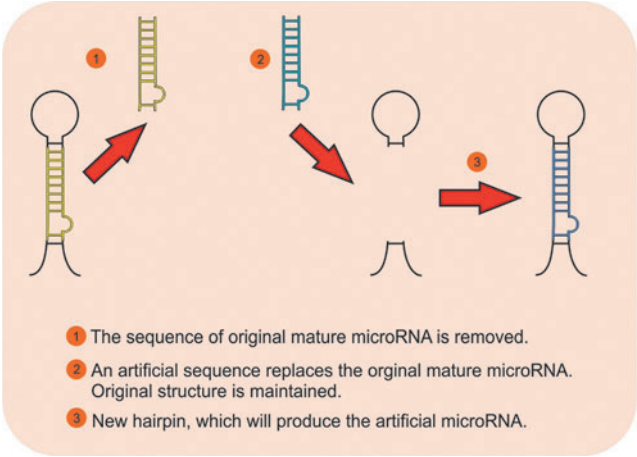


Figure 1. Simplified amiRNA construction scheme. Color images available online at www.liebertpub.com/hgtb

"miR-shRNA,"[18,23] "artificial miRNAs,"[24–27] and "shRNAmiR."[28] In this article, we chose to use "artificial miRNAs" (amiRNAs) as the terminology because it is the most common nomenclature found in the literature.

amiRNAs present several advantages over traditional silencing approaches. It is possible to co-express them with a reporter gene, such as *GFP* or *DsRed*, and the expression levels of the reporter gene frequently miRror the amiRNA expression levels.[26,29] The constructs are generally under the control of RNA polymerase (RNAP) II responsive promoters, which enables their expression to be regulated using inducers or cell-specific promoters.[23,30] Moreover, several layouts are possible when designing the constructs: (1) amiRNAs can be placed inside introns, mimicking the miRtrons expression,[17,31] (2) they can be co-expressed with a nonmarker gene, such as *p53*,[32] and (3) they can be grouped to form clusters.[27,29] For a comparison of the silencing efficiency of different amiRNA construct layouts, see ref.[19]

In addition to the above advantages, amiRNAs also seem to be a safer silencing approach. For example, although shRNAs are effective at reducing target gene mRNA and protein levels, these molecules cause side effects to cells.[22,25] shRNAs disrupt the endogenous miRNA pathway, impairing miRNA biogenesis and reducing cell viability both *in vivo* and *in vitro*.[25,33] Furthermore, shRNAs may trigger cellular immune responses, specifically the interferon response, which leads to an increase in apoptosis.[22] amiRNAs are as efficient as shRNAs at silencing genes, but they do not affect miRNA processing, even at a 10-fold higher vector dose, and do not elicit an interferon response.[22,25,33]

Although gene knockdown using amiRNAs was first described in animal cells, this procedure is not commonly used for gene expression studies in metazoan models and has thus been most largely applied to plant cells.[34] Therefore, in this article we compiled all of the publicly available information that we could find about applying amiRNAs to animal cells. Specifically, we describe different miRNA scaffolds that are already being used to promote amiRNA-based gene silencing, along with construction details of the vectors and the silencing efficiency associated with each. Moreover, both therapeutic application and further information about amiRNA libraries and the tools used to design these molecules are also described.

AMIRNA BACKBONES FOR ANIMAL CELLS

hsa-miR-26a

The miR-26a miRNA has already been identified in human, mouse, and zebrafish cells,[35] and it is the first miRNA whose scaffold was applied to induce amiRNA-based gene knockdown.[14] The miR-26a-based amiRNA effectively silenced the target when the loop was positioned at the 5′ end of the artificial antisense sequence, and a 2-nt-long stem bulge was placed at the 5′ end of the artificial sense sequence.[14] The amiRNA was placed under the control of the H1 promoter and was located upstream of the RNAP III termination sequence. When compared to an equivalent siRNA, the miR-26a amiRNA was shown to be slightly less effective at gene knockdown.[14]

hsa-miR-30

miR-30 is an miRNA family that has been documented to exist in humans, mice, and zebrafish.[36–39] It is composed of miR-30a, miR-30b, miR-30c, miR-30d, and miR-30e, which share high homology and the same seed sequence.[37] From this family, the human miR-30a stem–loop (generally reported just as miR-30) is the most widely utilized scaffold for the construction of amiRNA vectors.[6,25] Indeed, one of the first reports of the development of an amiRNA for animal cells used the miR-30 backbone; the sequence for the original mature miRNA in the stem–loop was replaced by a sequence based on the *Drosophila nxt* gene to generate what was later named miR-30-nxt.[15] miR-30-nxt was expressed as part of a long mRNA transcript and targeted an artificial construct that contained part of the *nxt* sequence. The mature amiRNA was produced by both the 5′ and 3′ arms of the hairpin, and a 7-fold reduction of the target mRNA levels was observed, which was similar to results obtained using a synthetic

shRNA.[15] Assays that inhibited the production of endogenous polypyrimidin tract binding protein (PTB) in 293T cells also revealed the production of the amiRNA by both the 5′ and 3′ arms of the hairpin and reported a 70–80% reduction of protein and mRNA.[15]

Following the publication of the Zeng et al.'s article,[15] several reports have been released describing different approaches that can be employed to promote silencing using miR-30-derived amiRNAs.[5,6,18,23,29] Variations of the human miR-30 scaffold structure were tested to check their knockdown potential. Those with major structural similarity to the original miR-30 hairpin were the most effective in silencing and had the highest expression levels.[6] The addition of 100 bp to each side of the pre-miRNA did not change the efficiency of the silencing.[6] Alternatively, the structures lacking the bulges distal to the loop in the original miR-30 had the least effective silencing.[6]

Moreover, different promoters for regulating amiRNA expression, and constructs mimicking natural miRNA clusters and miRNA expression from introns have been tested and described.[29,31] With regard to promoters, amiRNA expression has been primarily controlled by RNAP II responsive sequences.[6,29,31] However, promoters that are recognized by RNAP III, such as H1, U6, and tRNA-val, have also been tested.[5,39] The pSM2 vector (Fig. 2A) is controlled by the U6 promoter. In this case, the original miR-30 stem–loop was replaced with *Xho*I and *Eco*RI restriction sites to facilitate the insertion of the amiRNA. This construct was able to produce 12 times more small RNA than a previous shRNA that was reported by the same authors.[5,40]

Vectors permitting the regulated expression of amiRNAs have also been developed. The pPRIME (potent RNA interference using microRNA expression; Fig. 2B) vector permits the co-transcription of

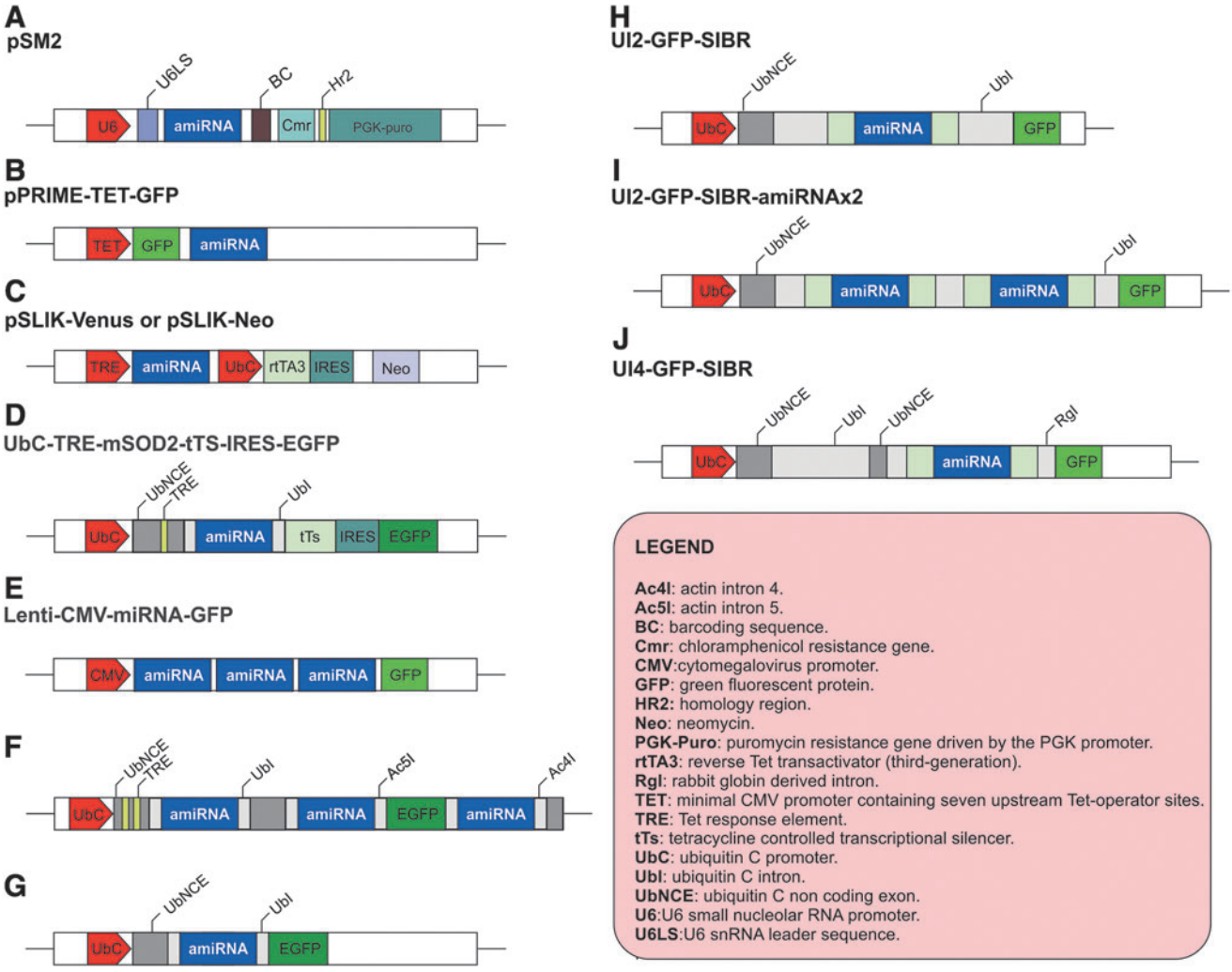


Figure 2. Schematic representation of different constructions to promote amiRNAs expression. Color images available online at www.liebertpub.com/hgtb

the reporter GFP with the amiRNA, and the expression is regulated by a tetracycline (Tet)-responsive promoter.[29] A total of three different Tet responsive promoters were evaluated in the construct: (1) TREX (Invitrogen, Carlsbad, EUA), a Tet-responsive promoter with two TET-repressor binding sites downstream of the CMV TATA box; (2) the TET-ON system, composed of a CMV promoter containing seven upstream Tet-operator sites; and (3) the TET-OFF system, in which the amiRNA is expressed only in the absence of doxicyclin (dox).[29] The highest degree of knockdown was observed using the construct regulated by TET-OFF, and the amount of GFP that was expressed was positively correlated to the degree of silencing.[29] Moreover, other reporter genes can be used instead of *GFP*, such as *dsRed*, *Neo*, and/or *LNGFR* (lacking the cytoplasmic signaling domain).[29] One drawback for the use of these constructs is the necessity of generating cells that stably express a Tet repressor in the case of TREX, or a reverse Tet-controlled transcriptional activator (rtTA) in the case of the TET-ON and TET-OFF systems.[29]

The pSLIK (single lentivector for inducible knockdown; Fig. 2C) vector is also based on a Tet-regulated promoter; however, unlike pPRIME, pSLIK expresses all of the required components for Tet-regulated promoter expression.[23] This construct is composed of an amiRNA controlled by a Tet-responsive promoter, followed by an UbC promoter to control the expression of rtTA3 and the marker gene. The rtTA3 sequence is separated from the marker gene (Venus or a neomycin resistance gene) by an IRES sequence.[23] Both TET-ON and TET-OFF were tested in the context of this construct and TET-ON produced more potent gene silencing. Knockdown assays in MEF cells revealed a 90% reduction in target gene expression levels after 2 days of dox treatment.[23] As an alternative design strategy, it is possible to locate the marker gene before the amiRNA sequence. The construct that included the GFP open reading frame between the Tet-responsive promoter and the amiRNA targeting the *Gβ2* gene was effective in silencing, whereas the depletion was accompanied by increased expression of GFP.[23] As with pSLIK, the UbC-TRE-mSOD2-tTS-IRES-EGFP construct (Fig. 2D) also contained all of the components necessary for the expression of the Tet-regulated promoter.[31] However, the amiRNA is expressed from an intron in this case.[31] The construct is regulated by the UbC promoter, followed by the first ubiquitin exon, the first ubiquitin intron, the *tTS* ORF, an IRES sequence, the *EGFP* coding sequence, and the poly (A) addition signal.[31] The first ubiquitin exon in this construct contained two tetracycline-responsive elements (TREs) and the first ubiquitin intron contained the amiRNA sequence.

Many miRNAs are grouped to form clusters in the genome.[41,42] In this case, the generation of mature miRNAs is achieved through the production of a single transcript from which all other miRNAs can be processed.[7] Inspired by this naturally occurring process, an miR-30-based polycistronic approach was proposed. Generally, amiRNAs are either concatenated to improve the knockdown of a single target or designed to silence either two or three different mRNAs. In the case of a single target, the inclusion of a second amiRNA led to a large increase in small RNA levels and a consistent reduction in target protein levels.[43] An improvement was also observed when three amiRNAs were clustered, but the gain was not as great as with the addition of the second amiRNA.[43] Interestingly, the gene knockdown was improved even when one of the hairpins was composed of an irrelevant sequence, indicating that the cluster approach could be more efficient at gene silencing.[43] Nevertheless, these observations are not a consensus. Two amiRNAs targeting the human *SOD1* gene that were placed in tandem under the control of the CMV promoter were not able to enhance the production of small RNA, and a decrease in the efficacy of RNAi was observed.[6]

The multi-hairpin design can also be applied to target two or three different mRNAs (Fig. 2E). In this case, the silencing efficiency can be equal to or higher than that observed for the single-hairpin vector and the order of the amiRNAs does not seem to affect the silencing rates.[18,43] Again, there is not yet a consensus regarding the efficiency of the clustering. In the case of using pSLIK to target two different mRNAs, the silencing potential was reported as slightly lower than that observed in the single amiRNA system.[23]

In addition to the above-discussed constructs, an alternative multi-hairpin approach, where each amiRNA is expressed from a different intron, has been described. The tetracycline-responsive construct controlled by the UbC promoter contains three introns: the first ubiquitin intron and the fourth and fifth introns of the human actin gene. In this case, each amiRNA is placed within a different intron[31] (Fig. 2F).

Using an intron to express amiRNA has also been proposed for constructs containing only one amiRNA. For example, the construct composed of the UbC promoter and UbC first intron (Fig. 2G) proposed by Zhou et al.[6] An ORF encoding *EGFP* was added downstream of the intron to provide a convenient method of monitoring amiRNA expression because both sequences are controlled by

the same promoter.[6] Thus, the levels of EGFP could be correlated to the silencing efficiency. Silencing efficiency and construct expression were similar to the values obtained with the original miR-30.[6]

hsa-miRR-31

has-miRR-31 was first described in 2001, being detected in human cells, where it is located on chromosome 9, but absent in mouse-, zebrafish-, *Xenopus*-, and *Drosophila*-derived cells.[35] For miR-31-based amiRNA construction, guide and complementary strands were replaced by sequences of interest, maintaining the structure of the original miRNA.[44] In addition, the amiRNA was flanked by 51 nt from the original miR-31 sequence each side and placed under the control of a U6 promoter or inserted within an exonic sequence of cytomegalovirus immediate-early promoter enhancer expression cassete.[44]

Constructions containing both RNAP II and RNAP III responsive promoters presented silencing rates reaching 95–98%, indicating that both systems are applicable for RNAi. Regarding to the amiRNA processing, both constructions lead to extra mature sequences with 20 and 22 nt together with the expected 21 nt fragment, evidencing an heterogeneous processing of the guide strand.[44] Moreover, the amiRNA processing was more effective in transcripts generated from the CMV promoter since larger-molecular-weight intermediates were detected only on the Northern blot analysis of the construction under the control of the U6 promoter.[44]

In addition, no activation of a type 1 interferon response and no endogenous miRNA pathway alteration were observed in cells expressing the miR-31-based amiRNA.[44]

hsa-miR-122

Conserved from mammalian species to fish, the miR-122 sequence is derived from the liver-specific noncoding RNA encoded by the *hcr* gene.[45] amiRNA construction based on this miRNA followed the same characteristics described for the miR-31-based amiRNA (see section hsa-miR-31). Similar to miR-31 construction, more efficient processing was observed in transcripts generated from the CMV promoter; however, no extra mature sequences other than the 21 nt were observed.[44] In addition, no activation of type 1 interferon response and no endogenous miRNA pathway alteration were observed in cells expressing the miR-122-based amiRNA.[44]

gga-miR-126

gga-miR-126 is an intron-expressed microRNA that has been detected in adult and embryonic chicken tissues.[46] Mature miRNAs are produced by both the 5′ and 3′ arms of the stem–loop and they are conserved in mice, humans, zebrafish, rats, fugu (*Fugu rubripes*), and pufferfish (*Tetraodon nigroviridis*).[46] The constructs that included the miR-126 backbone were composed of the stem–loop sequence plus an additional 200 bp flanking sequence on either side. The mature antisense sequence of gga-miR-126 was replaced with the sequence of interest and the final stem–loop structure mimicked the original gga-miR-126 structure.[27] The inhibition rate was measured via a luciferase assay and reached 95% with a transiently transfected vector and 85% with a stably integrated vector, each expressing only one miRNA.[27]

By clustering the constructs it becomes possible to express one to three different amiRNAs from the same vector.[27] In this case, the inhibition rate was measured at 80–90% with both the transiently transfected and the stably integrated vector.[27] Furthermore, the construct containing amiRNAs against three viral sequences was able to inhibit influenza virus production by 10-fold when compared with the control.[27]

Moreover, the constructs were shown to work well with promoters responsive to both RNAP II and III. The resulting amiRNAs were able to inhibit their targets in a variety of cell types, including chicken fibroblasts (DF-1), the mammalian epithelial 293T cells, Madin–Darby canine kidney (MDCK) cells, mouse embryonic fibroblasts (MEF), and African green monkey kidney (Vero) cells.[27]

mmu-miR-155

The coding sequence for mmu-miR-155 is found within the sequence of BIC,[47,48] a noncoding RNA that has been already described in humans, mice, and chickens.[49,50] Gene silencing via the miR-155 backbone was induced by using constructs that mimic intron-encoded miRNAs. Additionally, the expression of the amiRNA could be detected by the concomitant expression of marker genes, such as *GFP*/*EGFP* or the puromycin resistance gene.[17,24] In the case of EGFP/GFP, the fluorescence can be directly correlated with the amiRNA expression levels.[17,24]

There is another miR-155-based approach that consists of an artificial cassette inserted inside of an intron.[17] The synthetic inhibitory BIC-derived RNA (SIBR) cassette includes the region of the BIC sequence that spans from nt 134 to 283, which contains the pri-miR-155. In this fragment, the miR-155 precursor stem–loop sequence was replaced with two inverted *Bbs*I restriction sites, permitting the insertion of a 64-nt-long DNA duplex coding for the amiRNA of interest.[17] Synthetic

miRNAs were constructed by replacing the original mature miR-155 sequence with another 22-nt-long sequence and modifying the complementary strand to maintain the hairpin structure.[17]

The SIBR cassette was then inserted into an intron in two different types of construct. The one-intron model (Fig. 2H) consisted of the human UbC promoter, the first noncoding *UbC* exon and the first intron, followed by an exon encoding either GFP or the puromycin resistance protein.[17] The inhibition levels for the one-intron SIBR cassette ranged from 5- to 30-fold compared with the control. Moreover, the construct silenced the target as effectively as shRNA under the control of the RNAP III promoter. Additionally, more than one SIBR cassette could be inserted in tandem inside of the intron (Fig. 2I). Constructs were made to include from two to eight copies of an SIBR cassette that contained the same amiRNA, and upon testing they were found to produce successful inhibition. In this case of several equal cassettes, the inhibition levels were proportional to the SIBR copy number.[17]

The two-intron construction (Fig. 2J) comprised the UbC promoter, followed by a short noncoding exon, the first *UbC* intron, a second noncoding exon, a second intron derived from rabbit globin, and a third exon containing the coding sequence for either *GFP* or the puromycin resistance gene. In this case, the SIBR cassette was inserted into the second intron[17] and an enhanced inhibition was observed when compared to the construct containing only one intron. Additionally, in the two-intron model, the multimerization of the SIBR cassette improved the silencing of the target.[17]

Another report also demonstrated gene silencing following amiRNA expression from a one-intron construct. In this case, the artificial gene consisted of a portion of the CMV promoter, a synthetic exon, and a chimeric intron composed of the 5′ donor site from the first intron of the human β-globin gene and the branch and 3′ acceptor site from the intron of an immunoglobulin gene heavy chain variable region.[24] This construct was inserted into a pEGFP-N1 vector (Addgene, Cambridge, EUA); the *EGFP* coding sequence was placed in the exon following the artificial intron. The miR-155 backbone was placed into the artificial intron, and the original miR-155 stem–loop was replaced by two inverted *Bsm*BI restriction sites where synthetic hairpins could be inserted.[24]

In addition to the concurrent expression of a reporter gene together with the amiRNA, constructs that included the miR-155 backbone were developed to co-cistronically express with an endogenous gene. In this case, three different amiRNAs targeting p21 were placed upstream of the p53 open reading frame and under the control of a CMV promoter (for more detail, see "Potential Therapeutic Use of amiRNAs").[32]

hsa-miR-221

hsa-miR-221 is one of a collection of highly expressed miRNAs in hepatocellular carcinoma (HCC). For this reason, amiRNAs based on this miRNA were developed to induce gene silencing in corresponding cell lines. The inhibition potential of the miR-221-based amiRNA was tested in two HCC lineages using a luciferase reporter system. The miR-221-based amiRNA designed to target luciferase was expressed under the control of the CMV promoter and efficiently processed. The knockdown rates observed in the HCC cell lines were 64.04% and 80.48% compared with the controls.[21]

hsa-miR-223

hsa-miR-223 is highly expressed in myelomonocytic cells and barely detected in other cell types.[51] In addition to its high expression, hsa-miR-223 has a low passenger strand activity and, for this reason, it is a good alternative as an amiRNA scaffold.[52]

Two effective amiRNAs layouts were placed into the human immunodeficiency virus-1 intron under the control of a hybrid Tet-operator long-term repeat. In addition, the amiRNA expression was transcriptionally coupled to the Δ*LNGFR* (truncated human low-affinity nerve growth factor receptor) reporter and the whole construction was placed in a lentiviral platform (LV.miR). Alternatively, to the promoter controlled by doxicyclin, the miR-223 backbone can be placed downstream to an EF1α and-CMV promoter.

The first approach replaced the hsa-miR-223 stem–loop by the stem–loop of other miRNAs (hsa-miR-194, hsa-miR-155, hsa-miR-146, and hsa-miR-30), keeping unmodified miR-223 flanking sequences.[52] All the chimeric constructions, except for the one containing miR-146, were able to repress GFP expression similarly to miR-223 construction. The second approach replaced the upper miR-223 stem by a perfectly base-paired 21 bp siRNA against a gene of interest, maintaining the miR-223 flanking and loop sequences. This construction showed good results in gene silencing, reaching a 5- to 10-fold reduction in the protein levels compared with the control.[52] Moreover, this construction presented a strong strand-specific expression, with no unwanted passenger strand RISC loading.[52]

The second approach also showed to be an effective tool to induce gene silencing in primary

cells, which are the cell types where gene function studies and gene therapy are most relevant. Primary $CD4^+$ $CD25^{bright}$ T-cells were transduced with the vector carrying an amiRNA against *FOXP3* and a significant gene target downregulation was observed.[52] The silencing also induced significant biological effects in T-cells, which are commonly able to suppress $CD4^+$ cells proliferation. When exposed to T-cells transduced with the amiRNA against FOXP3, $CD4^+$ cells presented reduced proliferation suppression.[52]

In addition to single amiRNA constructions, a lentiviral vector containing two amiRNAs against the same target was also developed. However, to avoid recombination between sequences during reverse transcription, one siRNA was inserted in the miR-223 scaffold, which was placed in tandem with an siRNA inserted in the miR-155 scaffold. An additive effect in target downregulation was observed when this silencing approach was applied.[52]

bmo-miR-2764 and bmo-miR-279

miR-2764 and miR-279 are endogenous silkworm (*Bombyx mori*) miRNAs and are unique insect miRNAs whose use has already been reported for the construction of amiRNAs.[53] Two constructs are available to promote gene knockdown using miR-2764 and miR-279 as scaffolds. Both have the DsRed reporter gene upstream of the amiRNA hairpin and can either be induced or constitutively expressed.[53]

Constitutive expression is promoted by the *Orgyia pseudotsugata* multicapsid nucleopolyhedrovirus OpIE2 promoter. Only transient transfection is possible for this construct as stable transfection of the construct was unsuccessful, probably because of an oversaturation of the miRNA pathway.[53] The inhibition rates observed in the luciferase assay used to test the construct were 80–93%. However, when it was tested against *lef-11*, a gene from *Bombyx mori* nucleopolyhedrovirus (BmNPV), the mRNA levels were reduced by approximately 50%.[53]

miR-279-based amiRNA was also placed under the control of the baculovirus-induced homologous region 3 (hr3)-39K fusion promoter. This allowed the induced construct to only be expressed in cells infected with BmNPV, a silkworm pathogen. Silencing efficiency reached 50% as measured by luciferase assay.[53]

Additionally, the amiRNA cassettes were shown to be effective in other insect cell lines, such as BmN4 and BmE-SWU1 (*B. mori*), Sf9 (*Spodoptera frugiperda*), High Five (*Trichoplusia ni*), Spli221 (*Spodoptera litura*), and S2 (*Drosophila melanogaster*).[53]

hsa-miR-17-92 cluster

The cluster miR-17-92 is composed of six sequences arranged in tandem (miR-17, miR-18, miR-19a, miR-20, miR-19b, and miR-92), which code for seven mature miRNAs: miR-17-5p, miR-17-3p, miR-18, miR-19a, miR-20, miR-19b-1, and miR-92-1.[54] Although this cluster is mainly studied for its oncogenic characteristics, being the first oncomiR described, miR-17-92 also has the potential to be used as a tool for amiRNA-based viral inhibition.[54–56]

Related to the above, part of the original conformation of miR-17-92 was used as a backbone to create clustered amiRNAs.[55] Sequences from five miRNAs (hsa-miR-17, hsa-miR-18, hsa-miR-19a, hsa-miR-20, and hsa-miR-19b) were amplified by PCR and the original mature miRNA of each hairpin was replaced with sequences targeting different viral mRNAs, whereas the rest of original stem–loop plus 40 nt on each side of the hairpin were maintained.[55]

The best viral inhibition results were observed when a 19 nt antiviral sequence was inserted into the hairpin. Constructions based on the miR-20 and miR-19b backbones demonstrated individual inhibition efficiencies similar to what is observed when shRNAs are employed for silencing. However, constructs based on miR-18 were not able to silence their targets.[55] Moreover, amiRNA clustering seemed to improve the silencing activity. Northern blot analysis showed that, for the constructs based on the miR-17, miR-19a, and miR-19b backbones, the production of mature amiRNAs was improved when the hairpin was inside of a cluster. The position of the hairpin within the polycistronic construct did not affect the silencing potential.[55]

hsa-miR-106b cluster

The miR-106 cluster is located inside the 13th intron of the *MCM7* gene and it was originally composed of three miRNAs: miR-106b, miR-93, and miR-25, respectively.[41] The amiRNA construct that was engineered based on this cluster began with the introduction of restriction sites in the sequences flanking the original pre-miRNAs, permitting the insertion of amiRNAs in place of the naturally occurring hairpins. In cases where the basic pri-miRNA sequence was not maintained, a fully paired 23 bp stem replaced the original mature miRNA and the loop was substituted with a 5-base-long artificial one.[20] However, the second amiRNA of the cluster exactly mimics the structure of miR-93. This conformation was shown to avoid the off-target effects mediated by passenger strand incorporation into RISC. Additionally, the third

cluster element was substituted with an miR-93-based scaffold as the silencing promoted by the miR-25-based amiRNA was poor.[20]

It would appear that each amiRNA unit is processed independently and that the sequences that flank the original pre-miRNA are important for amiRNA expression. The removal of these fragments from the construct abolishes the amiRNA silencing capacity. Moreover, mimicking the original expression of the intron is beneficial to the functionality of the artificial cluster. The inclusion of original exon sequence fragments flanking the cluster (100 nt each side) also enhances the expression of the amiRNAs; when the exon–intron boundaries were removed, the silencing efficiency was reduced.[20]

POTENTIAL THERAPEUTIC USE OF AMIRNAS

Using gene silencing to study gene function is the primary application of amiRNAs and other RNAi molecules. However, RNAi may also potentially translate into a therapy that can be used to treat a variety disorders. Studies applying amiRNAs for use as a potential alternative therapy have been conducted, for instance, in the context of both cancer and antiviral therapy, and positive results have been observed.[20,32] The most common amiRNA construct that has been investigated is based on the murine miR-155 backbone, which was inserted into a pcDNA6.2-GW/EmGFP-miR-based miRNA expressive plasmid (Invitrogen).[55–62] We summarize several therapeutic applications of amiRNAs below.

Drug abuse

The use of amiRNAs to combat drug abuse is a very interesting therapeutic application for these molecules. After mice were injected in the brain with lentiviral constructs containing an amiRNA targeting NK1R (neurokinin-1 receptor), a protein associated with alcohol abuse,[63] their voluntary alcohol consumption rates decreased compared with controls.[60] The suppression of *NK 1R* was localized to the hippocampus and reached 40–50% of expression.[60]

Hepatic fibrogenesis

In this case, rno-miR-30a-based amiRNAs were designed to specifically express in activated HSCs, which are responsible for promoting liver fibrosis.[30] The specificity of expression was imparted by the SM-α actin promoter, which is active in HSCs but not in other hepatocyte-derived cell lines.[30] The amiRNA was expressed from an intron and targeted TGF-β1, a protein that is likely involved in fibrogenesis,[64] leading to an inhibition rate above 70%. As a consequence, the levels of collagen I, a marker of liver fibrogenesis,[65] were downregulated. Moreover, decreases in α-SMA and desmin, known HSC markers,[66] were also observed.

Hyperalgesia induced by muscle inflammation

Hyperalgesia is a state of increased pain sensitivity, and acid-sensing ion channels 3 (ASIC3) seems to be involved in this pathology. amiRNA-based reduction of ASIC3 in primary afferent fibers innervating muscle was shown to prevent the development of hyperalgesia *in vivo*.[67]

Cancer treatment

Interesting results have been observed in several different tumor-derived cell lines, such as those originating from melanoma, HCC, breast cancer, gastric cancer, and colorectal cancer.[32,57,59,61,62,68] Among the already explored targets are *CXCR4, Snai1, PLR-3, p21, PLP2*, heparanase, osteopontin, and *FAK* mRNAs.[32,57,59,61,62,68,69] Knockdown of these targets led to a reduction in the adhesion, proliferation, migration, and invasion capacity of the cells *in vitro*, compared with the controls.[57,59,61,68,69] Moreover, a significant decrease in the growth of the tumor and the inhibition of its metastatic capacity were observed *in vivo*.[59,62,68,69] Additional studies revealed that amiRNAs are not as toxic to the liver as shRNAs when used to block tumor progression.[68] In these studies, the reduction of target mRNA levels ranged from 80% for *PRL-3* in gastric tumor cells to almost 100% for *CXCR4* in breast tumor cells.[57,70]

Combining amiRNAs to target a gene encoding an antitumoral protein also showed to be an effective approach. Three amiRNAs targeting p21 were co-cistronically expressed within the p53 open reading frame.[32] This combination increased the population of malignant cells that underwent apoptosis. Furthermore, mice injected with tumoral cells and treated with viral particles containing the p53-p21 amiRNA construct presented with lower tumoral volumes compared with the controls.[32]

Viral replication inhibition

The use of amiRNAs offers an attractive alternative to combating viral infection. In general, reports investigating this approach showed both a decrease in viral replication and in infection efficiency (by $\sim$90%) for select constructs.[20,27,39,53,55,58] Additionally, amiRNAs appeared to more potently inhibit viral replication than shRNAs.[39] Indeed, amiRNAs have already been tested for their ability to inhibit the replication of hepatitis B, *B. mori* nucleopolyhedrovirus (BmNPV), HIV-1, and chicken influenza A.[20,27,40,44,52,54,57] In this sense, the constructs applied in viral replication inhibition

are usually based on human miR-30, miR-31, and miR-122, mouse miR-155, or silkworm miR-279 and –2764 scaffolds.[39,53,58] Additionaly, targeting more than one viral RNA is a way to circumvent the evolution of resistant viral strains. Therefore, clustered amiRNAs, such as those based on miR-17-92 and miR-106, have been developed to promote viral inhibition.[20,55] In the case of the miR-106b-based amiRNA construct, the efficiency of inhibition was greatly improved when a nucleolar localizing RNA TAR decoy (U16TAR) was placed downstream of the cluster and transcribed in the opposite orientation of the siRNAs.[20] *Pol*, *gag*, *tat*, and *rev* are among the viral genes that were targeted by these two approaches.[20,55] In addition, it is possible to induce amiRNA expression only when the cell suffers viral invasion. This is achieved by using promoters that are only active in infected cells, which provide an attractive alternative to avoid the disadvantages of constitutive amiRNA expression.[53]

Another possible approach to inhibit viral infection is to target RNAs that encode the cell surface proteins used by viruses to enter the cell (vs. targeting viral RNA directly). *CCR5*, one of the co-receptors used by HIV, was efficiently silenced using amiRNAs constructed with the mmu-miR-155 backbone.[71] Future directions of this work include evaluating the potential of the CCR5-targeted construct to inhibit the cellular entry of HIV.

Although most amiRNAs developed to inhibit viral replication were tested uniquely *in vitro*, the amiRNAs targeting hepatitis B virus proteins (based on hsa-miR-31 and hsa-miR-122) were also tested *in vivo*.[44] amiRNAs were delivered in infected mice by an hydrodynamic tail vein injection, and a significant reduction in the serum viral antigen concentration and in the number of circulating viral particles was observed.[44] Moreover, viral DNA was not detected in the liver of the animals that received amiRNAs injection.[44]

Sickle cells disease and β-thalassemia

Sickle cells disease (SCD) and β-thalassemia are hemoglobinopathies characterized, respectively, by β genes mutated to β^s type and decrease/absence of β globin production.[72] Clinical severity of both illnesses can be ameliorated by the increased expression of fetal hemoglobin (HbF),[73] whose expression is regulated by the BCL11A protein.[74] BCL11A is a potential target for SCD and β-thalassemia therapeutics as reduction of its expression in primary adult erythroid cells results in an elevated HbF expression.[74]

An amiRNA silencing approach was tested as a tool to promote *BCL11A* knockdown and consequently to promote *HbF* expression in G-CSF-mobilized peripheral blood $CD34^+$ hematopoietic stem and progenitor cells.[75] For some of the tested constructions, observed *BCL11A* expression reduction was similar to the shRNA approach together with a significant increase in the HbF transcription.[75] Moreover, amiRNAs did not lead to any deleterious effect on erythroid differentiation, in contrast to shRNAs, which caused a mild delay in the expression of later differentiation markers.[75]

amiRNAs DESIGN AND CONSTRUCTION

Researchers aiming to use amiRNAs to promote gene silencing have two options for obtaining these molecules: (1) design the sequence and outsource its synthesis to a supplier, or (2) in-house production of a construct of interest. To address the first option, a guide to designing miR-30-based amiRNAs has been developed, which includes tips on bulges and loop localization in addition to instructions regarding the insertion of appropriate restriction sites.[76] Software for designing amiRNAs is also available. AmiRzyn is a tool based on the PERL programming language and NET framework. This tool analyzes user input of the target DNA sequence to predict possible siRNAs sequences; an off-target verification is then performed through a BLAST facility. Following this step, the miRNA backbone and restriction sites can be chosen among five options. The AmiRzyn output includes the amiRNA sequence organized in the following order (5′–3′): sense restriction site, sense miRNA backbone, sense siRNA, loop, antisense siRNA, antisense miRNA backbone, and antisense restriction site.[77]

As an alternative to commercial synthesis, a method to construct miR-155-based amiRNAs using a one-step PCR has been developed. In this case, amiRNA is generated in a reaction that includes four primer sequences.[26] The first two act in the first set of PCR cycles and are the primers that specifically generate the amiRNA sequence that will act in gene silencing. The oligonucleotides partially anneal to each other through a short 19 bp complementary region, which corresponds to the hairpin loop. After annealing, the sequence is extended by a KOD polymerase. The resulting double-stranded fragment is then amplified by two universal primers, 155-5 and 155-3, which add the 5′ and 3′ ends of the original miR-155 to the construct.[26] The generated amiRNA cassettes are then flanked on each side with a 40-nt-long region of the original pri-miRNA. Restriction sites on both sides of the amiRNA enable it to be inserted into any vector and preserve its ability to form clusters. The above-described method also

worked for other backbones, such as miR-16, miR-206, and miR-331.[26]

Careful amiRNA design is important to improve the knockdown efficiency. It was shown that little differences in the sequences of mature guide strand generated by Drosha and Dicer cleavage may be the reason of amiRNA silencing efficiency lower than that observed for shRNAs.[75] For miR-223-based scaffolds, guide strand adjusting by the deletion of 4 bases in 5′ end and the addition of GCGC at the 3′ end resulted in higher target knockdown compared with nonmodified sequences.[75]

amiRNAs LIBRARIES

Validated amiRNAs covering known and predicted human and mouse genes are available in two different libraries, both based on the miR-30 backbone.[28] The first generation library amiRNAs are based on the pre-miRNA scaffold, whereas the second generation molecules are based on the pri-miRNA backbone. The expression of the amiRNAs is controlled by the U6 promoter and a 60 nt barcode sequence facilitates the identification of each different hairpin.[28] In the case of the second-generation molecules, it is possible to swap the region that encodes the miR-30 cassette, barcode, and microbial selection marker into recipient plasmids. Furthermore, this library includes six different amiRNA constructs that can target each known and predicted gene that is contained in either the mouse or human genome. For more details about these constructs, see ref.[28]

amiRNAs EXPRESSION AND OFF-TARGET EFFECTS

Mature miRNAs can be generated exclusively from one of the pre-miRna arms or from both 5p and 3p arms.[78] The construction of an amiRNA must be carefully done in order to reproduce the natural expression of the original miRNA,[76] as the generation

Table 1. *Natural miRNAs and amiRNAs expression along with amiRNAs off-target effects evaluation*

amiRNA scaffold	*Expression characteristics*		*Off-target effects*	*Reference*
	Natural miRNA[a]	*amiRNA*		
hsa-mir-26a	Preferentially expressed from the 5p arm	Not evaluated	Not evaluated	14
hsa-mir-30	Expressed from both 5p and 3p arms	Expressed from both 5p and 3p arms	Not evaluated	15
hsa-mir-31	Preferentially expressed from the 5p arm	Expression from 5p arm detected Expression from 3p arm not checked	amiRNA expression does not alter the silencing activity of endogenous miRNAs.	44
hsa-mir-122	Preferentially expressed from the 5p arm	Expression from 5p arm detected Expression from 3p arm not checked	amiRNA expression does not alter the silencing activity of endogenous miRNAs.	
gga-mir-126	Expressed from both 5p and 3p arms	Not evaluated	Not evaluated	27
mmu-miR-155	Preferentially expressed from the 5p arm	Preferentially expressed from the 5p arm	amiRNA expression does not affect nontarget genes expression (tested with endogenous genes).	17
hsa-mir-221	Preferentially expressed from the 3p arm	Not evaluated	Not evaluated	21
hsa-mir-223	Preferentially expressed from the 3p arm	Preferentially expressed from the 3p arm	amiRNA expression does not alter the expression and the silencing activity of endogenous miRNAs. Strand-specific RISC loading.	52
bmo-mir-279	Information not available	Not evaluated	Not evaluated	53
bmo-mir-2764	Expressed from the 3p arm	Not evaluated	Not evaluated	
hsa-mir-17-92 cluster				
hsa-mir-17	Preferentially expressed from the 5p arm	Not evaluated	amiRNA expression does not affect nontarget genes expression (tested with renilla reporter).	55
hsa-mir-18	Preferentially expressed from the 5p arm	No mature amiRNA is produced.	amiRNA expression does not affect nontarget genes expression (tested with renilla reporter).	
hsa-mir-19a	Preferentially expressed from the 3p arm	Higher silencing effectiveness when expressed from the 5p arm	amiRNA expression does not affect nontarget genes expression (tested with renilla reporter).	
hsa-mir-20	Preferentially expressed from the 5p arm	Higher silencing effectiveness when expressed from the 5p arm	amiRNA expression does not affect non-target genes expression (tested with renilla reporter).	
hsa-mir-19b	Preferentially expressed from the 3p arm.	Higher silencing effectiveness when expressed from the 3p arm.	amiRNA expression does not affect nontarget genes expression (tested with renilla reporter).	
hsa-mir-106b				
hsa-mir-106b	Preferentially expressed from the 5p arm	Negligible and marginal passenger strand activity	Not evaluated	20
hsa-mir-25	Preferentially expressed from the 3p arm	Negligible and marginal passenger strand activity	Not evaluated	
hsa-mir-93	Preferentially expressed from the 5p arm	Negligible and marginal passenger strand activity	Not evaluated	

[a]Information obtained from miRBase.

of a wrong mature sequence might lead to gene silencing of genes other than the interest gene. In addition, the expression of amiRNA might interfere in the expression of endogenous miRNAs. However, the majority of studies describing amiRNA backbones poorly evaluated the type of mature sequences expressed or the existence of off-target effects caused by the amiRNA expression.[5,14,15,21,22,27,53]

The ideal procedure is to check which mature sequences are being produced, the existence of wrong targets silencing, and alteration of endogenous miRNAs activity.[17,52,55] Off-target effects can be evaluated by measuring the expression of nontarget genes or by using a vector expressing luciferase or renilla reporters to verify any alteration in RNA expression.[17,55] Alteration of miRNAs pathway can be tested introducing a (1) vector coding for a miRNA, (2) a second vector containing the miRNA target, and (3) a third vector with the desired amiRNA in the cell studied. The transfection is followed by an analysis of the miRNA silencing rate in cells with and without amiRNA.[44] Another simpler approach is quantifying the expression of an endogenous miRNA and check its capacity to silence reporter genes containing its target sequence.[52] Table 1 summarizes how amiRNAs described in this review are expressed compared with their correspondent original miRNAs (from 5p, 3p, or both arms) along with RNAs off-target effects information.

CONCLUSIONS

amiRNAs are not employed to promote gene silencing in animal cells as often as shRNAs, the most commonly used knockdown method. However, several constructs designed to express amiRNAs have been developed and were shown to be efficient in a variety of cell types, from mammals to insects, in both basic research and therapeutic tests. Promoters can be chosen (including those that respond to RNAP II or III, in addition to other constitutive and inducible promoters) to express the amiRNA alone or as part of a cluster that includes additional amiRNAs, either derived from an intron or together with a marker gene. Moreover, tools and protocols are already available to aid in the design and production of amiRNAs, together with libraries of amiRNAs capable of targeting all known mouse and human genes. Because of the wide range of available options that can be used to control the expression of these molecules, the fact that they are safer than and as effective as shRNAs, and the promising results that have been observed in preclinical studies, the use of amiRNAs provides an attractive option for gene silencing endeavors.

ACKNOWLEDGMENTS

This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) through Grant 301149/2012-7, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) through Grant 004/12, and Fundação de Amparo à Pesquisa do Rio Grande do Sul (Fapergs) through Grant 11/2072-2.

AUTHOR DISCLOSURE

No competing financial interests exist.

REFERENCES

1. Martin SE, Caplen NJ. Applications of RNA interference in mammalian systems. Annu Rev Genomics Hum Genet 2007;8:81–108.
2. Agrawal N, Dasaradhi PVN, Mohmmed A, et al. RNA interference: Biology, mechanism, and applications. Microbiol Mol Biol Rev 2003;67:657–685.
3. Aagaard L, Rossi JJ. RNAi therapeutics: Principles, prospects and challenges. Adv Drug Deliv Rev 2007;59:75–86.
4. Sandy P, Ventura A, Jacks T. Mammalian RNAi: A practical guide. Biotechniques 2005;39:215–224.
5. Silva JM, Li MZ, Chang K, et al. Second-generation shRNA libraries covering the mouse and human genomes. Nat Genet 2005;37:1281–1288.
6. Zhou H, Xia XG, Xu Z. An RNA polymerase II construct synthesizes short-hairpin RNA with a quantitative indicator and mediates highly efficient RNAi. Nucleic Acids Res 2005;33:1–8.
7. Lee Y, Jeon K, Lee JT, et al. MicroRNA maturation: Stepwise processing and subcellular localization. EMBO J 2002;21:4663–4670.
8. Lee Y, Ahn C, Han J, et al. The nuclear RNase III Drosha initiates microRNA processing. Nature 2003;425:415–419.
9. Hutvágner G, McLachlan J, Pasquinelli AE, et al. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science 2001;293:834–838.
10. Lund E, Güttinger S, Calado A, et al. Nuclear export of microRNA precursors. Science 2004;303:95–98.
11. Maniataki E, Mourelatos Z. A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. Genes Dev 2005;19:2979–2990.
12. Bushati N, Cohen SM. microRNA functions. Annu Rev Cell Dev Biol 2007;23:175–205.
13. Eulalio A, Huntzinger E, Izaurralde E. Getting to the root of miRNA-Mediated Gene Silencing. Cell 2008;132:9–14.
14. McManus MT, Petersen CP, Haines BB, et al. Gene silencing using micro-RNA designed hairpins. RNA 2002;8:842–850.
15. Zeng Y, Wagner EJ, Cullen BR. Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. Mol Cell 2002;9:1327–1333.
16. Schwab R, Ossowski S, Riester M, et al. Highly specific gene silencing by artificial microRNAs in Arabidopsis. Plant Cell 2006;18:1121–1133.
17. Chung K-H, Hart CC, Al-Bassam S, et al. Polycistronic RNA polymerase II expression vectors for RNA interference based on BIC/miR-155. Nucleic Acids Res 2006;34:e53.

18. Zhu X, Santat LA, Chang MS, et al. A versatile approach to multiple gene RNA interference using microRNA-based short hairpin RNAs. BMC Mol Biol 2007;8:98.

19. Hu T, Chen P, Fu Q, et al. Comparative studies of various artificial microRNA expression vectors for RNAi in mammalian cells. Mol Biotechnol 2010;46:34–40.

20. Aagaard L a, Zhang J, Eije KJ von, et al. Engineering and optimization of the miR-106b cluster for ectopic expression of multiplexed anti-HIV RNAs. Gene Ther 2008;15:1536–1549.

21. Huang X, Jia Z. Construction of HCC-targeting artificial miRNAs using natural miRNA precursors. Exp Ther Med 2013;6:209–215.

22. Bauer M, Kinkl N, Meixner a, et al. Prevention of interferon-stimulated gene expression using microRNA-designed hairpins. Gene Ther 2009;16:142–147.

23. Shin K-J, Wall EA, Zavzavadjian JR, et al. A single lentiviral vector platform for microRNA-based conditional RNA interference and coordinated transgene expression. Proc Natl Acad Sci USA 2006;103:13759–13764.

24. Du G, Yonekubo J, Zeng Y, et al. Design of expression vectors for RNA interference based on miRNAs and RNA splicing. FEBS J 2006;273: 5421–5427.

25. McBride JL, Boudreau RL, Harper SQ, et al. Artificial miRNAs mitigate shRNA-mediated toxicity in the brain: Implications for the therapeutic development of RNAi. Proc Natl Acad Sci USA 2008; 105:5868–5873.

26. Hu T, Fu Q, Chen P, et al. Construction of an artificial microRNA expression vector for simultaneous inhibition of multiple genes in mammalian cells. Int J Mol Sci 2009;10:2158–2168.

27. Chen SCY, Stern P, Guo Z, et al. Expression of multiple artificial microRNAs from a chicken miRNA126-based lentiviral vector. PLoS One 2011;6:e22437.

28. Chang K, Elledge SJ, Hannon GJ. Lessons from nature: MicroRNA-based shRNA libraries. Nat Methods 2006;3:707–714.

29. Stegmeier F, Hu G, Rickles RJ, et al. A lentiviral microRNA-based system for single-copy polymerase II-regulated RNA interference in mammalian cells. Proc Natl Acad Sci USA 2005;102:13212–13217.

30. Chang Y, Jiang HJ, Sun XM, et al. Hepatic stellate cell-specific gene silencing induced by an artificial microRNA for antifibrosis *in vitro*. Dig Dis Sci 2010;55:642–653.

31. Xia XG, Zhou H, Xu Z. Multiple shRNAs expressed by an inducible pol II promoter can knock down the expression of multiple target genes. Biotechniques 2006;41:64–68.

32. Idogawa M, Sasaki Y, Suzuki H, et al. A single recombinant adenovirus expressing p53 and p21-targeting artificial microRNAs efficiently induces apoptosis in human cancer cells. Clin Cancer Res 2009;15:3725–3732.

33. Boudreau RL, Martins I, Davidson BL. Artificial microRNAs as siRNA shuttles: Improved safety as compared to shRNAs *in vitro* and *in vivo*. Mol Ther 2009;17:169–175.

34. Sablok G, Pérez-Quintero ÁL, Hassan M, et al. Artificial microRNAs (amiRNAs) engineering—On how microRNA-based silencing methods have affected current plant silencing research. Biochem Biophys Res Commun 2011;406:315–319.

35. Lagos-Quintana M, Rauhut R, Lendeckel W, et al. Identification of novel genes coding for small expressed RNAs. Science 2001;294:853–858.

36. Hand NJ, Master ZR, Eauclaire SF, et al. The microRNA-30 family is required for vertebrate hepatobiliary development. Gastroenterology 2009;136:1081–1090.

37. Ketley A, Warren A, Holmes E, et al. The mir-30 microRNA family targets smoothened to regulate Hedgehog signalling in zebrafish early muscle development. PLoS One 2013;8:e65170.

38. Ouzounova M, Vuong T, Ancey P-B, et al. MicroRNA miR-30 family regulates non-attachment growth of breast cancer cells. BMC Genomics 2013;14:139.

39. Boden D, Pusch O, Silbermann R, et al. Enhanced gene silencing of HIV-1 specific siRNA using microRNA designed hairpins. Nucleic Acids Res 2004;32:1154–1158.

40. Paddison PJ, Silva JM, Conklin DS, et al. A resource for large-scale RNA-interference-based screens in mammals. Nature 2004;428:427–431.

41. Altuvia Y, Landgraf P, Lithwick G, et al. Clustering and conservation patterns of human microRNAs. Nucleic Acids Res 2005;33:2697–2706.

42. Hayashita Y, Osada H, Tatematsu Y, et al. A polycistronic microRNA cluster, miR-17–92, is overexpressed in human lung cancers and enhances cell proliferation. Cancer Res 2005;65:9628–9632.

43. Sun D, Melegari M, Sridhar S, et al. Multi-miRNA hairpin method that improves gene knockdown efficiency and provides linked multi-gene knockdown. Biotechniques 2006;41:59–63.

44. Ely A, Naidoo T, Mufamadi S, et al. Expressed anti-HBV primary microRNA shuttles inhibit viral replication efficiently *in vitro* and *in vivo*. Mol Ther 2008;16:1105–1112.

45. Chang J, Nicolas E, Marks D, et al. miR-122, a mammalian liver-specific microRNA, is processed from hcr mRNA and may downregulate the high affinity cationic amino acid transporter CAT-1. RNA Biol 2004;1:106–113.

46. Xu H, Wang X, Du Z, et al. Identification of microRNAs from different tissues of chicken embryo and adult chicken. FEBS Lett 2006;580:3610–3616.

47. Lagos-Quintana M, Rauhut R, Yalcin A, et al. Identification of tissue-specific MicroRNAs from mouse. Curr Biol 2002;12:735–739.

48. Eis PS, Tam W, Sun L, et al. Accumulation of miR-155 and BIC RNA in human B cell lymphomas. Proc Natl Acad Sci USA 2005;102:3627–3632.

49. Tam W, Ben-Yehuda D, Hayward WS. bic, a novel gene activated by proviral insertions in avian leukosis virus-induced lymphomas, is likely to function through its noncoding RNA. Mol Cell Biol 1997;17:1490–1502.

50. Tam W. Identification and characterization of human BIC, a gene on chromosome 21 that encodes a noncoding RNA. Gene 2001;274:157–167.

51. Brown BD, Gentner B, Cantore A, et al. Endogenous microRNA can be broadly exploited to regulate transgene expression according to tissue, lineage and differentiation state. Nat Biotechnol 2007;25:1457–1467.

52. Amendola M, Passerini L, Pucci F, et al. Regulated and multiple miRNA and siRNA delivery into primary cells by a lentiviral platform. Mol Ther 2009;17:1039–1052.

53. Zhang J, He Q, Zhang CD, et al. Inhibition of BmNPV replication in silkworm cells using inducible and regulated artificial microRNA precursors targeting the essential viral gene lef-11. Antiviral Res 2014;104:143–152.

54. He L, Thomson JM, Hemann MT, et al. A microRNA polycistron as a potential human oncogene. Nature 2005;435:828–833.

55. Liu YP, Haasnoot J, Brake O, et al. Inhibition of HIV-1 by multiple siRNAs expressed from a single microRNA polycistron. Nucleic Acids Res 2008;36: 2811–2824.

56. Mogilyansky E, Rigoutsos I. The miR-17/92 cluster: A comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease. Cell Death Differ 2013;20:1603–1614.

57. Liang Z, Wu H, Reddy S, et al. Blockade of invasion and metastasis of breast cancer cells via targeting CXCR4 with an artificial microRNA. Biochem Biophys Res Commun 2007;363:542–546.

58. Gao Y-F, Yu L, Wei W, et al. Inhibition of hepatitis B virus gene expression and replication by artificial microRNA. World J Gastroenterol 2008;14:4684–4689.

59. Wang Z, He Y-L, Cai S-R, et al. Expression and prognostic impact of PRL-3 in lymph node metastasis of gastric cancer: Its molecular mechanism was investigated using artificial microRNA interference. Int J Cancer 2008;123:1439–1447.

60. Baek MN, Jung KH, Halder D, et al. Artificial microRNA-based neurokinin-1 receptor gene silencing reduces alcohol consumption in mice. Neurosci Lett 2010;475:124–128.

61. Guo HM, Zhang XQ, Xu CH, et al. Inhibition of invasion and metastasis of gastric cancer cells through snail targeting artificial microRNA interference. Asian Pacific J Cancer Prev 2011;12:3433–3438.

62. Ozawa H, Sonoda Y, Suzuki T, et al. Knockdown of proteolipid protein 2 or focal adhesion kinase with an artificial microRNA reduces growth and metastasis of B16BL6 melanoma cells. Oncol Lett 2012;3:19–24.

63. Seneviratne C, Ait-Daoud N, Ma JZ, C et al. Susceptibility locus in neurokinin-1 receptor gene associated with alcohol dependence. Neuropsychopharmacology 2009;34:2442–2449.

64. Gressner aM, Weiskirchen R. Modern pathogenetic concepts of liver fibrosis suggest stellate cells and TGF-α as major players and therapeutic targets. J Cell Mol Med 2006;10:76–99.

65. Seyer JM, Hutcheson ET, Kang H. Collagen polymorphism in normal and cirrhotic human liver. J Clin Invest 1977;59:241–248.

66. Moreira RK. Hepatic stellate cells and liver fibrosis. Arch Pathol Lab Med 2007;131:1728–1734.

67. Walder RY, Gautam M, Wilson SP, et al. Selective targeting of ASIC3 using artificial miRNAs inhibits primary and secondary hyperalgesia after muscle inflammation. Pain 2011;152:2348–2356.

68. Liu XY, Tang QS, Chen HC, et al. Lentiviral miR30-based RNA interference against heparanase suppresses melanoma metastasis with lower liver and lung toxicity. Int J Biol Sci 2013;9:564–577.

69. Sun B-S, Dong Q-Z, Ye Q-H, et al. Lentiviral-mediated miRNA against osteopontin suppresses tumor growth and metastasis of human hepatocellular carcinoma. Hepatology 2008;48:1834–1842.

70. Li Z, Zhan W, Wang Z, et al. Inhibition of PRL-3 gene expression in gastric cancer cell line SGC7901 via microRNA suppressed reduces peritoneal metastasis. Biochem Biophys Res Commun 2006;348:229–237.

71. Glazkova DV, Vetchinova S, Bogoslovskaya EV, et al. Downregulation of human CCR5 gene expression with artificial microRNAs. Mol Biol 2013;47:419–428.

72. Bank A, Mears JG, RamiRez F. Disorders of human hemoglobin. Science 1980;207:486–493.

73. Perrine SP, Pace BS, Faller DV. Targeted fetal hemoglobin induction for treatment of beta hemoglobinopathies. Hematol Oncol Clin North Am 2014;28:233–248.

74. Sankaran VG, Menne TF, Xu J, et al. Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. Science 2008;322:1839–1842.

75. Guda S, Brendel C, Renella R, et al. miRNA-embedded shRNAs for lineage-specific BCL11A knockdown and hemoglobin F induction. Mol Ther 2015;23:1465–1474.

76. Chang K, Marran K, Valentine A, et al. Creating an miR30-based shRNA vector. Cold Spring Harb Protoc 2013;8:631–635.

77. Baby J, Vrundha MN, Susheel GC. AmiRzyn: PERL centered artificial microRNA designing aid. Int Res J Biol Sci 2012;1:18–23.

78. Kozomara A, Griffiths-Jones S. miRBase: Integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res 2011;39: D152-D157.

Received for publication March 30, 2015;
accepted after revision August 21, 2015.

Published online: August 27, 2015.