UNIVERSIDADE FEDERAL DE UBERLÂNDIA

INSTITUTO DE BIOTECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA

ROBERTA MOURA RAMOS VARGAS

UTILIZAÇÃO DE FERRAMENTAS BIOINFORMÁTICAS PARA BUSCA, ANÁLISE E COMPREENSÃO DE INFORMAÇÕES ORIUNDAS DE *NEXT GENERATION SEQUENCING* E DESCOBERTA DE MIRNAS CONSERVADOS UTILIZANDO MIRDEEP2

PATOS DE MINAS – MG

DEZEMBRO DE 2021

ROBERTA MOURA RAMOS VARGAS

UTILIZAÇÃO DE FERRAMENTAS BIOINFORMÁTICAS PARA BUSCA, ANÁLISE E COMPREENSÃO DE INFORMAÇÕES ORIUNDAS DE *NEXT GENERATION SEQUENCING* E DESCOBERTA DE MIRNAS CONSERVADOS UTILIZANDO MIRDEEP2

Dissertação de mestrado apresentado ao Programa de Pós-graduação em Biotecnologia como requisito parcial para obtenção do título de Mestre em Biotecnologia.

Área de concentração: Bioinformática e Biologia Molecular aplicada à genômica, transcriptômica e proteômica

Nome do orientador Prof. Dr.: Laurence Rodrigues do Amaral

PATOS DE MINAS – MG

DEZEMBRO DE 2021

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

V297 Vargas, Roberta Moura Ramos, 1994-
2021 Utilização de ferramentas bioinformáticas para busca, análise e compreensão de informações oriundas de Next Generation Sequencing e descoberta de miRNAS conservados utilizando mirdeep2 [recurso eletrônico] / Roberta Moura Ramos Vargas. - 2021.

Orientador: Laurence Rodrigues do Amaral.
Dissertação (Mestrado) - Universidade Federal de Uberlândia, Pós-graduação em Biotecnologia.
Modo de acesso: Internet.
Disponível em: http://doi.org/10.14393/ufu.di.2022.36
Inclui bibliografia.
Inclui ilustrações.

1. Biotecnologia. I. Amaral, Laurence Rodrigues do, 1978-, (Orient.). II. Universidade Federal de Uberlândia. Pós-graduação em Biotecnologia. III. Título.

CDU: 60

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:

Gizele Cristine Nunes do Couto - CRB6/2091

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
Coordenação do Programa de Pós-Graduação em Biotecnologia
Av. Getúlio Vargas, 230, 3º andar, Sala 308 - Bairro Centro, Patos de Minas-MG, CEP 38700-128
Telefone: (34) 3823-3714 - Ramal 39 - www.ppgbiotec.ibtec.ufu.br - ppgbiotec@ibtec.ufu.br

ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Biotecnologia				
Defesa de:	Dissertação de Mestrado Acadêmico				
Data:	21 de dezembro de 2021	Hora de início:	09:00	Hora de encerramento:	12:00
Matrícula do Discente:	41812BTC012				
Nome do Discente:	Roberta Moura Ramos				
Título do Trabalho:	UTILIZAÇÃO DE FERRAMENTAS BIOINFORMÁTICAS PARA BUSCA, ANÁLISE E COMPREENSÃO DE INFORMAÇÕES ORIUNDAS DE NEXT GENERATION SEQUENCING E DESCOBERTA DE MIRNAS CONSERVADOS UTILIZANDO MIRDEEP2				
Área de concentração:	Biociência				
Linha de pesquisa:	Bioinformática e Biologia Molecular aplicada à genômica, transcriptômica e proteômica				
Projeto de Pesquisa de vinculação:					

Reuniu-se por Webconferência, Campus Patos de Minas, da Universidade Federal de Uberlândia, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Biotecnologia, assim composta: Professores Doutores: Gláucia Braga e Silva, Matheus de Souza Gomes e Laurence Rodrigues do Amaral, orientador da discente.

Iniciando os trabalhos o presidente da mesa, Dr. Laurence Rodrigues do Amaral apresentou a Comissão Examinadora e a candidata, agradeceu a presença do público, e concedeu à discente a palavra para a exposição do seu trabalho. A duração da apresentação da discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor(a) presidente concedeu a palavra, pela ordem sucessivamente, aos(às) examinadores(as), que passaram a arguir o(a) candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o(a) candidato(a):

Aprovado.

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre .

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.

Documento assinado eletronicamente por **Laurence Rodrigues do Amaral**, **Professor(a) do Magistério Superior**, em 21/12/2021, às 11:58, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Matheus de Souza Gomes**, **Membro de Comissão**, em 21/12/2021, às 11:58, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Gláucia Braga e Silva**, **Usuário Externo**, em 21/12/2021, às 12:00, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3272171** e o código CRC **1265B36C**.

Referência: Processo nº 23117.088499/2021-08 SEI nº 3272171

UTILIZAÇÃO DE FERRAMENTAS BIOINFORMÁTICAS PARA BUSCA, ANÁLISE E COMPREENSÃO DE INFORMAÇÕES ORIUNDAS DE *NEXT GENERATION SEQUENCING* E DESCOBERTA DE MIRNAS CONSERVADOS UTILIZANDO MIRDEEP2

Dissertação de mestrado apresentado ao Programa de Pós-graduação em Biotecnologia como requisito parcial para obtenção do título de Mestre em Biotecnologia.

Área de concentração: Bioinformática e Biologia Molecular aplicada à genômica, transcriptômica e proteômica

Aprovado em ___/___/___

BANCA EXAMINADORA:

_____

Laurence Rodrigues do Amaral – Professor Doutor (IBTEC/UFU)

_____

Matheus de Souza Gomes – Professor Doutor (IBTEC/UFU)

_____

Gláucia Braga e Silva– Professora Doutora (UFV/Florestal)

PATOS DE MINAS – MG

DEZEMBRO DE 2021

Dedico este trabalho à Deus por seu amor e por ter me concedido o dom da ciência; dedico este trabalho à Maria Santíssima e ao meu Anjo da Guarda pelo amparo e carinho; dedico este trabalho ao meu marido e aos meus pais por todo apoio e amor.

AGRADECIMENTOS

Agradeço à Deus por ter me amado a ponto de permitir que eu desempenhasse parte do seu plano aqui na Terra.

Agradeço à Nossa Senhora Maria por seu amor.

Agradeço ao meu Anjo da Guarda por sempre me manter sob sua proteção.

Agradeço ao meu marido Marcelo Fabiano por todo amor, compreensão e incentivo.

Agradeço aos meus pais Roberto e Edna por seus amores, suporte e preocupações.

Agradeço ao meu orientador Laurence pelos ensinamentos, trabalho e motivação nesta caminhada acadêmica.

Agradeço aos meus avós Ranulfo e Ana por seus amores e cuidados.

Agradeço a(o) meu (minha) irmã(o) (*in memoriam*); meus bisavós Sebastião (*in memoriam*) e Maria (*in memoriam*); meu padrinho João (*in memoriam*), meus avós Jesus (*in memoriam*) e Ana Rosa (*in memoriam*) por seus exemplos e iluminação.

Agradeço a minhas primas Talita, Xanina, Ana Isa, meu primo Marcelo, minhas tias Ângela e Lena por seus carinhos e que em algum momento da minha vida me apoiaram.

Agradeço aos meus sogros Nelson e Auxiliadora bem como minha cunhada Larissa pelas preocupações e carinho.

Agradeço as minhas madrinhas Juliana e Zeila pelos seus amores e incentivos.

Agradeço aos meus afilhados Maria Clara, Isabelly, Beattriz, Fernanda, Maria Valentina, Maria Paula, Leonardo Gustavo, Thaís e Maria Júlia pelos seus carinhos e por deixarem minha vida mais leve.

Agradeço aos meus tios, tias, primos, primas que sempre se fizeram presentes.

Agradeço a Francielle, Paula, Núbia por suas amizades.

Agradeço também aos meus pets Laica (desaparecida), Moema (*in memoriam*), Esmeralda (doada), Estrela (doada), Cravo (desaparecido), Peixes (*in memoriam*), Bibi (*in memoriam*), Kyara (*in memoriam*), Mel (*in memoriam*), Cacau (doada), Lila, Tuti (*in memoriam*), Tupi (desaparecido), Vaca (vendida) e Ragnar por terem me ajudado com minha ansiedade e pelos seus carinhos.

Agradeço a todos os colaboradores deste trabalho.

Agradeço a todos que torceram pelo meu sucesso.

"*Don't walk through the trodden path, as it only leads to where the others have been*". (Alexander Graham Bell).

RESUMO

Nos anos 70 foram criados os primeiros métodos de sequenciamento. Um tempo depois surgiram novos métodos automatizados, as técnicas de sequenciamento de última geração (NGS). O NGS é utilizado atualmente para a geração de dados genômicos, produzindo quantidades cada vez maiores de informações com baixo custo. Essas técnicas permitem sequenciar o RNA facilitando a aquisição de grandes conjuntos transcriptômicos. MicroRNAs são um subconjunto de moléculas pequenas de RNAs não codificantes, com aproximadamente 22 nucleotídeos, considerados reguladores pós-transcricional que podem clivar mRNAs alvo, reprimem sua tradução e levam à diminuição da estabilidade do mRNA, desempenham um papel importante na maioria dos processos biológicos que ainda não foi totalmente revelado. Uma quantificação precisa de miRNA, é muito importante como estratégia computacional para reduzir ou minimizar mapeamentos potencialmente falsos contra o genoma para sua identificação. O software mais utilizado na quantificação de miRNAs é o mirDeep2. Desta maneira, esta pesquisa tem por objetivo aplicar ferramentas computacionais de Bioinformática em dados oriundos de *Next Generation Sequencing* na busca por informações importantes que possam contribuir para o avanço da ciência, utilizando somente plataforma *in silico*. A partir da análise das amostras realizada pelo mirDeep2 pode-se constatar que há possíveis miRNAs presentes nos tecidos cancerosos. Houve a predição dos alvos dos miRNAs e também foram reunidas informações dos dados de sequenciamento brutos de provenientes de NGS (SRRs); sumarizou-se os 10 primeiros alvos de cada miRNA e mostrou também a ordem com a qual apareciam nos diferentes programas utilizados para esta análise.

Palavras-chave: Biotecnologia, Bioinformática, NGS, miRNAs, mirDeep2

ABSTRACT

In the 1970s, the first sequencing methods were created. A while later, new automated methods emerged, the next generation sequencing techniques (NGS). NGS is currently used for the generation of genomic data, producing increasing amounts of information at low cost. These techniques allow RNA sequencing, facilitating the acquisition of large transcriptomic sets. MicroRNAs are a subset of small molecules of non-coding RNAs, with approximately 22 nucleotides, considered post-transcriptional regulators that can cleave target mRNAs, repress their translation and lead to decreased mRNA stability, they play an important role in most biological processes that has not yet been fully revealed. An accurate quantification of miRNA is very important as a computational strategy to reduce or minimize potentially false mappings against the genome for its identification. The software most used in the quantification of miRNAs is mirDeep2. In this way, this research aims to apply Bioinformatics computational tools to data from Next Generation Sequencing in the search for important information that can contribute to the advancement of science, using only an in silico platform. From the analysis of samples performed by mirDeep2, it can be seen that there are possible miRNAs present in cancerous tissues. MiRNA targets were predicted and information was also gathered from raw sequencing data from NGS (SRRs); the first 10 targets of each miRNA were summarized and the order in which they appeared in the different programs used for this analysis was also shown.

Key words: Biotechnology, Bioinformatics, NGS, miRNAs, mirDeep2.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

A	Adenina
A3	Receptor de adenosina 3
ABI	*Life Technologies Applied Biosystems*
ADP	Difosfato de adenosina
AMP	Monofosfato de adenosina
AMPK	Proteína quinase ativada por monofosfato de adenosina
ATP	Trifosfato de adenosina
B2M	Gene β-2-microglobulina
C	Citosina
CD73	ecto-5'-nucleotidase
cDNA	DNA complementar
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
DCL1	Dicer-like 1
DDBJ	Banco de dados de DNA do Japão
DDL	Proteína *dawdle*
ddNTP	Didesoxinucleotídeo
dNTP	Desoxirribonucleotídeo Fosfatado
DNA	Ácido desoxirribonucleico
DGCR8	Proteína de ligação ao RNA DGCR8
EBI	Instituto Europeu de Bioinformática
EGF	Fator de crescimento epidérmico
ENTPDase1	Ectonucleosídeo trifosfato difosfohidrolase 1
ENTPDase3	Ectonucleosídeo trifosfato difosfohidrolase 3
G	Guanina

GC content ou RNA de G ou C	É a porcentagem de bases nitrogenadas em uma molécula de DNA
HYL1	*Hyponastic leaves 1*
ID	Identificação
INSDC nucleotídeos	Colaboração internacional de banco de dados de sequência de
kb	Quilobase
LBAM	Laboratório de Bioinformática e Análises Moleculares
LNCap	Linhagem celular de adenocarcinoma de próstata
$MgCl_2$	Cloreto de magnésio
miRBase	Banco de dados de microRNA
miRISC	Complexo Silenciador Mediado por miRNA
mRNA	Ácido ribonucléico mensageiro
miRNA	microRNA
miRNA-seq	Sequenciamento de microRNA
MiSeq	Sequenciador de alto rendimento para identificar miRNAs
Mbp	Milhões de pares de bases
N	Pode ser um dos quatro nucleotídeos (A T C G)
NCBI	Centro Nacional de Informações sobre Biotecnologia
NGS	Sequenciamento de Nova Geração
NK	Assassino natural
Nt	Nucleotídeos
OH	Hidroxila
P1	Receptor purinérgico 1
P2	Receptor purinérgico 2
pb	Pares de bases

PC3	Linhagem celular humana de tumor de próstata derivada de metástase óssea
PCa	Câncer de próstata
PCR	Reação em cadeia da polimerase
pH	potencial hidrogeniônico
Pri-miRNA	miRNA primário
Pré-miRNA	Precursor de micro Ácido ribonucléico
PTC	Controlador Térmico Programável
qPCR	PCR em Tempo Real
Ran-GTP	Proteína Nuclear Relacionada a Ras Trifosfato de Guanosina
RISC	Complexo de silenciamento induzido por RNA
RNA	Ácido ribonucléico
Rnase	Ribonuclease
RNAs-seq	Novas tecnologias de Sequenciamento de Nova Geração
RNA pol	Polimerase de Ácido ribonucleico
RPMI	Meio de crescimento usado na cultura de células
RT-qPCR	Reação em cadeia da polimerase de transcrição reversa quantitativa
T	Timina
U	Uracila
UFU	Universidade Federal de Uberlândia
UTR	Região não traduzida
ScRNA	RNA de célula única
SD	Desvio padrão
SE	Serrate
SOCS5	Supressor de sinalização de citocinas 5
SRR	Dados de sequenciamento brutos de provenientes de NGS

SRA *Sequence Read Archive* (é um banco de dados provenientes de sequenciamento de alto rendimento)

LISTA DE SÍMBOLOS

3'	Ligações que acontecem na hidroxila do carbono 3
-3p	Região que vai do *loop* e a extremidade 3'
5'	Ligações que acontecem na hidroxila do carbono 5
-5p	Região que vai da extremidade 5' até o *loop*
-a	Opção que leva ao relatório todos os alinhamentos válidos
.arf	Um arquivo com leituras mapeadas
best	Melhor alinhamento
bowtie	Mapeamento
bowtie-build	Uma ferramenta usada para construir um índice de transformação de *Burrows-Wheeler* dos precursores potenciais excisados
contig.txt	Arquivo de dados de sequenciamento de diferentes amostras processados e mapeados para o genoma juntos
dre.mature.dna.fa	Um arquivo FASTA de miRNAs conhecidos de espécies relacionadas
-e	Arquivo de entrada que está no formato FASTQ
-e80	É a soma máxima dos valores de qualidade em cada posição de incompatibilidade
esc0hr.fastq	Arquivo de entrada para análise
esc0hr.reads_collapsed.arf	Arquivo de alinhamento
esc0hr.reads_collapsed.fa	Leituras processadas
expression.html	Um arquivo que oferece uma visão geral de todos os miRNAs
extract_miRNAs.pl	Utilizado para extrair sequências de uma espécie
-f	Designa um arquivo FASTA como entrada
genome.fa	Um arquivo FASTA com o genoma de referência
-h	Analisar dado para o formato FASTA

hairpin.fa Contém sequências de formato FASTA de todos os *hairpins* de miRNA

html Arquivo *html* com uma visão geral de todos os miRNAs detectados nos dados de entrada de sequenciamento profundo

-i Converter RNA em letras de DNA para mapear contra o genoma

-j Remover todas as entradas que tem uma sequência que contém letras diferentes de a, c, g, t, u, n, A, C, G, T, U e N

-l Descartar *reads* menores que 18 bp

l18 Opção que corresponde aos primeiros 18 nt de uma sequência de leitura

lpa.hairpin.dna.fa Um arquivo FASTA com precursores de miRNA conhecidos das espécies em estudo

lpu.hairpin.dna.fa Um arquivo FASTA com precursores

lpu.mature.dna.fa Um arquivo FASTA com miRNAs conhecidos das espécies em estudo

lpg1 Arquivo de índice do genoma

lpg1.fa Um arquivo FASTA do genoma correspondente

-m Colapsar leituras

-m5 Opção que mantém apenas leituras que não mapeiam mais de cinco vezes contra o genoma

mature.fa Arquivo que contém sequências de formato FASTA de todas as sequências de miRNA maduras

-m mature.fa Sequências de miRNA maduras

mature_ref_this.fa Um arquivo FASTA com os miRNAs de referência maduros pertencente a espécie em estudo provenientes do miRBase

mature_ref_other.fa Um arquivo FASTA com os miRNAs de referência maduros das espécies relacionadas do miRBase

mirBase.mrd Um arquivo de assinatura

miRNAs_expressed_all_samples.csv Um arquivo separado por tabulação com identificadores de miRNA e sua contagem de *reads*

-n Sobrescrever arquivos existentes

-norc Aconselha o *bowtie* a não mapear leituras para o complemento reverso das sequências precursoras no índice de *bowtie*

-n0 Opção que mantém apenas os alinhamentos com nenhuma incompatibilidades na região semente de uma leitura mapeada contra o genoma

-p Mapear para o genoma, deve ser indexado por *bowtie-build*

pdfs Um diretório denominado *pdfs* que contém arquivos PDF

-p precursor.fa Sequências de precursores de miRNA

precursor_ref_this.fa Um arquivo FASTA com os miRNAs de referência precursores das espécies em estudo depositados no miRBase

-q Arquivo *.mrd*

reads.fa Um arquivo rápido com leituras de sequenciamento profundo

report.log Armazenamento da saída de progresso

results.html Arquivo que usa um navegador de internet para visualizar os resultados

rna2dna.pl Substitui u e U por T

-r reads.fa Pequenas sequências de RNA-Seq

-s Opção que designa os miRNAs maduros de referência no arquivo de formato FASTA como entrada

sanity_check_genome.pl Verifica o arquivo FASTA do genoma quanto à exatidão

script quantifier.pl Mapeia as leituras de sequenciamento profundo para os precursores de miRNA e determina os níveis de expressão dos miRNAs correspondentes

script mapper.pl	Execução do processamento da saída da sequência bruta da plataforma Illumina e mapeia as leituras para o genoma de referência
script mirDeep2.pl	Identifica o miRNA
-t	Gravar mapeamentos de leitura para um arquivo
-v	Gerar relatório de progresso
-v1	Relata os mapeamentos lidos com até uma incompatibilidade
-v50	Opção que mantém todos os precursores que têm uma pontuação acima de 50
-y	Opção que fornece um arquivo adicional com valores *randfold*.

SUMÁRIO

1 INTRODUÇÃO

1.1 PROBLEMAS DA PESQUISA E HIPÓTESES

- O uso de ferramentas de Bioinformática para análise de dados provenientes de Sequenciamento de Nova Geração podem trazer informações úteis para os pesquisadores?

Para saber quais os transcritos estão sendo expressos em um determinado momento é preciso utilizar de técnicas de Sequenciamento de Nova Geração (NGS), e com isso prosseguir os estudos para saber qual processo biológico um transcrito pode influenciar (Bartel, 2004; Lai, et al., 2003; Baskerville, Bartel, 2005; Lee, et al., 1993; Reinhart, et al., 2000; Reinhart, et al., 2002; apud CAI, et al., 2009; ORANG, SAFARALIZADEH, BAVILI, 2014).

A bioinformática contemporânea é amplamente utilizada como uma alternativa aos experimentos tradicionais, gerando dados que não podem ser obtidos por meio de experimentos de bancada. Ela tem a capacidade de analisar grandes conjuntos de dados nas escalas genômicas em tempo relativamente curto, e é capaz de prever genes de interesse que existem na natureza (WU, 2018).

Menon e colaboradores (2021) realizaram uma revisão onde um dos focos principais foram na análise de dados de sequenciamento de ScRNA por *pipelines* (ou fluxos de trabalhos) computacionais. Ocorreu um rápido desenvolvimento de ferramentas de NGS e tecnologias. Com isso percepções valiosas foram obtidas no sistemas biológicos complexos unidos a uma ampla gama de comunidades microbianas. O sequenciamento de ScRNA pode separar sistemas biológicos complexos de populações celulares raras, revelando o gene regulador, e traçar a progressão de linhagens celulares individuais

durante o desenvolvimento celular. Os pesquisadores concluíram que tecnologias de sequenciamento facilitaram muito o melhoramento biológico molecular devido à disponibilidade de ferramentas de bioinformática que são poderosas para enfrentar o problema de tempo de mapeamento e podem analisar informações biológicas, moleculares, celulares e genômicas.

A bioinformática é uma parte importante da ciência das últimas décadas que tem um enorme impacto sobre pesquisa biológica, pois possibilita a análise *in silico* de dados biológicos coletados, arquivados e que podem ser recuperados. Houve aumento de estudos de sequenciamento de genes envolvendo dados massivos e a bioinformática manifestou-se para desempenhar um papel maior na produção e aplicações da pesquisa. Na contemporaneidade, a biotecnologia, a biologia molecular, a medicina, a agricultura são exemplos de áreas que utilizam da bioinformática para melhor planejamento da pesquisa. Existem ferramentas, *softwares* e banco de dados que são utilizados para uma melhor compreensão e análise dos resultados, podendo estar localizados na web e disponibilizados também por meio de *softwares* ou algoritmos, melhores para analisar sequências em grande escala. O NGS possibilitou sequenciamento de alto rendimento de muitos genomas de seres de importância diversa e seus dados também estão sendo usados para abordar as mudanças climáticas causadas por mudanças abióticas (BARIK, et al., 2020).

O NGS permitiu a fácil identificação e mapeamento de mutações causais em um curto período e a um custo relativamente baixo. Identificar as mutações genéticas e os genes que estão relacionados às mudanças fenotípicas é essencial para a compreensão de uma ampla variedade de funções biológicas. Para ter sucesso na interpretação dos dados NGS, os métodos computacionais de bioinformática são elementos críticos na entrega de montagem precisa, alinhamento e detecção de variantes (SAHU, et al., 2020).

- É possível obter miRNAs conservados com o uso de ferramentas computacionais ligadas à área da Bioinformática, tal como o mirDeep2?

O objetivo do trabalho de Schmidt e colaboradores (2018) foi traçar o perfil de miRNAs na placenta de macacos, hipotetizando que a expressão é conservada entre o macaco e a placenta humana. Os primatas são modelos para estudar a biologia da gravidez humana. Um total de 1.064 sequências foram expressas em uma amostra, onde 607 sequências alinhadas com pré-miRNAs de macaco *rhesus* conhecidos e 457 novas sequências de miRNA candidatas foram previstas por meio de análise miRDeep2.

Dang e colaboradores (2019) utilizaram o sequenciamento de alto rendimento para investigar a população de miRNA em bactérias do gênero *Thalassospira* usando o miRDeep2. As bactérias desse gênero estão envolvidas na biodegradação de uma variedade de hidrocarbonetos. O *software* miRDeep2 foi utilizado para prever miRNAs com base em uma investigação da estrutura secundária das sequências precursoras de miRNA e sua interação com a Dicer, fornecendo os precursores e sequências de miRNA maduras. Usando miRDeep2, 86 prováveis sequências de premiRNA puderam ser identificadas. Os miRNAs maduros também foram identificados dentro de cada precursor.

- É possível utilizar ferramentas de Bioinformática web para buscar informações sobre dados oriundos de NGS?

O SRA (*Sequence Read Archive* ou Arquivo de leitura de sequência), é um recurso de arquivo público internacional para dados de sequência de próxima geração estabelecido sob a orientação do INSDC (*International Nucleotide Sequence Database Collaboration* ou Colaboração Internacional de Banco de Dados de Sequência de Nucleotídeos). A combinação significativamente de menor custo e maior velocidade de sequenciamento resultou em um crescimento explosivo de dados enviados para o arquivo de dados de sequência de próxima geração principal, o SRA. A preservação de dados experimentais é uma parte importante do registro científico, e um número crescente de periódicos e agências de financiamento exige que os dados da sequência de próxima

geração sejam depositados no SRA. Os parceiros do INSDC incluem o Centro Nacional de Informações sobre Biotecnologia (NCBI), o Instituto Europeu de Bioinformática (EBI) e o Banco de Dados de DNA do Japão (DDBJ). As leituras estão disponíveis para *download*. O objetivo na criação do SRA era ser um repositório de dados da fase primária de análise de sequenciamento (LEINONEN, et al., 2010).

Concomitante aos estudos de identificação de genes, houve trabalhos para prever e validar alvos de genes miRNA. O banco de dados miRBase foi criado e nele consta a nomenclatura de genes bem como suas funções, com o primeiro *pipeline* automatizado para prever genes alvo de miRNA. O miRBase contém sequências de miRNAs maduras publicadas, juntamente com seus precursores *hairpin* de origem previstos, anotações relacionadas à sua descoberta, estrutura e função. Os miRNAs maduros verificados experimentalmente são anotados com o método experimental usado para a descoberta. O banco de dados também contém sequências homólogas previstas de miRNAs verificados em um organismo relacionado (GRIFFITHS-JONES, et al., 2006).

Os miRNAs conservados podem ser agrupados em famílias, cada uma com uma região de semente exclusiva. Além disso, muitas interações não conservadas também funcionam para atuar em mRNAs e diminuir a produção de proteínas. Um ponto crítico para a compreensão da biologia do miRNA é a previsão das interações miRNA-alvo. Pesquisadores desenvolveram um modelo estatístico que prevê os efeitos da ligação de microRNAs. O modelo foi usado para alimentar a versão mais recente de um recurso disponível gratuitamente chamado TargetScan que pode ser um recurso valioso para pesquisadores que investigam os muitos papéis importantes dos microRNAs no controle da produção de proteínas (AGARWAL, et al., 2015).

Para caracterização das funções de miRNAs, Chen e Wang (2019) desenvolveram um banco de dados online, miRDB, para miRNA previsão de destino e anotações funcionais. O algoritmo proposto pelos autores foi aprimorado para a predição de alvo de miRNA, permitindo também a previsão de alvo personalizado com sequências fornecidas

pelo usuário. O miRDB agora hospeda os perfis de expressão de muitas linhagens celulares.

1.2 OBJETIVO GERAL

Aplicar ferramentas computacionais de Bioinformática em dados oriundos de *Next Generation Sequencing* na busca por informações importantes que possam contribuir para o avanço da ciência, utilizando somente plataforma *in silico*.

1.2.1 Objetivos específicos

- Selecionar miRNAs conservados em dados de NGS utilizando a ferramenta computacional mirDeep2
- Construir material de apoio sobre o mirDeep2
- Construir conhecimento utilizando ferramentas computacionais web a partir dos arquivos SRR provenientes de NGS

1.3 JUSTIFICATIVA

O que impulsionou a realização deste trabalho foi a necessidade de entender melhor quais e quantos miRNAs provenientes de uma grande quantidade de dados de NGS estão envolvidos nas vias de funcionamento de qualquer organismo em diferentes condições biológicas, sugerindo assim as identidades dos miRNAs que estão influenciando uma determinada situação consequentemente melhorando a anotação dos microRNAs.

Em 2012 foi lançado o *software* mirDeep2 que poderia identificar miRNAs novos e outros previamente validados dos organismos, realizar a quantificação e distinção dos mesmos quando comparados aos demais pequenos RNAs.

Como a identificação experimental é um processo demorado e de uso intensivo de recursos, isso levou os pesquisadores a usar a previsão computacional de miRNAs com base na informação da sequência genômica. Com o avanço das tecnologias de *microarray* e RNAseq, muitos estudos de expressão gênica foram realizados nos últimos anos. Os dados de expressão gênica mostraram promessas significativas em desvendar o comportamento dos genes sob condições biológicas variadas. Os dados de expressão gênica tem sido intensamente estudados para identificar o efeito de miRNAs (BANWAIT, BASTOLA, 2015).

Este estudo surgiu como um desdobramento da pesquisa da Christina Martins e colaboradores, na parte em que trata da análise dos miRNAs envolvidos no câncer de próstata.

Algumas importâncias de responder às questões de pesquisa deste trabalho e atingir os objetivos propostos são facilitar a leitura de algumas informações de amostras, de pessoas com câncer de próstata junto com o grupo controle, provenientes de uma análise de NGS; e com a seleção de miRNAs conservados verificar em quais vias eles provavelmente irão atuar primeiro de acordo com a ordem mostrada pelas ferramentas computacionais web.

2 REFERENCIAL TEÓRICO

2.1 A HISTÓRIA DO NGS E COMO ELE FUNCIONA

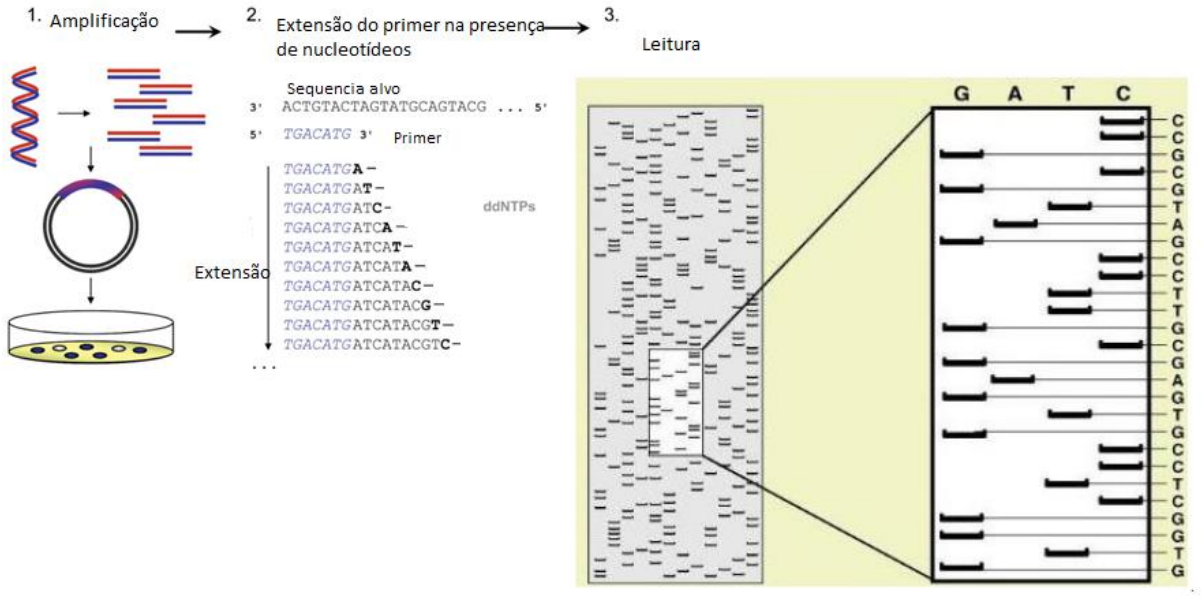
Os primeiros métodos de sequenciamento foram criados nos anos 70. Um dos mais importantes métodos criados foi o procedimento de Maxam e Gilbert que identifica a sequência das bases através da clivagem com produtos químicos para alteração de bases da ponta 5’ das sequências de DNA marcadas radioativamente. A metodologia de Sanger (Figura 1) e Coulon identifica a sequência de nucleotídeos através da síntese de uma fita simples de DNA com auxílio da DNA polimerase e dos didesoxinucleotídeos (ddNTPs). O uso da DNA polimerase é o diferencial entre os métodos (KIRCHER, KELSO, 2010; GONÇALVES, 2019; MARTINS, 2019).

Em 1977 o genoma de um bacteriófago foi completamente sequenciado pelo bioquímico Sanger. O sequenciamento de DNA era um processo pouco automatizado e muito tedioso, que envolvia a determinação de apenas algumas centenas de nucleotídeos de cada vez. A síntese da cadeia reversa é realizada nestas cópias usando uma sequência de iniciação conhecida, a montante da sequência a ser determinada e uma mistura de desoxirribonucleotídeos fosfatados (dNTP's) e didesoxinucleotídeos (ddNTP’s); que são nucleotídeos modificados que não possuem o grupo OH, ou grupo hidroxila constituído pelos átomos de hidrogênio e oxigênio, livre no carbono 3’ da pentose. Essa mistura causa a terminação não reversível da reação de extensão. Quando os ddNTP’s tentam se ligar a fita de DNA, na ausência do OH, o próximo nucleotídeo não tem onde se estabelecer e a replicação é finalizada. Após desnaturação e remoção de nucleotídeos livres, *primers* e enzima, as moléculas resultantes são classificadas pelo seu peso molecular, e o marcador anexado aos didesoxinucleotídeos de terminação é lida sequencialmente na ordem na qual está disposta (KIRCHER, KELSO, 2010; GONÇALVES, 2019; MARTINS, 2019).

A classificação por peso molecular foi originalmente realizada usando eletroforese em gel e marcação por composto radioativo em todos nucleotídeos sem diferenciação e a sequência era anotada manualmente com auxílio de uma régua onde se tinha escrito os nucleotídeos. No final dos anos 80, os sequenciadores semiautomáticos com maior rendimento tornaram-se disponíveis. Com isso foi substituída a marcação radioativa por

corantes em didesoxinucleotídeos que emitiam fluorescência quando estimulados por um comprimento de onda emitido na amostra com auxílio de um sequenciador acoplado a um computador, possibilitando as reações acontecerem no mesmo tubo. Um avanço no início dos anos 90 foi o desenvolvimento da eletroforese capilar, onde ocorria a migração de espécies carregadas eletricamente em um tubo capilar. A marcação fluorescente é lida após a emissão de um laser onde se projeta uma luz que é captada por um detector que transmite para um computador que a converte em picos coloridos diferenciando os quatro nucleotídeos. As novas tecnologias automatizadas possibilitaram o surgimento de novos métodos que o superam em questão de rendimento, reduzindo o custo de sequenciamento (KIRCHER, KELSO, 2010; GONÇALVES, 2019; MARTINS, 2019).

Figura 1- Sequenciamento de Sanger



Fonte: Adaptado de KIRCHER, KELSO, 2010; NAZARÉ, GONÇALVES, CAPELO, 2019. Representação esquemática do processo de sequenciamento Sanger. O DNA de entrada é fragmentado e clonado em vetores bacterianos para amplificação *in vivo*. A síntese da cadeia complementar é realizada nas cópias obtidas a partir de uma sequência de *primer* conhecida e usando uma mistura de desoxi-nucleotídeos (dNTPs) e didesoxinucleotídeos (ddNTPs). A mistura dNTP / ddNTP faz com que a extensão seja de forma

não reversível, criando moléculas estendidas de forma diferente. Subsequentemente, após a desnaturação, limpeza de nucleotídeos livres, *primers* e a enzima, as moléculas resultantes são classificadas pelo seu peso molecular através de eletroforese em gel.

Atualmente, as técnicas de sequenciamento de última geração (NGS) são utilizadas em larga escala para a geração de dados genômicos, produzindo quantidades cada vez maiores de informações com baixo custo (DAVEY, et al., 2011; ESCALONA, ROCHA, POSADA, 2016).

O NGS refere-se às plataformas ou métodos, de gerações diferentes, para sequenciar utilizando processos em locais onde ocorrem os sequenciamentos. Os métodos de sequenciamento de segunda geração são caracterizados pela preparação das bibliotecas antes de se realizar o sequenciamento dos clones de DNA amplificados. Já o sequenciamento molecular de terceira geração pode ser feito sem a necessidade de se criar bibliotecas de amplificação (KULSKI, 2016).

Essas técnicas permitem sequenciar o DNA e o RNA muito rapidamente, facilitando a aquisição de grandes conjuntos genômicos, transcriptômicos, de interação DNA-proteína e de dados epigenômicos. Os dados obtidos podem se tornar marcadores em quase qualquer genoma de interesse em uma única etapa, mesmo em populações em que pouca ou nenhuma informação genética está disponível (DAVEY, et al., 2011; ESCALONA, ROCHA, POSADA, 2016).

Embora haja vantagens nos métodos NGS, eles também oferecem taxas de erro mais altas devido aos seus comprimentos de leitura serem mais curtos (ESCALONA, M., ROCHA, S., POSADA, D., 2016). As tecnologias NGS mais populares são o sequenciamento da Illumina por síntese, o pirosequenciamento 454 da Roche, o sequenciamento "*Thermo Fisher*" por ligação e detecção de oligonucleotídeos (SOLiD), o sequenciamento de semicondutores "*IonTorrent*" da "*Thermo Fisher*", o sequenciamento simples em tempo real (SMRT) da "*Pacific Biosciences*" (PacBio) e sequenciamento de

cadeias de modelo de DNA de célula única da "*Oxford Nanopore Technologies*" (ESCALONA, ROCHA, POSADA, 2016).

Essas estratégias podem diferir, por exemplo, em relação ao tipo de leitura que produzem ou ao tipo de erros de sequenciamento que introduzem. Apenas duas das tecnologias atuais, Illumina e SOLiD*,* são capazes de produzir todos os três tipos de leitura de sequenciamento que são a extremidade única, extremidade emparelhada e *mate pair* (ESCALONA, ROCHA, POSADA, 2016).

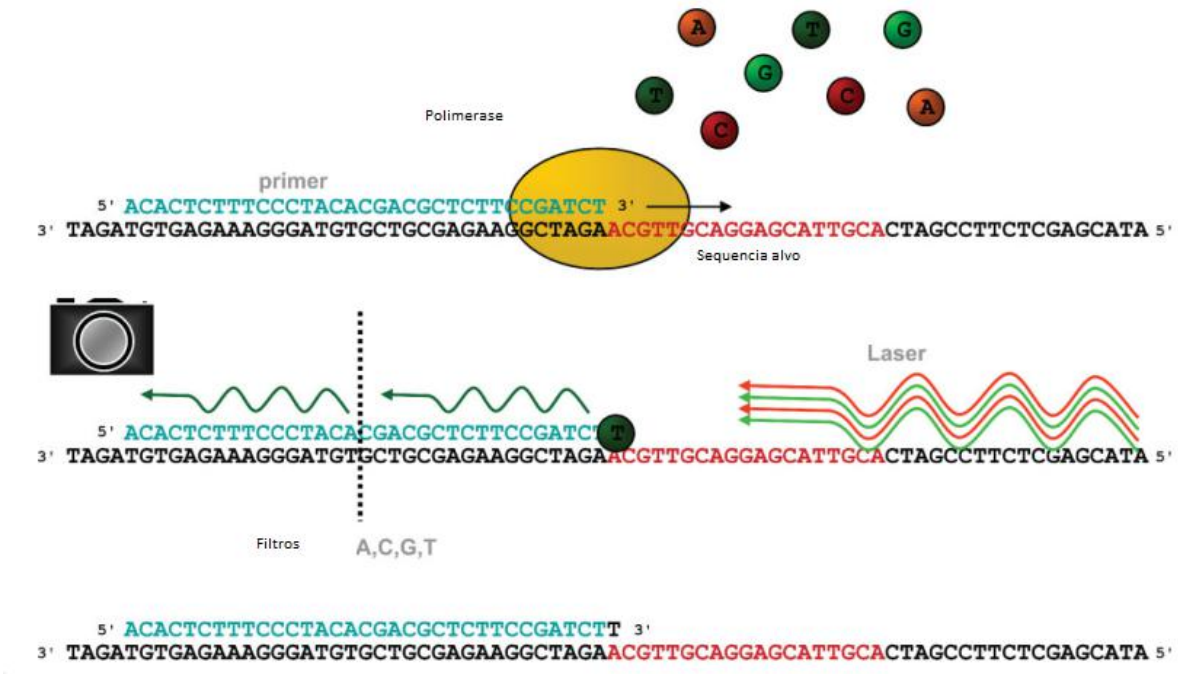
O comprimento de leitura também depende da máquina e do kit utilizado. Em plataformas como Illumina, SOLiD ou IonTorrent, é possível especificar o número de pares de bases desejados por leitura, sendo possível obter leituras com comprimentos máximos de 300 pb, 75 pb ou 400 pb, respectivamente. Por outro lado, em plataformas como 454, Nanopore ou PacBio, são fornecidas informações apenas sobre os comprimentos médios e máximos de leitura que podem ser obtidos, com comprimentos médios de 700 pb, 10 kb, e 15 kb e comprimentos máximos de 1 kb, 10 kb e 15 kb, por essa ordem. As taxas de erro variam dependendo da plataforma, de menor ou igual a 1% no Illumina variando até 30% no Nanopore (ESCALONA, ROCHA, POSADA, 2016).

A plataforma 454 foi a primeira dos novos sequenciamentos de última geração (lançada em outubro de 2005). Baseia-se na abordagem de pirosequenciamento com início do desenvolvimento em 1996 por Nyre e Mostafa Ronaghi no "*Royal Institute of Technology*". Em contraste com a tecnologia Sanger, o pirosequenciamento baseia-se na complementariedade iterativa de cadeias simples e, simultaneamente, na leitura do sinal emitido a partir do nucleotídeo a ser incorporado (também denominado sequenciamento por síntese). A eletroforese em gel não é mais necessária para gerar uma leitura ordenada dos nucleotídeos, já que a leitura é feita simultaneamente com a extensão da sequência. No processo de pirosequenciamento, um nucleotídeo é aproximado às várias cópias da sequência a ser determinada, fazendo com que as polimerases incorporem o nucleotídeo se ele for complementar à fita molde. A incorporação encerra se o maior trecho possível de

nucleotídeos complementares tenha sido sintetizado pela polimerase. No processo de incorporação, um pirofosfato por nucleotídeo é liberado e convertido em ATP . O ATP conduz a reação de luz e o sinal luminoso emitido é medido. Os nucleotídeos de desoxirribose padrão são utilizados. Após capturar a intensidade da luz, os nucleotídeos não incorporados são lavados e o ciclo reinicia (KIRCHER, KELSO, 2010).

A tecnologia de terminação reversível usada pela Illumina (Figura 2) emprega um conceito de sequenciamento por síntese que é semelhante ao usado no sequenciamento de Sanger. A base incorporada é corada com um componente fluorescente onde este sinal é lido por um computador (KIRCHER, KELSO, 2010).

Figura 2- Illumina



Fonte: Adaptado de KIRCHER, KELSO, 2010. Os adaptadores são ligados quimicamente aos terminadores da sequência e posteriormente fixados em uma base. *Primers* de sequenciamento são conduzidos para os adaptadores das sequências. Polimerases são utilizadas para estender o sequenciamento a partir dos *primers* por incorporação de nucleotídeos marcados com fluorescência. Posteriormente, as polimerases e os nucleotídeos livres são lavados. A sequência de bases incorporadas é lida com auxílio de diferentes filtros e

também com dois lasers diferentes (vermelho: para bases A e C; verde: para as bases G e T para iluminar os fluoróforos. Subsequentemente, os fluoróforos terminadores são removidos e o sequenciamento continua com a incorporação da próxima base.

O protótipo da plataforma de sequenciamento SOLiD que foi vendido pela *Life Technologies Applied Biosystems* (ABI), foi desenvolvido pela *Harvard Medical School* e pelo *Howard Hughes Medical Institute* e publicado em 2005. A reação de extensão de sequência não é realizada por polimerases, mas sim por ligases. Um iniciador é hibridizado para as cadeias simples da biblioteca a serem sequenciadas. Uma mistura de sondas contendo quatro marcadores fluorescentes distintos competem por ligação ao iniciador de sequenciamento. Após a clivagem de bases, um fosfato na 5' fica livre no *primer* estendido disponibilizando outra ligação. Após 10 ciclos de múltiplas ligações, as cadeias sintetizadas são fundidas e o produto de ligação é lavado. A partir do novo iniciador de sequenciamento, a reação de ligação é repetida. As leituras coradas podem ser convertidas em uma sequência. Essa conversão requer uma primeira base conhecida, que é a última base da sequência do adaptador da biblioteca utilizada. Este sistema de codificação permite a detecção de erros da máquina e a aplicação de uma correção. Em contraste com a tecnologia 454 da Roche, o sistema SOLiD não utiliza uma placa para fixação das esferas no processo de sequenciamento; em vez disso, as extremidades das sequências são modificadas de uma maneira que permite que elas sejam ligadas covalentemente a uma lâmina de vidro (KIRCHER, KELSO, 2010).

O *IonTorrent* é parecido com o pirosequenciamento pois utiliza microchip para detectar a incorporação do dNTP na sequência. Há adição de marcadores, dNTPs, DNA fragmentado, e cada fragmento liga-se a um *bead*. Todos esses elementos são posicionados em poços para replicação. Caso o dNTP ligue, um hidrogênio é liberado, ocasionando variação do pH que é detectada pelo chip (NAZARÉ, GONÇALVES, CAPELO, 2019).

O PacBio é um sistema baseado em uma nova tecnologia de sequenciamento de molécula única em tempo real (SMRT). Essa tecnologia possibilita a observação em tempo real da síntese de DNA através da enzima DNA polimerase, garantindo a esse método uma precisão superior a 99,999%. Além disso, permite que esse método possua uma faixa de leitura superior a qualquer outra tecnologia, com média de aproximadamente 8000 pares de bases e que todo o processo seja feito de forma rápida e simples. A biblioteca do DNA molde é composta por fragmentos de DNA dupla fita conectados à adaptadores em suas terminações. Esses fragmentos com adaptadores são chamados de "*SMRTbells*". Estes adaptadores serão capazes de transformar os fragmentos de DNA de fita dupla em moldes circulares, nos quais a enzima polimerase vai continuar a funcionar até que se inative. É realizada uma corrida com múltiplas passagens em torno desse molde circular, permitindo a condensação numa sequência consenso de maior precisão. Como desvantagem pode-se citar o grau de suscetibilidade a altas taxas de erro (PINTO, KREMER, 2019).

A "*Oxford Nanopore Technologies*" apresentou o seu próprio método para o sequenciamento. Incorporaram uma proteína cíclica em uma membrana de polímero sintético dentro do microchip. A cadeia de DNA a ser sequenciado passa pelo poro de proteína e, quando cada nucleotídeo atravessa, é registrado como um sinal elétrico diferente detectado pela máquina (SOARES, et al., 2019).

2.2 MICRORNAS

O genoma é o conjunto de genes de um organismo. Um gene pode transcrever RNAs funcionais que por sua vez podem traduzir proteínas ou participar de outros processos celulares. Como os transcritos de um genoma participam de diversos processos biológicos para o funcionamento de um organismo, a compreensão de como esse grupo de produtos gênicos está atuando é muito importante. As técnicas de sequenciamento

ajudam ao mostrar quais transcritos que estão sendo expressos naquele momento (Bartel, 2004; Lai, et al., 2003; Baskerville, Bartel, 2005; Lee, et al., 1993; Reinhart, et al., 2000; Reinhart, et al., 2002; apud CAI, et al., 2009; ORANG, SAFARALIZADEH, BAVILI, 2014).

MicroRNAs são um subconjunto de moléculas pequenas de RNAs não codificantes, com aproximadamente 22 nucleotídeos, considerados reguladores pós-transcricional que podem clivar RNAs mensageiros (mRNAs) alvo, reprimem sua tradução e levam à diminuição da estabilidade do mRNA. Estima-se que os miRNAs constituam quase 1% de todos os genes previstos em nematóides, moscas e mamíferos. A importância dos miRNAs foi negligenciada no passado devido à limitação de tecnologia e metodologia até sua descoberta inicial de que dois miRNAs eram responsáveis por controlar o tempo de desenvolvimento do nematóide (*Caenorbabditis elegans*) através do pareamento de bases incompleto dos mRNAs alvo para reprimir sua tradução. Pouco depois detectaram a existência de miRNAs em *Arabidopsis thaliana*, sugerindo que esse tipo de RNA não-codificador surgiu precocemente na evolução eucariótica. (Bartel, 2004; Lai, et al., 2003; Baskerville, Bartel, 2005; Lee, et al., 1993; Reinhart, et al., 2000; Reinhart, et al., 2002; apud CAI, et al., 2009; ORANG, SAFARALIZADEH, BAVILI, 2014).

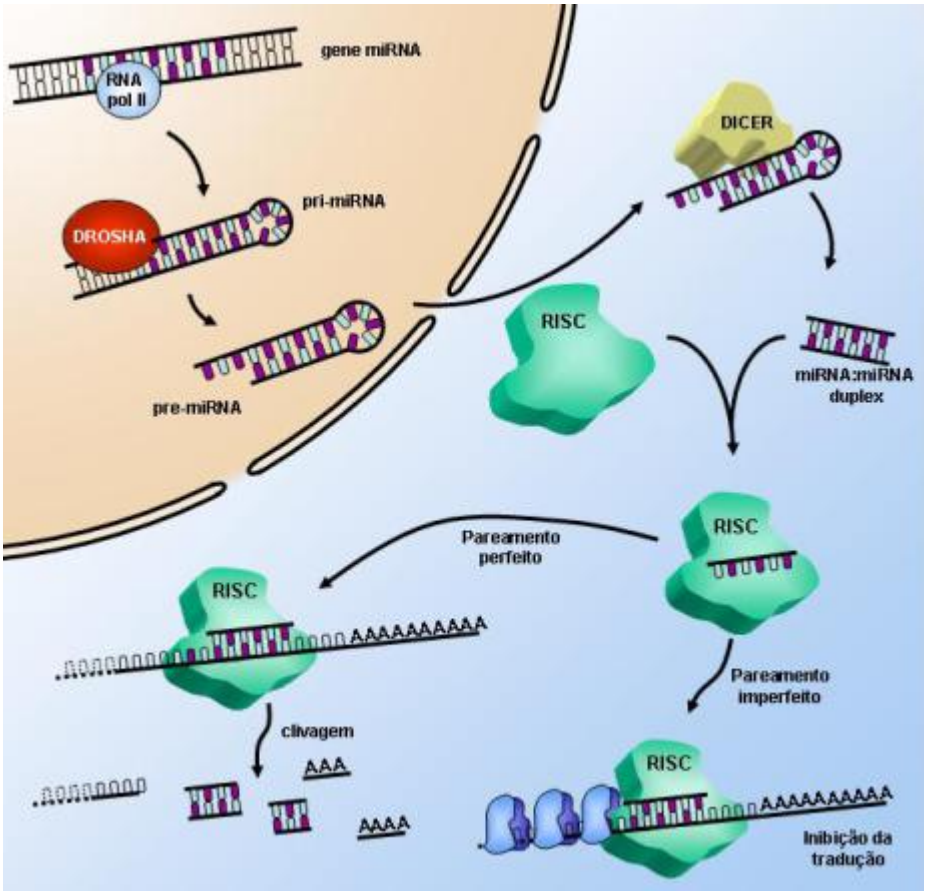
Ao combinar conjuntos de genoma atualizados e banco de dados de etiquetas de sequências expressas, ficou provado que os promotores de miRNAs podem ser divididos em duas grandes categorias: intergênicas e intragênicas. Os miRNAs do Reino Animalia intergênicos são transcritos principalmente por RNA pol III e estão localizados entre os genes; suas transcrições são independentes dos genes codificados. Portanto, miRNAs intergênicos foram relatados como mais conservados evolutivamente. Como miRNAs intragênicos são incorporados nas sequências exônicas ou intrônicas de genes codificadores de proteínas pode ocorrer a co-expressão na mesma orientação de seus genes hospedeiros pela RNA pol II (ORANG, SAFARALIZADEH, BAVILI, 2014). A RNA pol II é a mais atuante no Reino Plantae.

A biogênese dos miRNAs se inicia com a transcrição dependente de RNApol II ou III de um locus de gene miRNA gerando um longo RNA primário (ISSABEKOVA, BERILLO, REGNIER, 2012; GODNIC, et al., 2013; OZSOLAK, et al., 2008; SCHANEN, LI, 2011; LEE, et al., 2011; CAI, YU, HU, YU, 2009; apud ORANG, SAFARALIZADEH, BAVILI, 2014; ESTEVES, 2016; GULYAEVA, KUSHLINSKIY, 2016). Estes RNAs primários (pri-miRNAs) podem ser codificados em diferentes regiões como as intergênicas, exônicas e até de regiões intrônicas. Pri-miRNAs têm uma capa no 5'-7- metilguanosina e é poliadenilado na região 3' e têm uma capacidade de codificação para um ou mais miRNAs maduros, que é representado por uma estrutura em grampo (*hairpin*). Eles podem atingir várias quilobases de comprimento. Pri-miRNAs podem ser reconhecidos por um complexo microprocessador que contém uma proteína de ligação ao RNA DGCR8 e a RNase III Drosha. Este complexo corta o pri-miRNA e gera um precursor menor, o pré-miRNA de 60 à 70 nucleotídeos em animais (Figura 3) e em plantas (Figura 4) (ISSABEKOVA, BERILLO, REGNIER, 2012; GODNIC, et al., 2013; OZSOLAK, et al., 2008; SCHANEN, LI, 2011; LEE, et al., 2011; CAI, YU, HU, YU, 2009; apud ORANG, SAFARALIZADEH, BAVILI, 2014; ESTEVES, 2016; GULYAEVA, KUSHLINSKIY, 2016). Em plantas o pré-miRNA tem uma estrutura secundária característica, formando um *hairpin* através de uma complementaridade imperfeita, contendo 200-300 nucleotídeos de comprimento (CARDOSO, et al., 2018). O *hairpin* é representado por duas sequências em parte complementares entre si com espirais no meio das sequências e uma sequência de fita simples que tem a forma de uma alça (*loop*). Após o processamento nuclear, os pre-miRNAs são transportados pela Exportina-5 e seu co-fator Ran-GTP para o citoplasma. No citoplasma, este complexo é reconhecido pela RNase III Dicer, que corta o pré-miRNA em um dúplex de miRNA na forma madura. Isso libera um duplex de miRNA. Uma vez ocorrida a clivagem, o miRNA de cadeia dupla é liberado e integrado ao complexo efetor apropriado (miRISC), que é composto principalmente por proteínas da família Argonauta e proteínas de constituintes glicina e triptofano. O dúplex é desenrolado e uma fita de miRNA é selecionada enquanto a

outra se degrada e, finalmente, o microRNA desempenhará uma função junto com o complexo ao qual foi inserido (ISSABEKOVA, BERILLO, REGNIER, 2012; GODNIC, et al., 2013; OZSOLAK, et al., 2008; SCHANEN, LI, 2011; LEE, et al., 2011; CAI, YU, HU, YU, 2009; apud ORANG, SAFARALIZADEH, BAVILI, 2014; ESTEVES, 2016; GULYAEVA, KUSHLINSKIY, 2016).

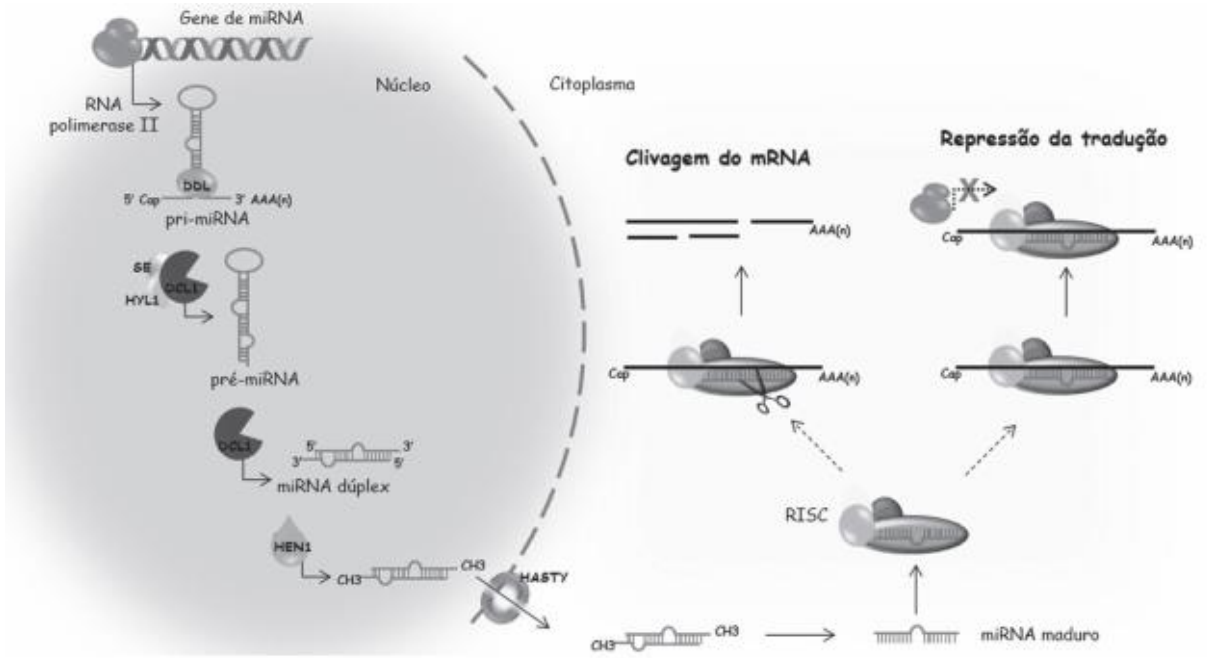
Em plantas, alguns passos são diferentes como uma das proteínas microprocessadoras do complexo que cliva o pri-miRNA não é representada pela Drosha, mas por uma RNase da família da Dicer; e a clivagem do pré-miRNA ocorre no núcleo e o transporte do duplex de miRNA ocorre com o auxílio da proteína HAST. A complementariedade dos miRNAs nem sempre é total, podendo apresentar nucleotídeos não pareados, ou seja, apresentam *mismatches*. As complementariedades mais importantes são entre os nucleotídeos 2 a 7 da ponta 5’ do miRNA, os quais são denominados região semente a qual é a região mais conservada entre organismos de diferentes. O acúmulo de evidências sugere que muitas sequências diferentes podem ser geradas a partir de um único locus de miRNA através do processo de maturação do miRNA. Essas sequências que podem ter extremidades 5′ e 3′ de comprimentos variados são chamadas de isomiRs, atuando na estabilidade do miRNA dentro do complexo RISC ou podem atuar associando-se a mRNAs (PAUL, et al., 2017).

Figura 3- Biogênese miRNA em animais



Fonte: Adaptado de FUZIWARA, 2010. Após a transcrição realizada pela polimerase, o pri-miRNA é clivado pela Drosha no núcleo dando origem ao segundo precursor de microRNA, que por sua vez é transportado para o citoplasma, processado pela Dicer, gerando um duplex de miRNA que é incorporado ao complexo RISC, que posteriormente libera uma das fitas. Quando o pareamento deste novo complexo miRISC é perfeito ocorrerá a degradação do mRNA alvo, quando o pareamento for imperfeito, provocará o impedimento da tradução.

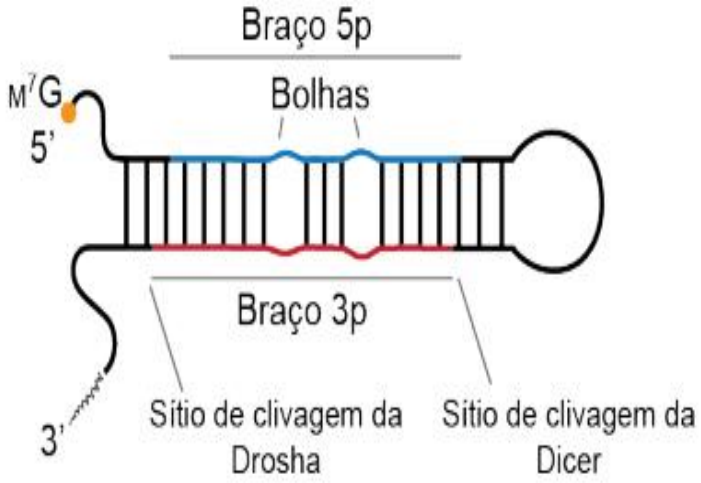
Figura 4- Biogênese miRNA em plantas



Fonte: Adaptado de KULCHESKI, MARGIS, 2017. Liga-se ao pri-miRNA a proteína DAWDLE (DDL), para estabilizar o pri-miRNA até sua transformação em pré-miRNA. O complexo microprocessador é constituído pelas proteínas SERRATE (SE), HYPONASTIC LEAVES 1 (HYL1) e DICER-LIKE 1 (DCL1). O pré-miRNA é levado para o citoplasma pela HASTY que é uma proteína de membrana. O duplex é metilado por HEN1, para proteção contra degradação antes do complexo ser incorporado ao RISC.

O nome de um miRNA é criado com três letras do nome científico do organismo ao qual foi identificado acrescido de hífen com as denominações “mir” caso a sequência seja de um precursor e “miR” indicando uma sequência madura. Posteriormente é incluída uma numeração que funciona como um código temporal de submissão no miRBase (http://www.mirbase.org/). O miRBase é um banco de dados público de miRNAs ou ortólogos. Para miRNAs maduros, há também os sufixos “-5p” ou “-3p” (Figura 5) que são atribuídos devido a origem do miRNA a partir de seu precursor (OLIVEIRA, K. P. V., 2016).

Figura 5- Pri-miRNA



Fonte: OLIVEIRA, 2016. Origem dos miRNAs (representados com as linhas azul e vermelha) a partir de seu precursor primário. Para miRNAs maduros com o sufixo “-5p” quer dizer que este se originou entre a extremidade 5' até o *loop,* e para miRNAs maduros com o sufixo “-3p” quer dizer que este se originou entre o *loop* e a extremidade 3'

2.3 MICRORNAS E NGS

Técnicas de sequenciamento de alto rendimento têm sido amplamente aplicadas em estudos de miRNA, e múltiplas pequenas sequências de RNA são mapeadas contra os pré-miRNAs e sequências do genoma. A sistemática da expressão gênica mediada por miRNA tem sido associada principalmente ao controle pós-transcricional da tradução de proteínas, porém desempenha papel importante na maioria dos processos biológicos incluindo desenvolvimento, proliferação, diferenciação, resposta imune, apoptose, transdução de sinal, desenvolvimento de órgãos, interações hospedeiro viral, adaptação ao estresse e determinação do destino celular bem como associações a vários tipos de doenças incluindo diabetes, doença cardiovascular, doença renal e câncer. Pesquisadores

demonstraram o potencial papel dos miRNAs como biomarcadores não invasivos de uma variedade de doenças (PAUL, et al., 2017).

Várias abordagens de alto rendimento foram estabelecidas para análise de miRNAs. O MiSeq foi o sequenciador de alto rendimento baseado em NGS utilizado por Kluiver e colaboradores (2019) para identificar miRNAs que estão envolvidos no controle do crescimento celular. Ele utiliza a plataforma Illumina, onde a identificação de bases é feita diretamente a partir das medições da intensidade do sinal em cada fase do sequenciamento, minimizando as taxas de erro quando contrastado com outras tecnologias.

Para que ocorra o sequenciamento de microRNAs e seus precursores, o procedimento requer uma separação do restante dos transcritos da amostra, por exemplo com Trizol uma solução química utilizada para extração de DNA, RNA e proteínas, para o próximo passo que é fazer uma biblioteca de cDNA, que se trata de uma construção de cópias reversas. Por transcrição reversa, o cDNA é obtido e através da técnica da reação em cadeia da polimerase (PCR) é amplificado. O RNA total é brevemente exposto à DNAase I livre de RNAase para degradação de DNA. O RNA é transcrito reversamente para cDNA usando um *primer* específico. Para amplificar um segmento de cDNA usando PCR, a amostra é primeiro aquecida para que o cDNA se desnature em dois pedaços de DNA de fita simples. Em seguida, uma enzima chamada Taq polimerase constrói duas novas fitas de DNA, usando as fitas originais como modelos. Esse processo resulta na duplicação do cDNA original, com cada uma das novas moléculas contendo uma velha e uma nova fita de DNA. Em seguida, cada uma dessas fitas podem ser usadas para criar duas novas cópias, e assim por diante. O ciclo de desnaturação e síntese de novo DNA é repetido muitas vezes, levando a muitas cópias exatas do segmento de cDNA original. Logo após, as bibliotecas são sequenciadas e depois de adquirir os dados brutos do processo de sequenciamento do NGS, os dados devem ser analisados através da bioinformática (HU, LAN, MILLER, 2017).

2.4 BIOINFORMÁTICA

A bioinformática inicialmente se referia ao uso de técnicas computacionais, em serviço de vários ramos da ciência para a compreensão de processos biológicos, contribuindo com ferramentas essenciais através do desenvolvimento de algoritmos para controle experimental, aquisição e análise de dados (WOOD, 2019). Mas, recentemente tornou-se um campo que combina várias disciplinas como biologia, estatística e ciência da computação para analisar dados biológicos (WU, 2018).

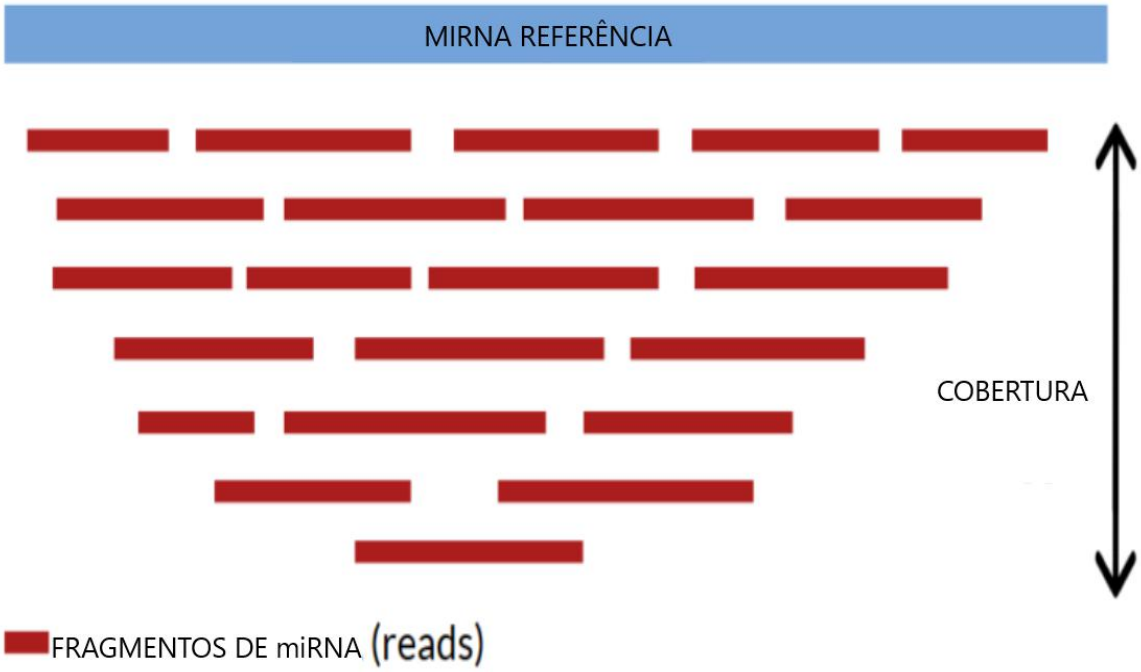
Existem duas abordagens diferentes de bioinformática. A primeira, preconiza o uso de ferramentas que estão na internet e a segunda utiliza ferramentas desenvolvidas para serem utilizadas através de linha de comando. Ferramentas depositadas na internet, não requerem conhecimento de programação de computadores e são imediatamente acessíveis. As ferramentas utilizadas através de linha de comando podem necessitar de um aprendizado prévio, mas quase sempre oferecem mais opções para executar os programas e são mais apropriadas para analisar conjuntos de dados em larga escala. Mesmo para conjuntos de dados menores, as abordagens de linha de comando podem oferecer mais flexibilidade e precisão na realização de suas tarefas (PEVSNER, 2015).

2.5 BIOINFORMÁTICA E MICRORNA

A importância dos miRNAs na saúde ainda não foi totalmente revelada. Um único miRNA pode regular centenas de mRNAs alvo simultaneamente. Isto oferece uma oportunidade emergente de desenvolver uma terapia eficaz baseada em miRNA. Avanços recentes em tecnologias de sequenciamento de última geração permitiram a obtenção de miRNAs expressos em larga escala e baixo custo. O *deep sequencing* (Figura 6) de miRNA é também chamado de miRNA-seq. O miRNA-seq é um método com a capacidade de mostrar um número elevado de leituras únicas de cada região de uma sequência,

proporcionando aos pesquisadores uma oportunidade de catalogar as leituras expressas de miRNA através de vários tecidos e modelos, bem como estudar sua desregulação (ZHAO, et al, 2017).

Figura 6- *Deep sequencing*



Fonte: Adaptado de <http://www.thetgmi.org/wp-content/uploads/2018/04/QSM.png>. Representação de dados de sequenciamento de *reads* (em vermelho) de uma amostra que corresponderam a parte da sequência de um miRNA já validado (em azul).

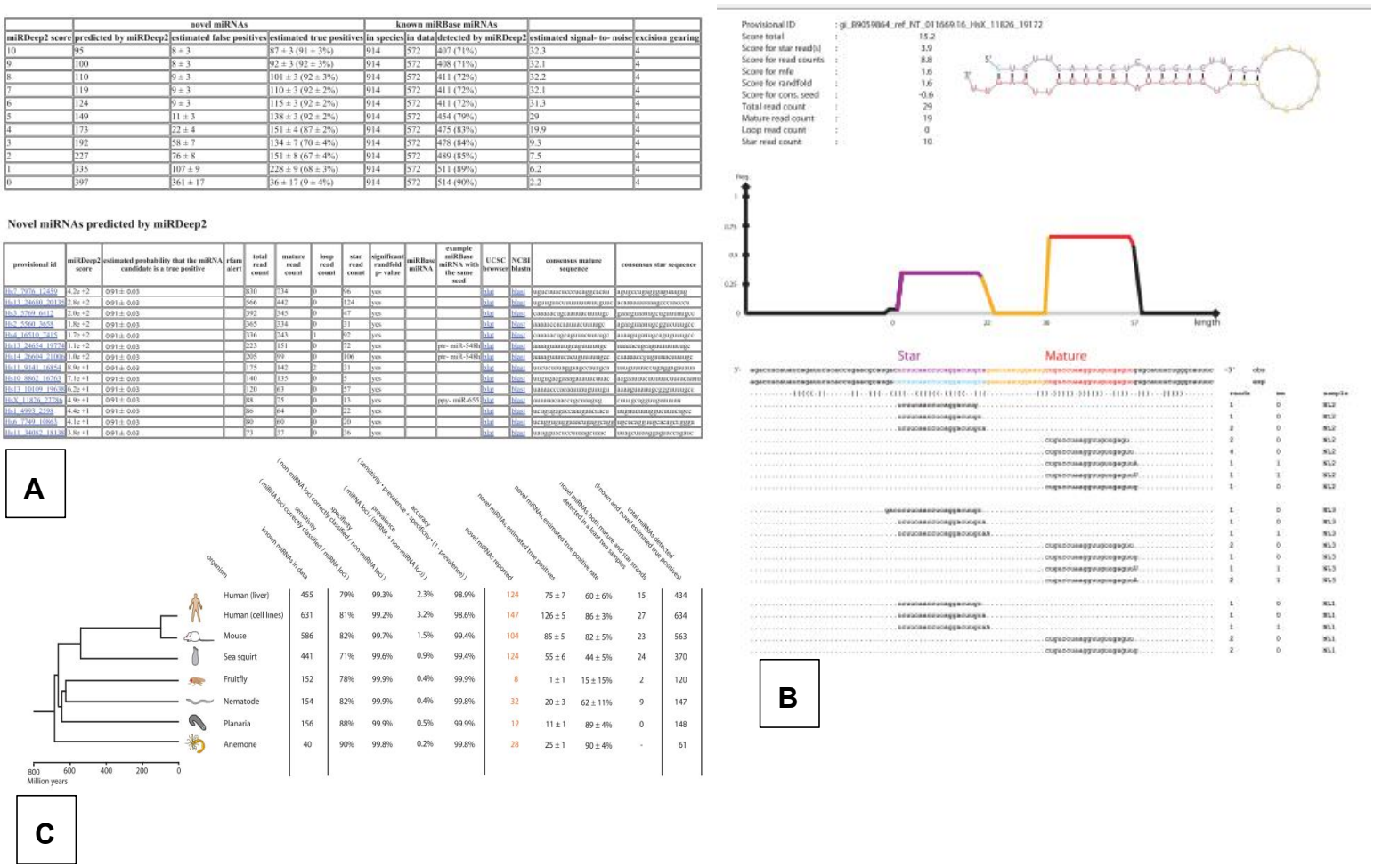
Muitos grupos desenvolveram ferramentas de código aberto para análise de dados de miRNA-seq. Essas ferramentas diferem nos métodos e algoritmos usados para várias etapas de processamento, como ajuste de adaptador e alinhamento de sequência. Apesar da disponibilidade dessas ferramentas, muitos desafios da bioinformática permanecem. Como, por exemplo, quando uma sequência de leitura pode mapear para mais de um miRNA. Esse problema se torna mais grave quando as leituras de miRNA-seq derivadas de múltiplos precursores são diretamente alinhadas a um genoma de referência. Portanto, para quantificação precisa de miRNA, é muito importante introduzir estratégias

computacionais para reduzir ou minimizar mapeamentos potencialmente falsos. Análises de conjuntos de dados miRNA-seq normalmente geram grandes quantidades de dados e uma variedade de arquivos de resultados que são difíceis de interpretar. Portanto, é crucialmente importante organizar e compartilhar os resultados da análise de dados miRNA-seq de uma maneira eficiente e amigável (ZHAO, et al., 2017).

As análises bioinformáticas envolvem arquivos que conduzem a uma série de transformações, chamadas de *pipeline* ou fluxo de trabalho. Normalmente, essas transformações são feitas por linhas de comando de *softwares* criados para sistemas operacionais compatíveis com Unix. Através do *Next Generation Sequencing*, no qual milhões de sequências curtas de DNA são usadas como fonte de entrada para interpretar uma série de fenômenos biológicos, intensificou-se a necessidade de *pipelines* mais robustos. As análises de NGS tendem a envolver etapas como alinhamento de sequência e anotação genômica que levam muito tempo para um grande número de parâmetros de entrada (LEIPZIG, 2017).

Um dos softwares mais utilizado na quantificação de miRNAs é o mirDeep2, desenvolvido por Friedlander e colaboradores (2012). Maiores detalhes sobre o mirDeep2 serão descritos na Seção 2.6 deste trabalho. Também em meio a estas necessidades de compreender o funcionamento do miRNA, Zhao e colaboradores (2017) implementaram o QuickMIRSeq para a quantificação rápida e precisa de miRNAs e isomiRs conhecidos de sequenciamento de pequenos RNAs em grande escala. Todo o *pipeline* consiste em três etapas principais: preparação de banco de dados, quantificação com anotação e integração com visualização. A primeira etapa prepara os bancos de dados necessários para a segunda etapa; a segunda etapa processa o conjunto de dados miRNA-seq e gera tabelas de contagem para miRNAs e isomiRs; e a terceira etapa produz um relatório de projeto integrado (Figura 7) e interativo para análise de dados (ZHAO, et al, 2017).

Figura 7– Resultados mirDeep2



Fonte: FRIEDLANDER, et al., 2012. No arquivo .html permite acesso as tabelas da figura 7A que por sua vez possui links para os demais resultados mostrados nas figuras 7B e 7C.

2.6 MIRDEEP2

Friedlander e colaboradores (2012) desenvolveram o miRDeep2, um software que identifica miRNAs canônicos e não canônicos, e informa sobre candidatos de alta confiança detectados em múltiplas amostras independentes.

A avaliação de algoritmos predecessores, como miRanalyzer e miRTRAP por exemplo, é enigmática de acordo com Friedlander e colaboradores (2012), pois cada um deles foi testado apenas em um número limitado de conjuntos de dados que representam

uma cobertura pouco ampla da árvore filogenética. Outro motivo está relacionado à validação dos novos miRNAs encontrados, devido à baixa abrangência utilizada. Para resolver esse problema de avaliação, Friedlander e colaboradores (2012), propuseram um método para identificar miRNAs em dados de sequenciamento em larga escala buscando atender três demandas: poder identificar com precisão miRNAs conhecidos e novos em todos os clados dos principais animais, poder distinguir miRNAs de outros pequenos RNAs ligados a argonautas, reportar miRNAs que podem ser validados pelo método de validação de alto rendimento, ser eficiente na memória e no consumo de tempo e ser fácil de utilizar.

2.6.1 Funcionamento

O miRDeep2 possui controles estatísticos internos que permitem estimar a precisão e a sensibilidade de seu desempenho. O pacote miRDeep2 consiste em três módulos:

- o módulo *miRDeep2* identifica miRNAs conhecidos e novos em dados de sequenciamento de alto rendimento,
- o módulo *mapper* processa a saída da sequência bruta da plataforma Illumina e mapeia as leituras para o genoma de referência,
- o módulo *quantifier* soma as contagens de leitura para miRNAs conhecidos no conjunto de dados de sequenciamento.

Os módulos funcionam de forma complementar. Por exemplo, a saída do *mapper* pode ser inserida diretamente no módulo *miRDeep2*. Como o módulo *miRDeep2* identifica miRNAs conhecidos e novos nos dados de sequenciamento de alto rendimento, ele forma o cerne do software (FRIEDLANDER, et al., 2012).

Em geral, cinco arquivos de entrada são necessários para iniciar a análise: *genome.fa*, um arquivo FASTA com o genoma de referência; *Mature_ref_this.fa*, um arquivo FASTA com os miRNAs de referência maduros pertencente a espécie em estudo

provenientes do miRBase; *Mature_ref_other.fa*, um arquivo FASTA com os miRNAs de referência maduros das espécies relacionadas do miRBase; *precursor_ref_this.fa*, um arquivo FASTA com os miRNAs de referência precursores das espécies em estudo depositados no miRBase e *reads.fa*, um arquivo rápido com leituras de sequenciamento profundo (LIU, LIU, 2017). Opcionalmente, miRNAs maduros, um da região 5' do dúplex do precursor e outro da região 3'do mesmo precursor, e precursores conhecidos das espécies analisadas e miRNAs maduros de espécies relacionadas podem ser inseridos. Um arquivo de relatório de progresso resumindo todas as etapas é gerado automaticamente para cada análise. O módulo *miRDeep2* integra todos os resultados em um arquivo *.html* que contém informações detalhadas sobre cada miRNA conhecido e novo nos dados (FRIEDLANDER, et al., 2012).

2.6.2 Construção dos Scripts

O miRDeep2 trabalha com três scripts: o *mapper.pl* para execução do processamento; o *quantifier.pl* para quantificação e perfil de expressão e o *miRDeep2.pl* para identificação do miRNA (WHAT IS MIRDEEP2, 2011).

Segundo Friedlander e colaboradores (2012) e Liu e Liu (2017), algumas etapas estão envolvidas na análise por meio do miRDeep2, além dos três *scripts* citados acima. Para a primeira etapa, a análise de miRDeep2 requer os miRNAs conhecidos como referências, que podem ser baixados de miRbase. O arquivo *mature.fa* contém sequências de formato FASTA de todas as sequências de miRNA maduras, e o arquivo *hairpin.fa* contém sequências de formato FASTA de todos os *hairpins* de miRNA. Tanto os miRNAs maduros quanto os *hairpins* estão no formato de RNA. O miRDeep2 requer que os miRNAs sejam de apenas uma única espécie referência e que estas sequências de RNA sejam convertidas em DNA. Alguns *scripts Perl* nos pacotes miRDeep2 podem ser usados

para converter esses arquivos para o formato FASTA. Usando *extract_miRNAs.pl* para extrair sequências de uma espécie, *$ perl extract_miRNAs.pl hairpin.fa ipu> hairpin.ipu.fa*, *$ perl extract_miRNAs.pl mature.fa ipu> mature.ipu.fa*. Usando *rna2dna.pl* para substituir u e U por T, *$ perl rna2dna.pl hairpin.ipu.fa> hairpin.ipu.dna.fa*, $ *perl rna2dna.pl mature.ipu.fa> mature.ipu.dna.fa*. Um arquivo com informações do genoma é produzido usando *bowtie-build*, *$ bowtie-build lpg1.fa lpg1*. Antes de construir o arquivo com informações do genoma, o arquivo FASTA do genoma precisa ser verificado quanto à exatidão usando o script *sanity_check_genome.pl*. As linhas de identificação do arquivo FASTA não podem conter espaços em branco e devem ser exclusivas. As linhas de sequência não podem conter caracteres diferentes de A, C, G, T, N, a, c, g, t, n.

Para a segunda etapa, o *script mapper.pl* processa as *reads* e as mapeia contra o genoma de referência, *$ mapper.pl esc0hr.fastq -e -h -i -j -l 18 -m -p lpg1 -s esc0hr.reads_collapsed.fa -t esc0hr.reads_collapsed.arf -v -n*. A entrada é um arquivo nos formatos FASTA, FASTQ ou *seq.txt*. A saída depende das opções usadas, um arquivo FASTA com leituras processadas ou um arquivo .*arf* com leituras mapeadas. Os arquivos de entrada são *esc0hr.fastq* (arquivo de entrada para análise) e *lpg1* (arquivo de índice do genoma). Os arquivos de saída são *esc0hr.reads_collapsed.fa* (leituras processadas) e *esc0hr.reads_collapsed.arf* (arquivo de alinhamento). O significado dos símbolos que definem o script são -*e* (o arquivo de entrada que está no formato FASTQ), -*h* (analisar dado para o formato FASTA), -*i* (converter RNA em letras de DNA para mapear contra o genoma), -*j* (remover todas as entradas que tem uma sequência que contém letras diferentes de a, c, g, t, u, n, A, C, G, T, U e N), -*l* (descartar *reads* menores que 18 bp), -*m* (colapsar leituras), -*p* (mapear para o genoma, deve ser indexado por *bowtie-build*), -*s* (gravar leituras processadas para este arquivo), -*t* (gravar mapeamentos de leitura para este arquivo), -*v* (gerar relatório de progresso) e - *n* (sobrescrever arquivos existentes). Os dados de sequenciamento de diferentes amostras, por exemplo, controle e tratamento, podem ser processados e mapeados para o genoma juntos, fornecendo um arquivo

contig.txt (usando a opção -*d*) em que cada linha designa os locais dos arquivos e um código específico de três letras. Esta etapa também corta adaptadores de 3' e colapsa as leituras. A sequência *read* é pesquisada em busca de correspondências com os primeiros 6 nt na sequência do adaptador. Esta pesquisa começa na posição 18 na leitura. Se não houver correspondências com os primeiros 6 nt, as correspondências com os primeiros 5 nt do adaptador serão pesquisadas nos últimos cinco nt da leitura, depois as correspondências dos primeiros quatro com as últimas quatro posições e assim por diante. Quando uma correspondência é encontrada pela primeira vez, a correspondência com a sequência do adaptador e todos os nucleotídeos a jusante são cortados da leitura e a próxima leitura é pesquisada. As leituras que não têm correspondências são retidas, mas não cortadas. Logo após há o mapeiamento das *reads* processadas para o genoma com *bowtie*, usando essas opções *bowtie –f –n 0 –e 80 –l 18 –a –m 5 –best –strata*. A opção -*n* 0 mantém apenas os alinhamentos com nenhuma incompatibilidades na região semente de uma leitura mapeada contra o genoma. A região semente é definida pela opção -*l* 18 que corresponde aos primeiros 18 nt de uma sequência de leitura. Ao usar a opção -*n*, é possível permitir que ocorram incompatibilidades após a região de semente de uma leitura em um alinhamento. Isso é determinado pela opção -*e* 80 e é a soma máxima dos valores de qualidade em cada posição de incompatibilidade. O valor de qualidade padrão para cada posição em um arquivo FASTA é definido como 40, o que significa que até duas incompatibilidades são permitidas na região de uma leitura após sua região de semente. A opção *-m 5* mantém apenas leituras que não mapeiam mais de cinco vezes contra o genoma. Opção *–best –strata* ordena os mapeamentos do melhor para o pior alinhamentos de acordo com a definição do *bowtie* (Liu, Liu, 2017; FRIEDLANDER, et al., 2012).

Para a terceira etapa, o *script quantifier.pl* mapeia as leituras de sequenciamento profundo para os precursores de miRNA e determina os níveis de expressão dos miRNAs correspondentes. Primeiro, as sequências de miRNA maduras são mapeadas contra os

precursores. Opcionalmente, as sequências de miRNAs localizadas nas porções 3' dos precursores também podem ser mapeadas contra os precursores. Com base no mapeamento, são determinadas as sequências de miRNAs maduros das regiões 5' e 3' nos precursores. Em segundo lugar, as leituras provenientes de sequenciamento profundo são mapeadas para os precursores. O número de leituras que estão em um intervalo 2 nt a montante e 5 nt a jusante das sequências maduras é determinado. *$ quantifier.pl -p lpu.hairpin.dna.fa -m lpu.mature.dna.fa -r esc0hr.reads_collapsed.fa -y esc0hr*. Arquivos de entrada *lpu.hairpin.dna.fa* (um arquivo FASTA com precursores), *lpu.mature.dna.fa* (um arquivo FASTA com miRNAs maduros) e *esc0hr.reads_collapsed.fa* (um arquivo FASTA com leituras de sequenciamento profundo). Os arquivos de saída consistem em um arquivo separado por tabulação denominado *miRNAs_expressed_all_samples.csv* com identificadores de miRNA e sua contagem de *reads*, um arquivo de assinatura denominado *miRBase.mrd*, um arquivo denominado *expression.html* que oferece uma visão geral de todos os miRNAs, os dados de entrada e um diretório denominado *pdfs* que contém arquivos PDF, cada um dos quais mostra uma assinatura e estrutura de miRNAs. Os significados das opções são *-p precursor.fa* (sequências de precursores de miRNA), *-m mature.fa* (sequências de miRNA maduras) e *-r reads.fa* (pequenas sequências de RNA-Seq). A ferramenta *bowtie-build* é usada com opções padrão para construir um índice de transformação *Burrows-Wheeler* dos precursores conhecidos. O mapeamento das leituras é feito com essas opções *bowtie –f –v 1 –a –best –strata –norc*. A opção *-f* designa um arquivo FASTA como entrada, a opção *-v* 1 relata os mapeamentos lidos com até uma incompatibilidade, a opção *-a* leva ao relatório todos os alinhamentos válidos, as opções *–best –strata* ordena os mapeamentos do melhor para o pior alinhamentos de acordo com a definição *strata* do *bowtie* e a opção *–norc* aconselha o *bowtie* a não mapear leituras para o complemento reverso das sequências precursoras no índice de *bowtie*. O mapeamento das sequências conhecidas de miRNAs maduros contra os miRNAs precursores conhecidos para as espécies de referência é feito com estas opções:

bowtie –f –v 0 –a –best –strata –norc. Aqui, nenhum número de incompatibilidades é permitida sendo definido como zero por meio da opção *-v* porque as sequências maduras anotadas devem estar contidas em suas sequências precursoras anotadas. Mapeamentos dos miRNAs maduros para precursores que não foram combinados são descartados e os dois arquivos de mapeamento são cruzados (Liu, Liu, 2017; FRIEDLANDER, et al., 2012).

Na quarta etapa, ocorre a identificação de miRNAs conhecidos e novos. O *script miRDeep2.pl* executa todos os *scripts* necessários do pacote miRDeep2 para realizar uma análise de detecção de miRNA com dados de sequenciamento profundo. *$ miRDeep2.pl esc0hr.reads_collapsed.fa lpg1.fa esc0hr.reads_collapsed_vs_genome.arf lpu.mature.rna2dna.fa dre.mature.rna2dna.fa lpu.hairpin.rna2dna.fa -q expression_analyses / expression_analyses_now /miRBase.mrd 2> report.log*. Arquivos de entrada *esc0hr.reads_collapsed.fa* (um arquivo FASTA com leituras de sequenciamento profundo), *lpg1.fa* (um arquivo FASTA do genoma correspondente), *esc0hr.reads_collapsed.arf* (um arquivo de leituras mapeadas contra o genoma em formato *miRDeep2 arf*), *lpu .mature.dna.fa* (um arquivo FASTA com miRNAs conhecidos das espécies em estudo), *dre.mature.dna.fa* (um arquivo FASTA de miRNAs conhecidos de espécies relacionadas) e *lpu.hairpin.dna.fa* (um arquivo FASTA com precursores de miRNA conhecidos das espécies em estudo). Arquivos de saída são uma planilha e um arquivo *html* com uma visão geral de todos os miRNAs detectados nos dados de entrada de sequenciamento profundo. Opções *-q* (arquivo *.mrd*) usa a saída do *quantifier.pl* para obter informações sobre os miRNAs, *2>*, canaliza toda a saída de progresso para o arquivo *report.log*. Depois disso, uma quantificação rápida de miRNAs conhecidos é feita se arquivos com precursores de miRBase e miRNAs maduros correspondentes forem fornecidos ao módulo. Em uma segunda etapa, potenciais precursores de miRNA são excisados do genoma usando os mapeamentos de leitura como diretrizes e são primeiro analisados de forma que apenas mapeamentos perfeitos de pelo menos 18 nt sejam retidos. Além disso, os mapeamentos de leitura de *reads* que mapeiam perfeitamente mais de cinco vezes

para o genoma são descartados. Em seguida, as duas fitas do genoma de cada contig de genoma são escaneadas separadamente, da extremidade 5' para a 3'. Se houver uma pilha de leitura dentro de 70 nt a jusante da pilha de leitura atual, ela será escolhida. Essa pesquisa *downstream* é repetida até que nenhuma pilha de leitura superior seja encontrada com 70 nt. Dessa forma, a pilha de leitura local mais alta é identificada. Em seguida, a sequência coberta pela pilha de leitura local mais alta é excisada duas vezes, uma incluindo 70 nt a montante e 20 nt a jusante a sequência de flanqueamento. Posteriormente, a varredura do genoma continua a partir da posição 1 nt a jusante da última sequência excisada. A próxima etapa do módulo é preparar o arquivo de assinatura. A ferramenta *bowtie-build* é usada com opções padrão para construir um índice de transformação de *Burrows-Wheeler* dos precursores potenciais excisados. Em seguida, o conjunto de leituras de sequenciamento é mapeado para o índice, usando o *bowtie* (versão 0.12.7) com as seguintes opções: *bowtie –f –v 1 –a –best –strata –norc*. A opção *-f* designa um arquivo FASTA como entrada, a opção *-v* 1 relata os mapeamentos lidos com até uma incompatibilidade com os precursores, a opção *-a* leva ao relatório todos os alinhamentos válidos, as opções *–best –strata* ordena os mapeamentos do melhor para o pior alinhamentos de acordo com a definição de *strata* do *bowtie*. Se as leituras mapearem perfeitamente para os precursores, os mapeamentos da mesma leitura com uma incompatibilidade não serão relatados. A opção *–norc* aconselha o *bowtie* a não mapear as leituras para o complemento reverso das sequências precursoras no índice do *bowtie*. O conjunto de miRNAs maduros conhecidos é mapeado contra as espécies de referência também com as seguintes opções: *bowtie –f –v 0 –a –best –strata –norc*. Aqui, não são permitimos qualquer incompatibilidade para os mapeamentos porque a sequência de miRNA madura e as sequências de precursores potenciais não foram sujeitas a qualquer fonte de ruído. Os dois arquivos de mapeamento são concatenados e todas as linhas são classificadas de acordo com os identificadores de precursores potenciais. Outra etapa é prever estruturas secundárias de RNA dos precursores potenciais. Isso é feito com

RNAfold. Opcionalmente, são calculados os valores P para um subconjunto dos precursores potenciais. Os precursores potenciais são descartados se menos de 60% dos nucleotídeos na parte madura candidata houver emparelhamento entre as bases. O algoritmo principal miRDeep2 é executado com estas opções: *-s –v -50 –y*. A opção -*s* designa os miRNAs maduros de referência no arquivo de formato FASTA como entrada, a opção -*v -50* mantém todos os precursores que têm uma pontuação acima de 50 e a opção -*y* fornece um arquivo adicional com valores *randfold*. O número de miRNAs conhecidos presentes nos dados é estimado como o número de miRNAs maduros conhecidos que mapeiam perfeitamente para um ou mais precursores potenciais excisados. O número de miRNAs conhecidos que são recuperados é estimado como o número de miRNAs maduros conhecidos que mapeiam perfeitamente para um ou mais hairpins que excedem o limite de pontuação determinado (Liu, Liu, 2017; FRIEDLANDER, et al., 2012).

Na quinta etapa consiste no arquivo *results.html* usando um navegador de internet para visualizar os resultados (Liu, Liu, 2017).

2.7 RESULTADOS

O estudo iniciou com a pesquisa da Christina Martins e colaboradores, onde o objetivo foi caracterizar o perfil de expressão da ectonucleotidase e investigar o envolvimento de purinas extracelulares no câncer de próstata. Duas das questões levantadas foram quais miRNAs poderiam estar envolvidos e qual sua quantidade presente em amostras de pessoas que foi diagnosticado câncer. E com esses dados em mãos descobrir quais os possíveis alvos e as funções relacionadas a estes alvos.

As questões de pesquisa propostas por este atual trabalho foram respondidas positivamente. Para dar continuidade, nosso trabalho teve como objetivo agregar informações a partir de duas planilhas de saída do mirDeep2 onde a primeira foi referente aos SRRs utilizados pelo programa, verificarmos informações de cada SRR (APÊNDICE 1-Tabela SRR) utilizado do trabalho no SRA. Para cada arquivo SRR da planilha *SRA.xlsx*, proveniente da análise do mirDeep2, foram obtidas as informações, do site <https://trace.ncbi.nlm.nih.gov/Traces/sra/>, que estão na aba *Run Browser* e sub aba *Metadata*. O *Run* é o código do SRR analisado; as *Bases* referem-se a quantidade de pares de bases; o *Size* sua extensão; *GC content* a percentagem de guanina e citosina presente no exemplar; *Experiment* um código do experimento; *Platform* refere ao método de sequenciamento utilizado na amostra; *Strategy* a técnica de sequenciamento; *Source* a estratégia de análise dos transcritos; *Organism* o organismo em estudo; *Bioproject* contem dados do trabalho como o código do trabalho (*Link*), o código do artigo científico (*Paper*), o título (Title). Na aba *Analisys* foram recuperadas as informações sobre a análise taxonômica (*Taxonomy analysis*), como as percentagens das leituras identificadas (*Identified reads*) e não identificadas (*Unidentified reads*). Também foi salvo o pdf do *paper*.

Na segunda planilha de saída do mirDeep2 que tratava da quantidade de cada miRNA encontrado em cada uma das amostras trabalhadas, primeiramente foi buscado

informações (APÊNDICE 2- miRNAs e alvos) através do portal do mirBase para cada miRNA da planilha. Selecionou o ID do miRNA e com este código fez uma consulta. Se o miRNA fosse -5p, selecionar os dados provenientes dos links que levavam a informações em outros sites, TargetScanVert e miRDB; e selecionar o ID de acesso do -3p presente no mirBase. Nos sites do TargetScanVert e miRDB selecionou o *Gene name* e *Gene description* e *gene symbol* dos alvos bem como o número de predição de alvos que cada site encontrou.

Como foram identificados muitos resultados, foram selecionados os dez primeiros alvos de cada miRNA no Target Scan Vert e também foram selecionados os dez primeiros alvos de cada miRNA no miRDB; para verificar as coincidências e as diferenças entre esses dois programas de bioinformática online e também se estes miRNAs agem de outras maneiras não só na via de câncer de próstata.

A tabela do Apêndice 2 foi dividida em três colunas, a primeira com dados de acesso das porções -5p e -3p do mirBase bem como nome e identificação do miRNA proveniente da análise do mirDeep2; na segunda coluna estão os dados do Target Scan Vert quanto ao nome do gene, descrição, símbolo e número de genes alvo preditos; na terceira coluna estão os dados do miRDB quanto ao nome do gene, descrição, símbolo e número de genes alvo preditos.

Um exemplo, para o hsa-miR-10-a-5p no Target Scan Vert o primeiro alvo identificado foi o BDNF, cuja função do gene pode desempenhar um papel na regulação da resposta ao estresse e na biologia dos transtornos do humor; e no miRDB o primeiro alvo identificado foi o CADM2 cuja função do gene codifica um membro da família da molécula de adesão de células sinápticas 1 que pertence à super família de imunoglobulinas. Para o hsa-miR-10-b-5p os primeiros alvos identificados foram os mesmos do hsa-miR-10-a-5p. Outro exemplo, para o hsa-miR-150-5p os primeiros alvos identificados foi MYB tanto no Target Scan Vert quanto no miRDB, cuja proteína desempenha um papel essencial na regulação da hematopoiese. Isso mostra que apesar das coincidências não há

unanimidade provavelmente porque o Target Scan Vert e o miRDB trabalham de formas diferentes.

3 ARTIGO

Adenosine in prostate cancer: functional importance and identification of potential biomarkers

MARTINS et al: THE ROLE OF ADENOSINE AS A POTENTIAL BIOMARKER IN PROSTATE CANCER

Christina Aparecida Martins[1], Sara Teixeira Soares Mota[2], Fabrícia Matos de Oliveira[3], Tamires Caixeta Alves[4], Laurence Rodrigues do Amaral[4], Roberta Moura Ramos Vargas[4], Matheus de Souza Gomes[4], Thaise Gonçalves de Araújo[2], Cristina Ribas Fürstenau[1*]

[1] Laboratório de Cultura de Células Animais, Instituto de Biotecnologia, Universidade Federal de Uberlândia, Patos de Minas-MG, Brazil.

[2] Laboratório de Genética e Biotecnologia, Instituto de Biotecnologia, Universidade Federal de Uberlândia, Uberlândia-MG, Brazil.

[3] Faculdade de Matemática, Universidade Federal de Uberlândia, Patos de Minas - MG, Brazil.

[4] Laboratório de Bioinformática e Análises Moleculares, Instituto de Biotecnologia, Universidade Federal de Uberlândia, Patos de Minas - MG, Brazil.

[*]Corresponding author: Cristina Ribas Fürstenau, PhD. Universidade Federal de Uberlândia. Avenida Getúlio Vargas, 230 - Centro. CEP 38128-000. Patos de Minas, Minas Gerais, Brasil. Phone number: +55 (34) 3823-3714 (Ext. 36). E-mail: cristinafurstenau@gmail.com

Abstract Extracellular nucleotides and nucleosides are involved in cell signalling processes, affecting the physiopathological responses of tumours such as prostate cancer (PCa). Ectonucleotidases modulate these responses, since they can regulate the bioavailability of these molecules. The altered activity of these enzymes may cause the progression and malignancy of tumours. The aim of this study was to characterise the profile of ectonucleotidase expression and to investigate the involvement of extracellular purines in PCa. The expression of ecto-nucleoside triphosphate diphosphohydrolase 1 and 3 (NTPDases 1 and 3) and ecto-5'-nucleotidase was quantified by qPCR in three prostate cell lines (RWPE, LNCap and PC3). Moreover, potential microRNAs were predicted using bioinformatic software. The expression of NTPDase1 and

ecto-5'-nucleotidase was significantly different in LNCap and PC3 when compared to RWPE cells. NTPDase1 was less expressed in tumoral line cells, presenting lower levels in LNCap. PC3 cells presented higher levels of ecto-5'-nucleotidase transcripts than LNCap, which is probably related to higher levels of adenosine. NTPDase3 expression increased with the tumour aggression, and was higher in PC3 than LNCap. Furthermore, we identified oncomiRNAs targeting A_3 adenosine receptors, suggesting the transcriptional crosstalk between miRNAs and NTPDase1 and ecto-5'-nucleotidase expression. To summarize, these results suggested adenosine as a key factor in PCa.

Keywords: Purinergic signalling. Ectonucleotidases. Prostate cancer. Adenosine. miRNA.

1. Introduction

Prostate cancer (PCa) is considered one of the main neoplasias diagnosed in males worldwide and represents the second biggest cause of death in this gender. The probability of metastasis increases with the progression of PCa. This stage is characterised by the interaction between different cells in the tumour microenvironment, which results in invasiveness and the migration of malignant cells. Signalling molecules are mediators of this interaction that are able to work actively in the biological pathways of cancer cells. The influence of these mediators is related to the modulation of the immune response, as well as the clearance of inflammatory factors and thromboregulation in tumours (1-3).

Many studies have investigated purinergic signalling as an important factor in the modulation of tumour processes (3-4). In fact, the involvement of adenosine triphosphate (ATP), adenosine diphosphate (ADP), adenosine monophosphate (AMP) and adenosine have been investigated in different types of tumours. These extracellular nucleotides and nucleoside interact with purinergic receptors in the control of many cellular pathways (3, 5, 6). Thus, the recognition of these processes will certainly help in the development of a therapeutic tool based on the potential of these signals in cancer control and treatment.

As described by Battisti et al. (1), the hydrolysis of extracellular nucleotides performs an important function in prostate cancer development. In this context, the concentration of these

nucleotides in the extracellular environment stimulates P1 and P2 receptors, which in turn induce different responses. These responses produce a cascade of physiological and pathological effects, such as cell proliferation, cytotoxicity, apoptosis, inflammation, phagocytosis, etc (6-8). All these processes are regulated by enzymes called ectonucleotidases. These enzymes are responsible for controlling nucleotide and nucleoside levels in the extracellular environment by using them as substrates (1, 9).

Since ectonucleotidases contribute to the homeostasis of ATP, ADP and adenosine, it is possible to suggest that the level of expression of these enzymes in the tumour environment is crucial for oncological effects. Furthermore, the level of expression of P1 and P2 receptors represents another decisive factor related to cancer progression. In fact, many studies have reported the altered expression of these enzymes and their receptors in the tumour environment (6, 9). Particularly, the activity of the ecto-5'-nucleotidase was found to be altered in different types of tumours, such as breast, neck, head and ovary carcinoma. The reason for this is that the ability of cancer cells to adhere, metastasise and resistance drug therapy is correlated to altered ecto-5'-nucleotidase activity (10-11).

In this sense, there is growing interest in the search for new potential molecular targets for alternative cancer treatments. An efficient method would be the inhibition of ectonucleotidase activity, in order to control the levels of these extracellular nucleotides and nucleoside. Another alternative is based on the blockade of the signalling cascade through purine receptor silencing (8). However, the purinergic signalling pathway comprises a complex system in cancer genesis, and the mechanism of action of this kind of signalling is still unknown in the prostate (8).

Micro RNAs (miRNAs) are short untranslated sequences of RNA (18-25 nucleotides) that play important roles in signalling processes, since they are a mechanism of post-transcriptional regulation. Basically, the process works by binding of the miRNA 5' region to the 3'untranslated region (3'-UTR) of the target mRNA, leading to its degradation or translational repression by the RNA-induced silencing complex (RISC) (12,13). A single molecule of miRNA is able to interact with multiple targets, and a single mRNA target may have many binding sites for different miRNAs.

Nowadays, it is possible to predict and to identify some targets of miRNAs by bioinformatic tools (12-14).

In this context, a huge number of miRNA sequences have been discovered in recent years. Each sequence is involved in a cellular pathway, modulating a range of biological responses, especially in diseases such as cancer (12, 14). According to Kong and co-workers (15), miRNAs are able to inhibit the mRNAs of tumour suppressor genes and/or genes associated with apoptosis, resulting in the modulation of biological pathways. On the other hand, miRNAs might inhibit oncogene expression, resulting tumour suppression. In this sense, miRNA expression is an important process to control the action of proteins with tumour promotion/suppression potential (14).

There is a need to investigate the role of ectonucleotidases in different types of cancer. Also, although well-characterised in other types of cancer, there are few studies reporting on the mechanisms and impacts of the altered expression and post-transcriptional regulation of these enzymes in prostate cancer.

The purpose of this work was to evaluate and to compare the gene expression of the ectonucleotidases ecto-nucleoside triphosphate diphosphohydrolase 1 and 3 (ENTPDase1 and ENTPDase3) and ecto-5'-nucleotidase in three prostate cell lines, i.e. one control (RPWE) and two cancer lines, PC3 (androgen unresponsive) and LNCap (androgen responsive) with different degrees of malignancy. Moreover, this study aimed to predict the involvement of miRNAs in the regulation of these enzymes as well as the adenosine receptor A3 in prostate cancer.

2. Materials and Methods

2.1 Cell culture

The PC3 and LNCap cell lines were cultured in RPMI 1640 medium (Gibco; Thermo Fisher Scientific, Inc., Waltham, MA, USA) supplemented with 10% foetal bovine serum (Gibco; Thermo

Fisher Scientific, Inc., Waltham, MA, USA). The RWPE control cell line was cultured in keratinocyte medium (Sigma-Aldrich; St. Louis, Missouri, USA) supplemented with EGF (Gibco; Thermo Fisher Scientific, Inc., Waltham, MA, USA), pituitary extract (Sigma-Aldrich; St. Louis, Missouri, USA) and 10% fatal bovine serum (Gibco; Thermo Fisher Scientific, Inc., Waltham, MA, USA). All cells were cultured at 37°C in a humidified atmosphere containing 5% CO_2 and 95% O_2. Culture medium was changed on alternate days until the cells had achieved 80% confluence.

2.2 Analysis of gene expression

2.2.1 RT-PCR

Total RNA was extracted from cells using the Trizol reagent (Invitrogen, Carlsbad, CA, EUA), following the manufacturer's instructions. For reverse transcription (RT-PCR) and cDNA transcription, 1 µg of the total RNA, 10 U of RNase inhibitor (Invitrogen; Carlsbad, CA, EUA), 40 U of Murine Moloney Leukaemia Virus Reverse Transcriptase (MMLV-RT) (AmershamBiosciences; Little Chalfont, United Kingdom), 200 µM of each deoxyribonucleotide triphosphate (dNTPs) and 126 pmol of random hexamer primers (AmershamBiosciences; Little Chalfont, United Kingdom) were incubated in PTC-100 thermal cycler (*StepOne Plus*, Thermo Scientific, EUA) at 37ºC for 1h, followed by 95ºC for 5 min.

The cDNA quality was analysed by the amplification of 536 pb of the β-2-microglobulin (*B2M*) gene which was flanked by the oligonucleotides: 5' AGCAGAGAATGGAAAGTCAAA 3' and 5' TGTTGATGTTGGATAAGAGAA 3'. The PCR reaction was performed with 1X PCR buffer (0.1 mM EDTA, 20 mM Tris-HCl pH 8.0, 1 mM DTT, 50% v/v glycerol) (Invitrogen, Carlsbad, CA, EUA), 5 pmol of primers (Invitrogen, Carlsbad, CA, EUA), 1U of Platinum® Taq DNA Polymerase (Invitrogen, Carlsbad, CA, EUA), 4 mM $MgCl_2$ and 200 µM dNTPs (Invitrogen, Carlsbad, CA, EUA). Cycling conditions were: 95°C for 5 min, 35 cycles at 94°C for 40 s, 59°C for 40 s, 72°C for 50 s, and a final extension at 72°C for 10 min. The quality of the reaction products was analysed using 1.5% agarose gels stained with GelRed.

2.2.2 Real-time qPCR

RT-qPCR reactions were performed on an ABI PRISM 7300 system (Applied Biosystems, Foster City, CA,USA) using SYBR Green PCR Master Mix (Invitrogen, Carlsbad, CA, EUA), 2 μL of cDNA (1:2), and 5 μM of specific primers (Table 1). The samples were run six times and all data were normalised by *B2M* expression. Relative quantification was calculated using the 2^-ΔΔCq method (17), after the construction of a relative standard curve.

2.3 Bioinformatic analysis

Potential miRNAs that could target the mRNA for ENTPDase1, ecto5'nucleotidase and adenosine receptor A_3 were predicted and analysed using the software TargetScan 7.1 available at http://www.targetscan.org. The gene IDs were obtained through NCBI platform, available at http://www.ncbi.nlm.nih.gov.

2.4 miRNAs in *H. sapiens* small RNA-seq libraries

Forty-four raw small data files of RNAs were retrieved from the NCBI Sequence Read Archive (SRA) with different access numbers. The library qualities were evaluated using FastQC software (18), adapters were removed with Trimmomatic (19) discard reads with quality score below 20 and length less than 17 nucleotides and more than 30 nucleotides. The filtered sequences were mapped and quantified using miRDeep2 (20). miRDeep2 and perl scripts were used in each sequence separately to generate the numbers of the reads for each identified miRNA.

2.5 Data analysis

Data were statistically analysed using GraphPad Prism 4 (GraphPad Software, Inc., La Jolla, CA, USA) and the computing environment R (R Development Core Team, 2018). Means and standard deviations (SD) were calculated for each measurement performed. Sample number (n)

represents the number of experiments runned with different prostate cell cultures. Analysis of variance (ANOVA) followed by the Tukey test was performed to establish differences among groups. Differences were considered significant when $p<0.05$.

3. Results

3.1 Gene expression by qPCR

Gene expression analysis showed that all prostate cell lines expressed the enzymes NTPDases 1, 3 and ecto-5'-nucleotidase. However, the expression of these enzymes was significantly different among the prostate cell lines ($p<0,05$). NTPDase1 and ecto-5'-nucleotidase mRNA expression was significantly different in LNCap and PC3 in comparison to RWPE ($p<0.05$). In fact, the tumour cell lines expressed lower levels of NTPDase1 mRNA, mainly in LNCap (Figure 1).

Similarly, the analysis of ecto-5'-nucleotidase mRNA expression followed the same pattern of NTPDase1, and was less expressed in tumour cell lines. However, we observed the overexpression of ecto-5'-nucleotidase mRNA in PC3 cells ($p<0.05$) in comparison with LNCap, whereas there was no difference observed in NTPDase1 transcripts between the tumour cell lines (Figure 1). Concerning NTPDase3 mRNA expression in tumour cell lines, there was no significant difference among the cell lines (Figure 1).

3.2 Identification of miRNAs involved in the PCa pathway

Bioinformatic analysis predicted around 641 miRNAs targeting the A_3 adenosine receptor. Most of them have not been characterised yet and have unknown functions. In spite of this, we identified 11 possible miRNAs with 3'UTRs with oncogenic function that may regulate the adenosine pathway in PCa via the A_3 receptor (Figure 2). In this sense, the results of miRNA prediction suggested that the A_3 adenosine receptor gene may be silenced by miR-301a-5p, miR-183-3p, miR-182-3p, miR-424-5p, miR-9-5p, miR-4534, miR-10, miR-367-5p, miR-208a-5p,

miR-130a-5p and miR-300. As a result of this, adenosine cannot activate apoptosis in PCa by the extrinsic pathway.

In relation to miRNAs possibly targeting NTPDase1 mRNA, our results show that the oncomiRNAs miR-150 and miR-9 may exert translational repression of this gene. Then, as in A_3 regulation, these miRNAs prevent the trigger of apoptosis in PCa by adenosine. On the other hand, miR-205, a tumour suppressor miRNA, may bind to the 3'UTR of NTPDase1 mRNA, thereby repressing its translation. Consequently, adenosine is not produced and it is not able to interact with P1 receptors on immune cells to cause immunosuppression.

Finally, the prediction of miRNA sites for ecto-5'-nucleotidase showed that miR-30 family (miR-30d, miR-30e, miR-30a, miR-30c, miR-30b) is able to inhibit the translation of this gene. The consequence of this process is the absence of adenosine in the tumour microenvironment, which affects apoptosis via A_3 receptors. In addition, adenosine is not able to trigger immunosuppression, highlighting the suppressor effect of this miRNA.

3.4 *In silico* validation of the miRNAs in PCa by small RNA-seq libraries

In order to validate the involvement of the miRNAs predicted in the bioinformatic analysis, we retrieved 44 small RNA libraries of blood samples from patients with PCa available in the Sequence Read Archive (SRA) and verified the presence of mature miRNAs in each library (Table SI and SII). The analysis of the different mature miRNA sequences confirmed the presence of the members of the miR-10, miR-150, miR-182, miR-183, miR-205, miR-301, miR-30, miR-424 and miR-9 families in the libraries related to PCa patients. Only miR-4534, miR-367 and miR-208 were not found in the analysed libraries.

4. Discussion

Purinergic signalling is involved in many metabolic pathways, in normal physiology as well as under pathological conditions. This signalling pathway has important roles in cell proliferation, differentiation, apoptosis and motility, i.e. processes that are essential for cancer progression (4, 5).

In this sense, ectonucleotidases are critical in cancer, since they are able to alter the responses of the system depending on the concentration of extracellular purines (23,24).

Stella et al. (10) showed the relationship between the altered activity of ectonucleotidases and the occurrence of malignant cells, especially with regard to tumour aggression and invasiveness. This theory was supported by Schmid and co-works (6), as their study reported an increase in ectonucleotidase ENTPDase1 expression in patients with metastatic tumours, compared to subjects with non-metastatic lung cancer. For this reason, we investigated the expression of these enzymes, aiming to characterise their involvement in PCa.

We analysed ENTPDase1 and 3 as well as ecto-5'-nucleotidase expression in three prostate cell lines, one non-tumorigenic (RPWE) and two cancer lines (PC3 and LNCap). Our results suggested that high levels of extracellular nucleotides in tumour cells are associated to tumour aggression. In fact, analysing tumour prostate cell lines, we observed high expression of ecto-5'-nucleotidase mRNA in PC3. This result is consistent with the results of Burnstock (5), who reported that prostate tumours have elevated concentrations of ecto-5'-nucleotidase (CD73) in response to the high sensitivity of these cell lines to extracellular nucleotides. The elevated expression of ecto-5'-nucleotidase mRNA in PC3 cells suggests that this cell line may have higher concentrations of adenosine in its microenvironment. This may explain the malignancy of PC3, since adenosine is strongly related to tumour growth and invasiveness.

According to Di Virgilio (7), E-NTPDase 1 and ecto-5'-nucleotidase are very important to cancer progression since the coordinated action of both enzymes represents the main pathway for extracellular adenosine production. In fact, NTPDase1 is responsible for the hydrolysis of ATP and ADP into AMP, which is degraded by ecto-5'-nucleotidase into adenosine. This nucleoside is fundamental since it is associated to immunosuppression. Moreover, ecto-5'-nucleotidase is involved in cell-cell interactions, which could affect the invasiveness and metastasis of prostate cancer (6, 24).

According to Schmid et al. (6), the tumour microenvironment is a crucial factor regulating tumour invasion and growth. Naturally, the bioavailability of oxygen to cancer cells is limited. Thus,

there is a preference for the anaerobic glycolytic pathway to meet the metabolic needs of these cells. As a result of this process, the quick production and accumulation of ATP leads to an excess of adenosine. A high concentration of these purines in the tumour microenvironment promotes the diffusion of these molecules into the extracellular space, where they concentrate and interact with purine receptors on immune cells, resulting in the inhibition of cytotoxic T cell and natural killer (NK) cell activity. Moreover, they act as neoplastic growth factors by facilitating tumour invasion and metastasis. Hence, the biggest implication of this pathway is the selection of malignant cells that are more resistant and adapted to these adverse conditions, which is correlated with the occurrence of more aggressive tumours (5-8, 11).

On the other hand, evidence suggests that adenosine exerts dual roles since it may also act as tumour suppressor in the prostate. This fact might be supported by our results on gene expression, since the highest levels of ENTPDase 1 and ecto-5'nucleotidase were found in the RWPE cell line. In this sense, Virtanen and co-workers (25) reported that adenosine may regulate p53 to induce apoptosis in prostate cell lines. Furthermore, adenosine can activate apoptosis mechanisms via AMP protein kinase (AMPK) in these cells. This evidence is consistent with the findings of Aghaei et al. (26), in that adenosine inhibits the proliferation of PCa by arresting the cell cycle at G1 phase, followed by apoptosis.

The observed effects of adenosine are mediated by two different pathways in LNCap and PC3 cell lines: the intrinsic and the extrinsic pathways. In the intrinsic pathway, adenosine is uptaken into cells by specific transporters, where it is converted into AMP through adenosine kinase activity. Then, this intraconverted AMP is responsible for activating AMPK, which promotes p53 expression. The extrinsic pathway is based on the interaction between extracellular adenosine and its receptors A_1, A_{2a}, A_{2b} and A_3. The upregulation of p53 expression occurs because of this interaction, which in turn induces apoptosis (27). Aghaei et al. (26) suggest that the cell death induced by adenosine in LNCap and PC3 is mediated by the extrinsic pathway, via A_3 adenosine receptor activation.

These contradictory findings led to two hypotheses to explain prostate cancer progression: 1) gene expression may undergo post-transcriptional modulation, which represses NTPDase1 and ecto-5'-nucleotidase mRNA translation or 2) adenosine A_3 receptors are being silenced to block apoptosis in prostate cancer cells. Thus, some cell mechanisms such as miRNAs may be involved in these processes, silencing transcripts and altering molecular pathways in PCa. In this sense, we investigated miRNAs that may target the mRNA of A_3 receptors.

Considering our qPCR data, together with data regarding the predicted miRNAs targeting A_3 adenosine receptors, miR-301a-5p binds to the 3'UTR of A_3 mRNA, thereby repressing this receptor. Consequently, we inferred that apoptosis by the extrinsic adenosine pathway would probably be inhibited. As in other types of cancer, like breast, gastric, colorectal and pancreatic cancer, miR-301a-5P works as an oncomiRNA in PCa. In fact, according to Xie et al. (28), miR-301a may increase cell invasion and metastasis via regulation of the androgen receptor (AR) in prostate.

We also found that miR-183-3p could probably be involved in A_3 silencing, since it is an oncomiRNA targeting this receptor. In fact, the tumour promoting activity of miR-183 in PCa was first reported by Ueno et al. (29), as this miRNA inhibits the tumour suppressorsDkk-3 and SMAD4. In addition, the expression of miR-183 in PC3 and LNCap is upregulated in comparison to RWPE, which reinforces the oncogenic effect in PCa with the promotion of proliferation, migration and invasion (29, 30). Likewise, miR-182-3p directly targets and possibly suppresses A_3 due to its oncogenic function. According to Yao et al. (31), miR-182 promotes prostate cancer growth by inhibiting apoptosis via targeting the ARRDC3/ITGB4 pathway.

Our results suggest that the A_3 adenosine receptor is also a potential target of miR-424-5p. This miRNA exerts oncogenic roles in prostate and breast cancer. Specifically, in PCa, miR-424 acts by silencing ubiquitination and activating tumour progression (32). Furthermore, miR-9-5p can bind to A_3, modulating the adenosine pathway in PCa. This regulation is mediated by the repression of A_3, resulting in the oncogenic potential of miR-9. As described by Seashols-Williams et al. (33), miR-9 is a oncomiRNA that is overexpressed in PCa, with roles in tumour progression and

metastasis via multiple pathways, such as E-cadherin and suppressor of cytokine signalling 5 (SOCS5). In addition, our findings suggest that miR-10 blocks apoptosis in PC3 and LNCap cells by targeting A_3. In fact, miR-10 is an oncomiRNA which is overexpressed in breast and prostate tumours. In PCa, miR-10 alters the expression of adhesion proteins to promote metastasis (34, 35).

In relation to possible miRNAs targeting ecto-5'-nucleotidase mRNA, the bioinformatic analysis found the miR-30 family (miR-30d, miR-30e, miR-30a, miR-30c, miR-30b) as a possible candidate (Figure 3). According to Kobayashi et al. (13), miR-30 is overexpressed in prostate cancer cell lines such as PC3 and LNCap, showing oncogenic activity in these cell lines by targeting SOCS1, a tumour suppressor in PCa. In this sense, miR-30 can increase tumour proliferation and invasion through the downregulation of cytokine signalling pathways. Similarly, we suggest that miR-30 binds to the 3'UTR of ecto-5'-nucleotidase, thus suppressing the purinergic signalling by adenosine in PCa. Consequently, adenosine is not taken up or interacts with the A_3 receptor to promote apoptosis.

On the other hand, Kao and co-workers (36), reported that miR-30 may be associated with tumour suppressor functions in PCa. Considering that adenosine might play dual functions in PCa, we suggest that miR-30 may regulate ecto-5'-nucleotidase mRNA expression in order to inhibit the adenosine immunosuppressive pathway. Likewise, NTPDase1 is a target of miR-205 that is downregulated in PCa and exerts tumour suppressor role by blocking tumour growth (37). Therefore, miR-205 may modulate NTPDase1 mRNA translation, causing the blockade of adenosine-mediated immunosuppression.

NTPDase1 also can be targeted by miR-150 (Figure 4), which is overexpressed in PCa and works by suppressing the p27 gene to promote the prostate cancer development (38). In this case, miR-150 may target NTPDase1, inhibiting the translation of this enzyme. As a result of this process, the hydrolysis of ATP and ADP may be impaired, also impacting purinergic signalling by adenosine. Another miRNA targeting NTPDase1 is miR-9 (Figure 4) which may act silencing NTPDase1 due to its oncogenic nature.

Finally, all of these miRNA's regulation hypothesis were supported by the validation miRNA-mediated effects on the purinergic signalling pathway in PCa via the libraries available from the SRA. In fact, the analysis of these libraries confirmed the involvement of miR-10, miR-150, miR-182, miR-183, miR205, miR301, miR-30, miR-424 and miR-9 in the expression of the investigated enzymes and receptor. This evidence shows the success of the methodology used for miRNA prediction. However, it is important to emphasise the need to better investigate the levels of adenosine and enzyme activity in PCa.

5. Conclusions

Adenosine is a key factor in tumour development since it plays a range of physiological and pharmacological effects in the prostate. Moreover, the results regarding miRNA prediction by bioinformatics analysis clarified some aspects of the modulation of purinergic signalling in normal and tumour-derived prostate cell lines. Altogether, these data helped us to understand the molecular mechanisms of PCa

and provided possible candidates as biomarkers for this cancer.

Signalling via A_3 adenosine receptors is crucial for tumour suppression in PCa, since it activates apoptotic mechanisms, and our results suggest that some miRNAs are possibly silencing this receptor in order to promote cancer development. Moreover, our results indicate that extracellular adenosine production is being affected by miRNA-mediated regulation of the ENTPDases 1 and ecto-5' nucleotidase. Consequently, immunosuppression and apoptotic pathways may be repressed, which affects the development of PCa.

Finally, our study provides new insights into the involvement of purines and purinergic signalling in PCa. However, further investigations are necessary to validate the participation of ectonucleotidases in PCa, such as the analysis of the hydrolysis of nucleotides and the protein expression analysis of these enzymes, as well as the expression of identified miRNAs in these cell lines.

Acknowledgements

Not applicable.

Funding

This study was supported by grants from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - Process: 446747/2014-9.

Availability of data and materials

All data generated or analysed during the present study are included in this published article.

Author contributions

CRF designed, supervised and managed the entire study. TGA participated in the discussion, data interpretation and in the study coordination. CAM and STSM performed the experiments and drafted the manuscript. TCA, LRA and MSG collaborated on the *in silico* validation of the miRNAs by small RNA-seq libraries analysis. FMO contributed to the statistical analysis. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Conflict of interest

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

References

1. Battisti V., Maders L.D.K., Bagatini M.D., Battisti I.E, Belle´ L.P., Santos K.F., Maldonado P.A., Thomé G.R., Schetinger M.R.C. and Morsch V.M.: Ectonucleotide pyrophosphatase/phosphodiesterase (E-NPP) and adenosine deaminase (ADA) activities in prostate cancer patients: influence of Gleason score, treatment and bone metastasis. Biomedicine & Pharmacotherapy, 67: 203-208, 2013.

2. Janssens R. and Boeynaems J.M.: Effects of extracellular nucleotides and nucleosides on prostate carcinoma cells. British Journal of Pharmacology, 132: 536-546, 2001.

3. Li W-H., Qiu Y., Zhang H-Q, Liu Y, You JF, Tian XX and Fang WG: P2Y2 receptor promotes cell invasion and metastasis in prostate cancer cells. British Journal of Cancer, 109: 1666-75, 2013.

4. Burnstock G.: Purinergic signalling: pathophysiology and therapeutic potential. The Keio Journal of Medicine, 62: 63-73, 2013.

5. Stagg J., Thompson L. and Dwyer, K.: Ectonucleotidases in cancer and inflammation. Journal of Biomedicine and Biotechnology, 2012: 1-3, 2012.

6. Schmid S., Ayata K., Kübler M., Lazar Z, Haager B, Hoßfeld M, Meyer A, Cicko S, Elze M, et al: Altered purinergic signalling in the tumour associated immunologic microenvironment in metastasized non-small-cell lung cancer. Lung Cancer, 3: 516-521, 2015.

7. Di Virgilio F: Purines, purinergic receptors, and cancer. The Journal of Cancer Research, 72: 5441-7, 2012.

8. Feng L., Tapper E.B., Sun X., Gehring M., Robson S.C. and Wu Y.: Purinergic modulation and CD39/ENTPD1 in cancer. Frontiers in Anticancer Drug Discovery, 64: 229-292, 2014.

9. Beckenkamp A., Santana D.B., Bruno A.N., Calil LN, Casali EA, Paccez JD, Zerbini LF, Lenz G, Wink MR et al: Ectonucleotidase expression profile and activity in human cervical cancer cell lines. Biochemistry Cell Biology, 2: 95-104, 2014.

10. Stella J., Bavaresco L., Braganhol E., Rockenbach L, Farias PF, Wink MR, Azambuja AA, Barrios CH, Morrone FB, et al: Differential ectonucleotidase expression in human bladder cancer cell lines. Urologic Oncology: Seminars and Original Investigations, 28: 260-7. 2010.

11. Young A., Mittal D., Sttag J. and Smyth MJ.: Targeting cancer-derived adenosine: new therapeutic approaches. American Association for Cancer Research Journal, 4: 1-10, 2014.

12. Jackson B.L., Grabowska A. and Ratan H.L. MicroRNA in prostate cancer: functional importance and potential as circulating biomarkers. BMC Cancer, 14: 1-10, 2014.

13. Kobayashi N., Uemura H., Nagahama K., Okudela K, Furuya M, Ino Y, Ito Y, Hirano H, Inayama Y, et al: Identification of miR-30d as a novel prognostic marker of prostate cancer. Oncotarget, 3: 1455-71, 2012.

14. Shi X-B., Tepper C.G. and White R.W.: MicroRNAs and prostate cancer. J.Cell.Mol. 12: 1456-65, 2008.

15. Kong Y.W., Ferland-McCollough D., Jackson T.J. and Bushell M.: MicroRNAs in cancer management. The Lancet Oncology, 13: 249-58, 2012.

16. White N. and Burnstock G.: P2 receptors and cancer. Trends in Pharmacological Sciences, 27: 211-7, 2006.

17. Livak KJ and Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCq method. Methods 25: 402-408, 2001.

18. Kroll, K. W. et al.: Quality control for RNA-Seq (QuaCRS): an integrated quality control pipeline. Cancer Inform., 13: 7-14, 2014.

19. Bolger, A. M., Lohse, M., Usadel, B.: Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30: 2114-2120, 2014.

20. Friedländer, M.R., MacKowiak, S.D., Li, N., Chen, W., Rajewsky, N.: miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res., 40: 37-52, 2012.

21. Graphpad Software. CEO: GraphPad Prism. 2016. http://www.graphpad.com/scientific-software/prism/. Accessed 25 May 2016.

22. R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. Accessed 14 December 2018.

23.Burnstock G., Vaughn B. and Robson S.C.: Purinergic signalling in the liver in health and disease. Purinergic Signalling, 10: 51-70, 2013.

24. Cappellari A.R., Rockenbach L., Dietrich F., Clarimundo V, Glaser T, Braganhol E, Abujamra AL, Roesler R, Ulrich H, et al: Characterization of ectonucleotidases in human medulloblastoma cell lines: ecto-5' NT/CD73 in metastasis as potential prognostic factor. Plos One, 7: 1-11, 2012.

25. Virtanen S.S., Kukkonen-Macchi A., Vainio M., Elima K., Härkönen P.L., Jalkanen S. and Yegutkin G.G.: Adenosine inhibits tumour cell invasion via receptor-independent mechanisms. Molecular Cancer Research, 12:1863-74, 2014.

26. Aghaei M., Karami-Tehrani F., Panjehpour M., Salami S. and Fallahian F.: Adenosine induces cell cycle arrest and apoptosis in androgen-dependent and independent prostate cancer cell lines, LNCap-FGC-10, DU-145, and PC3. The Prostate, 72: 361-75, 2012.

27. Nogi Y., Kanno T., Nakano T., Fujita Y., Tabata C., Fukuoka K., Gotoh A. and Nishizaki T.: AMP converted from intracellularly transported adenosine upregulates p53 expression to induce malignant pleural mesothelioma cell apoptosis. Cellular Physiology and Biochemistry, 30:61-74, 2012.

28. Xie H., Li L., Zhu G., Dang Q., Ma Z., He D., Chang L., Song W., Chang H-C. et al: Infiltred pre-adipocytes increase prostate cancer metastasis via modulation of the miR-301a/ androgen receptor (AR)/TGF-β1/Smad/ MMP9 signals. Oncotarget, 6: 12326-12339, 2015.

29. Ueno K., Hirata H., Shahryari V., Deng G., Tanaka Y., Tabatabai Z.L., Hinoda Y. and Dahiya R.: MicroRNA-183 is an oncogene targeting Dkk-3 and SMAD4 in prostate cancer. British Journal of Cancer, 8: 1659-67, 2013.

30. Mihelich B.L., Dambal S., Lin S. and Nonn L.: miR-182, of the miR-183 cluster family, is packaged in exosomes and is detected in human exosomes from serum, breast cells and prostate cells. Oncology Letters, 12: 1197-1203, 2016.

31. Yao J., Xu C., Fang Z., Li Y., Liu H., Wang Y., Xu C., Sun Y. et al: Androgen receptor regulated microRNA miR-182-5p promotes prostate cancer progression by targeting the ARRDC3/ITGB4 pathway. Biochemical and Biophysical Research Communication, 474: 213-9, 2016.

32. Dallavalle C., Albino D., Civenni G., Merulla J., Ostano P., Mello-Grand M., Rossi S., Losa M., D'Ambrosio G. et al: MicroRNA-424 impairs ubiquitination to activate STAT3 and promote prostate tumour progression. The Journal of Clinical Investigation, 126: 1-18, 2016.

33. Seashols-Williams S.J., Budd W., Clark G.C., Wu Q., Daniel1 R., Dragoescu E., and Zehner Z.E.: miR-9 acts as an OncomiR in prostate cancer through multiple pathways that drive tumour progression and metastasis. PLoS One, 11: 1-20, 2016.

34. Ma L., Teruya-Feldstein J. and Weinberg, R.A.: Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. Nature, 449: 682-688, 2007.

35. Prueitt R.L., Yi M., Hudson R.S., Wallace T.A., Howe T.M., Yfantis H.G., Lee D.H., Stephens R.M. et al: Expression of microRNAs and protein-coding genes associated with perineural invasion in prostate cancer. The Prostate, 68: 1152-64, 2008.

36. Kao C-J., Martiniez A., Shi X-B., Yang J., Evans C.P., Dobi A., deVere White RW and Kung H-J: miR-30 as a tumour suppressor connects EGF/Src signal to ERG and EMT. Oncogene, 33: 2495-2503, 2013.

37. Wang N., Li Q., Feng N-H., Cheng G., Guan Z-L., Wang Y., Qin C., Yin C-J.
and Hua L-X.: miR-205 is frequently downregulated in prostate cancer and acts as a tumour suppressor by inhibiting tumour growth. Asian Journal of Andrology, 15:735-41, 2013.

38. Liu D.Z, Zhang H.Y., Long X.L., Zou S.L., Zhang X.Y, Han G.Y. and Cui Z.G: MIR-150 promotes prostate cancer stem cell development via suppressing p27Kip1. European Review for Medical and Pharmacological Sciences, 19:4344-52, 2015.

4 CONCLUSÃO

A partir da análise das amostras realizadas pelo mirDeep2 pode-se constatar que há possíveis miRNAs presentes nos tecidos cancerosos. Através de uma das duas planilhas de saída do mirDeep2, foi possível recuperar cada ID de miRNA identificado pelo programa que estava envolvido com as enzimas ENTPDase1 e ecto-5'-nucleotidase, bem como do receptor de adenosina A3 no câncer de próstata.

E para verificar qual destino destes miRNAs encontrados utilizando somente ferramentas de bioinformática, complementando o estudo anteriormente iniciado houve a predição dos alvos dos miRNAs e também foram reunidas informações dos SRRs que foram selecionados para o mirdeep2.

O conjunto de dados dos SRRs referente ao trabalho do qual se originaram precisou ser reunido para ficar mais adequado de ser compreendido de onde e o que continham as amostras utilizadas para análise computacional. Estas informações estão presentes no Apêndice 1 Tabela SRR.

Este trabalho encontrou outros alvos dos miRNAs mostrados na tabela do Apêndice 2. Nela estão reunidas informações pesquisadas por este trabalho sobre a porção de origem do miRNA proveniente da análise do mirDeep2 e suas identificações; foram procurados os alvos de cada miRNA no Target Scan Vert e miRDB; os alvos foram sumarizados e a tabela mostrou algumas similaridades bem como as variedades na ordem de apresentação dos mesmos.

Este trabalho faz parte de uma pesquisa maior e uma sugestão para compor os dados mostrados neste trabalho, seria pesquisar os sinônimos dos nomes das enzimas ENTPDase 1, ENTPDase 3 e ecto-5'nucleotidase e as evidenciar nos resultados gerados por Target Scan Vert e miRDB em uma tabela, para cada miRNA identificado pelo mirDeep2, para ficar mais fácil de entender em qual local do *ranking* também estão presentes como alvos.

REFERÊNCIAS

1. AGARWAL, V.; BELL, J. W.; NAM, J.W.; BARTEL, D. P. Predicting effective microRNA target sites in mammalian mRNAs. **eLife**, Germany, v. 4, p. 1-38, 2015. https://doi.org/10.7554/eLife.05005

2. BANWAIT, J., K.; BASTOLA, D. R. Contribution of bioinformatics prediction in microRNA-based cancer therapeutics. **Advanced Drug Delivery Reviews Elsevier B.V.**, v. 81, p. 94 – 103, 2015.

https://doi.org/10.1016/j.addr.2014.10.030

3. BARIK, S.; RAI, N.; MISHRA, P.; SINGH S. K.; GAUTAM, V. Bioinformatics: how it helps to boost modern biological research. **CURRENT SCIENCE**, India, v. 118, n. 5, p. 698-699, 2020.

4. CAI, Y.; YU, X.; HU, S.; YU, J. A Brief Review on the Mechanisms of miRNA Regulation. **Genomics Proteomics Bioinformatics**, v. 7, n. 4, p. 147 - 154, dec., 2009.

https://doi.org/10.1016/S1672-0229(08)60044-3

5. CARDOSO, T. C. DE S.; ALVES, T. C.; CANESCHI, C. M; SANTANA, D. DOS R. G.; FERNANDES-BRUM, C. N.; REIS, G. L. DOS; DAUDE, M. M.; RIBEIRO, T. H. C.; GÓMEZ, M. M. D.; LIMA, A. A.; GOMES, L. A. A.; GOMES, M. DE S.; GANDOLF, P. E.; AMARAL, L. R. DO; CHALFUNJÚNIOR, A.; MALUF, W. R.; GOMES, M. DE S. New insights into tomato microRNAs. **Scientific Reports**, p. 1–22, 2018.

https://doi.org/10.1038/s41598-018-34202-3

6. CHEN, Y.; WANG, X. miRDB: an online database for prediction of functional microRNA targets. **Nucleic Acids Research**, USA, v. 48, p. 127-131, 2019.

https://doi.org/10.1093/nar/gkz757

7. DANG, T. H. Y.; TYAGI, S.; D'CUNHA, G.; BHAVE, M.; CRAWFORD, R.; IVANOVA, E. P. Computational prediction of microRNAs in marine bacteria of the genus *Thalassospira*. **PLOS ONE**, Australia, v. 14, n. 3, p. 1-17, 2019.

https://doi.org/10.1371/journal.pone.0212996

8. DAVEY, J. W.; HOHENLOHE, P. A.; ETTER, P. D.; BOONE||, J. Q.; CATCHEN; J. M.; BLAXTER, M. L. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. **NATURE REVIEWS GENETICS**, UK, v.12, p. 499-510, 2011.

https://doi.org/10.1038/nrg3012

9. ESCALONA, M.; ROCHA, S.; POSADA, D. A comparison of tools for the simulation of genomic next-generation sequencing data. **NATURE REVIEWS GENETICS**, Spain, v. 17, p. 459-469, 2016.

https://doi.org/10.1038/nrg.2016.57

10. ESTEVES, J. V. D. C. **microRNAs 29b, 29c, 199a e 532-3p são potenciais repressores da expressão de GLUT4 e HK2 em músculo esquelético de ratos diabéticos**. 2016. Tese (Doutorado em Ciências) - Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2016.

11. FRIEDLANDER, M., R.; MACKOWIAK, S., D.; LI, N.; CHEN, W.; RAJEWSKY, N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. **Nucleic Acids Research Oxford University Press.**, v.40, n.1, p. 37 – 52, 2012.

https://doi.org/10.1093/nar/gkr688

12. FUZIWARA, C. S. **Influência do microRNA let-7 e miR-17-92 como oncomiRs no câncer**. 2010. Dissertação (Mestre em Ciências) – Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2010.

13. GONÇALVES, M. W. **O Sequenciamento genômico**. Disponível em:<http://professor.pucgoias.edu.br/SiteDocente/admin/arquivosUpload/18497/material/0

5.%20M%C3%A9todos%20de%20Sequ%C3%AAnciamento%20do%20DNA.pdf>. Acesso em: mar. 2019.

14. GRIFFITHS-JONES, S.; GROCOCK, R. J.; VAN DONGEN, S.; BATEMAN, A.; ENRIGHT'A. J. miRBase: microRNA sequences, targets and gene nomenclature. **Nucleic Acids Research**, UK, v. 34, p. 140-144, 2006.

https://doi.org/10.1093/nar/gkj112

15. GULYAEVA, L. F.; KUSHLINSKIY, N. E. Regulatory mechanisms of microRNA expression. **Journal of Translational Medicine**, v. 14, e143, 2016.

https://doi.org/10.1186/s12967-016-0893-x

16. HU, Y., LAN, W., MILLER, D., 2017. Next-Generation Sequencing for MicroRNA Expression Profile. In: Huang, J., et al. **Bioinformatics in MicroRNA Research**. 2017. Cap. 12, p. 169 – 177.

https://doi.org/10.1007/978-1-4939-7046-9_12

17. JACKSON, B.L.; GRABOWSKA, A.; RATAN, H.L. MicroRNA in prostate cancer: functional importance and potential as circulating biomarkers. **BMC Cancer**, v. 14: p. 1-10, 2014.

https://doi.org/10.1186/1471-2407-14-930

18. KIRCHER; M.; KELSO, J. High-throughput DNA sequencing –concepts and limitations. **Bioessays WILEY Periodicals**, Germany, v. 32: p. 524–536, 2010.

https://doi.org/10.1002/bies.200900181

19. KLUIVER, J.; NIU, F.; YUAN, Y.; KOK, K.; BERG, A. V. D.; DZIKIEWICZ-KRAWCZYK, A. NGS-Based High-Throughput Screen to Identify MicroRNAsRegulating Growth of B-Cell Lymphoma. In: Kuppers, R. **Lymphoma: MethodsandProtocols**. Springer Nature, 2019, Cap. 12, p. 269 – 282.

https://doi.org/10.1007/978-1-4939-9151-8_12

20. KONG, Y.W.; FERLAND-MCCOLLOUGH, D.; JACKSON, T.J.; BUSHELL, M.: MicroRNAs in cancer management. **The Lancet Oncology**, v. 13: p. 249-58, 2012.

https://doi.org/10.1016/S1470-2045(12)70073-6

21. KULCHESKI., F. R., MARGIS, R. Pequenos RNAs e aplicações de RNAi em plantas. In: **Biotecnologia Aplicada à Agro & Indústria**. 2017. Cap. 19, p. 675 – 708.

https://doi.org/10.5151/9788521211150-19

22. KULSKI, J. K. Next-Generation Sequencing - An Overview of the History, Tools, and "Omic" Applications. In: INTECH OPEN SCIENCE. **We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists**. Japan, 2016. Cap. 1, p.1.

https://doi.org/10.5772/61964

23. LEINONEN, R.; SUGAWARA, H.; SHUMWAY, M.; INSDC. The Sequence Read Archive. **Nucleic Acids Research**, UK-Japan-USA, v. 39, p. 19-21, 2010.

https://doi.org/10.1093/nar/gkq1019

24. LEIPZIG, J. A review of bioinformatic pipeline frameworks. **Briefings in Bioinformatics Oxford**, v. 18, n. 3, p. 530 – 536, 2017.

25. LIU, S; LIU, Z. Bioinformatics in Aquaculture: Principles and Methods. Liu, Z. UK: John Wiley & Sons Ltd, 2017.

https://doi.org/10.1002/9781118782392

26. MARTINS, C. **Sequenciamento de genomas: princípios e métodos clássicos**. Disponível em:<http://www.ibb.unesp.br/Home/Departamentos/Morfologia/Laboratorios/LaboratoriodeGenomicaIntegrativa/ag2013-3-sequenc-dna-cmartins.pdf>. Acesso em: mar. 2019.

27. MENON, S.; LUI, V. C. H.; TAM, P. K. H. Bioinformatics Tools and Methods to Analyze Single-Cell RNA Sequencing Data. **INTERNATIONAL JOURNAL OF INNOVATIVE SCIENCE AND RESEARCH TECHNOLOGY**, China, v. 6, n.8 , p. 282-288, 2021.

28. NAZARÉ, E.; GONÇALVES, F.; CAPELO, R. **Sequenciamento de DNA e Aplicações**. Disponível em:<https://edisciplinas.usp.br/pluginfile.php/2346804/mod_resource/content/1/Sequenciamento%20de%20DNA%20e%20Aplica%C3%A7%C3%B5es.pdf>. Acesso em: mar. 2019.

29. OLIVEIRA, K. P. V. **Análise comparativa da evolução de miRNAs utilizando insetos como modelo**. 2016. Dissertação (Mestrado em Bioinformática) – Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, jan. 2016.

30. ORANG, A.V.; SAFARALIZADEH, R.; BAVILI, M.K. Mechanisms of miRNA-Mediated Gene Regulation from Common Downregulation to mRNA-Specific Upregulation. **Hindawi Publishing Corporation International Journal of Genomics**, v. 2014, p. 1 - 15, aug., 2014.

https://doi.org/10.1155/2014/970607

31. PAUL, P.; CHAKRABORTY, A.; SARKAR, D.; LANGTHASA, M.; RAHMAN, M.; BARI, M.; SINGHA, R. K. S.; MALAKAR, A. K., CHAKRABORTY, S. Interplay between miRNAs and human diseases. **Wiley Periodicals, Inc.**, v. 2018, n. e233, p. 2007 - 2018, feb., 2017.

https://doi.org/10.1002/jcp.25854

32. PINTO, L. S.; KREMER, F. S. **Plataformas de Sequenciamento de Nova Geração e Pré-processamento de dados**. Disponível em:<http://labbioinfo.ufpel.edu.br/aulas_2016/Capitulo%203.pdf>. Acesso em: mar. 2019.

33. SAHU, P. K.; SAO, R.; MONDAL,S.; VISHWAKARMA , G.; GUPTA, S. K.; KUMAR, V.; SINGH, S.; SHARMA, D.; DAS, B. K. Next Generation Sequencing Based Forward Genetic Approaches for Identifification and Mapping of Causal Mutations in Crop Plants: A Comprehensive Review. **Plants**, India, v. 9, n. 1355, p. 1-47, 2020.

https://doi.org/10.3390/plants9101355

34. SCHMIDT, J. K.; BLOCK, L. N.; GOLOS, T. G. Defining the rhesus macaque placental miRNAome: conservation of expression of placental miRNA clusters between the macaque and human. **HHS Public Access**, USA, v. 65,n. , p. 1-21, 2018. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6007866/pdf/nihms960226.pdf>. DOI: 10.1016/j.placenta.2018.04.003. Acesso em: 2021.

https://doi.org/10.1016/j.placenta.2018.04.003

35. SHI, X-B.; TEPPER, C.G.; WHITE, R.W.: MicroRNAs and prostate cancer. **J.Cell.Mol**. v. 12: p. 1456-65, 2008.

https://doi.org/10.1111/j.1582-4934.2008.00420.x

36. SOARES, E.; SOARES, M.; SIQUEIRA, J.; LOPES, R.; MISAEL, B.; PRELLWITZ, I.; REIS, L.; FERREIRA, G. **ANÁLISE EXPERIMENTAL E *IN SILICO* DE VIROMAS: DA DESCOBERTA DE NOVOS VÍRUS À MONTAGEM DE GENOMAS VIRAIS.** Disponível em:<https://quelopes.github.io/files/courses/XSummerINCA/Apostila_2018_curso1.pdf>. Acesso em: mar. 2019.

37. WHAT IS MIRDEEP2. **mdc-berlin**, 2011. Disponível em: <https://www.mdc-berlin.de/content/mirdeep2-documentation>. Acesso em: 29 de ago. de 2020.

38. WOOD, C. C. The computational stance in biology. **Philosophical Transactions Royal Society Publishing**, v. 374, p. 1- 5, jan., 2019.

https://doi.org/10.1098/rstb.2018.0380

39. WU, L. **Bioinformatic and phylogenetic analyses of retroelements in bacteria**. 2018. Tese (Doctor of Philosophy) - Program in Biological Sciences, University of Calgary. Alberta, 2018.

40. PEVSNER, J. **Bioinformatics and functional genomics**. 3 ed. Maryland: Wiley Blackwell, 2015.

41. ZHAO, S.; GORDON, W.; DU, S.; ZHANG, C.; HE, W.; XI, L.; MATHUR, S.; AGOSTINO, M.; PARADIS, T.; SCHACK, D. V.; VINCENT, M.; ZHANG, B. QuickMIRSeq: a pipeline for quick and accurate quantification of both known miRNAs and isomiRs by jointly processing multiple samples from microRNA Sequencing. **BMC Bioinformatics**, v. 2017, n.18, e180, mar., 2017.

https://doi.org/10.1186/s12859-017-1601-4

APÊNDICE 1- Tabela SRR

Run	Bases	Size	GC content	Experiment	Platform	Strategy	Source	Organism	Bioproject			Taxonomy analysis	
									Link	Paper	Title	Identified reads	Unidentified reads
SRR5453989	710.0Mbp	437.0M	48.6%	SRX2742155	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	97.38%	2.62%
SRR5453990	632.8Mbp	391.2M	46.6%	SRX2742156	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	96.92%	3.08%
SRR5453991	732.5Mbp	449.1M	47%	SRX2742157	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	96.88%	3.12%
SRR5453992	584.2Mbp	356.9M	45.9%	SRX2742158	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	96.98%	3.02%
SRR54	752.2Mb	461.0M	46.6%	SRX274215	Illumina	RNA-Seq	Transcripto	*Homo sapien*	PRJNA383	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate	93.9%	6.1%

53993	p			9			mic	*s*	159		Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)		
SRR53994	598.6Mbp	366.8M	47.8%	SRX274216 0	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	96.79%	3.21%
SRR53995	808.0Mbp	493.7M	47.5%	SRX274216 1	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	95.92%	4.08%
SRR53996	630.2Mbp	382.4M	46.3%	SRX274216 2	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	93.76%	6.24%
SRR53997	516.9Mbp	313.5M	45.9%	SRX274216 3	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	97.21%	2.79%
SRR53998	527.9Mbp	323.7M	47.3%	SRX274216 4	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	97.53%	2.47%
SRR53999	547.0Mbp	339.8M	47.4%	SRX274216 5	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	94.75%	5.25%
SRR540	563.7Mbp	348.2M	49%	SRX274216 6	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	90.42%	9.58%

00													
SRR54539888	893.4Mbp	548.1M	48.1%	SRX2742154	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	93.32%	6.68%
SRR54539887	592.7Mbp	359.5M	47.3%	SRX2742153	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	93.62%	6.38%
SRR54539886	832.9Mbp	512.9M	47.8%	SRX2742152	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	96.39%	3.61%
SRR54539885	746.1Mbp	459.8M	49.2%	SRX2742151	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	93.82%	6.18%
SRR54539884	835.4Mbp	509.7M	46.6%	SRX2742150	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	97.4%	2.6%
SRR54539883	452.5Mbp	273.6M	46.2%	SRX2742149	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	96.24%	3.76%
SRR54539882	650.9Mbp	399.3M	47.7%	SRX2742148	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	93.14%	6.86%

SRR54539381	202.2Mbp	139.4M	47.2%	SRX2742147	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	92.59%	7.41%
SRR54539380	139.5Mbp	95.6M	47.9%	SRX2742146	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	95.38%	4.62%
SRR54539379	525.2Mbp	359.6M	48.3%	SRX2742145	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	96.54%	3.46%
SRR54539378	216.7Mbp	146.7M	48.5%	SRX2742144	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	70.44%	29.56%
SRR54539377	260.8Mbp	177.1M	50.1%	SRX2742143	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	92.67%	7.33%
SRR54539376	427.4Mbp	290.9M	48.6%	SRX2742142	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	94.74%	5.26%
SRR54539375	1.1 Gbp	739.6M	50.7%	SRX2742141	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	18.11%	81.89%
SR	707.	481	50.3	SRX2	Illumi	RNA-	Trans	*Homo*	PRJN	SRP1	Next Generation Sequencing Facilitates Differential Expression Analysis of	94.1%	5.9%

R54539774	4Mbp	.0M	%	742140	na	Seq	criptomic	*sapiens*	A383159	04148	miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)		
SRR54539773	682.3Mbp	463.8M	48.4%	SRX2742139	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	89.11%	10.89%
SRR54539772	548.5Mbp	373.6M	48%	SRX2742138	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	93.94%	6.06%
SRR54539771	771.9Mbp	526.3M	49.1%	SRX2742137	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	71.4%	28.6%
SRR54539770	526.7Mbp	357.2M	49.3%	SRX2742136	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	96.44%	3.56%
SRR54539769	201.3Mbp	137.2M	7.3%	SRX2742135	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	95.82%	4.18%
SRR54539768	190.3Mbp	129.5M	50.2%	SRX2742134	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	95.87%	4.13%
SRR54	226.1Mb	153.7M	46.7%	SRX2742	Illumina	RNA-Seq	Transcripto	*Homo sapien*	PRJNA383	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate	94.44%	5.56%

53967	p			3			mic	*s*	159		Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)		
SRR53966	156.9Mbp	106.6M	51.3%	SRX2742132	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	90.88%	9.12%
SRR53965	228.1Mbp	156.9M	49.4%	SRX2742131	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	92.89%	7.11%
SRR53964	142.7Mbp	97.5M	49.1%	SRX2742130	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	92.29%	7.71%
SRR53963	87.6Mbp	51.5M	54.2%	SRX2742129	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	16.2%	83.8%
SRR53962	104.9Mbp	60.7M	53.7%	SRX2742128	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	17.91%	82.09%
SRR53961	80.8Mbp	46.8M	53.5%	SRX2742127	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	20.99%	79.01%
SRR539	73.7Mbp	42.7M	54.2%	SRX2742126	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	18.04%	81.96%

60													
SRR54539559	79.3 Mbp	45.9M	53.3%	SRX2742125	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	19.39%	80.61%
SRR54539558	82.3 Mbp	48.2M	53.8%	SRX2742124	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	19.71%	80.29%
SRR54539557	84.1 Mbp	49.2M	54.3%	SRX2742123	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	13.55%	86.45%
SRR54539556	71.9 Mbp	41.6M	54.4%	SRX2742122	Illumina	RNA-Seq	Transcriptomic	*Homo sapiens*	PRJNA383159	SRP104148	Next Generation Sequencing Facilitates Differential Expression Analysis of miRNA Expression In the Whole Blood Samples Obtained From Prostate Cancer Patients vs. Controls (DOI:10.3390/ijms18061281)	17.07%	82.93%

Fonte: Própria.

APÊNDICE 2- miRNAs e alvos

mirbase				Target Scan Vert					miRDB				
miRNA	**ID miRNA**	**5p Accession**	**3p Accession**	**Ranking of predicted miRNA's targets**	**Gene name**	**Descripition**	**Symbol**	**Number of predicted target genes**	**Ranking of predicted miRNA's targets**	**Gene name**	**Descripition**	**Symbol**	**Number of predicted target genes**
hsa-miR-10-a-5p	MIMAT0000253	MIMAT0000253	MIMAT0004555	1	Brain-derived neurotrophic factor	This gene may play a role in the regulation of the stress response and in the biology of mood disorders	BDNF	338	1	Cell adhesion molecule 2	This gene encodes a member of the synaptic cell adhesion molecule 1 (SynCAM) family which belongs to the immunoglobulin (Ig) superfamily	CADM2	352
				2	Arylsulfatase family, member J	They are involved in hormone biosynthesis, modulation of cell signaling, and degradation of macromolecul	ARSJ		2	Transcription factor AP-2 gamma	The protein encoded by this gene is a sequence-specific DNA-binding transcription factor involved in the activation of several developmental	TFAP2C	

						es					genes		
				3	Cytokine receptor-like factor 3	This gene is one of several genes located in the neurofibromatosis type I tumor suppressor region on the q arm of chromosome 17, a region that is subject to microdeletions, duplications, chromosomal breaks and rearrangements	CRLF3		3	CCR4-NOT transcription complex subunit 6	The CCR4-NOT complex plays a role in many cellular processes, including miRNA-mediated repression, mRNA degradation, and transcriptional regulation	CNOT6	
				4	Transcription factor AP-2 gamma	The protein encoded by this gene is a sequence-specific DNA-binding transcription factor involved	TFAP2C		4	RAR related orphan receptor A	It can bind as a monomer or as a homodimer to hormone response elements upstream of several genes to enhance the expression of those	RORA	

						in the activation of several developmental genes					genes		
				5	Homeobox A3	This gene is part of the A cluster on chromosome 7 and encodes a DNA-binding transcription factor which may regulate gene expression, morphogenesis, and differentiation	HOXA3		5	Polypeptide N-acetylgalactosaminyltransferase 1	GalNAc-Ts initiate mucin-type O-linked glycosylation in the Golgi apparatus by catalyzing the transfer of GalNAc to serine and threonine residues on target proteins	GALNT1	
				6	Von Willebrand factor C domain containing protein 2-like	Restricted expression toward brain	VWC2L		6	E2F transcription factor 7	Play an essential role in the regulation of cell cycle progression	E2F7	
				7	Solute carrier family 6 (neutral amino acid transporter), member 19	This gene encodes a system B(0) transmembrane protein that actively	SLC6A19		7	Cytokine receptor-like factor 3	This gene is one of several genes located in the neurofibromatosis type I tumor suppressor region	CRLF3	

						transports most neutral amino acids across the apical membrane of epithelial cells					on the q arm of chromosome 17, a region that is subject to microdeletions, duplications, chromosomal breaks and rearrangements		
				8	Ring finger protein 186		RNF186		8	Potassium voltage-gated channel subfamily A member 6	Their diverse functions include regulating neurotransmitter release, heart rate, insulin secretion, neuronal excitability, epithelial electrolyte transport, smooth muscle contraction, and cell volume	KCNA6	
				9	Sine oculis binding protein homolog	The protein encoded by this gene is a nuclear zinc finger protein that is involved in development of the cochlea.	SOBP		9	Sine oculis binding protein homolog	The protein encoded by this gene is a nuclear zinc finger protein that is involved in development of the cochlea. Defects in this gene have also been linked to	SOBP	

						Defects in this gene have also been linked to intellectual disability					intellectual disability		
				10	transferrin receptor	This gene encodes a cell surface receptor necessary for cellular iron uptake by the process of receptor-mediated endocytosis	TFRC		10	c		ELOVL2	
hsa-miR-10-b-5p	MIMAT0000254	MIMAT0000254	MIMAT0004556	1	Brain-derived neurotrophic factor	This gene may play a role in the regulation of the stress response and in the biology of mood disorders	BDNF	338	1	Cell adhesion molecule 2	This gene encodes a member of the synaptic cell adhesion molecule 1 (SynCAM) family which belongs to the immunoglobulin (Ig) superfamily	CADM2	352
				2	Arylsulfatase family, member J	They are involved in hormone biosynthesis,	ARSJ		2	Transcription factor AP-2 gamma	The protein encoded by this gene is a sequence-specific	TFAP2C	

						modulation of cell signaling, and degradation of macromolecules					DNA-binding transcription factor involved in the activation of several developmental genes		
				3	Cytokine receptor-like factor 3	This gene is one of several genes located in the neurofibromatosis type I tumor suppressor region on the q arm of chromosome 17, a region that is subject to microdeletions, duplications, chromosomal breaks and rearrangements	CRLF3		3	CCR4-NOT transcription complex subunit 6	The CCR4-NOT complex plays a role in many cellular processes, including miRNA-mediated repression, mRNA degradation, and transcriptional regulation	CNOT6	
				4	Transcription factor AP-2 gamma	The protein encoded by this gene is a	TFAP2C		4	polypeptide N-acetylgalactosaminyltransferase 1	GalNAc-Ts initiate mucin-type O-linked glycosylation in the	GALNT1	

						sequence-specific DNA-binding transcription factor involved in the activation of several developmental genes					Golgi apparatus by catalyzing the transfer of GalNAc to serine and threonine residues on target proteins		
				5	Homeobox A3	This gene is part of the A cluster on chromosome 7 and encodes a DNA-binding transcription factor which may regulate gene expression, morphogenesis, and differentiation	HOXA3		5	RAR related orphan receptor A	It can bind as a monomer or as a homodimer to hormone response elements upstream of several genes to enhance the expression of those genes	RORA	
				6	Von Willebrand factor C domain containing protein 2-like	Restricted expression toward brain	VWC2L		6	E2F transcription factor 7	Play an essential role in the regulation of cell cycle progression	E2F7	
				7	Solute carrier family 6	This gene	SLC6A1		7	Cytokine	This gene is one of	CRLF3	

					(neutral amino acid transporter), member 19	encodes a system B(0) transmembrane protein that actively transports most neutral amino acids across the apical membrane of epithelial cells	9			receptor-like factor 3	several genes located in the neurofibromatosis type I tumor suppressor region on the q arm of chromosome 17, a region that is subject to microdeletions, duplications, chromosomal breaks and rearrangements		
				8	Ring finger protein 186		RNF186		8	Sine oculis binding protein homolog	The protein encoded by this gene is a nuclear zinc finger protein that is involved in development of the cochlea. Defects in this gene have also been linked to intellectual disability	SOBP	
				9	sine oculis binding protein homolog	The protein encoded by this gene is a nuclear zinc finger protein	SOBP		9	Potassium voltage-gated channel subfamily A member 6	Their diverse functions include regulating neurotransmitter release, heart rate,	KCNA6	

						that is involved in development of the cochlea. Defects in this gene have also been linked to intellectual disability					insulin secretion, neuronal excitability, epithelial electrolyte transport, smooth muscle contraction, and cell volume		
				10	Transferrin receptor	This gene encodes a cell surface receptor necessary for cellular iron uptake by the process of receptor-mediated endocytosis	TFRC		10	ELOVL fatty acid elongase 2		ELOVL 2	
hsa-miR-150-5p	MIMAT0000451	MIMAT0000451	MIMAT0004610	1	V-myb avian myeloblastosis viral oncogene homolog	This protein plays an essential role in the regulation of hematopoiesis. This gene may be	MYB	351	1	MYB proto-oncogene, transcription factor	This protein plays an essential role in the regulation of hematopoiesis. This gene may be aberrently expressed or rearranged or	MYB	902

						aberrently expressed or rearranged or undergo translocation in leukemias and lymphomas, and is considered to be an oncogene					undergo translocation in leukemias and lymphomas, and is considered to be an oncogene		
				2	Mdm4 p53 binding protein homolog (mouse)	Bind the p53 tumor suppressor protein and inhibit its activity, and have been shown to be overexpressed in a variety of human cancers	MDM4		2	Synaptic vesicle glycoprotein 2B	This protein and other members of the family are localized to synaptic vesicles and may function in the regulation of vesicle trafficking and exocytosis	SV2B	
				3	Endosulfine alpha	This protein was identified as an endogenous ligand for the sulfonylurea	ENSA		3	MDM4, p53 regulator	Bind the p53 tumor suppressor protein and inhibit its activity, and have been shown to be overexpressed in a	MDM4	

						receptor, ABCC8/SUR1. ABCC8 is the regulatory subunit of the ATP-sensitive potassium (KATP) channel, which is located on the plasma membrane of pancreatic beta cells and plays a key role in the control of insulin release from pancreatic beta cells					variety of human cancers		
				4	Adiponectin receptor 2	The adiponectin receptors, ADIPOR1 and ADIPOR2, serve as receptors for globular and	ADIPOR 2		4	Adiponectin receptor 2	The adiponectin receptors, ADIPOR1 and ADIPOR2, serve as receptors for globular and full-length adiponectin and	ADIPO R2	

						full-length adiponectin and mediate increased AMPK and PPAR-alpha ligand activities, as well as fatty acid oxidation and glucose uptake by adiponectin					mediate increased AMPK and PPAR-alpha ligand activities, as well as fatty acid oxidation and glucose uptake by adiponectin		
				5	Hypoxia inducible lipid droplet associated		HILPDA		5	Transcriptional adaptor 1	TADA1L is a protein subunit of the human STAGA complex, which is a chromatin-modifying multiprotein complex	TADA1	
				6	Transcriptional adaptor 1	TADA1L is a protein subunit of the human STAGA complex, which is a chromatin-modifying multiprotein	TADA1		6	Chromodomain helicase DNA binding protein 2	CHD genes alter gene expression possibly by modification of chromatin structure thus altering access of the transcriptional apparatus to its chromosomal DNA	CHD2	

						complex					template		
				7	mitochondrial carrier 2	This gene encodes a member of the SLC25 family of nuclear-encoded transporters that are localized in the inner mitochondrial membrane. This gene is one such locus, which is highly expressed in white adipose tissue and adipocytes, and thought to play a regulatory role in adipocyte differentiation and biology	MTCH2		7	Neuronal cell adhesion molecule	This ankyrin-binding protein is involved in neuron-neuron adhesion and promotes directional signaling during axonal cone growth. This gene is also expressed in non-neural tissues and may play a general role in cell-cell communication via signaling from its intracellular domain to the actin cytoskeleton during directional cell migration	NRCAM
				8	DDB1 and CUL4 associated factor 6	The encoded protein and	DCAF6		8	Pappalysin 1	This gene encodes a secreted	PAPPA	

						DNA damage binding protein 2 may act as tumor promoters and tumor suppressors, respectively, by regulating the level of androgen receptor in prostate tissues					metalloproteinase which cleaves insulin-like growth factor binding proteins		
				9	ELOVL fatty acid elongase 3	This gene encodes a protein that belongs to the GNS1/SUR4 family. Members of this family play a role in elongation of long chain fatty acids to provide precursors for synthesis of	ELOVL3		9	Brain abundant membrane attached signal protein 1	This gene encodes a membrane bound protein with several transient phosphorylation sites and PEST motifs	BASP1	

						sphingolipids and ceramides							
				10	IKBKB interacting protein		IKBIP		10	homeobox A2	In vertebrates, the genes encoding the class of transcription factors called homeobox genes are found in clusters named A, B, C, and D on four separate chromosomes. Expression of these proteins is spatially and temporally regulated during embryonic development	HOXA2	
hsa-miR-182-5p	MIMAT0000259	MIMAT0000259	MIMAT0000260	1	Protein phosphatase 1, regulatory (inhibitor) subunit 1C	Protein phosphatase-1 is a major serine/threonine phosphatase that regulates a variety of cellular functions	PPP1R1C	1310	1	Protein kinase cAMP-activated catalytic subunit beta	The encoded protein is a catalytic subunit of cAMP. cAMP signaling is important to a number of processes, including cell proliferaton and differentiation	PRKACB	1266
				2	Regulator of G protein signaling 17	This gene encodes a member of the	RGS17		2	Regulator of G protein signaling 17	This gene encodes a member of the regulator of	RGS17	

						regulator of G-protein signaling family					G-protein signaling family		
				3	LIM homeobox 1	The encoded protein is a transcription factor important for the development of the renal and urogenital systems	LHX1		3	Basonuclin 2	This gene encodes a conserved zinc finger protein. The encoded protein functions in skin color saturation	BNC2	
				4	Cortactin	This gene has two roles: (1) regulating the interactions between components of adherens-type junctions and (2) organizing the cytoskeleton and cell adhesion structures of epithelia and	CTTN		4	sorting nexin family member 30		SNX30	

						carcinoma cells							
				5	Neurocalcin delta	This gene encodes a member of the neuronal calcium sensor family of calcium-bindin g proteins	NCALD		5	LIM domain containing preferred translocation partner in lipoma	The encoded protein localizes to the cell periphery in focal adhesions and may be involved in cell-cell adhesion and cell motility	LPP	
				6	Hyaluronan synthase 2	It serves a variety of functions, including space filling, lubrication of joints, and provision of a matrix through which cells can migrate	HAS2		6	melanocyte inducing transcription factor	The encoded protein regulates melanocyte development and is responsible for pigment cell-specific transcription of the melanogenesis enzyme genes	MITF	
				7	ZFP36 ring finger protein-like 1	This putative nuclear transcription factor most likely functions in regulating the response to growth	ZFP36L 1		7	Fibroblast growth factor receptor substrate 2		FRS2	

						factors							
				8	Flotillin 1	This gene encodes an protein that localizes to the caveolae, which are small domains on the inner cell membranes	FLOT1		8	Calmodulin regulated spectrin associated protein family member 2		CAMSAP2	
				9	Paternally expressed 10	It is expressed in adult and embryonic tissues and reported to have a role in cell proliferation, differentiation and apoptosis	PEG10		9	Hyaluronan synthase 2	It serves a variety of functions, including space filling, lubrication of joints, and provision of a matrix through which cells can migrate	HAS2	
				10	Solute carrier family 35, member G1	The encoded protein may play a role in the regulation of calcium levels inside the cell	SLC35G1		10	Proline rich and Gla domain 3	The encoded protein is a member of a family of vitamin K-dependent transmembrane proteins which contain a glutamate-rich	PRRG3	

											extracellular domain		
hsa-miR-183-5p	MIMAT0000261	MIMAT0000261	MIMAT0004560	1	Nudix (nucleoside diphosphate linked moiety X)-type motif 4	The protein encoded by this gene regulates the turnover of diphosphoinositol polyphosphates. The turnover of these high-energy diphosphoinositol polyphosphates represents a molecular switching activity with important regulatory consequences. Molecular switching by diphosphoinositol polyphosphates may contribute to	NUDT4	502	1	profilin 2	The protein encoded by this gene is a ubiquitous actin monomer-binding protein belonging to the profilin family. It is thought to regulate actin polymerization in response to extracellular signals	PFN2	565

						regulating intracellular trafficking							
				2	Guanine nucleotide binding protein (G protein), gamma 5	G proteins are trimeric (alpha-beta-gamma) membrane-associated proteins that regulate flow of information from cell surface receptors to a variety of internal metabolic effectors	GNG5		2	Potassium two pore domain channel subfamily K member 10	The protein encoded by this gene belongs to the family of potassium channel proteins containing two pore-forming P domains. This channel is an open rectifier which primarily passes outward current under physiological K+ concentrations, and is stimulated strongly by arachidonic acid and to a lesser degree by membrane stretching, intracellular acidification, and general anaesthetics	KCNK10	
				3	Presenilin 2 (Alzheimer	Presenilins are	PSEN2		3	Myocyte enhancer	This protein may	MEF2C	

					disease 4)	postulated to regulate amyloid precursor proteinprocessing through their effects on gamma-secretase, an enzyme that cleaves amyloid precursor protein				factor 2C	play a role in maintaining the differentiated state of muscle cells		
				4	Profilin 2	The protein encoded by this gene is a ubiquitous actin monomer-binding protein belonging to the profilin family. It is thought to regulate actin polymerization in response to extracellular	PFN2		4	forkhead box N2	This gene encodes a forkhead domain binding protein and may function in the transcriptional regulation of the human T-cell leukemia virus long terminal repeat	FOXN2	

						signals							
				5	Potassium two pore domain channel subfamily K member 10	The protein encoded by this gene belongs to the family of potassium channel proteins containing two pore-forming P domains. This channel is an open rectifier which primarily passes outward current under physiological K+ concentrations , and is stimulated strongly by arachidonic acid and to a lesser degree by membrane stretching,	KCNK10		5	guanine nucleotide binding protein (G protein), gamma 5	G proteins are trimeric (alpha-beta-gamma) membrane-associated proteins that regulate flow of information from cell surface receptors to a variety of internal metabolic effectors	GNG5	

						intracellular acidification, and general anaesthetics							
				6	Reticulocalbin 2, EF-hand calcium binding domain	The protein encoded by this gene is a calcium-bindin g protein located in the lumen of the ER.This gene maps to the same region as type 4 Bardet-Biedl syndrome, suggesting a possible causative role for this gene in the disorder	RCN2		6	unc-13 homolog B	This gene is expressed in the kidney cortical epithelial cells and is upregulated by hyperglycemia.Hype rglycemia increases the levels of diacylglycerol, which has been shown to induce apoptosis in cells transfected with this gene and thus contribute to the renal cell complications of hyperglycemia	UNC13 B	
				7	SLAIN motif family member 1		SLAIN1		7	SLAIN motif family member 1		SLAIN1	
				8	6-pyruvoyltetrahydropt erin synthase	The enzyme encoded by this gene catalyzes the elimination of	PTS		8	stanniocalcin 1	The protein may play a role in the regulation of renal and intestinal calcium and	STC1	

						inorganic triphosphate from dihydroneopterin triphosphate, which is the second and irreversible step in the biosynthesis of tetrahydrobiopterin from GTP					phosphate transport, cell metabolism, or cellular calcium/phosphate homeostasis		
				9	Integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)	Integrin family members are membrane receptors involved in cell adhesion and recognition in a variety of processes including embryogenesis, hemostasis, tissue repair, immune response and metastatic	ITGB1		9	SREK1 interacting protein 1		SREK1 IP1	

						diffusion of tumor cells							
				10	Protein phosphatase 2, catalytic subunit, alpha isozyme	This gene encodes the phosphatase 2A catalytic subunit. Protein phosphatase 2A is one of the four major Ser/Thr phosphatases, and it is implicated in the negative control of cell growth and division	PPP2CA		10	ATPase plasma membrane Ca2+ transporting 4	Remove bivalent calcium ions from eukaryotic cells against very large concentration gradients and play a critical role in intracellular calcium homeostasis	ATP2B4	
hsa-miR-205-5p	MIMAT0000266	MIMAT0000266	MIMAT0009197	1	Bicaudal C homolog 1 (Drosophila)	This gene encodes an RNA-binding protein that is active in regulating gene expression by modulating protein	BICC1	580	1	Modulator of smoothened		MOSMO	737

						translation during embryonic development							
				2	Tumor necrosis factor, alpha-induced protein 8		TNFAIP 8		2	Protein phosphatase 2 regulatory subunit Bdelta		PPP2R 2D	
				3	Chimerin 1	This gene encodes GTPase-activating protein for ras-related p21-rac and a phorbol ester receptor	CHN1		3	BicC family RNA binding protein 1	This gene encodes an RNA-binding protein that is active in regulating gene expression by modulating protein translation during embryonic development	BICC1	
				4	sialidase 1 (lysosomal sialidase)	The protein encoded by this gene is a lysosomal enzyme that cleaves terminal sialic acid residues from substrates such as glycoproteins	NEU1		4	RAB11 family interacting protein 1	This gene encodes one of the Rab11-family interacting proteins, which play a role in the Rab-11 mediated recycling of vesicles. The encoded protein may be involved in endocytic sorting, trafficking of	RAB11 FIP1	

						and glycolipids					proteins including integrin subunits and epidermal growth factor receptor, and transport between the recycling endosome and the trans-Golgi network		
				5	Chromosome 11 open reading frame 86		C11orf86		5	Cyclin dependent kinase 19	This gene encodes a protein that is one of the components of the Mediator co-activator complex. The Mediator complex is a multi-protein complex required for transcriptional activation by DNA binding transcription factors of genes transcribed by RNA polymerase II	CDK19	
				6	Hydroxysteroid (17-beta) dehydrogenase 11	Short-chain alcohol dehydrogenases, such as HSD17B11,	HSD17B11		6	Chimerin 1	This gene encodes GTPase-activating protein for ras-related p21-rac and a phorbol ester	CHN1	

						metabolize secondary alcohols and ketones					receptor		
				7	Chromosome 10 open reading frame 131, also known as coiled-coil and C2 domain containing 2B		C10orf131		7	RNA binding motif protein 47		RBM47	
				8	Translocase of inner mitochondrial membrane 17 homolog A (yeast)		TIMM17A		8	Lysophosphatidylcholine acyltransferase 1		LPCAT1	
				9	Heparan sulfate (glucosamine) 3-O-sulfotransferase 1	It possesses both heparan sulfate glucosaminyl 3-O-sulfotransferase activity, anticoagulant heparan sulfate conversion activity, and is a rate limiting enzyme for synthesis of anticoagulant heparan	HS3ST1		9	AP2 associated kinase 1	The protein interacts with and phosphorylates a subunit of the AP-2 complex, which promotes binding of AP-2 to sorting signals found in membrane-bound receptors and subsequent receptor endocytosis	AAK1	

				10	Transmembrane protein 144		TMEM144		10	BTB domain containing 3		BTBD3	
hsa-miR-301-a-5p	MIMAT0022696	MIMAT0022696	MIMAT0000688	1	Zinc finger protein 286B		ZNF286B	4447	1	Nuclear receptor subfamily 2 group C member 2	The protein encoded by this gene plays a role in protecting cells from oxidative stress and damage induced by ionizing radiation	NR2C2	457
				2	Phosphatase and actin regulator 2		PHACTR2		2	DExH-box helicase 9	The encoded protein is an enzyme that catalyzes the ATP-dependent unwinding of double-stranded RNA and DNA-RNA complexes	DHX9	
				3	Zinc finger protein 286A		ZNF286A		3	Basonuclin 2	The encoded protein functions in skin color saturation	BNC2	
				4	Zinc finger protein 487		ZNF487		4	Peptidylprolyl isomerase like 3	This gene encodes a member of the cyclophilin family. Cyclophilins catalyze the cis-trans isomerization of peptidylprolyl imide	PPIL3	

											bonds in oligopeptides. They have been proposed to act either as catalysts or as molecular chaperones in protein-folding events		
				5	Periaxin	This gene encodes a protein involved in peripheral nerve myelin upkeep	PRX		5	Heparan sulfate-glucosamine 3-sulfotransferase 4		HS3ST4	
				6	Ubiquitin-conjugating enzyme E2L 6	This gene encodes a member of the E2 ubiquitin-conjugating enzyme family	UBE2L6		6	Nuclear transcription factor Y subunit beta	The protein encoded by this gene is one subunit of a trimeric complex, forming a highly conserved transcription factor that binds with high specificity to CCAAT motifs in the promoter regions in a variety of genes	NFYB	
				7	GRIK1 antisense RNA		GRIK1-		7	Zinc finger E-box	This gene encodes	ZEB1	

					2		AS2			binding homeobox 1	a zinc finger transcription factor. The encoded protein likely plays a role in transcriptional repression of interleukin 2		
				8	Peptidylprolyl isomerase like 3	This gene encodes a member of the cyclophilin family. Cyclophilins catalyze the cis-trans isomerization of peptidylprolyl imide bonds in oligopeptides. They have been proposed to act either as catalysts or as molecular chaperones in protein-folding events	PPIL3		8	Nicotinamide phosphoribosyltransferase	The protein belongs to the nicotinic acid phosphoribosyltransferase family and is thought to be involved in many important biological processes, including metabolism, stress response and aging	NAMPT	

				9	Potassium voltage-gated channel, subfamily H (eag-related), member 1	Diverse functions include regulating neurotransmitter release, heart rate, insulin secretion, neuronal excitability, epithelial electrolyte transport, smooth muscle contraction, and cell volume.Overexpression of the gene may confer a growth advantage to cancer cells and favor tumor cell proliferation	KCNH1		9	Forkhead box J2		FOXJ2	
				10	Peptidylprolyl isomerase H	Accelerate the folding of	PPIH		10	PNMA family member 1	This gene encodes a neuron- and	PNMA1	

					(cyclophilin H)	proteinsand may act as a protein chaperone that mediates the interactions between different proteins inside the spliceosome					testis-specific protein that is also expressed in some paraneoplastic syndromes affecting the nervous system		
hsa-miR-30-a-3p	MIMAT0000088	MIMAT0000087	MIMAT0000088	1	Dystonin	This gene encodes a member of the plakin protein family of adhesion junction plaque proteins	DST	6117	1	Cell division cycle 73	This gene encodes a tumor suppressor that is involved in transcriptional and post-transcriptional control pathways	CDC73	1390
				2	Immunoglobulin superfamily, member 6		IGSF6		2	Zinc finger E-box binding homeobox 2	It is located in the nucleus and functions as a DNA-binding transcriptional repressor that interacts with activated SMADs (which then accumulate in the	ZEB2	

											nucleus and regulate the transcription of target genes)		
				3	Integrin, alpha 1	This gene encodes the alpha 1 subunit of integrin receptors. This protein heterodimerizes with the beta 1 subunit to form a cell-surface receptor for collagen and laminin. The heterodimeric receptor is involved in cell-cell adhesion and may play a role in inflammation and fibrosis. The alpha 1 subunit	ITGA1		3	Nuclear FMR1 interacting protein 2		NUFIP 2	

						contains an inserted von Willebrand factor type I domain which is thought to be involved in collagen binding							
				4	Canopy FGF signaling regulator 2		CNPY2		4	Piccolo presynaptic cytomatrix protein	The protein encoded by this gene is part of the presynaptic cytoskeletal matrix, which is involved in establishing active synaptic zones and in synaptic vesicle trafficking	PCLO	
				5	Chromosome 9 open reading frame 170		C9orf170		5	Leucine rich repeat transmembrane neuronal 2		LRRTM2	
				6	Centrosomal protein 19kDa	The protein encoded by this gene localizes to centrosomes and primary cilia and	CEP19		6	Neurobeachin like 1		NBEAL1	

						co-localizes with a marker for the mother centriole. This gene resides in a region of human chromosome 3 that is linked to morbid obesity							
				7	Glucosamine-6-phosphate deaminase 2	The protein encoded by this gene is an allosteric enzyme that catalyzes the reversible reaction converting D-glucosamine-6-phosphate into D-fructose-6-phosphate and ammonium	GNPDA2		7	Pantothenate kinase 3	This gene encodes a protein belonging to the pantothenate kinase family. Pantothenate kinase is a key regulatory enzyme in the biosynthesis of coenzyme A	PANK3	
				8	Tumor necrosis factor (ligand) superfamily, member 13b	This cytokine is expressed in B cell lineage cells, and acts	TNFSF13B		8	Abhydrolase domain containing 5	The protein encoded by this gene belongs to a large family of	ABHD5	

						as a potent B cell activator					proteins defined by an alpha/beta hydrolase fold, and contains three sequence motifs that correspond to a catalytic triad found in the esterase/lipase/thio esterase subfamily		
				9	Transmembrane protein 30C		TMEM30C		9	cAMP responsive element binding protein 1	This gene encodes a transcription factor that is a member of the leucine zipper family of DNA binding proteins	CREB1	
				10	HCG1996301; Uncharacterized protein	It is not a protein (lncRNA)	RP11-321M21.3		10	Solute carrier family 12 member 6	This gene is a member of the K-Cl cotransporter (KCC) family. K-Cl cotransporters are integral membrane proteins that lower intracellular chloride concentrations below the electrochemical equilibrium potential	SLC12A6	
hsa-	MIMAT00	MIMAT00	MIMAT00	1	Sorting nexin 16	This gene	SNX16	1566	1	Defective in cullin		DCUN1	1539

miR-30-a-5p	00087	00087	00088			encodes a member of the sorting nexin family. Members of this family contain a phox domain, which is a phosphoinositide binding domain, and are involved in intracellular trafficking				neddylation 1 domain containing 3		D3	
				2	Ankyrin repeat, family A (RFXANK-like), 2		ANKRA2		2	Nuclear factor of activated T cells 5	The product of this gene is a member of the nuclear factors of activated T cells family of transcription factors	NFAT5	
				3	Makorin ring finger protein 3	The protein encoded by this gene contains a RING zinc finger motif	MKRN3		3	Twinfilin actin binding protein 1	This gene encodes twinfilin, an actin monomer-binding protein conserved from yeast to mammals	TWF1	
				4	LIM homeobox 8	The protein encoded by	LHX8		4	Kelch like family member 20	The protein encoded by this	KLHL20	

						this gene is a member of the LIM homeobox family of proteins, which are involved in patterning and differentiation of various tissue types					gene is a member of the kelch family of proteins, appears in many different polypeptide contexts and contains multiple potential protein-protein contact sites		
				5	Kelch like family member 20	The protein encoded by this gene is a member of the kelch family of proteins, appears in many different polypeptide contexts and contains multiple potential protein-protein contact sites	KLHL20		5	PPARG coactivator 1 beta	The protein encoded by this gene stimulates the activity of several transcription factors and nuclear receptors, including estrogen receptor alpha, nuclear respiratory factor 1, and glucocorticoid receptor	PPARGC1B	
				6	Embryonic ectoderm development	This gene encodes a member of the Polycomb-gro	EED		6	Makorin ring finger protein 3	The protein encoded by this gene contains a RING zinc finger	MKRN3	

						up (PcG) family. PcG family members form multimeric protein complexes, which are involved in maintaining the transcriptional repressive state of genes over successive cell generations					motif		
				7	Discs, large (Drosophila) homolog-associated protein 1		DLGAP 1		7	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 5	This enzyme is a type II membrane protein	B3GNT 5	
				8	Metadherin		MTDH		8	Ankyrin repeat family A member 2		ANKRA 2	
				9	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 5	This enzyme is a type II membrane protein	B3GNT5		9	Lin-28 homolog B	This gene is highly expressed in testis, fetal liver, placenta, and in primary human tumors and	LIN28B	

											cancer cell lines. It is negatively regulated by microRNAs that target sites in the 3' UTR, and overexpression of this gene in primary tumors is linked to the repression of let-7 family of microRNAs and derepression of let-7 targets, which facilitates cellular transformation		
				10	Transmembrane protein 181	The TMEM181 gene encodes a putative G protein-coupled receptor expressed on the cell surface	TMEM181		10	Storkhead box 2	This protein is differentially expressed in decidual tissue and may be involved in the susceptibility to pre-eclampsia with fetal growth restriction	STOX2	
hsa-miR-30-b-5p	MIMAT0000420	MIMAT0000420	MIMAT0004589	1	Sorting nexin 16	This gene encodes a member of the sorting nexin	SNX16	1566	1	Twinfilin actin binding protein 1	This gene encodes twinfilin, an actin monomer-binding protein conserved	TWF1	1545

						family and it is involved in intracellular trafficking					from yeast to mammals		
				2	Ankyrin repeat, family A (RFXANK-like), 2		ANKRA2		2	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 5	This enzyme is a type II membrane protein	B3GNT5	
				3	Makorin ring finger protein 3	The protein encoded by this gene contains a RING zinc finger motif	MKRN3		3	Embryonic ectoderm development	This gene encodes a member of the Polycomb-group that the members form multimeric protein complexes, which are involved in maintaining the transcriptional repressive state of genes over successive cell generations	EED	
				4	LIM homeobox 8	The protein encoded by this gene is a member of the LIM homeobox family of proteins, which	LHX8		4	WD repeat domain 7	This gene encodes a member of the WD repeat protein family. Members of this family are involved in a variety of cellular	WDR7	

						are involved in patterning and differentiation of various tissue types					processes, including cell cycle progression, signal transduction, apoptosis, and gene regulation		
				5	Kelch like family member 20	The protein encoded by this gene is a member of the kelch family of proteins, appears in many different polypeptide contexts and contains multiple potential protein-protein contact sites	KLHL20		5	Sodium voltage-gated channel alpha subunit 2	Voltage-gated sodium channels function in the generation and propagation of action potentials in neurons and muscle	SCN2A	
				6	Embryonic ectoderm development	This gene encodes a member of the Polycomb-gro up that the members form multimeric protein	EED		6	Bromodomain and WD repeat domain containing 3	It is thought to have a chromatin-modifying function, and may thus play a role in transcription	BRWD 3	

						complexes, which are involved in maintaining the transcriptional repressive state of genes over successive cell generations							
				7	Discs, large (Drosophila) homolog-associated protein 1	DLG associated protein 1	DLGAP 1		7	Prostaglandin F2 receptor inhibitor		PTGFR N	
				8	Metadherin		MTDH		8	Defective in cullin neddylation 1 domain containing 3		DCUN1 D3	
				9	UDP-GlcNAc:betaGal beta-1,3-N-acetylgluco saminyltransferase 5	This enzyme is a type II membrane protein	B3GNT5		9	Nuclear factor of activated T cells 5		NFAT5	
				10	Transmembrane protein 181	The TMEM181 gene encodes a putative G protein-couple d receptor expressed on the cell surface	TMEM1 81		10	Kelch like family member 20	The protein encoded by this gene is a member of the kelch family of proteins, appears in many different polypeptide	KLHL2 0	

											contexts and contains multiple potential protein-protein contact sites		
hsa-miR-30-c-5p	MIMAT0000244	MIMAT0000244	MIMAT0004550	1	Sorting nexin 16	This gene encodes a member of the sorting nexin family. Members of this family contain a phox domain, which is a phosphoinositide binding domain, and are involved in intracellular trafficking	SNX16	1566	1	Twinfilin actin binding protein 1	This gene encodes twinfilin, an actin monomer-binding protein conserved from yeast to mammals	TWF1	1545
				2	Ankyrin repeat, family A (RFXANK-like), 2		ANKRA2		2	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 5	This enzyme is a type II membrane protein	B3GNT5	
				3	Makorin ring finger protein 3	The protein encoded by this gene	MKRN3		3	Embryonic ectoderm development	This gene encodes a member of the Polycomb-group	EED	

						contains a RING zinc finger motif					(PcG) family. PcG family members form multimeric protein complexes, which are involved in maintaining the transcriptional repressive state of genes over successive cell generations		
				4	LIM homeobox 8	The protein encoded by this gene is a member of the LIM homeobox family of proteins, which are involved in patterning and differentiation of various tissue types	LHX8		4	WD repeat domain 7	This gene encodes a member of the WD repeat protein family. Members of this family are involved in a variety of cellular processes, including cell cycle progression, signal transduction, apoptosis, and gene regulation	WDR7	
				5	Kelch like family member 20	The protein encoded by this gene is a member of the kelch family of	KLHL20		5	Sodium voltage-gated channel alpha subunit 2	Voltage-gated sodium channels function in the generation and propagation of	SCN2A	

						proteins, appears in many different polypeptide contexts and contains multiple potential protein-protein contact sites					action potentials in neurons and muscle		
				6	Embryonic ectoderm development	This gene encodes a member of the Polycomb-gro up (PcG) family. PcG family members form multimeric protein complexes, which are involved in maintaining the transcriptional repressive state of genes over	EED		6	Bromodomain and WD repeat domain containing 3	It is thought to have a chromatin-modifying function, and may thus play a role in transcription	BRWD 3	

						successive cell generations							
				7	Discs, large (Drosophila) homolog-associated protein 1		DLGAP1		7	Prostaglandin F2 receptor inhibitor		PTGFRN	
				8	Metadherin		MTDH		8	Defective in cullin neddylation 1 domain containing 3		DCUN1D3	
				9	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 5	This enzyme is a type II membrane protein	B3GNT5		9	Nuclear factor of activated T cells 5	The product of this gene is a member of the nuclear factors of activated T cells family of transcription factors	NFAT5	
				10	Transmembrane protein 181	The TMEM181 gene encodes a putative G protein-coupled receptor expressed on the cell surface	TMEM181		10	Kelch like family member 20	The protein encoded by this gene is a member of the kelch family of proteins, appears in many different polypeptide contexts and contains multiple potential protein-protein contact sites	KLHL20	
hsa-miR-	MIMAT0000245	MIMAT0000245	MIMAT0004551	1	Sorting nexin 16	This gene encodes a	SNX16	1566	1	Twinfilin actin binding protein 1	This gene encodes twinfilin, an actin	TWF1	1539

30-d-5p						member of the sorting nexin family. Members of this family contain a phox domain, which is a phosphoinositide binding domain, and are involved in intracellular trafficking					monomer-binding protein conserved from yeast to mammals		
				2	Ankyrin repeat, family A (RFXANK-like), 2		ANKRA2		2	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 5	This enzyme is a type II membrane protein	B3GNT5	
				3	Makorin ring finger protein 3	The protein encoded by this gene contains a RING zinc finger motif	MKRN3		3	WD repeat domain 7	This gene encodes a member of the WD repeat protein family. Members of this family are involved in a variety of cellular processes, including cell cycle progression, signal	WDR7	

											transduction, apoptosis, and gene regulation		
				4	LIM homeobox 8	The protein encoded by this gene is a member of the LIM homeobox family of proteins, which are involved in patterning and differentiation of various tissue types	LHX8		4	Sodium voltage-gated channel alpha subunit 2	Voltage-gated sodium channels function in the generation and propagation of action potentials in neurons and muscle	SCN2A	
				5	Kelch like family member 20	The protein encoded by this gene is a member of the kelch family of proteins, appears in many different polypeptide contexts and contains multiple potential protein-protein	KLHL20		5	Bromodomain and WD repeat domain containing 3	It is thought to have a chromatin-modifying function, and may thus play a role in transcription	BRWD 3	

						contact sites							
				6	Embryonic ectoderm development	This gene encodes a member of the Polycomb-group (PcG) family. PcG family members form multimeric protein complexes, which are involved in maintaining the transcriptional repressive state of genes over successive cell generations	EED		6	Prostaglandin F2 receptor inhibitor		PTGFRN	
				7	Discs, large (Drosophila) homolog-associated protein 1		DLGAP1		7	Defective in cullin neddylation 1 domain containing 3		DCUN1D3	
				8	Metadherin		MTDH		8	Nuclear factor of activated T cells 5	The product of this gene is a member of the nuclear factors	NFAT5	

											of activated T cells family of transcription factors		
				9	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 5	This enzyme is a type II membrane protein	B3GNT5		9	Kelch like family member 20	The protein encoded by this gene is a member of the kelch family of proteins, appears in many different polypeptide contexts and contains multiple potential protein-protein contact sites	KLHL20	
				10	Transmembrane protein 181	The TMEM181 gene encodes a putative G protein-coupled receptor expressed on the cell surface	TMEM181		10	PPARG coactivator 1 beta	The protein encoded by this gene stimulates the activity of several transcription factors and nuclear receptors, including estrogen receptor alpha, nuclear respiratory factor 1, and glucocorticoid receptor	PPARGC1B	
hsa-miR-	MIMAT0000693	MIMAT0000692	MIMAT0000693	1	Dystonin	This gene encodes a	DST	6117	1	Cell division cycle 73	This gene encodes a tumor suppressor	CDC73	1396

30-e-3p						member of the plakin protein family of adhesion junction plaque proteins					that is involved in transcriptional and post-transcriptional control pathways	
				2	Immunoglobulin superfamily, member 6		IGSF6		2	Zinc finger E-box binding homeobox 2	It is located in the nucleus and functions as a DNA-binding transcriptional repressor that interacts with activated SMADs (which then accumulate in the nucleus and regulate the transcription of target genes)	ZEB2
				3	Integrin, alpha 1	This gene encodes the alpha 1 subunit of integrin receptors. This protein heterodimerizes with the beta	ITGA1		3	Nuclear FMR1 interacting protein 2		NUFIP2

						1 subunit to form a cell-surface receptor for collagen and laminin. The heterodimeric receptor is involved in cell-cell adhesion and may play a role in inflammation and fibrosis. The alpha 1 subunit contains an inserted von Willebrand factor type I domain which is thought to be involved in collagen binding							
				4	Canopy FGF signaling regulator 2		CNPY2		4	Piccolo presynaptic cytomatrix protein	The protein encoded by this gene is part of the presynaptic	PCLO	

											cytoskeletal matrix, which is involved in establishing active synaptic zones and in synaptic vesicle trafficking		
				5	Chromosome 9 open reading frame 170		C9orf170		5	Leucine rich repeat transmembrane neuronal 2		LRRTM2	
				6	Centrosomal protein 19kDa	The protein encoded by this gene localizes to centrosomes and primary cilia and co-localizes with a marker for the mother centriole. This gene resides in a region of human chromosome 3 that is linked to morbid obesity	CEP19		6	Neurobeachin like 1		NBEAL1	
				7	Glucosamine-6-phosphate deaminase 2	The protein encoded by this gene is an	GNPDA2		7	Pantothenate kinase 3	This gene encodes a protein belonging to the pantothenate	PANK3	

						allosteric enzyme that catalyzes the reversible reaction converting D-glucosamine -6-phosphate into D-fructose-6-p hosphate and ammonium					kinase family. Pantothenate kinase is a key regulatory enzyme in the biosynthesis of coenzyme A		
				8	Tumor necrosis factor (ligand) superfamily, member 13b	This cytokine is expressed in B cell lineage cells, and acts as a potent B cell activator	TNFSF1 3B		8	Abhydrolase domain containing 5	The protein encoded by this gene belongs to a large family of proteins defined by an alpha/beta hydrolase fold, and contains three sequence motifs that correspond to a catalytic triad found in the esterase/lipase/thio esterase subfamily	ABHD5	
				9	Transmembrane protein 30C		TMEM3 0C		9	cAMP responsive element binding protein 1	This gene encodes a transcription factor that is a member of	CREB1	

											the leucine zipper family of DNA binding proteins		
				10	HCG1996301; Uncharacterized protein	It is not a protein (lncRNA)	RP11-321M21.3		10	Solute carrier family 12 member 6	This gene is a member of the K-Cl cotransporter (KCC) family. K-Cl cotransporters are integral membrane proteins that lower intracellular chloride concentrations below the electrochemical equilibrium potential	SLC12A6	
hsa-miR-30-e-5p	MIMAT0000692	MIMAT0000692	MIMAT0000693	1	Sorting nexin 16	This gene encodes a member of the sorting nexin family. Members of this family contain a phox domain, which is a phosphoinositide binding domain, and are involved in	SNX16	1566	1	Twinfilin actin binding protein 1	This gene encodes twinfilin, an actin monomer-binding protein conserved from yeast to mammals	TWF1	1541

						intracellular trafficking							
				2	Ankyrin repeat, family A (RFXANK-like), 2		ANKRA2		2	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 5	This enzyme is a type II membrane protein	B3GNT5	
				3	Makorin ring finger protein 3	The protein encoded by this gene contains a RING zinc finger motif	MKRN3		3	WD repeat domain 7	This gene encodes a member of the WD repeat protein family. Members of this family are involved in a variety of cellular processes, including cell cycle progression, signal transduction, apoptosis, and gene regulation	WDR7	
				4	LIM homeobox 8	The protein encoded by this gene is a member of the LIM homeobox family of proteins, which are involved in patterning and	LHX8		4	Sodium voltage-gated channel alpha subunit 2	Voltage-gated sodium channels function in the generation and propagation of action potentials in neurons and muscle	SCN2A	

						differentiation of various tissue types							
				5	Kelch like family member 20	The protein encoded by this gene is a member of the kelch family of proteins, appears in many different polypeptide contexts and contains multiple potential protein-protein contact sites	KLHL20		5	Bromodomain and WD repeat domain containing 3	It is thought to have a chromatin-modifying function, and may thus play a role in transcription	BRWD 3	
				6	Embryonic ectoderm development	This gene encodes a member of the Polycomb-gro up (PcG) family. PcG family members form multimeric protein complexes,	EED		6	Prostaglandin F2 receptor inhibitor		PTGFR N	

						which are involved in maintaining the transcriptional repressive state of genes over successive cell generations							
				7	Discs, large (Drosophila) homolog-associated protein 1		DLGAP 1		7	Defective in cullin neddylation 1 domain containing 3		DCUN1 D3	
				8	Metadherin		MTDH		8	Nuclear factor of activated T cells 5	The product of this gene is a member of the nuclear factors of activated T cells family of transcription factors	NFAT5	
				9	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 5	This enzyme is a type II membrane protein	B3GNT5		9	Kelch like family member 20	The protein encoded by this gene is a member of the kelch family of proteins, appears in many different polypeptide contexts and contains multiple	KLHL2 0	

											potential protein-protein contact sites		
				10	Transmembrane protein 181	The TMEM181 gene encodes a putative G protein-coupled receptor expressed on the cell surface	TMEM181		10	PPARG coactivator 1 beta	The protein encoded by this gene stimulates the activity of several transcription factors and nuclear receptors, including estrogen receptor alpha, nuclear respiratory factor 1, and glucocorticoid receptor	PPARGC1B	
hsa-miR-424-3p	MIMAT0004749	MIMAT0001341	MIMAT0004749	1	Heterogeneous nuclear ribonucleoprotein A0	These proteins are associated with pre-mRNAs in the nucleus and appear to influence pre-mRNA processing and other aspects of mRNA metabolism and transport	HNRNPA0	1769	1	Heterogeneous nuclear ribonucleoprotein A0	These proteins are associated with pre-mRNAs in the nucleus and appear to influence pre-mRNA processing and other aspects of mRNA metabolism and transport	HNRNPA0	187

				2	Rhox homeobox family member 2	Encodes a transcriptional repressor and is highly expressed in a variety of cancers. The encoded protein associates with the cell membrane and with microtubules, and is concentrated at the leading edge of migratory cells	RHOXF2		2	Odd-skipped related transcription factor 1	Protein	OSR1	
				3	Thioredoxin-like 4A	It is involved in pre-mRNA splicing	TXNL4A		3	Retention in endoplasmic reticulum sorting receptor 1	The protein encoded by this gene is a multi-pass membrane protein that is localized to the golgi apparatus	RER1	
				4	Uncharacterized protein; cDNA FLJ58069		AL078585.1		4	Tankyrase 2		TNKS2	
				5	Transmembrane	This gene	TMEM7		5	SMAD family	The protein	SMAD7	

					protein 70	likely encodes a mitochondrial membrane protein	0			member 7	encoded by this gene is a nuclear protein that binds the E3 ubiquitin ligase SMURF2		
				6	Testis expressed 30		TEX30		6	Thrombospondin 4	The protein encoded by this gene belongs to the thrombospondin protein family. Thrombospondin family members are adhesive glycoproteins that mediate cell-to-cell and cell-to-matrix interactions	THBS4	
				7	cAMP responsive element binding protein 3-like 4	The encoded protein is a transcriptional activator which contains a dimerization domain, and this protein may function in a number of processing pathways	CREB3L 4		7	Rhox homeobox family member 2	Encodes a transcriptional repressor and is highly expressed in a variety of cancers. The encoded protein associates with the cell membrane and with microtubules, and is concentrated at the leading edge of	RHOX F2	

						including protein processing					migratory cells		
				8	EP300 interacting inhibitor of differentiation 3		EID3		8	NIMA related kinase 6	The protein encoded by this gene is a kinase required for progression through the metaphase portion of mitosis	NEK6	
				9	Ring finger protein 13	The protein encoded by this gene contains a RING zinc finger, a motif known to be involved in protein-protein interactions	RNF13		9	Rhox homeobox family member 2B		RHOX F2B	
				10	Uncharacterized protein		AC027763.2		10	Eukaryotic translation initiation factor 5A		EIF5A	
hsa-miR-10-b-3p	MIMAT0004556	MIMAT0000254	MIMAT0004556	1	Coenzyme Q4 homolog (*S. cerevisiae*)	This gene encodes a component of the coenzyme Q biosynthesis pathway.	COQ4	4109	1	MYC associated factor X	The protein encoded by this gene is a member of the basic helix-loop-helix leucine zipper family	MAX	463

						Coenzyme Q, an essential component of the electron transport chain, shuttles electrons between complexes I or II to complex III of the mitochondrial transport chain					of transcription factors. It is able to form homodimers and heterodimers with other family members,they compete for a common DNA target site and rearrangement among these dimer forms provides a complex system of transcriptional regulation		
				2	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 1, 7kDa		NDUFB1		2	Chromosome 11 open reading frame 87		C11orf87	
				3	Killer cell immunoglobulin-like receptor, three domains, long cytoplasmic tail, 1	Killer cell immunoglobulin-like receptors are transmembrane glycoproteins expressed by natural killer cells and subsets of T cells	KIR3DL1		3	RAB4A, member RAS oncogene family	This gene is a member of the largest group in the Ras superfamily of small GTPases, which regulate membrane trafficking	RAB4A	

				4	PYD and CARD domain containing	This gene encodes an adaptor protein that is composed of two protein-protein interaction domains. The PYD and CARD domains are members of the six-helix bundle death domain-fold superfamily that mediates assembly of large signaling complexes in the inflammatory and apoptotic signaling pathways via the activation of caspase	PYCARD		4	Adenylate cyclase 9	Adenylate cyclase is a membrane bound enzyme that catalyses the formation of cyclic AMP from ATP	ADCY9	
				5	RAB4A, member RAS	This gene is a	RAB4A		5	Serine/threonine	This gene encodes	STK39	

					oncogene family	member of the largest group in the Ras superfamily of small GTPases, which regulate membrane trafficking				kinase 39	a serine/threonine kinase that is thought to function in the cellular stress response pathway		
				6	Uncharacterized protein	lncRNA	AC023590.1		6	Apoptosis inhibitor 5	This gene encodes an apoptosis inhibitory protein whose expression prevents apoptosis after growth factor deprivation	API5	
				7	MYC associated factor X	The protein encoded by this gene is a member of the basic helix-loop-helix leucine zipper family of transcription factors. It is able to form homodimers and	MAX		7	LanC like 3		LANCL3	

						heterodimers with other family members,they compete for a common DNA target site and rearrangement among these dimer forms provides a complex system of transcriptional regulation							
				8	Olfactory receptor, family 4, subfamily D, member 5	Olfactory receptors interact with odorant molecules in the nose, to initiate a neuronal response that triggers the perception of a smell	OR4D5		8	Mannan binding lectin serine peptidase 1	This gene encodes a serine protease that functions as a component of the lectin pathway of complement activation. The complement pathway plays an essential role in the innate and adaptive immune response	MASP1	
				9	Major histocompatibility complex, class I, B	HLA-B belongs to the HLA	HLA-B		9	Required for meiotic nuclear division 5		RMND 5A	

						class I heavy chain paralogues. This class I molecule is a heterodimer consisting of a heavy chain and a light chain. The heavy chain is anchored in the membrane. Class I molecules play a central role in the immune system by presenting peptides derived from the endoplasmic reticulum lumen.				homolog A			
				10	Chromosome 14 open reading frame 182		C14orf182		10	Frizzled class receptor 3	The function of this protein is unknown, although it may play a role in mammalian	FZD3	

											hair follicle development		
hsa-miR-150-3p	MIMAT0004610	MIMAT0000451	MIMAT0004610	1	Ferritin, heavy polypeptide 1	A major function of ferritin is the storage of iron in a soluble and nontoxic state	FTH1	2558	1	5'-3' exoribonuclease 1	The encoded protein may be involved in replication-dependent histone mRNA degradation, and interacts directly with the enhancer of mRNA-decapping protein 4	XRN1	413
				2	Amelogenin, Y-linked	This gene encodes a member of the amelogenin family of extracellular matrix proteins. Amelogenins are involved in biomineralization during tooth enamel development	AMELY		2	Solute carrier family 38 member 2		SLC38A2	
				3	Tubulin polymerization-promoting protein family		TPPP3		3	Ferritin heavy chain 1	A major function of ferritin is the storage of iron in a soluble	FTH1	

					member 3						and nontoxic state		
				4	Mitochondrial ribosomal protein L35	This gene encodes a 39S subunit protein	MRPL35		4	Cadherin 2	This protein plays a role in the establishment of left-right asymmetry, development of the nervous system and the formation of cartilage and bone	CDH2	
				5	Atypical chemokine receptor 4	This receptor has been shown to bind dendritic cell- and T cell-activated chemokines	ACKR4		5	Dual specificity tyrosine phosphorylation regulated kinase 1B	The encoded protein participates in the regulation of the cell cycle	DYRK1B	
				6	Protein tyrosine phosphatase, receptor type, O	This gene contains two distinct promoters, and alternatively spliced transcript variants encoding multiple isoforms have been observed. The	PTPRO		6	Cathepsin C	This gene encodes a member of the peptidase C1 family and lysosomal cysteine proteinase that appears to be a central coordinator for activation of many serine proteinases in cells of the immune system	CTSC	

						encoded proteins may have multiple isoform-specific and tissue-specific functions, including the regulation of osteoclast production and activity, inhibition of cell proliferation and facilitation of apoptosis							
				7	Beta-1,4-N-acetyl-galactosaminyl transferase 2	B4GALNT2 catalyzes the last step in the biosynthesis of the human Sd(a) antigen through the addition of an N-acetylgalactosamine residue via a beta-1,4	B4GALNT2		7	Sodium voltage-gated channel alpha subunit 1	Voltage-dependent sodium channels are heteromeric complexes that regulate sodium exchange between intracellular and extracellular spaces and are essential for the generation and propagation of action potentials in	SCN1A	

						linkage to a subterminal galactose residue substituted with an alpha-2,3-linked sialic acid. B4GALNT2 also catalyzes the last step in the biosynthesis of the Cad antigen					muscle cells and neurons		
				8	Solute carrier family 38 member 2		SLC38A2		8	Regulatory factor X3	This gene is a member of the regulatory factor X gene family, which encodes transcription factors that contain a highly-conserved winged helix DNA binding domain	RFX3	
				9	Keratin associated protein 4-12	This protein is a member of the keratin-associ	KRTAP4-12		9	Zinc finger protein 217		ZNF217	

						ated protein (KAP) family. The KAP proteins form a matrix of keratin intermediate filaments which contribute to the structure of hair fibers							
				10	Transmembrane protein 179B		TMEM179B		10	cAMP regulated phosphoprotein 21	This gene encodes a cAMP-regulated phosphoprotein. The encoded protein is enriched in the caudate nucleus and cerebellar cortex. A similar protein in mouse may be involved in regulating the effects of dopamine in the basal ganglia	ARPP21	
hsa-miR-30-c	MIMAT0004674	MIMAT0000244	MIMAT0004674	1	Chemokine (C-C motif) ligand 22	The cytokine encoded by this gene	CCL22	6424	1	A-kinase anchoring protein 6	The A-kinase anchor proteins (AKAPs) are a	AKAP6	737

-1-3p						displays chemotactic activity for monocytes, dendritic cells, natural killer cells and for chronically activated T lymphocytes					group of structurally diverse proteins, which have the common function of binding to the regulatory subunit of protein kinase A (PKA) and confining the holoenzyme to discrete locations within the cell		
				2	Signal sequence receptor, beta (translocon-associated protein beta)	The signal sequence receptor is a glycosylated endoplasmic reticulum membrane receptor associated with protein translocation across the endoplasmic reticulum membrane	SSR2		2	MARCKS like 1	This gene encodes a member of the myristoylated alanine-rich C-kinase substrate (MARCKS) family. Members of this family play a role in cytoskeletal regulation, protein kinase C signaling and calmodulin signaling	MARCKSL1	
				3	G antigen 12B		GAGE12B		3	Ras association domain family member 8	This gene encodes a member of the Ras-assocation	RASSF8	

											domain family of tumor suppressor proteins		
				4	Membrane-spanning 4-domains, subfamily A, member 14		MS4A14		4	Keratin 222		KRT222	
				5	Shisa family member 6		SHISA6		5	MAM domain containing glycosylphosphatidylinositol anchor 2		MDGA2	
				6	MOB kinase activator 3A		MOB3A		6	Chloride intracellular channel 5	The encoded protein associates with actin-based cytoskeletal structures and may play a role in multiple processes including hair cell stereocilia formation, myoblast proliferation and glomerular podocyte and endothelial cell maintenance	CLIC5	
				7	Zinc finger protein 880		ZNF880		7	Pellino E3 ubiquitin protein ligase 1		PELI1	
				8	G antigen 2B		GAGE2B		8	Vestigial like family member 3		VGLL3	
				9	G antigen 12J		GAGE1		9	DPY30 domain	This gene encodes	DYDC2	

							2J			containing 2	a member of a family of proteins that contains a DPY30 domain. It is about 40 residues long and is probably formed of two alpha-helices		
				10	G antigen 13		GAGE13		10	Autophagy related 16 like 2		ATG16L2	
hsa-miR-424-5p	MIMAT0001341	MIMAT0001341	MIMAT0004749	1	ADP-ribosylation factor-like 2	This gene encodes a small GTP-binding protein of the RAS superfamily which functions as an ADP-ribosylation factor	ARL2	1508	1	Pappalysin 1	This gene encodes a secreted metalloproteinase which cleaves insulin-like growth factor binding proteins	PAPPA	1349
				2	Cyclin E1	Different cyclins exhibit distinct expression and degradation patterns which contribute to	CCNE1		2	Fatty acid synthase	Its main function is to catalyze the synthesis of palmitate from acetyl-CoA and malonyl-CoA, in the presence of NADPH, into	FASN	

						the temporal coordination of each mitotic event. This cyclin forms a complex with and functions as a regulatory subunit of CDK2, whose activity is required for cell cycle G1/S transition. This protein accumulates at the G1-S phase boundary and is degraded as cells progress through S phase					long-chain saturated fatty acids. In some cancer cell lines, this protein has been found to be fused with estrogen receptor-alpha, in which the N-terminus of Fas cell surface death receptor is fused in-frame with the C-terminus of ER-alpha		
				3	Sprouty-related, EVH1 domain containing 1	The protein encoded by this gene is a member of the Sprouty family of proteins and	SPRED 1		3	Unc-80 homolog, NALCN channel complex subunit	The protein encoded by this gene is a component of a voltage-independent 'leak' ion-channel	UNC80	

						is phosphorylated by tyrosine kinase in response to several growth factors					complex. Leak channels play an importnat role in establishment and maintenance of resting membrane potentials in neurons		
				4	Phospholipid scramblase 4		PLSCR4		4	Fibroblast growth factor 2	This protein has been implicated in diverse biological processes, such as limb and nervous system development, wound healing, and tumor growth	FGF2	
				5	DnaJ (Hsp40) homolog, subfamily B, member 4	The protein encoded by this gene is a molecular chaperone, tumor suppressor, and member of the heat shock protein-40 family	DNAJB4		5	Trinucleotide repeat containing 6B		TNRC6B	
				6	Wingless-type MMTV	The WNT	WNT3A		6	Protein tyrosine	The protein	PTPN4	

					integration site family, member 3A	gene family consists of structurally related genes which encode secreted signaling proteins. These proteins have been implicated in oncogenesis and in several developmental processes, including regulation of cell fate and patterning during embryogenesis				phosphatase, non-receptor type 4	encoded by this gene is a member of the protein tyrosine phosphatase (PTP) family. PTPs are known to be signaling molecules that regulate a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation		
				7	Tumor necrosis factor (ligand) superfamily, member 13b	This cytokine is expressed in B cell lineage cells, and acts as a potent B cell activator	TNFSF13B		7	PHD finger protein 19		PHF19	
				8	Zinc finger and BTB		ZBTB34		8	Ubiquitin	This gene encodes	UBE2Q	

					domain containing 34					conjugating enzyme E2 Q1	a member of the E2 ubiquitin-conjugating enzyme family	1	
				9	LSM11, U7 small nuclear RNA associated		LSM11		9	LSM11, U7 small nuclear RNA associated		LSM11	
				10	Cytochrome b561 family, member D1		CYB561D1		10	Ankyrin repeat and ubiquitin domain containing 1		ANKUB1	
hsa-miR-183-3p	MIMAT0004560	MIMAT0000261	MIMAT0004560	1	Neurobeachin like 1		NBEAL1	4675	1	Coiled-coil domain containing 91		CCDC91	774
				2	MT-RNR2-like 13 (pseudogene)		MTRNR2L13		2	GLI pathogenesis related 1	The protein has proapoptotic activities in prostate and bladder cancer cells	GLIPR1	
				3	Uncharacterized protein	This gene encodes a component of the spliceosome complex and is one of several retinitis pigmentosa-causing genes. When the gene product is added to the	RP11-67H2.1		3	Leucine rich repeat transmembrane neuronal 3		LRRTM3	

						spliceosome complex, activation occurs							
				4	Proline-rich protein BstNI subfamily 4	The encoded preproprotein undergoes proteolytic processing to generate one or more mature peptides before secretion from the parotid glands	PRB4		4	Nuclear transcription factor Y subunit beta	The protein encoded by this gene is one subunit of a trimeric complex, forming a highly conserved transcription factor that binds with high specificity to CCAAT motifs in the promoter regions in a variety of genes	NFYB	
				5	MT-RNR2-like 4		MTRNR 2L4		5	Heparan sulfate-glucosamine 3-sulfotransferase 3B1	The protein encoded by this gene is a type II integral membrane protein that belongs to the 3-O-sulfotransferases family. These proteins catalyze the addition of sulfate groups at the 3-OH position of	HS3ST 3B1	

											glucosamine in heparan sulfate		
				6	MT-RNR2-like 3		MTRNR2L3		6	Catenin beta 1	The protein encoded by this gene is part of a complex of proteins that constitute adherens junctions, adherens junctions are necessary for the creation and maintenance of epithelial cell layers by regulating cell growth and adhesion between cells	CTNNB1	
				7	Up-regulated during skeletal muscle growth 5 homolog (mouse	Ubiquitous expression in heart colon	USMG5		7	Akirin 1		AKIRIN1	
				8	MT-RNR2-like 7	Note: It is unclear if this is a transcribed protein-coding gene, or if it is a nuclear pseudogene of the	MTRNR2L7		8	Zinc finger protein 492		ZNF492	

						mitochondrial MT-RNR2 gene							
				9	Sulfotransferase family, cytosolic, 6B, member 1		SULT6B1		9	Zinc finger protein 99		ZNF99	
				10	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 12	This gene encodes a protein which is part of mitochondrial complex 1, part of the oxidative phosphorylation system in mitochondria. Complex 1 transfers electrons to ubiquinone from NADH which establishes a proton gradient for the generation of ATP	NDUFA12		10	Fucosyltransferase 9	It is localized to the golgi, and catalyzes the last step in the biosynthesis of Lewis X antigen, the addition of a fucose to precursor polysaccharides	FUT9	
hsa-	MIMAT00	MIMAT00	MIMAT00	1	Dystonin	This gene	DST	6117	1	Cell division cycle	This gene encodes	CDC73	1390

miR-30-d-3p	04551	00245	04551			encodes a member of the plakin protein family of adhesion junction plaque proteins				73	a tumor suppressor that is involved in transcriptional and post-transcriptional control pathways		
				2	Immunoglobulin superfamily, member 6		IGSF6		2	Zinc finger E-box binding homeobox 2	It is located in the nucleus and functions as a DNA-binding transcriptional repressor that interacts with activated SMADs (which then accumulate in the nucleus and regulate the transcription of target genes)	ZEB2	
				3	Integrin, alpha 1	This gene encodes the alpha 1 subunit of integrin receptors. This protein heterodimerize	ITGA1		3	Nuclear FMR1 interacting protein 2		NUFIP 2	

						s with the beta 1 subunit to form a cell-surface receptor for collagen and laminin. The heterodimeric receptor is involved in cell-cell adhesion and may play a role in inflammation and fibrosis. The alpha 1 subunit contains an inserted von Willebrand factor type I domain which is thought to be involved in collagen binding							
				4	Canopy FGF signaling regulator 2		CNPY2		4	Piccolo presynaptic cytomatrix protein	The protein encoded by this gene is part of the	PCLO	

											presynaptic cytoskeletal matrix, which is involved in establishing active synaptic zones and in synaptic vesicle trafficking		
				5	Chromosome 9 open reading frame 170		C9orf170		5	Leucine rich repeat transmembrane neuronal 2		LRRTM2	
				6	Centrosomal protein 19kDa	The protein encoded by this gene localizes to centrosomes and primary cilia and co-localizes with a marker for the mother centriole. This gene resides in a region of human chromosome 3 that is linked to morbid obesity	CEP19		6	Neurobeachin like 1		NBEAL1	
				7	Glucosamine-6-phosphate deaminase 2	The protein encoded by	GNPDA2		7	Pantothenate kinase 3	This gene encodes a protein belonging	PANK3	

						this gene is an allosteric enzyme that catalyzes the reversible reaction converting D-glucosamine -6-phosphate into D-fructose-6-p hosphate and ammonium					to the pantothenate kinase family. Pantothenate kinase is a key regulatory enzyme in the biosynthesis of coenzyme A		
				8	Tumor necrosis factor (ligand) superfamily, member 13b	This cytokine is expressed in B cell lineage cells, and acts as a potent B cell activator	TNFSF1 3B		8	Abhydrolase domain containing 5	The protein encoded by this gene belongs to a large family of proteins defined by an alpha/beta hydrolase fold, and contains three sequence motifs that correspond to a catalytic triad found in the esterase/lipase/thio esterase subfamily	ABHD5	
				9	Transmembrane protein 30C		TMEM3 0C		9	cAMP responsive element binding	This gene encodes a transcription factor	CREB1	

										protein 1	that is a member of the leucine zipper family of DNA binding proteins		
				10	HCG1996301; Uncharacterized protein	It is not a protein (lncRNA)	RP11-321M21.3		10	Solute carrier family 12 member 6	This gene is a member of the K-Cl cotransporter (KCC) family. K-Cl cotransporters are integral membrane proteins that lower intracellular chloride concentrations below the electrochemical equilibrium potential	SLC12A6	
hsa-miR-301-a-3p	MIMAT0000688	MIMAT0022696	MIMAT0000688	1	SLAIN motif family member 1		SLAIN1	1029	1	SLAIN motif family member 1		SLAIN1	916
				2	Mdm4 p53 binding protein homolog (mouse)	Bind the p53 tumor suppressor protein and inhibit its activity, and have been shown to be overexpressed in a variety of human	MDM4		2	SKI/DACH domain containing 1		SKIDA1	

						cancers							
				3	Lysine (K)-specific demethylase 2A	This gene encodes a member of the F-box protein family which is characterized by an approximately 40 amino acid motif, the F-box. The F-box proteins constitute one of the four subunits of ubiquitin protein ligase complex called SCFs, which function in phosphorylation-dependent ubiquitination	KDM2A		3	Gap junction protein alpha 1	This gene is a member of the connexin gene family. The encoded protein is a component of gap junctions, which are composed of arrays of intercellular channels that provide a route for the diffusion of low molecular weight materials from cell to cell	GJA1	
				4	Kruppel-like factor 7 (ubiquitous)	The protein encoded by this gene is a member of the Kruppel-like	KLF7		4	Cytoplasmic polyadenylation element binding protein 1	This gene encodes a member of the cytoplasmic polyadenylation element binding	CPEB1	

						transcriptional regulator family. Members in this family regulate cell proliferation and differentiation					protein family. It is involved in a role in cell proliferation and tumorigenesis		
				5	PAN3 poly(A) specific ribonuclease subunit homolog (*S. cerevisiae*)		PAN3		5	CAP-Gly domain containing linker protein 1	The protein encoded by this gene links endocytic vesicles to microtubules	CLIP1	
				6	Ectonucleotide pyrophosphatase/phosphodiesterase 5 (putative)	This gene encodes a type-I transmembrane glycoprotein. Studies in rat suggest the encoded protein may play a role in neuronal cell communications	ENPP5		6	Activin A receptor type 1	Activins are dimeric growth and differentiation factors which belong to the transforming growth factor-beta (TGF-beta) superfamily of structurally related signaling proteins	ACVR1	
				7	F-box protein 28	Members of the F-box	FBXO28		7	TSC complex subunit 1	This gene is a tumor suppressor gene	TSC1	

						protein family, such as FBXO28, interact with SKP1 through the F box, and F box interact with ubiquitination targets through other protein interaction domains					that encodes the growth inhibitory protein hamartin		
				8	V-myb avian myeloblastosis viral oncogene homolog-like 1		MYBL1		8	Acyl-CoA synthetase long chain family member 4	The protein encoded by this gene is an isozyme of the long-chain fatty-acid-coenzyme A ligase family. Although differing in substrate specificity, subcellular localization, and tissue distribution, all isozymes of this family convert free long-chain fatty acids into fatty acyl-CoA esters,	ACSL4	

											and thereby play a key role in lipid biosynthesis and fatty acid degradation		
				9	Activin A receptor type 1	Activins are dimeric growth and differentiation factors which belong to the transforming growth factor-beta (TGF-beta) superfamily of structurally related signaling proteins	ACVR1		9	Ribosomal protein S6 kinase A5		RPS6KA5	
				10	Cytoplasmic polyadenylation element binding protein 1	This gene encodes a member of the cytoplasmic polyadenylation element binding protein family. It is involved in a	CPEB1		10	Mdm4 p53 binding protein homolog (mouse)	Bind the p53 tumor suppressor protein and inhibit its activity, and have been shown to be overexpressed in a variety of human cancers	MDM4	

						role in cell proliferation and tumorigenesis							
hsa-miR-30-c-2-3p	MIMAT0004550	MIMAT0000244	MIMAT0004550	1	Chemokine (C-C motif) ligand 22	The cytokine encoded by this gene displays chemotactic activity for monocytes, dendritic cells, natural killer cells and for chronically activated T lymphocytes	CCL22	6424	1	A-kinase anchoring protein 6	The A-kinase anchor proteins (AKAPs) are a group of structurally diverse proteins, which have the common function of binding to the regulatory subunit of protein kinase A (PKA) and confining the holoenzyme to discrete locations within the cell	AKAP6	743
				2	Signal sequence receptor, beta (translocon-associated protein beta)	The signal sequence receptor is a glycosylated endoplasmic reticulum membrane receptor associated with protein translocation	SSR2		2	MARCKS like 1	This gene encodes a member of the myristoylated alanine-rich C-kinase substrate (MARCKS) family. Members of this family play a role in cytoskeletal regulation, protein kinase C signaling	MARCKSL1	

						across the endoplasmic reticulum membrane					and calmodulin signaling		
				3	G antigen 12B		GAGE12B		3	MAM domain containing glycosylphosphatidylinositol anchor 2		MDGA2	
				4	Membrane-spanning 4-domains, subfamily A, member 14		MS4A14		4	Keratin 222		KRT222	
				5	Shisa family member 6		SHISA6		5	Ras association domain family member 8	This gene encodes a member of the Ras-assocation domain family of tumor suppressor proteins	RASSF8	
				6	MOB kinase activator 3A		MOB3A		6	Pellino E3 ubiquitin protein ligase 1		PELI1	
				7	Zinc finger protein 880		ZNF880		7	Chloride intracellular channel 5	The encoded protein associates with actin-based cytoskeletal structures and may play a role in multiple processes including hair cell stereocilia formation, myoblast	CLIC5	

											proliferation and glomerular podocyte and endothelial cell maintenance		
				8	G antigen 2B		GAGE2B		8	Vestigial like family member 3		VGLL3	
				9	G antigen 12J		GAGE12J		9	Autophagy related 16 like 2		ATG16L2	
				10	G antigen 13		GAGE13		10	DPY30 domain containing 2	This gene encodes a member of a family of proteins that contains a DPY30 domain. It is about 40 residues long and is probably formed of two alpha-helices	DYDC2	
hsa-miR-30-b-3p	MIMAT0004589	MIMAT0000420	MIMAT0004589	1	Phosphatidylinositol-specific phospholipase C, X domain containing 1	This gene is the most terminal protein-coding gene in the pseudoautosomal region on chromosomes X and Y	PLCXD1	6956	1	Immunity related GTPase Q		IRGQ	1207
				2	Chemokine (C-C motif) ligand 28	This antimicrobial gene belongs	CCL28		2	Chromobox 5	The encoded product is involved in the formation of	CBX5	

						to the subfamily of small cytokine CC genes. Cytokines are a family of secreted proteins involved in immunoregulatory and inflammatory processes. The cytokine encoded by this gene displays chemotactic activity for resting CD4 or CD8 T cells and eosinophils					functional kinetochore through interaction with essential kinetochore proteins		
				3	Chromosome 10 open reading frame 91	Long intergenic non-protein coding RNA 2870 (ncRNA)	C10orf91		3	Phosphatase and actin regulator 4	This gene encodes a member of the phosphatase and actin regulator family. Other phosphatase and	PHACTR4	

											actin regulator family members have been shown to inhibit protein phosphatase 1 activity		
				4	Coiled-coil domain containing 163, pseudogene		CCDC163P		4	ELK1, ETS transcription factor	This gene is a member of the Ets family of transcription factors and of the ternary complex factor subfamily. Proteins of the ternary complex factor subfamily form a ternary complex by binding to the the serum response factor and the serum response element in the promoter of the c-fos proto-oncogen	ELK1	
				5	Atypical chemokine receptor 2	This gene encodes a beta chemokine receptor,	ACKR2		5	Poly(A) polymerase gamma	This gene encodes a member of the poly(A) polymerase family which catalyzes	PAPOLG	

						which is predicted to be a seven transmembran e protein similar to G protein-couple d receptors. Chemokines and their receptor-medi ated signal transduction are critical for the recruitment of effector immune cells to the inflammation site					template-independe nt extension of the 3' end of a DNA/RNA strand		
				6	Mdm4 p53 binding protein homolog (mouse)	This gene encodes a nuclear protein that contains a p53 binding domain at the N-terminus and a RING finger domain	MDM4		6	Hook microtubule tethering protein 3	Hook proteins are cytosolic coiled-coil proteins that contain conserved N-terminal domains, which attach to microtubules, and more divergent C-terminal domains,	HOOK 3	

						at the C-terminus, and shows structural similarity to p53-binding protein MDM2. Both proteins bind the p53 tumor suppressor protein and inhibit its activity					which mediate binding to organelles		
				7	Chemokine (C-C motif) ligand 22	The cytokine encoded by this gene displays chemotactic activity for monocytes, dendritic cells, natural killer cells and for chronically activated T lymphocytes	CCL22		7	AT-rich interaction domain 2	This gene encodes a member of the AT-rich interactive domain (ARID)-containing family of DNA-binding proteins, that it have roles in embryonic patterning, cell lineage gene regulation, cell cycle control, transcriptional regulation and	ARID2	

											chromatin structure modification		
				8	Uncharacterized protein		AL133373.1		8	Zinc finger protein 544		ZNF544	
				9	Uncharacterized protein		AC140061.12		9	RNA binding motif single stranded interacting protein 2	It has diverse functions as DNA replication, gene transcription, cell cycle progression and apoptosis	RBMS2	
				10	Zinc finger protein 880		ZNF880		10	Myosin X		MYO10	
hsa-miR-9-5p	MIMAT0000441	MIMAT0000441	MIMAT0000442	1	One cut homeobox 2	This gene encodes a member of the onecut family of transcription factors, which are characterized by a cut domain and an atypical homeodomain. The protein binds to specific DNA sequences and stimulates expression of	ONECUT2	1377	1	One cut homeobox 2	This gene encodes a member of the onecut family of transcription factors, which are characterized by a cut domain and an atypical homeodomain. The protein binds to specific DNA sequences and stimulates expression of target genes, including genes involved in melanocyte and hepatocyte	ONECUT2	1236

						target genes, including genes involved in melanocyte and hepatocyte differentiation					differentiation		
				2	Y box binding protein 3		YBX3		2	Suppressor of cytokine signaling 5	The protein thus belongs to the suppressor of cytokine signaling family, also known as STAT-induced STAT inhibitor protein family. suppressor of cytokine signaling family members are known to be cytokine-inducible negative regulators of cytokine signaling. The specific function of this protein has not yet been determined	SOCS5	
				3	Lymphatic vessel endothelial hyaluronan receptor 1	The encoded protein acts as a receptor and	LYVE1		3	POU class 2 homeobox 1	Members of this family contain the POU domain, a	POU2F 1	

						binds to both soluble and immobilized hyaluronan. This protein may function in lymphatic hyaluronan transport and have a role in tumor metastasis					160-amino acid region necessary for DNA binding to the octameric sequence ATGCAAAT		
				4	POU class 2 homeobox 1	Members of this family contain the POU domain, a 160-amino acid region necessary for DNA binding to the octameric sequence ATGCAAAT	POU2F1		4	ADP ribosylation factor guanine nucleotide exchange factor 2	ADP-ribosylation factors play an important role in intracellular vesicular trafficking. The protein encoded by this gene is involved in the activation of ADP-ribosylation factors by accelerating replacement of bound GDP with GTP and is involved in Golgi transport	ARFGEF2	
				5	One cut homeobox 1	This gene may	ONECU		5	Protogenin 2	The encoded	PRTG	

						influence a variety of cellular processes including glucose metabolism, cell cycle regulation, and it may also be associated with cancer	T1				transmembrane protein has been associated with the development of various tissues, especially neurogenesis. It has been suggested that this gene may be associated with attention deficit hyperactivity disorder		
				6	Solute carrier family 50 (sugar efflux transporter), member 1		SLC50A1		6	PR/SET domain 6	The encoded protein is involved in regulation of vascular smooth muscle cells contractile proteins	PRDM6	
				7	POU class 6 homeobox 2	This gene is a tumor suppressor involved in Wilms tumor predisposition	POU6F2		7	Potassium voltage-gated channel subfamily J member 2	The encoded protein, which has a greater tendency to allow potassium to flow into a cell rather than out of a cell, probably participates in establishing action potential waveform	KCNJ2	

											and excitability of neuronal and muscle tissues		
				8	Transient receptor potential cation channel, subfamily M, member 7	The encoded protein is involved in cytoskeletal organization, cell adhesion, cell migration and organogenesis	TRPM7		8	Syndecan 2	The syndecan-2 protein functions as an integral membrane protein and participates in cell proliferation, cell migration and cell-matrix interactions via its receptor for extracellular matrix proteins	SDC2	
				9	Adaptor-related protein complex 1, sigma 2 subunit	Adaptor protein complex 1 is found at the cytoplasmic face of coated vesicles located at the Golgi complex, where it mediates both the recruitment of clathrin to the membrane	AP1S2		9	MAM domain containing glycosylphosphatidylinositol anchor 2		MDGA2	

						and the recognition of sorting signals within the cytosolic tails of transmembrane receptors. The protein encoded by this gene serves as the small subunit of this complex and is a member of the adaptin protein family						
				10	Pyruvate dehydrogenase kinase, isozyme 4	This protein is located in the matrix of the mitrochondria and inhibits the pyruvate dehydrogenase complex by phosphorylating one of its subunits,	PDK4		10	Solute carrier family 50 member 1		SLC50 A1	

						thereby contributing to the regulation of glucose metabolism							
hsa-miR-10-a-3p	MIMAT0004555	MIMAT0000253	MIMAT0004555	1	Brain expressed, X-linked 1		BEX1	3450	1	IKAROS family zinc finger 2	This gene encodes a member of the Ikaros family of zinc-finger proteins	IKZF2	188
				2	Tumor-associated calcium signal transducer 2	This intronless gene encodes a carcinoma-associated antigen. This antigen is a cell surface receptor that transduces calcium signals	TACSTD2		2	Zinc finger DHHC-type containing 21		ZDHHC21	
				3	Chromosome 1 open reading frame 145		C1orf145		3	Cyclin dependent kinase 15		CDK15	
				4	Deoxyuridine triphosphatase	The encoded protein forms a ubiquitous, homotetrameric enzyme that hydrolyzes	DUT		4	Contactin 3		CNTN3	

						dUTP to dUMP and pyrophosphate							
				5	SH3 and cysteine rich domain 3	The protein encoded by this gene is a component of the excitation-contraction coupling machinery of muscles	STAC3		5	POC1B-GALNT4 readthrough	This locus represents naturally occurring transcripts that splice the 5' exons of the POC1 centriolar protein homolog B gene on chromosome 12 to the UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 4 gene, which is located within a POC1 centriolar protein homolog B intron	POC1B-GALNT4	
				6	Transmembrane protein 156		TMEM156		6	Polypeptide N-acetylgalactosaminyltransferase 4	his gene encodes a member of the UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase	GALNT4	

											(GalNAc-T) family of enzymes. GalNAc-Ts initiate mucin-type O-linked glycosylation in the Golgi apparatus by catalyzing the transfer of GalNAc to serine and threonine residues on target proteins		
				7	Exosome component 7		EXOSC 7		7	Latent transforming growth factor beta binding protein 3	The protein encoded by this gene forms a complex with transforming growth factor beta proteins and may be involved in their subcellular localization.This protein also may play a structural role in the extracellular matrix	LTBP3	
				8	COBW domain containing 5		CBWD5		8	Brain expressed X-linked 1		BEX1	
				9	Poly(rC) binding protein 1	This intronless gene is	PCBP1		9	Lysophosphatidic acid receptor 4	The encoded protein may play a	LPAR4	

						thought to have been generated by retrotransposition of a fully processed PCBP-2 mRNA. The protein encoded by this gene appears to be multifunctional. It along with PCBP-2 and hnRNPK corresponds to the major cellular poly(rC)-binding protein.					role in monocytic differentiation		
				10	Integral membrane protein 2A	This gene encodes a type II membrane protein that belongs to the ITM2 family. Studies in	ITM2A		10	EPH receptor A7	This gene belongs to the ephrin receptor subfamily of the protein-tyrosine kinase family. EPH and EPH-related receptors have been	EPHA7	

						mouse suggest that it may be involved in osteo- and chondrogenic differentiation					implicated in mediating developmental events, particularly in the nervous system		
hsa-miR-182-3p	MIMAT0000260	MIMAT0000259	MIMAT0000260	1	GIPC PDZ domain containing family, member 3	The protein encoded by this gene belongs to the GIPC family	GIPC3	3329	1	Bradykinin receptor B2	This gene encodes a receptor for bradykinin. The 9 aa bradykinin peptide elicits many responses including vasodilation, edema, smooth muscle spasm and pain fiber stimulation	BDKRB2	388
				2	Bradykinin receptor B2	This gene encodes a receptor for bradykinin. The 9 aa bradykinin peptide elicits many responses including vasodilation,	BDKRB2		2	Ral GEF with PH domain and SH3 binding motif 1		RALGPS1	

						edema, smooth muscle spasm and pain fiber stimulation							
				3	Transferrin receptor	This gene encodes a cell surface receptor necessary for cellular iron uptake by the process of receptor-mediated endocytosis	TFRC		3	Tetratricopeptide repeat domain 13		TTC13	
				4	Mitochondrial ribosomal protein L13	This gene encodes a 39S subunit protein	MRPL13		4	Cell adhesion molecule 1		CADM 1	
				5	Fasciculation and elongation protein zeta 2 (zygin II)	This gene is an ortholog of the C. elegans unc-76 gene, which is necessary for normal axonal bundling and elongation within axon	FEZ2		5	Interleukin 6 signal transducer	This protein functions as a part of the cytokine receptor complex. The activation of this protein is dependent upon the binding of cytokines to their receptors	IL6ST	

						bundles							
				6	Zinc finger protein 622		ZNF622		6	Latent transforming growth factor beta binding protein 2	It has thus been suggested that it may have multiple functions: as a member of the TGF-beta latent complex, as a structural component of microfibrils, and a role in cell adhesion	LTBP2	
				7	Small leucine-rich protein 1		SMLR1		7	cAMP responsive element modulator	This gene encodes a bZIP transcription factor that binds to the cAMP responsive element found in many viral and cellular promoters	CREM	
				8	Uncharacterized protein	This transcript is not in the current gene set	AC020922.1		8	One cut homeobox 2	This gene encodes a member of the onecut family of transcription factors, which are characterized by a cut domain and an atypical homeodomain. The	ONECUT2	

											protein binds to specific DNA sequences and stimulates expression of target genes, including genes involved in melanocyte and hepatocyte differentiation		
				9	GINS complex subunit 3 (Psf3 homolog)	This gene encodes a protein subunit of the GINS heterotetrameric complex, which is essential for the initiation of DNA replication and replisome progression in eukaryotes	GINS3		9	PWWP domain containing 3B	This gene encodes a protein which contains a mutated melanoma-associated antigen 1 domain	PWWP 3B	
				10	BCL2/adenovirus E1B 19kD interacting protein like	It may function as a bridge molecule between BCL2 and	BNIPL		10	Alport syndrome, mental retardation, midface hypoplasia and elliptocytosis chromosomal region	The exact function of this gene is not known, however, submicroscopic deletion of the X	AMMECR1	

						ARHGAP1/CDC42 in promoting cell death				gene 1	chromosome including this gene, result in a contiguous gene deletion syndrome		
hsa-miR-301-b-3p	MIMAT0004958	MIMAT0032026	MIMAT0004958	1	SLAIN motif family member 1		SLAIN1	1029	1	Gap junction protein alpha 1	This gene is a member of the connexin gene family. The encoded protein is a component of gap junctions, which are composed of arrays of intercellular channels that provide a route for the diffusion of low molecular weight materials from cell to cell	GJA1	913
				2	Mdm4 p53 binding protein homolog (mouse)	Bind the p53 tumor suppressor protein and inhibit its activity, and have been shown to be overexpressed	MDM4		2	SLAIN motif family member 1		SLAIN1	

						in a variety of human cancers							
				3	Lysine (K)-specific demethylase 2A	This gene encodes a member of the F-box protein family which is characterized by an approximately 40 amino acid motif, the F-box. The F-box proteins constitute one of the four subunits of ubiquitin protein ligase complex called SCFs, which function in phosphorylation-dependent ubiquitination	KDM2A		3	CAP-Gly domain containing linker protein 1	The protein encoded by this gene links endocytic vesicles to microtubules	CLIP1	
				4	Kruppel-like factor 7 (ubiquitous)	The protein encoded by this gene is a	KLF7		4	SKI/DACH domain containing 1		SKIDA 1	

						member of the Kruppel-like transcriptional regulator family. Members in this family regulate cell proliferation and differentiation							
				5	PAN3 poly(A) specific ribonuclease subunit homolog (*S. cerevisiae*)		PAN3		5	Cytoplasmic polyadenylation element binding protein 1	This gene encodes a member of the cytoplasmic polyadenylation element binding protein family. It is involved in a role in cell proliferation and tumorigenesis	CPEB1	
				6	Ectonucleotide pyrophosphatase/phosphodiesterase 5 (putative)	This gene encodes a type-I transmembrane glycoprotein. Studies in rat suggest the encoded protein may	ENPP5		6	TSC complex subunit 1	This gene is a tumor suppressor gene that encodes the growth inhibitory protein hamartin	TSC1	

						play a role in neuronal cell communications							
				7	F-box protein 28	Members of the F-box protein family, such as FBXO28, interact with SKP1 through the F box, and F box interact with ubiquitination targets through other protein interaction domains	FBXO28		7	Estrogen receptor 1	This gene encodes an estrogen receptor, a ligand-activated transcription factor composed of several domains important for hormone binding, DNA binding, and activation of transcription	ESR1	
				8	V-myb avian myeloblastosis viral oncogene homolog-like 1		MYBL1		8	Kruppel-like factor 7 (ubiquitous)	The protein encoded by this gene is a member of the Kruppel-like transcriptional regulator family. Members in this family regulate cell proliferation and differentiation	KLF7	

				9	Activin A receptor type 1	Activins are dimeric growth and differentiation factors which belong to the transforming growth factor-beta (TGF-beta) superfamily of structurally related signaling proteins	ACVR1		9	RAP2C, member of RAS oncogene family	The protein encoded by this gene is a member of the Ras-related protein subfamily of the Ras GTPase superfamily. Members of this family are small GTPases that act as molecular switches to regulate cellular proliferation, differentiation, and apoptosis	RAP2C	
				10	Cytoplasmic polyadenylation element binding protein 1	This gene encodes a member of the cytoplasmic polyadenylation element binding protein family. It is involved in a role in cell proliferation and tumorigenesis	CPEB1		10	Ribosomal protein S6 kinase A5		RPS6KA5	
hsa-	MIMAT00	MIMAT00	MIMAT00	1	Zinc finger protein 99		ZNF99	4304	1	Sestrin 3	The encoded	SESN3	991

miR-9-3p	00442	00441	00442								protein reduces the levels of intracellular reactive oxygen species induced by activated Ras downstream of RAC-alpha serine/threonine-protein kinase (Akt) and FoxO transcription factor. The protein is required for normal regulation of blood glucose, insulin resistance and plays a role in lipid storage in obesity		
				2	Chromosome 10 open reading frame 11	This gene encodes a leucine-rich repeat protein. The encoded protein is thought to play a role in melanocyte differentiation	C10orf11		2	One cut homeobox 2	This gene encodes a member of the onecut family of transcription factors, which are characterized by a cut domain and an atypical homeodomain. The protein binds to specific DNA sequences and	ONECUT2	

											stimulates expression of target genes, including genes involved in melanocyte and hepatocyte differentiation		
				3	Chromosome 12 open reading frame 36		C12orf36		3	Collagen type IV alpha 3 binding protein	This gene encodes a kinase that specifically phosphorylates the N-terminal region of the non-collagenous domain of the alpha 3 chain of type IV collagen, known as the Goodpasture antigen	COL4A3BP	
				4	Actin-like 6A	his gene encodes a family member of actin-related proteins, which share significant amino acid sequence identity to conventional	ACTL6A		4	Integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)	Integrin family members are membrane receptors involved in cell adhesion and recognition in a variety of processes including embryogenesis, hemostasis, tissue repair, immune	ITGB1	

						actins.The actin-related proteins are involved in diverse cellular processes, including vesicular transport, spindle orientation, nuclear migration and chromatin remodeling					response and metastatic diffusion of tumor cells		
				5	DNA replication and sister chromatid cohesion 1	It is component of an alternative replication factor Ccomplex that loads PCNAonto DNA during S phase of the cell cycle	DSCC1		5	YOD1 deubiquitinase	Deubiquitinating enzymes are cysteine proteases that specifically cleave ubiquitin from ubiquitin-conjugated protein substrates. The protein encoded by this gene belongs to a DUB subfamily characterized by an ovarian tumor	YOD1	

											domain		
				6	Mitotic spindle organizing protein 1		MZT1		6	Discoidin, CUB and LCCL domain containing 2		DCBLD2	
				7	N(alpha)-acetyltransferase 20, NatB catalytic subunit	It is a component of N-acetyltransferase complex B. Human N-acetyltransferase complex B performs cotranslational N(alpha)-terminal acetylation of methionine residues when they are followed by asparagine	NAA20		7	Nuclear transport factor 2 like export factor 2	The protein encoded by this gene contains a nuclear transport factor 2 domain, which plays an important role in the trafficking of macromolecules, ions, and small molecules between the cytoplasm and nucleus	NXT2	
				8	General transcription factor IIH, polypeptide 5	This gene encodes a subunit of transcription/repair factor TFIIH, which functions in gene transcription	GTF2H5		8	Recombination signal binding protein for immunoglobulin kappa J region	The encoded protein acts as a repressor when not bound to Notch proteins and an activator when bound to Notch proteins. It is thought to function	RBPJ	

						and DNA repair					by recruiting chromatin remodeling complexes containing histone deacetylase or histone acetylase proteins to Notch signaling pathway genes		
				9	GTP-binding protein 8 (putative)		GTPBP 8		9	PHD finger protein 3	This gene may function as a transcription factor and may be involved in glioblastomas development	PHF3	
				10	MT-RNR2-like 8	Note: It is unclear if this is a transcribed protein-coding gene, or if it is a nuclear pseudogene of the mitochondrial MT-RNR2 gene	MTRNR 2L8		10	Killin, p53 regulated DNA replication inhibitor	The protein encoded by this intronless gene is found in the nucleus, where it can inhibit DNA synthesis and promote S phase arrest coupled to apoptosis	KLLN	

hsa-miR-301-b-5p	MIMAT0032026	MIMAT0032026	MIMAT0004958	1	Zinc finger protein 286B		ZNF286B	4447	1	Nuclear receptor subfamily 2 group C member 2	The protein encoded by this gene plays a role in protecting cells from oxidative stress and damage induced by ionizing radiation	NR2C2	457
				2	Phosphatase and actin regulator 2		PHACTR2		2	DExH-box helicase 9	The encoded protein is an enzyme that catalyzes the ATP-dependent unwinding of double-stranded RNA and DNA-RNA complexes	DHX9	
				3	Zinc finger protein 286A		ZNF286A		3	Basonuclin 2	The encoded protein functions in skin color saturation	BNC2	
				4	Zinc finger protein 487		ZNF487		4	Peptidylprolyl isomerase like 3	This gene encodes a member of the cyclophilin family. Cyclophilins catalyze the cis-trans isomerization of peptidylprolyl imide bonds in oligopeptides. They	PPIL3	

											have been proposed to act either as catalysts or as molecular chaperones in protein-folding events		
				5	Periaxin	This gene encodes a protein involved in peripheral nerve myelin upkeep	PRX		5	Heparan sulfate-glucosamine 3-sulfotransferase 4		HS3ST4	
				6	Ubiquitin-conjugating enzyme E2L 6	This gene encodes a member of the E2 ubiquitin-conjugating enzyme family	UBE2L6		6	Nicotinamide phosphoribosyltransferase	The protein belongs to the nicotinic acid phosphoribosyltransferase family and is thought to be involved in many important biological processes, including metabolism, stress response and aging	NAMPT	
				7	GRIK1 antisense RNA 2		GRIK1-AS2		7	Forkhead box J2		FOXJ2	
				8	Peptidylprolyl isomerase like 3	This gene encodes a member of the	PPIL3		8	Zinc finger E-box binding homeobox 1	This gene encodes a zinc finger transcription factor.	ZEB1	

						cyclophilin family. Cyclophilins catalyze the cis-trans isomerization of peptidylprolyl imide bonds in oligopeptides. They have been proposed to act either as catalysts or as molecular chaperones in protein-folding events					The encoded protein likely plays a role in transcriptional repression of interleukin 2		
				9	Potassium voltage-gated channel, subfamily H (eag-related), member 1	Diverse functions include regulating neurotransmitter release, heart rate, insulin secretion, neuronal excitability,	KCNH1		9	Nuclear transcription factor Y subunit beta	The protein encoded by this gene is one subunit of a trimeric complex, forming a highly conserved transcription factor that binds with high specificity to CCAAT motifs in the promoter regions in	NFYB	

						epithelial electrolyte transport, smooth muscle contraction, and cell volume.Overe xpression of the gene may confer a growth advantage to cancer cells and favor tumor cell proliferation					a variety of genes		
				10	Peptidylprolyl isomerase H (cyclophilin H)	Accelerate the folding of proteinsand may act as a protein chaperone that mediates the interactions between different proteins inside the spliceosome	PPIH		10	Hematopoietically expressed homeobox	This gene encodes a member of the homeobox family of transcription factors, many of which are involved in developmental processes. Expression in specific hematopoietic lineages suggests that this protein may	HHEX	

											play a role in hematopoietic differentiation		

Fonte: Própria.

ANEXO 1 - Instruções para autores da revista "Oncology Letters"

Manuscript submission process

The content of this webpage is available for download as a PDF.

The content of this webpage is available for download in PDF.

Please visit "Information for authors" for each of our journals:

International Journal of Molecular Medicine	**International Journal of Oncology**
Experimental and Therapeutic Medicine	**Molecular and Clinical Oncology**
World Academy of Sciences Journal	**Molecular Medicine Reports**
Oncology Reports	**Biomedical Reports**
Oncology Letters	**International Journal of Functional Nutrition**
International Journal of Epigenetics	**Medicine International**

Online Submission

Spandidos Publications journals utilize an online submission and tracking system designed to provide a faster, more efficient service to authors.

Contents

Submission guidelines

- The principal aim of Spandidos Publications is to publish promptly original works of high quality in English.
- Manuscripts will be considered on the understanding that they report original work, or are review articles summarizing and interpreting progress in a thematic area and are not under consideration for publication by another journal.
- Manuscripts should be written in clear, concise English and should contain all essential data in order to make the presentation clear and the results of the study replicable.
- Authors who are not native speakers of English but would like to improve the English to ensure the meaning is clear can make use of the English Language Editing service offered by Spandidos Publications by accessing the following link: https://www.spandidos-publications.com/languageediting.
- All material submitted will be subject to review by appropriate referees selected by the Editorial Office and will be examined to detect inappropriate use of previously published material without attribution.
- Spandidos Publications utilizes iThenticate to screen submitted manuscripts against published studies and other relevant sources. iThenticate may also be utilized by authors to screen a manuscript prior to submission.
- Figures submitted will be subject to checks using the MOTUIN image authenticity detector.
- The Editors reserve the right to improve the grammar and style of manuscripts.
- The corresponding author is responsible for the submission on behalf of all authors.

Prior to submitting your manuscript, please ensure that it has been prepared according to the guidelines below.

1. Submission method

- Manuscripts may be submitted only by using the online submission system accessible via our website.
- Create a user account, log in and follow the onscreen directions.

2. Cover letter

Summarize briefly the important points of the submitted work including a brief description of the study to be submitted, that it is an original study presenting novel work, that it has not been previously submitted to or accepted by any other journal, that is has been approved by all authors, that ethics approval and written informed consent have been obtained, and explain whether any author has a conflict of interest.

3. Format of articles and reviews

3.1 General style

- Times New Roman. Font size 12. Spacing 1.5. Alignment Justified.
- Use a single tab on the first line of each new paragraph.
- Do not use page breaks or multiple returns between sections (one section should directly follow the previous one on the page).
- Do not insert page numbers or line numbers.
- Sub-headings and general headings should be presented in lower case letters (not capitals).
- Use British English or American English spellings throughout your manuscript, but not both.

3.2 Manuscripts

The first page should include:

- The title of the manuscript in sentence case. No abbreviations other than gene names or in common use.
- Full names and full postal addresses, but not including street names, of all authors and ORCID if desired.
- Affiliations of the authors indicated by numbers (not symbols).
- Equal contribution indicated by asterisk.
- Name, full postal address, including street number and name, and e-mail address of the corresponding author(s).
- Abbreviations, if relevant.
- Key words (5-10).
- Running title preceded by the first author's name (maximum 100 characters with spaces, including the author's name). For example: PEARSON *et al*: REGULATION OF HER2 EXPRESSION BY NASCENT GROWTH FACTORS.

Manuscripts reporting experimental results must be divided into the following sections:

- Abstract. This section should have 150-300 words, be continuous (not structured) and without reference numbers. Abbreviations that appear once only, should be defined in full, unless they correspond to a gene name. If abbreviations appear more than once, the definition should be provided once, and then subsequently used throughout the Abstract.
- Introduction. The information in this section should always be referenced.

- Materials and methods
  - This section should include sufficient technical information to allow the experiments to be repeated. This implies that a full description of all the experiments described in Results and presented in the Figures/Tables is expected in this section. For each experiment, all steps (e.g., DNA and protein extraction, quantification, cloning, PCR and microscopy) need to be mentioned, along with instruments the analyses were performed on, reagents and methods (e.g., BCA method for protein quantification, ΔΔCq method for qPCR), and relevant citations. For specific details on our standards of reporting for individual techniques, please click here.
  - For steps performed with commercialized kits, provide the full name of the kit, along with the full name and location (city, province or state if USA/Canada, and country) of the supplier, and state whether the protocol of the manufacturer was followed or explain any modifications made to the standard protocol. For PCRs, provide the name of the kit used, 5'-3' sequence of the primers, final concentration of all reagents in the reaction, and cycling conditions. Carefully review your text to ensure that the type of PCR, quantitative or semi-quantitative, is clearly explained. If the PCR is performed using cDNA synthesized from RNA samples by reverse transcription (RT), make sure that all steps are described, and refer to the method as RT-PCR or, if quantitative, as RT-qPCR. In relative quantification, ΔΔCt is referred to as ΔΔCq. When using the ΔΔCq method, this **must** be referenced. One suitable reference is: Livak and Schmittgen: Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCt method. Methods 25: 402-408, 2001. Manufacturers/suppliers/software details need to be provided for all reagents used (including chemicals), instruments (e.g. thermal cyclers, microscopes) and software, ideally accompanied by the corresponding kit number/model/version. For antibodies, include the type (monoclonal/polyclonal), species in which they were raised and targeted species (e.g., mouse anti-human). Please explain any antigen retrieval steps, mention the dilution used, and state the catalogue number and supplier. Please also state the temperature and duration of incubation. For centrifugation steps, provide centrifugal force units in x g rather than revolutions per minute (rpm).
  - For bioinformatic analyses: state the software used along with the relevant citation, unless the software is not published, in which case a website link can be provided. For microarray/RNA sequences, data downloaded from GEO or other databases, this needs to be clarified in the text, along with the corresponding accession number of the dataset. The use of software should be described with regards to the parameters (default, study-specific) and the applied thresholds; please explicitly name the parameters, e.g. 'association value' or 'false-positive rate'. For all software analysis of data from public databases, cite the database (along with date of access for databases as these are constantly updated), and species (e.g., human). If figures/tables contain data from a public database (e.g., Gene Ontology/KEGG), cite the source in the legend/title explicitly. For publically available sequences, provide the accession number.
  - For flow cytometry experiments authors are encouraged to adhere to MIFlowCyt guidelines (Lee J et al. (2008) MIFlowCyt: The minimum information about a flow cytometry experiment. Cytometry 73A: 926-930. doi: 10.1002/cyto.a.20623). Axis labels should include the marker and the dye used (rather than instrument-specific parameter descriptions such as FL-1H). The scaling (log/lin) should be clearly displayed. If most events are "piled up" on the plot axes, adjust scale or provide a different scale if necessary. If statistical analysis is provided, please clarify if it is the fluorescence intensity of the gated population (mean, median, Geo mean) or the proportion of cells within a specific gate that is examined.

- The source of material used and relevant ethical framework for all experiments should be clearly identified (ethics approval and/or written informed consent). For tissues, explain how these were collected, handled and stored, and where they were from. For bacterial strains or cells, provide the name and supplier. For studies on humans, a minimum of information is required: number of subjects, age range, gender ratio, health status, matching between controls and disease patients with regards to the above parameters. Please note that 'normal' should be avoided for controls; rather, the precise health status needs to be described, e.g., 'healthy', or 'individuals with no recorded tumor complication'. For manuscripts presenting studies on humans and animals see 3.8 below.
- For statistical analyses: when statistical analyses have been performed, the following information should be provided: the name of the statistical test used, the n number for each analysis, the comparisons of interest, the alpha level and the actual P-value for each test. It should be clear which statistical test was used to generate every P-value. Error bars on graphs should be clearly labeled, and it should be stated whether the number following the ± sign is a standard deviation or a standard error. The word 'significant' should only be used when referring to statistically significant results and should be accompanied by the relevant P-value. Significance indicators should be used on graphs and tables, and should be described in the figure or table legend, clearly indicating which groups are being compared.

  If your study involves performing statistical analysis of datasets containing 3 or more groups, please note that the use of Student's t-test, Mann-Whitney U test or similar two-sample tests is not considered appropriate, as when these tests are used to perform multiple comparisons, the familywise error rate is raised to unacceptably high levels. These data should be analyzed using tests designed to control for type I error; the most popular tests for such comparisons are analysis of variance (ANOVA) followed by a post hoc test (e.g. Tukey, Bonferroni, Dunnett, etc.) for parametric data, or Kruskal-Wallis test followed by Dunn's test for non-parametric data.
- Please note that figure legends are not expected to contain information already described in Materials and methods, except for image-specific information, for example, for microscopy, mention the type of image, e.g., fluorescence and the original magnification if scale bars are not used. Legends should provide information concerning what is shown in the figure(s)/figure parts. The x- and y-axes of the graphs must be clearly explained in the legends, and when P-values are provided to indicate probability, the comparison to which these P-values refer must be clearly stated.
- If cell lines are used, authors are strongly encouraged to include the following information in the materials and methods section of their manuscript: i) Confirm that mycoplasma testing has been done for the cell lines used; ii) confirm that the cell lines used have been authenticated and state what method was used for the authentication; and iii) provide the source, supplier and, if available, catalogue number of all specific cell lines used in the study. The authors are strongly encouraged to submit a detailed methodology stating the maintenance and culture of cell lines according to international guidelines on good cell culture practice (fundamental techniques, mycoplasma contamination, passage number, etc.). Furthermore, information regarding misidentified or cross-contaminated cell lines must be provided and cross-checked from the International Cell Line Authentication Committee and ExPASy Cellosaurus databases in order to exclude their contamination with other cell lines or their incorrect identification. If a cell line has been previously reported

to be contaminated or misidentified, an STR profile of the cell line used in the study must be available for evaluation by the journal's editor.

- Results
- Discussion
- Acknowledgements
- Funding
- Availability of data and materials
- Authors' contributions
- Ethics approval and consent to participate
- Patient consent for publication
- Competing interests
- Authors' information (optional)
- References

Footnotes should not be used.

For Review articles:

- Abstract. This section should have 150-300 words, be continuous (not structured) and without reference numbers.
- May have different sections and sub-headings according to the subject matter.
- The main headings of the review should be summarized as a numbered Contents section immediately following the Abstract.

3.3 Figures

Submission of figures to us implies that the images or parts thereof have not been published elsewhere (unless mentioned and/or cited in the text and permission has been obtained and provided to us).

Images showing any patient or patient's scans should not contain information that might identify them, unless you provide written permission from the patient allowing use of the specific image.

We accept that figures in our journals are rarely simple, and that certain adjustments are acceptable to help show experimental results clearly. The guiding principle when preparing digital artwork should be to ensure that the version submitted to us is an honest and accurate representation of the original observation(s) and will not lead to possible misinterpretation of what was done experimentally.

The Editors may assess submitted images for unacceptable manipulation using forensic tools and other means. This might delay progress of your manuscript and/or lead to further investigations and action to preserve the integrity of the scientific record, such as not accepting or revoking a manuscript. We may request the original unmanipulated source files and may contact the author's institution for assistance with enquiries to establish probity. Our guidance builds on that described by Rossner and Yamada (1).

- If brightness, contrast or color balance is altered, the change should apply to the entire image shown and not a selected part. For images from gels or filters, ensure that details are not lost from bright areas or obscured in dark areas.

- No feature of a data image should be selectively enhanced, obscured, removed or added. If a composite image shows gels or blots with tracks from groups of samples analyzed separately (or from different exposures) then make the grouping obvious using black or white lines and explain this in the figure legend. The boundaries of individual panels of a tiled image should be marked.
- In the Methods or individual figure legends, outline the changes you made to images and how. For example, "Figure 99. Light microscopy of a frozen section of a lesion stained with toluidine blue. Original magnification x100. Uneven illumination was corrected using a control image as described (2)".

(1) Rossner M and Yamada KM: What's in a picture? The temptation of image manipulation. J Cell Biol 166: 11-15, 2004.

(2) Marty GD: Blank-field correction for achieving a uniform white background in brightfield digital photomicrographs. BioTechniques 42: 716-720, 2007.

3.3.1 File format

- Acceptable
  - TIFF without layers and preferably using Lempel-Ziv-Welch (LZW) compression as it does not reduce image quality.
  - JPEG (only if originally saved at the highest quality).
- Unacceptable
  - Images imported or copy pasted into Word or PowerPoint.
  - BMP, GIF, PCT, PNG or low quality JPEG files originally saved at low quality.

3.3.2 Color mode

- Acceptable:
  - Color figures: Use RGB as this will offer the best reproduction of your data in the final PDF version of your article on screen. CMYK mode is also acceptable. Fluorescence images must be submitted for publication in color.
  - Black and white figures and line art: grey scale mode or RGB mode.
  - Combination figures with color images and line art: RGB mode.
- PLEASE NOTE
  - Color figures are welcome but must be submitted only if reproduction in color is intended (a charge will apply).
  - There is a charge of Euro 390 per each published page containing color.
  - Changing color figures to black and white following evaluation is NOT possible.

3.3.3 Image size

- Image size is measured in centimeters or inches
- Create your figures at the size (width) at which they will be printed:
  - 8.00 cm (3.15 in) wide for a single-column figure
  - 17.00 cm (6.70 in) maximum for a double-column (full page width) figure
  - Maximum height 20.00 cm (7.87 in)

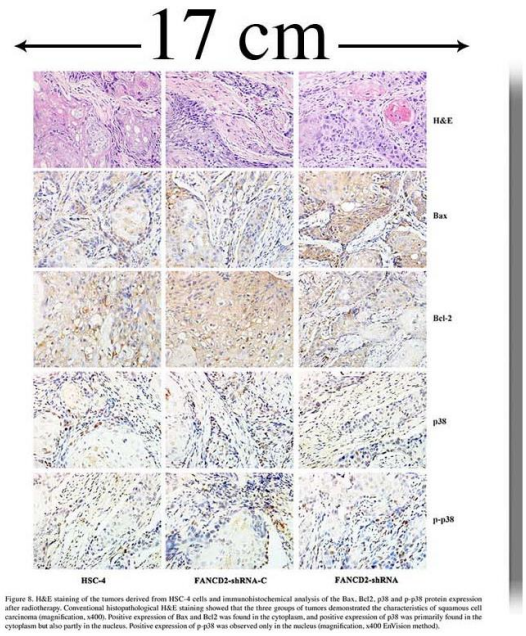


Figure 8. H&E staining of the tumors derived from HSC-4 cells and immunohistochemical analysis of the Bax, Bcl2, p38 and p-p38 protein expression after radiotherapy. Conventional histopathological H&E staining showed that the three groups of tumors demonstrated the characteristics of squamous cell carcinoma (magnification, x400). Positive expression of Bax and Bcl2 was found in the cytoplasm, and positive expression of p38 was primarily found in the cytoplasm but also partly in the nucleus. Positive expression of p-p38 was observed only in the nucleus (magnification, x400 EnVision method).

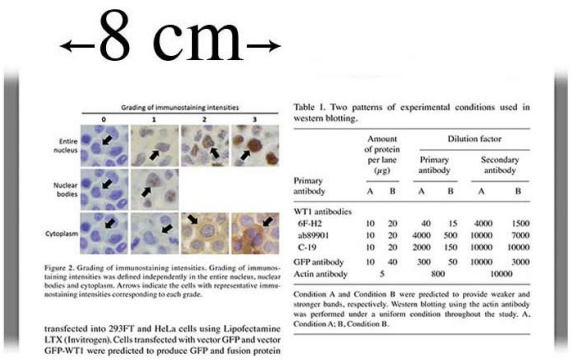


Figure 2. Grading of immunostaining intensities. Grading of immunostaining intensities was defined independently in the entire nucleus, nuclear bodies and cytoplasm. Arrows indicate the cells with representative immunostaining intensities corresponding to each grade.

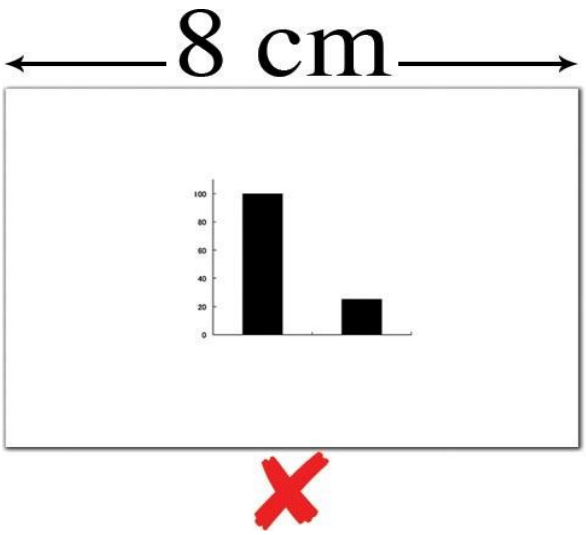
transfected into 293FT and HeLa cells using Lipofectamine LTX (Invitrogen). Cells transfected with vector GFP and vector GFP-WT1 were predicted to produce GFP and fusion protein

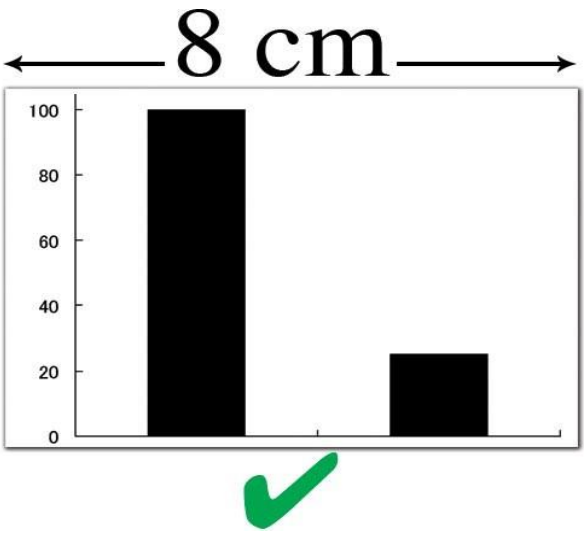
Table 1. Two patterns of experimental conditions used in western blotting.

Primary antibody	Amount of protein per lane (μg)		Dilution factor: Primary antibody		Dilution factor: Secondary antibody	
	A	B	A	B	A	B
WT1 antibodies						
6F-H2	10	20	40	15	4000	1500
ab89901	10	20	4000	500	10000	7000
C-19	10	20	2000	150	10000	10000
GFP antibody	10	40	300	50	10000	3000
Actin antibody	5		800		10000	

Condition A and Condition B were predicted to provide weaker and stronger bands, respectively. Western blotting using the actin antibody was performed under a uniform condition throughout the study. A, Condition A; B, Condition B.

Empty white space surrounding a figure should NOT be included when calculating image size. Images should, therefore, be cropped (cut) as close to the outside edges of the figure as possible.





If a figure is too wide or contains too much information to be fit within 17 cm while keeping details clearly visible, figures must be divided into several clearly labeled separate parts.

3.3.4 Image resolution

- Image resolution in this context is simply a measure of the number of pixels per inch (also called dots per inch, dpi) defining the image and does not relate to the quality of an image in terms of focus, contrast and legibility.
- Images must be clear, of good contrast and legible at the size they are to appear in the journal.
- Images should be AT LEAST 300 dpi, at the size at which they will be printed (8 or 17 cm wide).
- Insufficient image size and/or resolution (dpi) will result in poor quality (blurred) printed figures if they are upscaled.

3.3.5 Exporting/capturing/saving figures

Figures may be produced by scanning, digital photography, or exporting from scientific software or a program such as PowerPoint.

- Scanning
  - Use a good quality scanner set to scan in RGB for color images or grey scale for line art or to scan gel images, at a resolution of at least 300 dpi and with the output file type set preferably to TIFF or JPG with the highest quality (lowest compression).
- Digital photographs
  - Set simple cameras to a 'fine' or 'extra fine' setting to help ensure that images have sufficient pixels.
  - Exporting
    - When exporting from scientific graphing software, choose settings to ensure the highest possible final size and resolution with lines of sufficient thickness to be seen at final printed size.
    - When exporting from PowerPoint, DO NOT choose 'Save as TIFF' from the Save as dialogue box as this will NOT result in an image of sufficiently high resolution. Instead, save the individual slide image as a PDF (from the Print dialogue box), THEN open the PDF with image editing software, such as Photoshop or GIMP, and when prompted specify 300 dpi resolution. Finally, save the resulting image as a TIFF (with LZW compression).
    - Note: figures initially scanned, photographed or exported at an insufficient size and resolution cannot be improved by upscaling, i.e., artificially increasing the resolution of a low-quality figure. Using image-editing software to keep the figure size the same while raising the dpi will NOT improve its quality.

3.3.6 File size

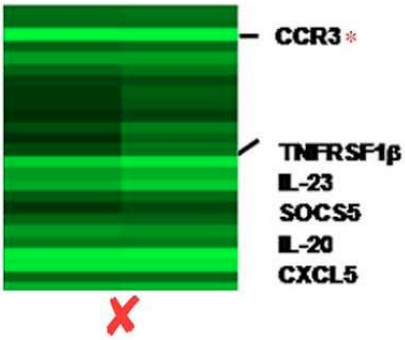
- If saved according to our guidelines, files will rarely exceed 10 MB.
- To reduce the file size of images:
  - Ensure figures are the exact width and height they should be for publication (not smaller), make sure the figures are saved at no more than 300 dpi.
  - Ensure that layers in the image have been flattened.
  - Save black and white figures as grey scale.
  - Ensure that TIFF files are saved with LZW compression.
  - Consider saving files as highest quality JPEGs. These may be smaller files than TIFF with LZW compression, but will lose some detail.
  - Try using a compression or stuffing utility, such as WinZip or StuffIt.

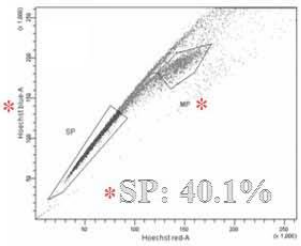
3.3.7 Figure labels

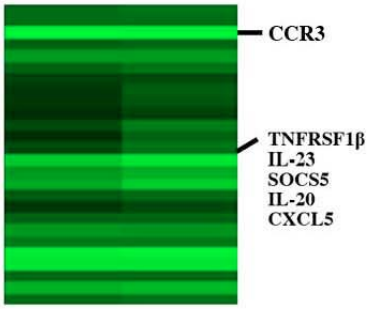
- Font size
  - Labels must be sized in proportion to the image, sharp, and clearly legible.
  - When figures are prepared at the correct size (8 or 17 cm at 300 dpi), the font size for labels should be 8-10 points.
  - If the figure is saved at a size larger than that needed for printing, the font size of labels must also be larger to maintain the correct proportions.

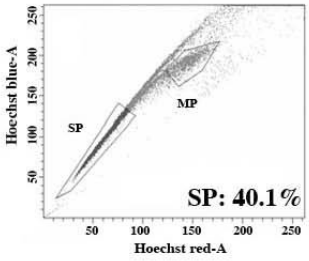
- If labels cannot fit on an 8-cm-wide page unless the font size is smaller than 8 points, the figure must be prepared as a double column figure (14-17 cm wide). If labels cannot fit on the 17-cm-wide page unless the font size is smaller than 8 points, the figure must be split into several parts.

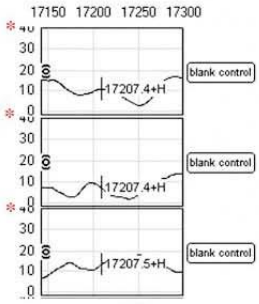
- Font style and appearance
  - Labels must be saved using standard fonts (Times New Roman, Times, Arial, Helvetica or Symbol font).
  - The labels should be of the same font and size in all figures. Also, the numbering should be of the same font and size in all figures.
  - Labels should be evenly spaced and aligned, easy to see (including exponential numbers around figure axes), and NOT faded, broken, or distorted by JPG compression artifact. Do NOT use light grey color lines or labels.
  - There must be strong contrast between labels and their background (e.g., labels placed over shaded bar graphs should be in a color that stands out against the shading, NOT blend in with it). Whenever possible, labels should be placed in black font on a white background. Consider using a black label with a white stroke applied to create contrast.
  - Letters of labels must NOT be overlapping, condensed, expanded, have unnecessary gaps between them or be otherwise irregularly spaced, and must NOT be stretched (distorted) horizontally or vertically.
  - Labels must NOT overlap or be concealed by other parts of the image, or be cropped (cut off) by the edge of the figure.

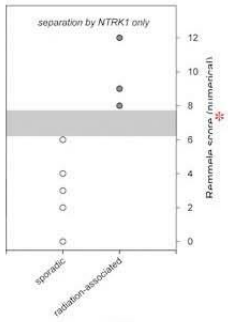

CCR3
TNFRSF1β
IL-23
SOCS5
IL-20
CXCL5

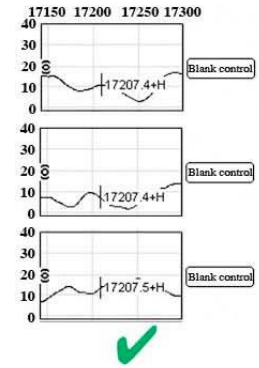

SP: 40.1%

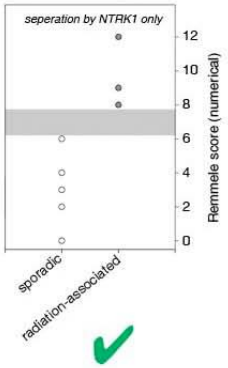

CCR3
TNFRSF1β
IL-23
SOCS5
IL-20
CXCL5


Hoechst blue-A
Hoechst red-A
SP
MP
SP: 40.1%

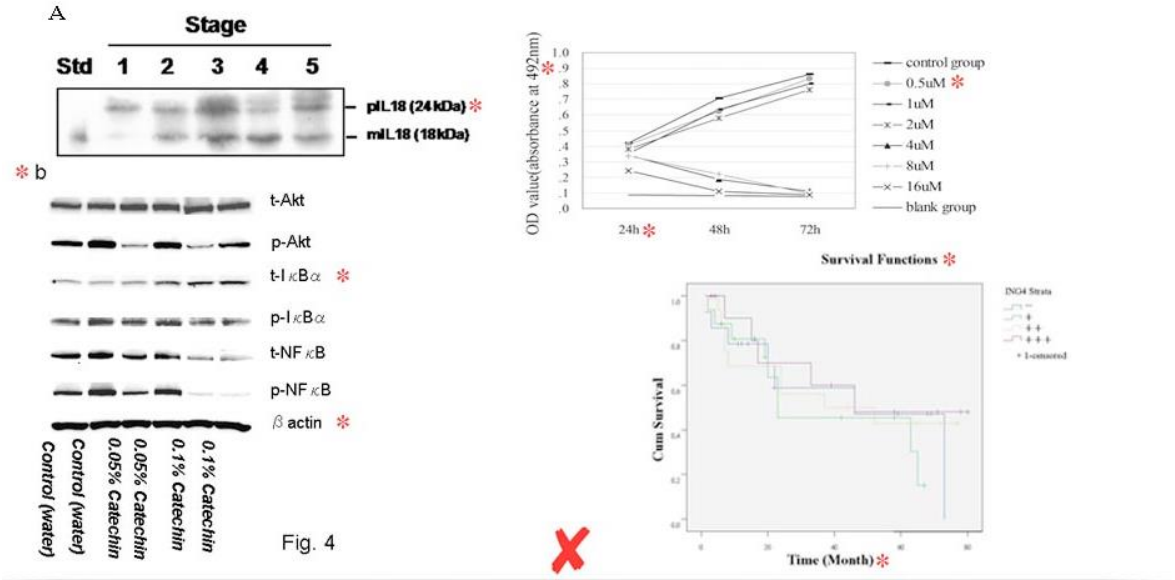

17150 17200 17250 17300
blank control
17207.4+H
blank control
17207.4+H
blank control
17207.5+H

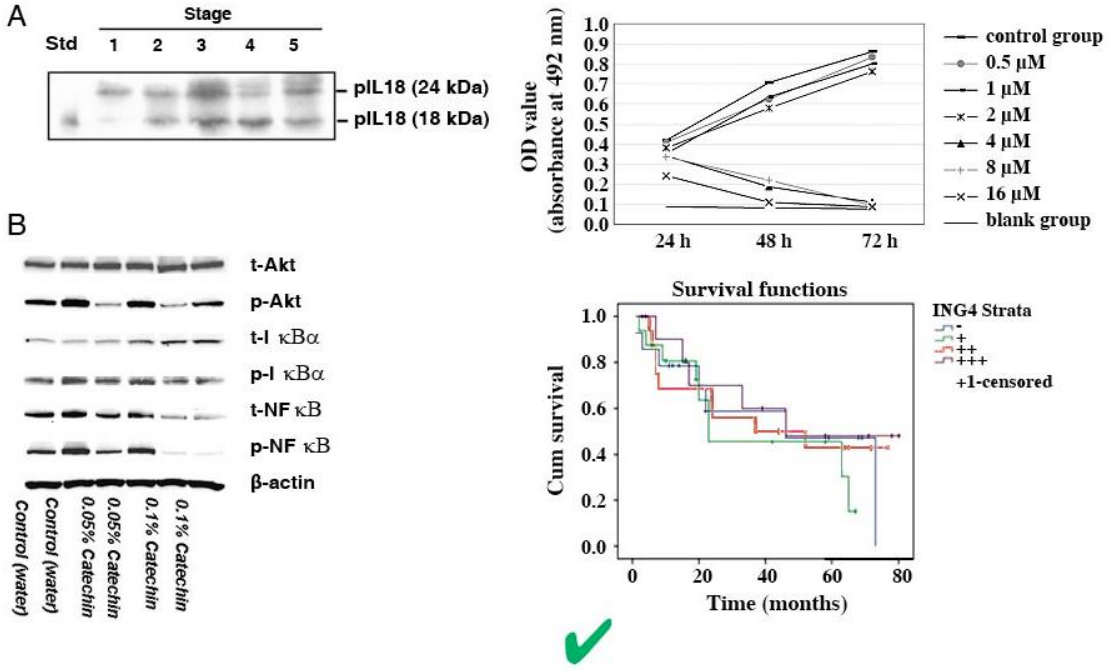

separation by NTRK1 only
Remmele score (numerical)
sporadic
radiation-associated


17150 17200 17250 17300
Blank control
17207.4+H
Blank control
17207.4+H
Blank control
17207.5+H

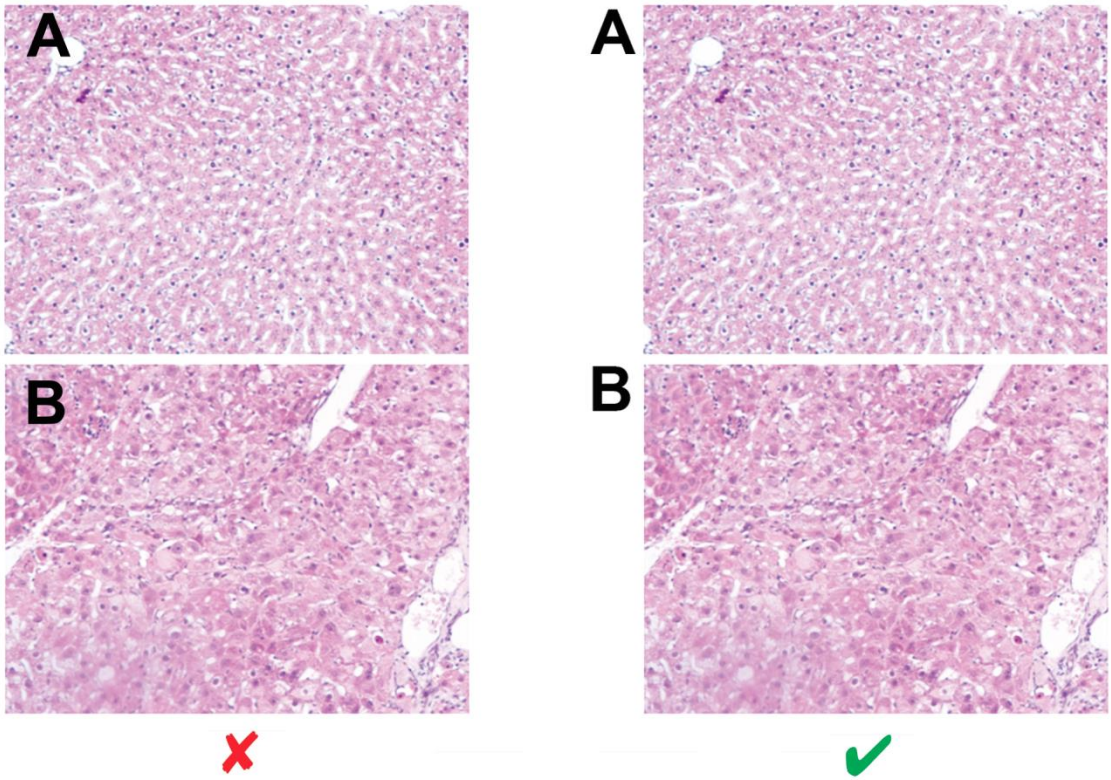

seperation by NTRK1 only
Remmele score (numerical)
sporadic
radiation-associated

- Label styles and language
  - Labels must be prepared according to our in-house style, be phrased in accordance to the manuscript, and free of spelling and other language errors.
  - The first letter of each phrase, NOT each word, must be capitalized [e.g., 'Overall survival (months)' not 'Overall Survival (Months)' and not 'overall survival (months)]'.
  - Always use a leading zero (0) before decimal points: 0.5 NOT .5.
  - Decimal points must use a full stop/period (.) NOT a comma (,).
  - A space must be inserted before measurement units: 132 bp NOT 132bp, 5 mm NOT 5mm, 1 h NOT 1h.
  - Measurements must be written as:
    - second(s): sec
    - minute(s): min
    - hour(s): h
    - day(s): day(s)
    - week(s): week(s)
    - month(s): month(s)
    - micro: μ, µ (available in Times and Helvetica) NOT u
    - liter(s): l NOT L
    - kilo Dalton: kDa NOT kD, Da, bp, kb
    - 5 units BUT 5 U/ml
      - Greek letters must be inserted using the correct Greek symbol (using Times, Helvetica or Symbol font), NOT written in full, i.e., alpha: α; beta: β, ß, (available in Times and Helvetica); and gamma: γ, etc.





- Figures may be divided into separate sections. Each section may be saved as a separate file (clearly indicated in file name) or included together in one file (with parts clearly labeled).
- Separate parts of a figure should be labeled using just A, B, C, NOT 1A, 1B, 1C.
- Figure sections may be divided and subdivided as follows:
  - A, B, C
  - A a,b,c; B a,b,c; C a,b,c
  - A a-1, a-2, b-1, b-2; B, a-1, a-2, b-1, b-2
- The number of the figure must NOT be included in the image, especially if placed on the overlapping part of the image. Instead, the file itself should be named using the figure number.
- A, B, Cs must be placed to the top left of each section of the figure, NOT overlapping the image.



3.3.8 Figure appearance

- Figure backgrounds must be white. Grey backgrounds (or backgrounds of any other color) are NOT acceptable.
- White space surrounding figures should be cropped so that the image is as close to the edges of the page as possible.
- Figures and specific sections of figures should NOT be surrounded by borders (frames).
- Figures should NOT be stretched out of proportion (distorted) horizontally or vertically.
- Yellow must NOT be used for lines in diagrams. Any darker color may be used instead.
- Line art should be dark, and lines and labeling thick enough to be clearly visible, even at small sizes.
- Charts, graphs and diagrams should NOT use more than 5 shades of grey. Patterns are acceptable.
- In charts, graphs and diagrams, unnecessary colors should be avoided (e.g., color that does not impart any additional information and is used for slight emphasis only, or color that can be replaced by shades of grey, patterns or shapes).

3.3.9 Copyright

- If a figure or table has been published previously (even if you were the author of the manuscript), copyright permission for re-use of the figure or table will often be required.
- You must acknowledge the original source and submit written permission from the copyright holder to reproduce the material where necessary.
- As an author of your manuscript, you are responsible for obtaining permissions to use material owned by others.

3.4 Figure legends

- Figure legends should be listed one after the other, as part of the text document and separate from the figure files.
- Figure legends must begin with a brief title for the whole figure and continue with a short description of each panel or part.
- All symbols (e.g. asterisks, hashtags) used to indicate significant differences in the figures must be defined accordingly in the figure legend.
- All error bars must be defined in the figure legend.
- Legends should not contain any details of the methods.

3.5 Tables

- Each table should be submitted on a separate Word file.
- Times New Roman. Font size 12. Spacing 1.5.
- Label using Roman numerals, i.e. Table I, Table II, etc.
- Include a short title.
- All symbols and abbreviations should be defined immediately below the table.

3.6 Supplementary material

Supplementary data and other materials can now be submitted to all of our journals to support and enhance research manuscripts. The material should be directly relevant to your paper and can include information in the form of audio, video, tables and figures. Supplementary data should be submitted together with the original manuscript, as these data will undergo the peer review process as well.

Acceptable file types are: JPG, JPEG, EPS, TIFF, TIF, DOC, DOCX, ODT, ODF and for video/audio: MKV, MOV, AVI, MPG, MPEG, MP4.

Please note that supplementary materials should be referenced in the text as: 'Fig. S1', 'Table SI, Table SII e.t.c', 'Data S1' or 'Appendix S1'. Supplementary video/audio clips should be called 'Supplementary_Data1.mp4'.

3.7 Nomenclature and abbreviations

- Naming of chemicals should follow that given in Chemical Abstracts Service.
- Use standard abbreviations where possible. Use the generic name of any drug unless making claims about a specific brand or formulation.
- New abbreviations must be defined at first usage.
- When reporting sequence variants and phenotypes, please follow the recommendations of the Human Variome Project Consortium for describing sequence variants (Human Genome Variation Society) and phenotypes (Human Phenotype Ontology).

3.8. Characterization of chemical and biomolecular materials

Manuscripts submitted to *Spandidos publications* must contain adequate data to support their assignment of identity and purity for each new compound described in the manuscript. Authors should provide a statement confirming the source, identity and purity of known

compounds that are key in their study, even if they are purchased or resynthesized using published methods.

3.8.1 Chemical identity

Chemical identity for organic and organometallic compounds should be established through spectroscopic analysis. Standard peak listings for ^{1}H NMR and proton-decoupled ^{13}C NMR spectra should be provided for all new compounds. Other NMR data such as ^{31}P NMR or ^{19}F NMR should be reported when appropriate. For new materials, authors should also provide mass spectral data to support molecular weight identity. UV or IR spectral data may be reported for the identification of characteristic functional groups, when appropriate. Melting-point ranges should be provided for crystalline materials. Specific rotations may be reported for chiral compounds. For known compounds, references rather than detailed procedures should be provided, unless the authors followed a modification of the published methods.

3.8.2 Combinatorial compound libraries

Standard characterization data for a diverse panel of library components should be included in manuscripts describing the preparation of combinatorial libraries.

3.8.3 Biomolecular identity

If direct structural analysis of new biopolymeric materials (e.g. oligosaccharides, peptides, nucleic acids) by NMR spectroscopy is not possible, authors must provide evidence of identity based on sequence (when appropriate) and mass spectral characterization.

3.8.4 Biological constructs

Authors should be able to provide sequencing or functional data that validates the identity of their biological constructs (plasmids, fusion proteins, site-directed mutants) upon request.

3.8.5 Sample purity

Evidence of sample purity must be shown for each new compound. For organic and organometallic compounds, purity may be demonstrated by high-field ^{1}H NMR or ^{13}C NMR, while elemental analysis is encouraged for small molecules. Quantitative analytical methods, including chromatographic (e.g. GC, HPLC) or electrophoretic analyses may be used for small molecules and polymeric materials.

3.8.6 Spectral data

Detailed spectral data for new compounds should be provided in the Materials and methods section. Figures containing spectra must be made available to the Editor upon request. The authors should explain how specific, unambiguous NMR assignments were made in the Materials and methods section.

3.8.7 Crystallographic data for small molecules

Authors reporting new structures of small molecules from crystallographic analysis must be able to provide a standard crystallographic information file (.cif); structure factors for each structure; and a structural figure with probability ellipsoids upon request. The structure factors and structural output should be checked using International Union of Crystallography checkCIF. Crystallographic data for small molecules should be submitted

to the Cambridge Structural Database, and the deposition number must be referenced in the manuscript.

3.8.8 Macromolecular structural data

Manuscripts reporting new structures should contain a table summarizing structural and refinement statistics, and the different programs used in the analysis should be mentioned and referenced. To assess the quality of the structural data, a stereo image of a portion of the electron density map (for crystallography papers); of the superimposed lowest energy structures (>10; for NMR papers); or of the entire structure (as a backbone trace) if the reported structure represents a novel overall fold should be provided upon request. For cryo-EM structures, a representative micrograph showing individual particles should be provided at submission. Protein structures should be deposited in the Protein Data Bank PDB (https://www.rcsb.org/) and the deposition number must be referenced in the manuscript.

Table I. Data collection and refinement statistics.

	Crystal name
Data collection	
Space group	
Cell dimensions	
a, *b*, *c* (Å)	
α, β, γ (˚)	
Resolution (Å)	##()[a]
R_{sym} or R_{merge}	##()
$I/\sigma I$	##()
Completeness (%)	##()
Redundancy	##()
Refinement	
Resolution (Å)	
No. reflections	
R_{work}/R_{free}	
No. atoms	
Protein	
Ligand/ion	
Water	
B-factors	
Protein	
Ligand/ion	
Water	
R.m.s. deviations	
Bond lengths (Å)	
Bond angles (˚)	

[a]Values in parentheses are for the highest-resolution shell. The number of crystals for each structure should be indicated in the table legend.

Ramachandran statistics should be included in the Materials and methods section, at the end of Refinement subsection. The wavelength of data collection, temperature and beamline should be specified in the Materials and methods sections.

3.8.9 Chemical structures

Structures of compounds should be prepared using a drawing program, such as ChemDraw. Please ensure to use the following settings (ACS Style sheet in ChemDraw):

- Chain angle: 120°
- Bond spacing: 18% of width
- Fixed length: 14.4 pt (0.508 cm, 0.2 in.)
- Bold width: 2.0 pt (0.071 cm, 0.0278 in.)
- Line width: 0.6 pt (0.021 cm, 0.0084 in.)
- Margin width: 1.6 pt (0.056 cm, 0.0222 in.)
- Hash spacing: 2.5 pt (0.088 cm, 0.0347 in.)
- Font: Arial/Helvetica
- Size: 10 pt

3.9 References

- Spandidos Publications has updated its services by incorporating Edifix, a new program that automatically links and corrects bibliographic references.
- Based on our particular style, the first 9 authors will be listed as they appear. When more than 10 authors are listed, Edifix will automatically include only the first 10 authors, followed by *et al*.
- Cite journal titles using NLM Title Abbreviation found online in the U.S. National Library of Medicine Catalog.
- References must be numbered consecutively in the order mentioned.
- Do NOT use full stop after initials or abbreviations.
- In the text, cite references by number in parentheses e.g., (1-3) (1,2).
- Inclusive page numbers should be given.
- The following are examples of order and style, which should be strictly adhered to:
    - Spandidos DA and Wilkie NM: Malignant transformation of early passage rodent cells by a single mutated human oncogene. Nature 310: 469-475, 1984.
    - Spandidos A, Wang X, Wang H and Seed B: PrimerBank: A resource of human and mouse PCR primer pairs for gene expression detection and quantification. Nucleic Acids Res 38: D792-D799, 2010.
    - Barbu CG, Arsene AL, Florea S, Albu A, Sirbu A, Martin S, Nicolae AC, Burcea-Dragomiroiu GTA, Popa DE, Velescu BS, *et al*: Cardiovascular risk assessment in osteoporotic patients using osteoprotegerin as a reliable predictive biochemical marker. Mol Med Rep 16: 6059-6067, 2017.
    - Hall A, Morris JDH, Price B, Lloyd A, Hancock JF, Gardener S, Houslay MD, Wakelam MJO and Marshall CJ: The function of the mammalian Ras proteins. In: Ras oncogenes. Spandidos DA (ed.) Plenum Publ. Corp., New York, pp99-104, 1989.
- Personal communications should be avoided.
- Manuscripts in preparation or submitted (but not yet accepted) and abstracts, may be cited in the text, but should NOT be included in the list of references.
- Non-English references should not be included in the Reference list. The entire manuscript cited must be in English.
- Output Style files are available to download in Endnote format from Spandidos Publications at http://www.spandidos-publications.com/pages/outputStyles or from Thomson Reuters and also

in CSL 1.0.1 format from https://zotero.org/styles to work with software such as Mendeley, Papers and Zotero.

3.10 Declarations

All manuscripts must contain the following sections at the end of the manuscript, prior to the Reference list:

- Acknowledgements
- Funding
- Availability of data and materials
- Authors' contributions
- Ethics approval and consent to participate
- Patient consent for publication
- Competing interests
- Authors' information (optional)

If any of the sections are not relevant to your manuscript, please include the heading and write 'Not applicable' for that section.

Acknowledgements

All contributors who do not meet the criteria for authorship should be listed in an 'Acknowledgements' section. Examples of those who might be acknowledged include a person who provided purely technical help or writing assistance, or a department chair who provided only general support. The involvement of scientific (medical) writers or anyone else who assisted with the preparation of the manuscript content should be explicitly acknowledged, along with their source of funding, in the 'Acknowledgements' or 'Authors' contributions' section as appropriate. Authors should obtain permission to acknowledge from all those mentioned in the Acknowledgements section. If you do not have anyone to acknowledge, please write "Not applicable" in this section. Please note that company names and generic statements should not be mentioned in this section.

Funding

All sources of funding for the research reported should be declared. The role of the funding body in the design of the study, collection, analysis and interpretation of data, and manuscript writing should be declared. If a funder name is not applicable for your manuscript, please state so in your manuscript (e.g., "Funding information is not applicable"/"No funding was received").

Availability of data and materials

Submission of a manuscript to Spandidos Publications implies that the materials included in the manuscript, including all relevant raw data, may be made freely available to any researchers who wish to use them for non-commercial purposes, while preserving any necessary confidentiality and anonymity.

For our journals, we strongly encourage that all datasets on which the conclusions of the paper depend should be available to readers, unless they are already provided as part of the submitted

article. Where datasets are included in the study, authors should include an "Availability of data and materials" section in the article, stating where the data supporting their findings may be found. Authors who do not wish to share their data must state this fact, and also provide an explanation as to why the data are unavailable.

Availability of data and materials statements can take one of the following forms (or a combination of more than one if required for multiple datasets):

- The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.
- The datasets generated and/or analyzed during the current study are available in the [NAME] repository, [PERSISTENT WEB LINK TO DATASETS].
- All data generated or analyzed during this study are included in this published article.
- The datasets generated and/or analyzed during the current study are not publicly available due to [REASON WHY DATA ARE NOT PUBLIC] but are available from the corresponding author on reasonable request.
- Data sharing is not applicable to this article, as no datasets were generated or analyzed during the current study.
- The data that support the findings of this study are available from [THIRD PARTY NAME] but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of [THIRD PARTY NAME].
- Not applicable. If your manuscript does not contain any data, please state 'Not applicable' in this section.

Authors are required to deposit their sequencing, X-ray crystallography and microarray datasets in public repositories, unless there is a compelling reason for them not to do so (such as protection of patient privacy, pending or approved patents, biosecurity reasons or any other legislation prohibiting public data sharing). Authors who are unable to share their data must state this fact and also provide an explanation in the manuscript as to why the data are unavailable. Publicly available datasets must be described in the Availability of data and materials section of the manuscript, which must include the accession number and the name and URL of the repository in which the datasets are available. We leave the selection of the repository entirely at the authors' discretion, although note that the data must be freely available to readers. A list of recommended repositories is featured below; consulting the Registry of Research Data Repositories (http://www.re3data.org/) may also be useful in this regard.

List of suggested repositories

Type of data	**Repositories**
Research output in general (including supplementary figures and data, and multimedia adjuncts)	Figshare SimTK
Protein sequences	UniProt
Macromolecular structure	Worldwide Protein Data Bank (wwPDB) Biological Magnetic Resonance Data Bank (BMRB)
Crystallographic data for small molecules	The Cambridge Structural Database

DNA and RNA sequences	DNA DataBank of Japan (DDBJ) European Nucleotide Archive (ENA) Gene Expression Omnibus (GEO) Trace Archive
Sanger sequencing	GenBank
Deep sequencing	NCBI Sequence Read Archive
Genetic polymorphisms	dbSNP dbVAR European Variation Archive (EVA)
Linked genotype and phenotype data	dbGAP European Genome-phenome Archive (EGA)
Microarray data (must be compliant with Minimum Information About a Microarray Experiment, or MIAME)	Gene Expression Omnibus (GEO) ArrayExpress
Flow cytometry	FlowRepository

Authors' contributions

The individual contributions of authors to the manuscript should be specified in this section. Please use initials to refer to each author's contribution in this section, for example: "FC analyzed and interpreted the patient data regarding the hematological disease and the transplant. RH performed the histological examination of the kidney, and was a major contributor in writing the manuscript. FC and RH confirm the authenticity of all the raw data. All authors read and approved the final manuscript."

Please note that an 'author' is generally considered to be someone who has made substantive intellectual contributions to a published study. According to the ICMJE guidelines, to qualify as an author one should have i) made substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data; ii) been involved in drafting the manuscript or revising it critically for important intellectual content; iii) given final approval of the version to be published. Each author should have participated sufficiently in the work to take public responsibility for appropriate portions of the content, and iv) agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. Acquisition of funding, the collection of data, or general supervision of the research group alone, does not usually justify authorship. Please note that changes to authorship cannot be made after acceptance of a manuscript.

If any changes to the list of authors of a manuscript are necessary after the initial submission but before publication, the corresponding author must contact the journal staff and provide a clear reason for the change. If the change to the authorship list is appropriate and in keeping with the guidelines given above, the corresponding author will be asked to provide written confirmation that

all other authors listed on the manuscript at that time consent. Spandidos Publications will individually inform anyone who is added or removed from the author list. The authenticity of all the raw data must have been assessed by at least two authors to ensure its legitimacy. The authors who are responsible for confirming the authenticity must be mentioned in the Authors' contributions section of the declarations.

Ethics approval and consent to participate

Manuscripts reporting studies involving human participants, identifiable human data, human primary cells or human tissue must include a statement on ethics approval and consent (even where the need for approval was waived), as well as the name of the ethics committee that approved the study and the committee's reference number if appropriate. Similarly, studies involving animals must include a statement on ethics approval. In addition, studies involving preimplantation stages of human development, human embryos, or embryo-derived cells or production of human gametes *in vitro* when such gametes are tested by fertilization or used for the creation of embryos should be subject to review, approval and ongoing monitoring by a specialized human embryo research oversight (EMRO) process capable of evaluating the unique aspects of the science, according to the guidelines of the International Society for Stem Cell Research. Of note, research studies belonging to category 3 of these guidelines (including *in vitro* culture or gestation ex utero/non-human uterus of any intact human preimplantation embryo or organized embryo-like cellular structure with human organismal potential beyond 14 days or formation of the primitive streak; implantation of human embryos produced by reprogramming of nuclei from somatic cells; implantation or gestation of human embryos that have undergone modification of their nuclear genome; and experiments involving animal chimeras incorporating human cells with the potential to form human gametes are bred to each other) are prohibited.

If a study was granted an exemption from requiring ethics approval, this should also be detailed in the manuscript (including the name of the ethics committee that granted the exemption). Further information and documentation to support this should be made available to the Editor on request. Please note that manuscripts may be rejected if the Editor considers that the research has not been carried out within an appropriate ethical framework. Furthermore, please bear in mind that, if a study was not granted ethics committee approval prior to commencing, retrospective ethics approval usually cannot be obtained and it may not be possible to consider the manuscript for peer review. The decision on whether to proceed to peer review in such cases is at the Editor's discretion.

In the case of clinical trials, which are defined as "any research study that prospectively assigns human participants or groups of humans to one or more health-related interventions to evaluate the effects on health outcomes", the trial registration number (TRN) and date of registration should be included as the last line of the manuscript abstract. Suitable publicly available registries are those listed on the ICMJE website and the WHO International Clinical Trials Registry Platform. For clinical trials that have not been registered prospectively, Spandidos Publications encourages retrospective registration to ensure the complete publication of all results. In such cases, the TRN, date of registration and the words 'retrospectively registered' should be included as the last line of the manuscript abstract.

Research that is performed on humans should follow international and national regulations in accordance with the Declaration of Helsinki, or any other relevant set of ethical principles. For all research involving human subjects or tissues, the manuscript should include a statement declaring

that informed consent for participation in the study or use of their tissue was obtained from all participants (or their parent or legal guardian in the case of children under 18, or patients otherwise considered minors under local legislation). Consent is also required for the procurement of biomaterials for stem cell research and translation, including gamete donors in IVF studies.

With regard to the use of experimental animals, any research performed must follow internationally recognized guidelines on animal welfare, as well as local and national regulations, in accordance with the U.K. Animals (Scientific Procedures) Act and associated guidelines, the EU Directive 2010/63/EU for animal experiments, or the National Institutes of Health guide for the care and use of laboratory animals. All animal studies should also comply with the ARRIVE guidelines and the AVMA euthanasia guidelines 2020. A statement must be included in the manuscript, identifying the institutional and/or licensing committee that approved the experiments undertaken, including the reference number where appropriate. The ethics committee from which approval was obtained must be transparent in its functioning, independent of the researcher, the sponsor and any other unwarranted influence and duly qualified. If a study was granted an exemption from requiring ethics approval, this should also be detailed in the manuscript (including the name of the ethics committee that granted the exemption and the reasons for the exemption). The Editor will take into account animal welfare issues and reserves the right to reject a manuscript, especially if the research involves protocols that are inconsistent with commonly accepted norms of animal research. When rodents are used as in vivo cancer models, the tumor burden should not exceed the recommendations of the University of Pennsylvania Institutional Animal Care and Use Committee guidelines.

Patient consent for publication

Patients have a right to privacy that should not be violated without informed consent. Identifying information, including names, initials, date of birth or hospital numbers, images or statements should not be included in the manuscript unless the information is essential for scientific purposes and the patient (or parent or guardian) has provided written informed consent for publication. A statement must be included in the manuscript declaring that the patient or parent, guardian or next of kin provided written informed consent for the publication of any associated data and accompanying images. The consent form must be made available to the Editor if requested and will be treated confidentially.

Furthermore, in case reports or other studies in which case details, personal information or images are included that may enable an individual to be identified, the individual or a parent, guardian or next of kin must consent to its publication, and this consent should be declared in the manuscript. Authors should disclose to patients that personally identifiable material would be available via the Internet as well as in print after publication.

Publication without written consent may be considered if all identifying information is removed, public interest considerations outweigh the potential harm, it is impossible to obtain permission and a reasonable individual would be unlikely to object to publication. The final decision on whether consent to publish is required lies with the Editor.

Publication of clinical datasets

For datasets containing clinical data, authors have an ethical and legal responsibility to respect participants' rights to privacy and to protect their identity. Ideally, authors should gain informed consent for publication of the dataset from participants at the point of recruitment to the trial. If this is not possible, authors must demonstrate that publication of such data does not compromise anonymity or confidentiality or breach local data protection laws, for the dataset to be considered for publication. Authors must consider whether the dataset contains any direct or indirect identifiers and consult their local ethics committee or other appropriate body before submission if there is any possibility that participants will not be fully anonymous. Authors must state in their manuscript on submission whether informed consent was obtained for publication of patient data. If informed consent was not obtained, authors must state the reason for this, and which body was consulted in the preparation of the dataset.

Competing interests

Authors, reviewers and editors must declare whether there are any competing interests with regard to the publication of a study. A competing interest exists when the authors' interpretation of data or presentation of information may be influenced by, or may be perceived to be influenced by, their personal or financial relationship with other people or organizations, such as reimbursement for salaries, equipment or supplies, or a personal belief that may influence their objectivity and motivation, and consequently affect the data interpretation. This can include competing patents, grants, funding, employment, personal relationships and strong ethical beliefs, among other factors. Such conflicts must be declared, as they may affect the integrity or reliability of the science in the study, as well as that of otherwise unassociated studies in the same journal. Competing interest statements for public funding sources, including government agencies, charitable or academic institutions, need not be included. For example, if a charitable foundation sponsored the study and a pharmaceutical company provided the drugs, only the pharmaceutical company should be mentioned.

Full disclosure of the competing interests is to be made in the cover letter and manuscript at the time of submission, even if the author judges that it has not influenced the work. If no conflict exists, this must also be stated clearly in the manuscript as follows: 'Competing interests. The authors declare that they have no competing interests', and all authors should confirm its accuracy. If there is a conflict, please include it in a 'Competing interests' section. Examples of conflict of interest statements include 'The present study was supported by Jones Women's University, grant no. 12345', 'XY University provided a graduate scholarship to Dr Jones', 'The compound xyz was kindly provided by ABC Company, city, country'. Authors may be asked to confirm or update, or provide further details regarding such disclosure statements following acceptance of the manuscript. Further details regarding requirements for conflict of interest statements are provided in http://www.icmje.org/conflicts-of-interest/.

External peer reviewers must disclose any conflicts of interest that could bias their opinions of the manuscript, and they should disqualify themselves from reviewing specific manuscripts if they believe it appropriate. Should any such competing interest be declared, the journal editor will judge whether the reviewer's comments should be recognized or will interpret the reviewer's comments in the context of any such declaration.

Authors' information (optional)

This section is optional. You may choose to use this section to include any relevant information about the author(s) that may aid the reader's interpretation of the article, and understand the standpoint of the author(s). This may include details about the authors' qualifications, current positions they hold at institutions or societies, or any other relevant background information. Please refer to authors using their initials. Note this section should not include authors' email addresses or competing interests.

3.11 Standards of reporting

Spandidos Publications recommend that authors refer to the minimum reporting guidelines for health research hosted by the EQUATOR Network when preparing their manuscript, and FAIRsharing.org for reporting checklists for biological and biomedical research, where applicable. Authors should adhere to these guidelines when drafting their manuscript.

Checklists are available for a number of study designs, including:

- Randomized controlled trials (CONSORT) and protocols (SPIRIT)
- Systematic reviews and meta-analyses (PRISMA) and protocols (PRISMA-P)
- Observational studies (STROBE)
- Case reports (CARE)
- Qualitative research (COREQ)
- Diagnostic/prognostic studies (STARD and TRIPOD)
- Economic evaluations (CHEERS)
- Pre-clinical animal studies (ARRIVE)
- Describing sequence variants (Human Genome Variation Society)
- Describing phenotypes (Human Phenotype Ontology)

Spandidos Publications supports the prospective registration of systematic reviews and encourages authors to register their systematic reviews in a suitable registry (such as PROSPERO), including the registration number as the last line of the manuscript abstract. Authors of systematic reviews should also provide a file describing all details of the search strategy. For an example of how a search strategy should be presented, see the Cochrane Reviewers' Handbook.

In addition, authors are encouraged to follow the Sex and Gender Equity in Research (SAGER) guidelines (https://doi.org/10.1186/s41073-016-0007-6) in their studies. The general principles are as follows:

- Authors should use the terms sex (biological) and gender (social or cultural) carefully in order to avoid confusing both terms.

- Where the subjects of research comprise organisms capable of differentiation by sex, the research should be designed and conducted in a way that can reveal sex-related differences in the results, even if these were not initially expected.

- Where subjects can also be differentiated by gender (shaped by social and cultural circumstances), the research should be conducted similarly at this additional level of distinction.

Both positive and negative results of sex and/or gender-based analyses should be included. If only one sex is included in the study, or if the results of the study are to be applied to only one sex or gender, the title and the abstract should specify the sex of animals or any cells, tissues and other material derived from these and the sex and gender of human participants; implications of the lack of sex- and/or gender-based analysis on the interpretation of the results should be addressed in the Discussion.

4. Proofs

Authors must check their manuscript carefully before submission, after responding to peer-review, and when answering queries raised at the proof stage. Any errors will be faithfully transferred into the final PDF and print versions.

5. Reprints

A reprint order form will accompany notification of acceptance of a manuscript and should be completed and returned immediately. For further information, please email contact@spandidos-publications.com.

6. Article charges

After final acceptance of manuscripts, authors can choose between one of the two following options to cover for publications costs.

Option 1: The publication charge is calculated based on the number of pages and color figures (per page) of the final accepted manuscript. All Spandidos Publications journals have the same fees for option 1.

	EUR	GBP	USD
Basic charge (total amount for up to 5 pages)	€250	£230	$325
Each additional page	€110	£100	$160
Reproduction in color, per page	€390	£340	$520

Color figures should only be submitted if intended for publication. Fees for color reproduction are calculated based on the number of pages with color at the time of manuscript acceptance. Figures may not be changed to black and white once a manuscript has been accepted.

Articles become freely available 12 months after their publication online.

Authors requiring immediate open access of articles published in our other journals must use Option 2. Option 2: A fixed charge is applied irrespective of the total number of pages or color figures. In addition, the articles are made immediately available as open access and appear online at PubMed, PubMed Central and Europe PMC. This option complies with a number of funding bodies (see Open Access section).

	EUR	GBP	USD
International Journal of Oncology	€2,500	-	$2,600
Oncology Reports	€2,300	-	$2,400
International Journal of Molecular Medicine	€1,390	£1,320	$1,860
Molecular Medicine Reports	€1,090	£1,030	$1,450
Oncology Letters	€890	£840	$1,190
Experimental and Therapeutic Medicine	€890	£840	$1,190
Molecular and Clinical Oncology	€890	£840	$1,190
Biomedical Reports	€890	£840	$1,190

If your institute has a Gold Membership subscription with Spandidos Publications, then your article may be published free of charge. Please check with your institute librarian or subscription manager.

7. Open Access

- "Open access" articles published online can be freely accessed at the journal's website by all users. The articles are made immediately available as open access and appear online at PubMed, PubMed Central and Europe PMC.
- Articles may be licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, a Creative Commons Attribution-NonCommercial 4.0 International License or a Creative Commons Attribution 4.0 International License.
- Spandidos Publications' publishing policies enable authors to comply fully with the public access requirements of the major worldwide funding agencies (visit http://www.sherpa.ac.uk for more information). Authors must take the necessary actions to obtain this compliance, including self-archiving, utilization of Spandidos Publications manuscript deposition service and selection of open access publication under the correct license.
- Spandidos Publications makes articles freely available 12 months after publication.
- Authors wishing to purchase immediate open access for International Journal of Oncology, Oncology Reports, International Journal of Molecular Medicine, Molecular Medicine Reports, Oncology Letters, Experimental and Therapeutic Medicine, Molecular and Clinical Oncology and Biomedical Reports must select option 2.
- Spandidos Publications will refund an article processing charge (APC) if an error on our part has resulted in a failure to publish an article under the open access terms selected by the authors. This may include a failure to make an article openly available on the journal platform, or publication of an article under a different Creative Commons license from that selected by the authors. If you become aware of an error in the open access status or licensing of your article, please contact the appropriate journal immediately. Note that APCs will not be refunded in instances where articles are retracted or removed as a consequence of author error or misconduct. Neither we will be liable if delays in the processing of the article result from changes made to the authorship of the article, or in cases where delays result from editorial decisions.

8. Compliance with funding agencies

All manuscripts that are agency-funded (e.g. NIH, HHMI, Cancer Research UK, Wellcome Trust, the Chinese Academy of Sciences (CAS), the National Natural Science Foundation of China (NSFC), etc.) and paid for under the Open Access Option will be deposited automatically by the publisher and become publicly available in PubMed Central and Europe PMC. Alternatively, agency-funded articles without

Open Access purchase will be deposited and made publicly available in PubMed Central and Europe PMC 6-12 months following publication, depending on the public access policy of the agency.

9. Funding body policies and agreements

- Spandidos Publications has obtained agreements and constructed policies to enable authors to comply with manuscript archiving requirements of the following funding bodies specified in conditions of researcher grant awards:

    - Arthritis Research Campaign (UK)
    - Austrian Science Fund (FWF, Austria)
    - Biotechnology and Biological Sciences Research Council (BBSRC, UK)
    - British Heart Foundation (UK)
    - Cancer Research (UK)
    - Chief Scientist Office (UK)
    - Chinese Academy of Sciences (CAS, China)
    - Department of Health (UK)
    - Dunhill Medical Trust (UK)
    - Economic and Social Research Council (ESRC, UK)
    - Howard Hughes Medical Institute (US)
    - Medical Research Council (UK)
    - National Institutes of Health (US)
    - National Natural Science Foundation of China (NSFC, China)
    - Research Councils (UK)
    - Telethon (Italy)
    - Wellcome Trust (UK)

- This enables authors to follow their funding body's policies on archiving without breaching their publishing agreements with Spandidos Publications.

10. Author self-archiving

Authors are encouraged to submit the final publisher’s version PDF of their manuscript to their institution’s repository 6 months following publication, as well as to their funding body’s archive. A link to the published version on the Spandidos Publications website must be included with full citation details and acknowledgement of the journal as the original source. Authors that have purchased open access can add the final publisher's PDF to their institutional repository and funding body's archive immediately. In addition, authors are allowed to archive their articles pre-print (i.e., pre-refereeing). Pre-print can be on author's personal website or pre-print server.

11. English language editing, manuscript formatting and figure preparation

If English is not your first language, you may decide to have your manuscript proofread or edited by an English speaker, prior to submission. A colleague or a professional service may do this. Clear and concise language enables editors and reviewers to concentrate on the scientific content of your manuscript. This facilitates the peer review process and ensures that your article will not be rejected based on English language alone. Spandidos Publications offers a number of Manuscript Services including an English Language Editing service, through which the use of English in your manuscript

may be checked and refined. The editors of this service will focus on the spelling, grammar, clarity and style of your manuscript.

For more information on this and other services, please access http://www.spandidos-publications.com/languageediting.

Please be aware that an author's use of this or any other service in no way guarantees that his or her submission will be accepted by a Spandidos journal or any other. Any arrangement an author enters into will be exclusively between the author and their chosen company. Using an editing service is neither a requirement for nor a guarantee of acceptance for publication. For further details, please refer to our disclaimer on our affiliated website.