

Universidade de Ribeirão Preto
Doutorado em Biotecnologia

WEDSON GOMES DA SILVEIRA JUNIOR

ANÁLISE DE MICRORNAS ASSOCIADOS À ILHAS DE CPG E SEUS
POTENCIAIS GENES ALVOS NO GENOMA CANINO E HUMANO

RIBEIRÃO PRETO - SP
2017

Wedson Gomes da Silveira Junior

ANÁLISE DE MICRORNAS ASSOCIADOS A ILHAS DE CPG E SEUS
POTENCIAIS GENES ALVOS NO GENOMA CANINO E HUMANO

Tese apresentada à Universidade de Ribeirão Preto, UNAERP, como requisito parcial para obtenção do título de Doutor em Biotecnologia.

Orientador: Prof. Dr. Mozart de Azevedo Marins

RIBEIRÃO PRETO

2017

Ficha catalográfica preparada pelo Centro de Processamento Técnico
da Biblioteca Central da UNAERP

- Universidade de Ribeirão Preto -

Silveira Junior, Wedson Gomes da, 1985-

S587a Análise de micrnas associados à ilhas de CPG e seus potenciais
genes alvos no genoma canino e humano / Wedson Gomes da Silveira
Junior. - - Ribeirão Preto, 2017.

82 f.: il. color.

Orientador: Prof. Dr. Mozart de Azevedo Marins.

Tese (doutorado) - Universidade de Ribeirão Preto, UNAERP,
Biotecnologia. Ribeirão Preto. 2017.

WEDSON GOMES DA SILVEIRA JUNIOR

ANÁLISE DE MICRO RNAs ASSOCIADOS À ILHAS DE CPG E SEUS
POTENCIAIS GENES ALVOS NO GENOMA CANINO E HUMANO

Tese de Doutorado apresentada ao Programa de Pós-
Graduação em Biotecnologia da Universidade de
Ribeirão Preto, para obtenção do título de Doutor em
Biotecnologia.

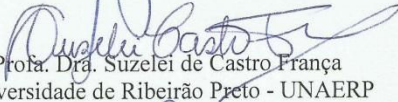
Área de Concentração: Biotecnologia Aplicada à Saúde

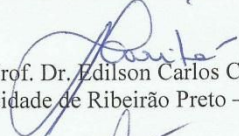
Data da defesa: 11 de dezembro de 2017


Resultado: Aprovado

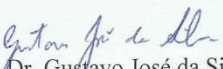
BANCA EXAMINADORA


Prof. Dr. Mozart de Azevedo Marins
Universidade de Ribeirão Preto - UNAERP


Profa. Dra. Suzelêi de Castro França
Universidade de Ribeirão Preto - UNAERP


Prof. Dr. Edilson Carlos Caritá
Universidade de Ribeirão Preto - UNAERP


Prof. Dr. Vinícius Alves Silva
Instituto Federal do Sul de Minas - IFSuldeminas


Prof. Dr. Gustavo José da Silva
Instituto Federal do Sul de Minas - IFSuldeminas

RIBEIRÃO PRETO
2017

Dedico este trabalho aos meus pais Wedson e Irene pelo amor, motivação e apoio que sempre me deram no curso de toda a minha vida.

Obrigado, sem vocês isto não seria possível

AGRADECIMENTOS

Ao meu orientador, Prof. Dr. Mozart de Azevedo Marins, pelas diretrizes de pesquisa, pela amizade e paciência durante a realização deste trabalho.

À minha família que sempre me incentivou e me apoiou.

Ao Centro Federal de Educação Tecnologia de Minas Gerais – CEFET-MG, pelo incentivo a mim concedido.

Aos professores, funcionários e colegas de pós-graduação (mestrado e doutorado) da Universidade de Ribeirão Preto - UNAERP.

À UNAERP.

“O começo de todas as ciências é o espanto
de as coisas serem o que são.”

(Aristóteles)

RESUMO

Os agrupamentos de dinucleotídeos CpG, denominados ilhas de CpG (ICG), estão presentes em aproximadamente 70% dos genes de mamíferos sobrepondo-se com as regiões de início de transcrição, caracterizando-se inclusive como elementos definidores de promotores. As alterações do perfil de metilação do DNA nestas regiões são um importante elemento de regulação da expressão gênica e estão associados ao desenvolvimento de diversas doenças, inclusive o câncer. Os microRNAs (miRNAs) são um tipo de RNA não codificante que também exercem papel importante na regulação da expressão gênica. Por sua vez, a regulação da expressão de miRNAs pode também envolver a metilação do DNA em regiões de ICGs. Existem atualmente 1883 miRNAs catalogados em humanos e 502 miRNAs em cães, sendo que para muitos a associação com processos celulares normais ou patológicos ainda não é totalmente conhecida. O objetivo deste estudo foi realizar uma análise comparativa de miRNAs humanos e caninos associados à ICGs, com intuito de contribuir para o conhecimento dos elementos de regulação de sua expressão. Foram adquiridas sequências do genoma humano e canino de bancos de dados. Estas sequências foram submetidas a análise *in silico* com várias ferramentas computacionais, inclusive uma ferramenta desenvolvida. Foram identificados respectivamente 178 e 67 miRNAs humanos e caninos associados à ICGs. Um total de 113 microRNAs apresentam similaridade entre as sequências para as duas espécies e os seus potenciais genes alvos foram também identificados. Destaca-se também a identificação de 5 potenciais novos miRNAs caninos não catalogados. Conclui-se que as análises realizadas indicam a possibilidade de detecção por similaridade de miRNAs caninos que, em conjunto com dados já descritos para seres humanos, podem ser explorados como alvos para o desenvolvimento de novas terapias de tratamento de doenças caninas, incluindo o câncer.

Palavras-chave: MicroRNA. Ilhas de CpG. Câncer.

ABSTRACT

CpG Dinucleotides groupings, called CpG Islands (CGI), are present in approximately 70% of mammalian genes, overlapping the beginning of transcription regions, and can be considered as defining elements of promoters. The changes in the DNA methylation profile in these regions are an important element of gene expression regulation and are associated with the development of various diseases, including cancer. MicroRNAs (miRNAs) are a type of non-coding RNA, which also perform an important role in regulating gene expression. Regulation of miRNAs expression, on the other hand, may also involve DNA methylation in CGI regions. There are currently 1883 miRNAs in humans and 502 miRNAs in dogs, and for many its association with normal or pathological cellular processes are not been fully known yet. The objective of this study was to conduct a comparative analysis between human and canine miRNAs associated with CGIs, aiming to contribute to the knowledge of the regulatory elements of its expression. These sequences were been subjected to analysis *in silico* with computational tool developed in this work. 178 and 67, respectively, human and canine miRNAs associated with CGIs were been identified 113 showed similarity between the sequences for the two species and their potential target genes were also been identified. The results also highlight the identification of 5 potential new canines miRNAs so far not catalogued. The analyses conducted indicate the possibility of detection by similarity of canine miRNAs that, along with data already described for humans, can be exploited as targets for the development of new treatment therapies for canine diseases, including cancer.

Keywords: MicroRNA. CpG Island. Cancer.

SUMÁRIO

1 Introdução	14
2 Revisão Bibliográfica.....	16
2.1 Câncer	16
2.2 Metilação do DNA.....	17
2.3 MicroRNA	20
2.4 Bioinformática	23
3 Motivação	29
4 Objetivos.....	30
4.1 Objetivo geral	30
4.2 Objetivos específicos	30
5 Materiais e métodos.....	31
5.1 Linguagem de Programação	31
5.2 Aquisição dos dados	32
6 Resultados e Discussão	39
6.1 Ilhas de CpG	41
7 Conclusão	64

LISTA DE ILUSTRAÇÕES

Figura 1 - Incidência de tipos específicos de câncer em cães.	16
Figura 2 - Mecanismo de metilação do DNA.....	18
Figura 3A- Dogma central da biologia B - novo fluxo com as novas descobertas.	22
Figura 4 - Ciclo do microRNA.....	23
Figura 5 - Pipeline de targets em organismos vertebrados.....	28
Figura 6 - Script para aquisição dos genomas.....	32
Figura 7 - microRNAs Canis familiaris miRBase.....	33
Figura 8 - Arquivo de miRNAs com sequências.....	33
Figura 9 - Diagrama de Caso de Uso.....	34
Figura 10 - Interface do Software.....	35
Figura 11- Funções em Python para converter nucleotídeos.....	36
Figura 12 - Algoritmo para detecção de Ilhas de CpG.....	37
Figura 13 - Padrão de desenvolvimento do software.....	38
Figura 14 - Quantidade de microRNAs por cromossomo canino.	39
Figura 15 - Quantidade de miRNAs por cromossomo Humano.	40
Figura 16 - Quantidade de regiões de ICG por cromossomo em Humanos.....	42
Figura 17 - Quantidade de regiões de ICG por cromossomo em Cães.....	43
Figura 18- Modelo Entidade-Relacionamento(MER) do banco de dados mySQL.....	49
Figura 19 - Fluxograma dos dados após análises.....	54
Figura 20 - Quantidade de miRNAs que interagem com genes alvos.....	60

LISTA DE TABELAS E QUADROS

Tabela 1 - Casos de novos câncer no Brasil	17
Tabela 2- Exemplos de ncRNAs	21
Tabela 3 - Comparação das funções dos softwares de ICG	26
Tabela 4 - Comparação entre os softwares de ICG	41
Tabela 5 - Regiões de miRNAs com ICG próximas em cães.....	43
Tabela 6 - Regiões de miRNAs com ICG próximas em humanos	45
Tabela 7 - Domínios de Sequências miRNAs Humano geradas pelo Homer	50
Tabela 8 - Domínios de Sequências miRNAs Canina geradas pelo Homer.....	52
Tabela 9 - miRNAs associados aos tipos de câncer.	54
Tabela 10- Análise dos genes e regiões dos miRNAs com ICG.....	56
Tabela 11 - Possíveis miRNAs Caninos não catalogados	61
Tabela 12 - Resultado cluster de regiões no mirBase.....	61
Tabela 13 – Genes alvos dos miRNAs	62

LISTA DE ABREVIATURAS E SIGLAS

DNA	Ácido desoxirribonucléico
RNA	Ácido ribonucléico
DNMT	DNA metiltransferases
ICG	Ilhas de CpG
ncRNA	RNA não codificante
miRNAs	Micro RNAs
RNA_m	RNAs mensageiros
RNA_{nc}	RNAs não codificantes
NCBI	<i>National Center for the Biotechnology Information</i>
CpG	Citosina – Fosfato - Guanina
Pb	Pares de base
MVC	<i>Model View Controller</i>
LLC	Leucemia Linfocíta Crônonica
INCA	Instituto Nacional de Câncer
RISC	RNA-Induced Silencing Complex
INSDC	<i>International Nucleotide Sequence Database Collaboration</i>
XML	eXtensible Markup Language
DNMTs	DNA Metiltransferases
HATs	Histonas Acetiltransferases
HDACs	Histonas Deacetilases

1 Introdução

A relação entre os seres humanos e os cães (*Canis lupus familiaris*) é de longa data e teve influência na evolução cultural e psicológica da humanidade (LOPES; SILVA, 2012, p. 177). Os cães estão presentes nas mais diversas atividades humanas, tanto como força de trabalho na zona rural, vigilância ou como animal de estimação de companhia.

Segundo o IBGE, estima-se que 44,3% dos lares brasileiros possuíam pelo menos um cão em 2013. A população de cães foi estimada em 52,2 milhões, o que representa uma média de 1,8 cães por domicílio (IBGE, 2015). Esse número de animais cresce cerca de 5% ao ano, segundo estimativas da Associação Brasileira da Indústria de Produtos para Animais de Estimação (ABRINPET, 2012).

Com o crescimento do número de cães, demandam-se cada vez mais estudos sobre diversos tipos de patologias, como por exemplo, o câncer. Dentre as patologias, o câncer é classificado como a principal causa de morte em cães nos países desenvolvidos, e a segunda maior causa no Brasil (HORTAL; COSTA, 2012).

O comportamento de tumores em humanos e cães possuem várias características em comum. Por este motivo, nos estudos em oncologia comparativa, os camundongos estão sendo substituídos pelos cães domésticos devido a sua similaridade genética, principalmente, nos genes associados ao câncer (PAOLONI; KHANNA, 2008). Outro motivo que justifica esse tipo de estudo comparativo é que os humanos e os cães estão expostos aos mesmos fatores externos.

Os tipos de cânceres mais comuns em humanos são o de mama e o de próstata. Em cães, o câncer de mama também possui alta incidência, porém, pela castração de cadelas ainda jovens, a incidência deste câncer é reduzida, o que indica uma associação com a produção hormonal. Já no caso de câncer de próstata, também relacionado com a produção hormonal, é pouco comum o aparecimento em cães, sendo estes castrados ou não (JANE, 2013).

As causas do câncer são variadas, podendo ser externas ou internas ao organismo, estando ambas inter-relacionadas. No nosso estudo iremos dar ênfase nas causas internas que são, na maioria das vezes, geneticamente pré-determinadas e estão ligadas à capacidade do organismo de se defender das agressões externas.

Uma das causas internas está relacionada com os micros RNAs (miRNAs), estes foram descritos em 1993 por Victor Ambros e Gary Kuvkun's em *Caenorhabditis elegans*, porém, a função dos microRNAs em cânceres humanos só foi descoberta em 2002, quando

George Callin e Carlo Croce clonaram os miR-15a/16-1 em uma região de recorrente deleção cromossômica em pacientes com Leucemia Linfocita Crônica (LLC) (VINCENT; PICHLER, 2014).

Além da influência dos miRNAs, o câncer pode estar associado também à metilação do DNA, que por meio dos dinucleotídeos CpG geralmente apresentam uma alta frequência de metilação. Dessa maneira, as regiões do genoma que apresentam concentração de dinucleotídeos CpG são denominados ilhas de CpG (ICG), desempenhando uma importante função na regulação da expressão gênica.

Este trabalho se propõe a realizar uma análise computacional para caracterizar e analisar regiões do genoma canino e humano que possuem microRNAs com ICGs próximas, com a intenção de verificar e classificar a relação dessas regiões com genes alvos associados ao câncer.

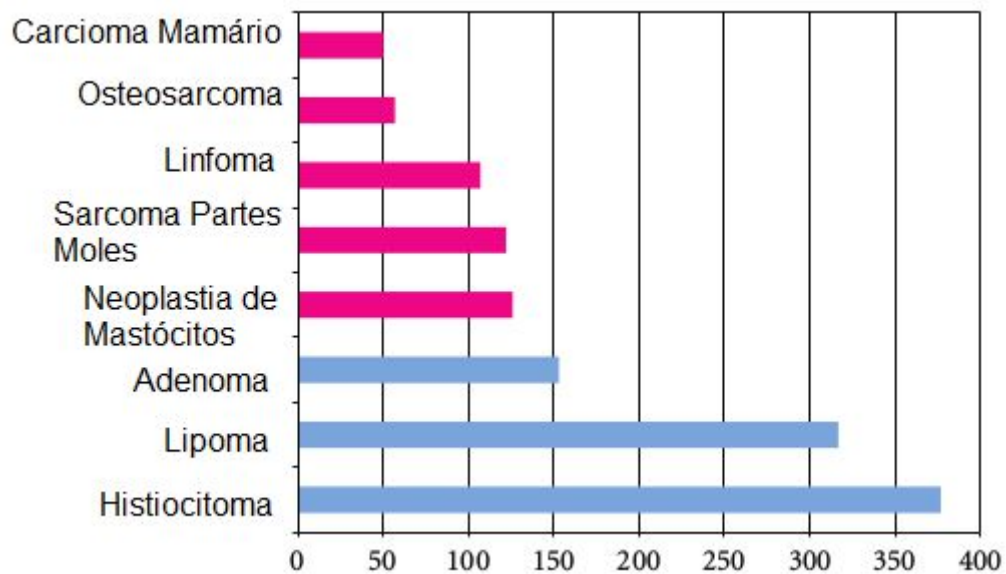
2 Revisão Bibliográfica

2.1 Câncer

Nos Estados Unidos da América (EUA) são diagnosticados cerca de 1 milhão de casos de câncer em cães por ano. O câncer é responsável por 27% das mortes caninas no Reino Unido, dentre todas as doenças caninas. A ausência de registros sobre tumores caninos dificulta saber se o câncer é crescente em cães ou não, bem como se os cães vivem mais e se o diagnóstico da doença melhorou (JANE, 2013, p. 23).

A Figura 1 apresenta a incidência de cada tipo de câncer nos cães pesquisados, podendo ser observado quais os tipos que mais se destacam.

Figura 1 - Incidência de tipos específicos de câncer em cães.





Fonte: Adaptado de JANE (2013, p. 23)

Em humanos o câncer também é considerado um problema de saúde pública. Segundo dados do *World Cancer Report 2014* é esperado que nas próximas décadas o impacto do câncer na população corresponda a 80% dos 20 milhões de novos casos estimados para 2025, nos países desenvolvidos, sendo os tipos mais comuns o câncer de próstata, mama e pulmão (INCA, 2015).

No Brasil a estimativa é de 600 mil novos casos de câncer em 2016 e 2017, onde os tipos de câncer mais comuns são o de próstata e de mama, que correspondem a 61 mil e 58 mil, respectivamente, conforme Tabela 1.

Tabela 1 - Casos de novos câncer no Brasil

Localização Primária	Casos	%			Localização Primária	Casos	%
Próstata	61.200	28,6%	Homens 	Mulheres 	Mama feminina	57.960	28,1%
Traqueia, Brônquio e Pulmão	17.330	8,1%			Cólon e Reto	17.620	8,6%
Cólon e Reto	16.660	7,8%			Colo do útero	16.340	7,9%
Estômago	12.920	6,0%			Traqueia, Brônquio e Pulmão	10.890	5,3%
Cavidade Oral	11.140	5,2%			Estômago	7.600	3,7%
Esôfago	7.950	3,7%			Corpo do útero	6.950	3,4%
Bexiga	7.200	3,4%			Ovário	6.150	3,0%
Laringe	6.360	3,0%			Glândula Tireoide	5.870	2,9%
Leucemias	5.540	2,6%			Linfoma não Hodgkin	5.030	2,4%
Sistema Nervoso Central	5.440	2,5%			Sistema Nervoso Central	4.830	2,3%

*Números arredondados para múltiplos de 10.

Fonte: INCA (2015)

O câncer é uma doença genética que necessita da inativação de genes supressores tumorais e a ativação de proto-oncogenes. Dessa forma, sequências de DNA mutadas são transcritas em RNA mensageiro e, posteriormente, em proteínas com funções desordenadas (AMARAL; NONAKA, 2010, p. 105).

Uma célula normal pode sofrer mutação genética no DNA, processo em que o material genético alterado das células passa a receber instruções erradas para as suas atividades. A regulação da expressão gênica por mecanismos epigenéticos está envolvida nos processos de diferenciação celular e desenvolvimento de organismos eucarióticos (MINÁROVITS, 2009).

Estes mecanismos não dependem diretamente da sequência primária do DNA e estão mais relacionados às alterações estruturais como a compactação da fita de DNA, a metilação de nucleotídeos, as modificações pós-tradicionais de histonas, o posicionamento de nucleossomos e a interação entre RNAs mensageiros e RNAs não codificantes (ncRNA) (BALLESTAR, 2011, p. 711).

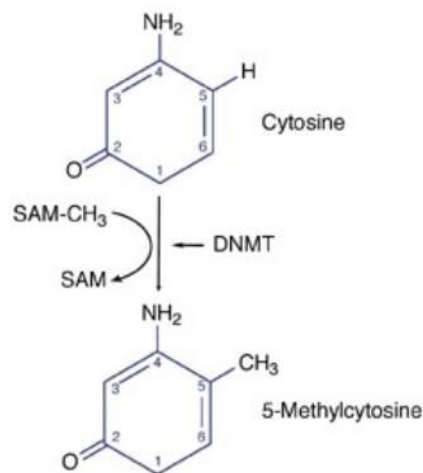
2.2 Metilação do DNA

A epigenética é uma área que tem como característica a alteração herdada na expressão gênica das outras gerações, sem que haja mudança na sequência primária de DNA (WOLFFE; GUSCHIN, 2000). A epigenética dessa maneira tem seu efeito efetivo na

regulação da expressão gênica em nível transcricional. As alterações correspondem a uma série de modificações moleculares no DNA e na cromatina (RICHARDS, 2006; TALBERT; HENIKOFF, 2006). O processo de ligação de histonas e condensação da cromatina é regulada por padrões epigenéticos de metilação do DNA e modificações químicas das histonas por enzimas como DNA Metiltransferases (DNMTs), Histonas Acetiltransferases (HATs), Histonas Deacetilases (HDACs), proteínas de ligação a grupo metis (MECP2), e histonas metiltransferases (KIEFER, 2007; WAGGONER, 2007; WEIDMAN; DOLINOY, 2007).

A metilação do DNA e a modificação de histonas têm chamado a atenção de pesquisadores a inibidores destes mecanismos e já estão em uso como drogas para o tratamento de alguns tipos de câncer (MANN; JOHNSON, 2007)(KAMINSKAS; FARRELL, 2005)(MANN; JOHNSON, 2007). Este processo consiste na adição de um radical metil (CH_3) no carbono 5 de Citosina, geralmente seguido por Guanina (dinucleotídeo CpG), conforme apresentado na Figura 2, catalisada por enzimas DNA metiltransferases (DNMTs) (PAULSEN & FERGUSON-SMITH, 2001). Existem várias isoformas de DNMT, porém a DNMT1, a DNAMT3a e a DNMT3b são consideradas como as de maior relevância para a metilação global do genoma (LIANGFENG; WITMER, 2007).

Figura 2 - Mecanismo de metilação do DNA



Fonte: Adaptado de STRATHDEE; BROWN (2002).

A DNMT1 tem preferência por DNA hemimetilado e é apontada como a mais importante para a manutenção do padrão de metilação na replicação do DNA, enquanto as outras duas são associadas à metilação de novo do DNA durante a embriogênese, a qual se torna o padrão de metilação de células somáticas. Os genes transcionalmente ativos

apresentam normalmente hipometilação de citosinas, enquanto a hipermetilação está normalmente associada a uma baixa atividade transcricional (KINNEY; PRADHAN, 2011, p. 311).

O processo de metilação possui importantes funções no desenvolvimento embrionário normal, inativação do cromossomo X, regulação gênica, *imprinting* genômico e modificações da cromatina. Alterações nos padrões de metilação do DNA podem aparecer em organismos com idade avançada e nos estágios iniciais da carcinogênese (HERMAN; BAYLIN, 2003).

Outro fator que pode influenciar a perda da expressão gênica é a hipermetilação de regiões promotoras com muitos dinucleotídeos CG. Tipicamente a hipometilação do DNA desencadeia um aumento na expressão gênica, enquanto a hipermetilação diminui a expressão dos genes-alvos (LIDDLE; JIRTLE, 2006).

Algumas regiões do DNA são ricas em sequências dinucleotídicas CpG, formando as chamadas ilhas de CpG (ICG). Essas regiões da cromatina são fortemente acetiladas e, frequentemente, a acetilação ocorre no resíduo de lisina 9 da histona H3, deixando a cromatina em sua configuração ativa (CAIAFA; ZAMPIERI, 2005; WILSON; POWER, 2007).

As ilhas de CpG correspondem as regiões genômicas com mais de 500 pares de base de comprimento e com um grande número de dinucleotídeos GC, sendo que cerca de 55% dessas ilhas estão localizadas nas regiões promotoras de aproximadamente 40% dos genes dos mamíferos. Normalmente as ilhas de CpG são mantidas não-metiladas, exceto em genes submetidos ao *imprinting* ou situados no cromossomo X inativo, o que permite a ligação de proteínas e enzimas que iniciam a transcrição (HERMAN; BAYLIN, 2003). E ainda existem as ilhas de CpG metiladas que estão relacionadas com o silenciamento transcricional.

Aproximadamente 70-80% dos dinucleotídeos CpG do genoma de mamíferos apresentam metilação da citosina, porém regiões de alta concentração de dinucleotídeos CpG, constituindo um arranjo denominado ilhas de CpG (ICG), não apresentam esta modificação. No genoma de mamíferos, aproximadamente 70% dos genes possuem ICGs sobrepondo-se com as regiões de início de transcrição, caracterizando-se como elementos definidores de promotores, e que podem se expandir além da porção 5' dos genes (primeiros exons) (COSTA; PACHECO, 2013).

As ilhas de CpG desempenham um papel importante em cerca de 70% dos genes nas regiões promotoras (BABENKO; CHADAEVA, 2017). Essas regiões promotoras determinam uma estrutura de cromatina propícia para a transcrição, porém são também sítios

para a ligação de fatores de transcrição que regulam a expressão gênica por mecanismos genéticos e epigenéticos (ILLINGWORTH; BIRD, 2009).

Estes fatores de transcrição podem ser agrupados em conjuntos de famílias, sendo que essas famílias são identificadas por meio de sequências, que acredita-se estarem relacionadas biologicamente, seja por sua estrutura, função ou evolução (EIDHAMMER; JONASSEN, 2004).

Buscam-se, nessas sequências, características em comum baseadas somente em um padrão das sequências, que podemos chamar de *motifs*. Essas sequências podem ser classificadas por meio de padrões probabilísticos e/ou padrões determinísticos, dependendo do método e da situação a ser analisada, sendo uma forma importante de encontrar padrões entre várias sequências diferentes.

Em geral, todos os tipos de câncer possuem ganhos de metilação nos promotores gênicos associados a ilhas CpG (GALM; HERMAN; BAYLIN, 2006). Um estudo com 38 amostras indicou hipermetilação da região promotora dos genes BRCA1 (14/38, 37%), CDKN2A (13/38, 34%) e SFN (33/38, 87%), apresentando 95% das regiões sendo hipermetiladas para, pelo menos, um dos genes (JING; ZHANG, 2007). Outro exemplo é a hipometilação, normalmente associada à atividade transcricional, da região promotora do gene *Shh* e relacionada ao câncer de mama. Um aumento de expressão deste gene, acompanhada de hipometilação da região promotora, foi observada em 70,5% (43/61) das amostras de tecido tumoral analisadas (CUI; WANG, 2010).

As ilhas de CpG podem estar associadas também com microRNAs (miRNAs), sugerindo que estes microRNAs controlam o processo de regulação da expressão epigenética (PAVICIC; PERKIÖ, 2011). Esses miRNAs podem contribuir para o estudo de tumores, agindo como oncogenes ou genes supressores de tumor e podem ser importantes no diagnóstico, prognóstico e tratamento de câncer.

2.3 MicroRNA

Anteriormente acreditava-se que os RNAs eram apenas intermediários entre o genoma e as proteínas, conforme o dogma central da biologia. Descobertas apresentam vários RNAs não codificantes (ncRNA), os ncRNAs não participam de maneira direta da síntese de proteínas e conforme sua estrutura podem ser classificados como *small nuclear RNA*

(snRNA), *small nucleolar RNAs* (snoRNA), *small interfering RNA* (siRNA) e em destaque os microRNAs (CHAMORRO; AMBROSIO-ALBUQUERQUE; ROSA, 2015).

Os microRNAs possuem, em média, de 17 a 25 nucleotídeos distribuídos em um filamento único. Muitos deles exercem um papel regulatório do mecanismo de expressão gênica, controlando a atividade de diversos genes, codificantes ou não. Eles são provenientes de regiões não-codificantes (íntrons) e têm um papel importante na regulação gênica, como a diferenciação tecidual, ciclo celular, proliferação e apoptose (CHAMORRO; AMBROSIO-ALBUQUERQUE; ROSA, 2015).

Sabe-se que os ncRNAs desempenham importantes funções (CHAMORRO; AMBROSIO-ALBUQUERQUE; ROSA, 2015), dentre as quais algumas estão destacadas na Tabela 2.

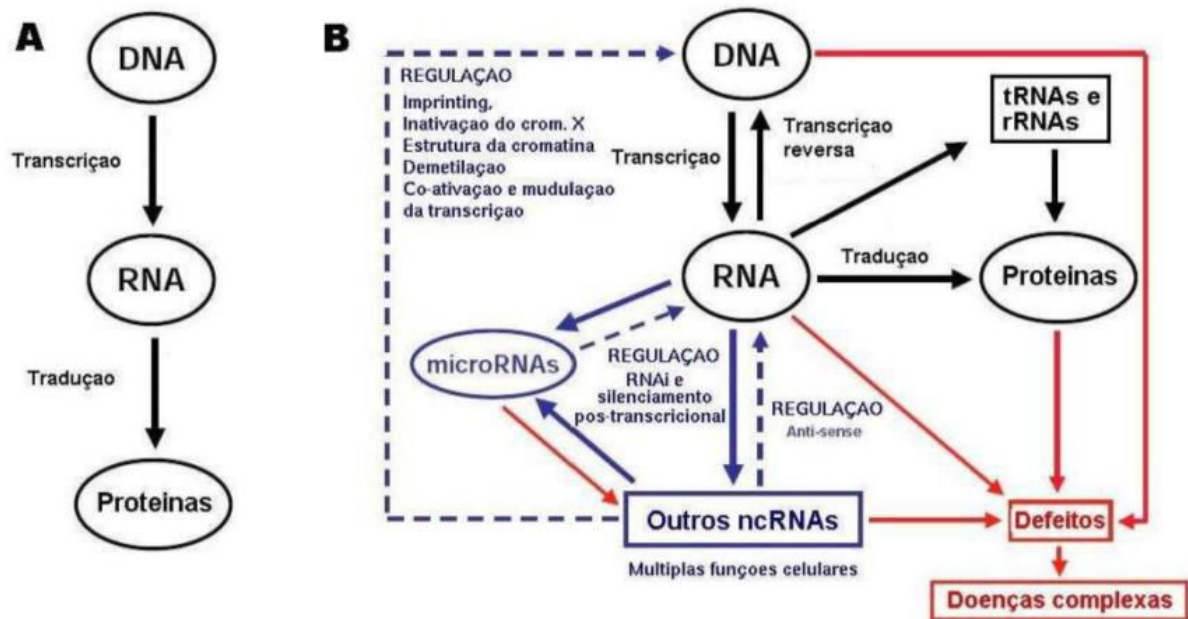
Tabela 2- Exemplos de ncRNAs

RNA	Função
6S	Inibição de uso de promotores específicos
7H4	Regulação gênica em sinapses
7SL, 4.5S	Transporte de proteínas (reconhecimento de sinal)
BC1	Ativação de proteínas repressoras
KHPS1a	Demetilação de DNA
OxyS	Repressão de tradução
miRNA	Regulação gênica pós-transcricional
snoRNA	Modificação de outros RNAs
XIST, TSIX e roX	Inativação do cromossomo X
	Co-ativação seletiva de receptores de esteróides

Fonte: adaptado de LIMA (2006).

Acredita-se que existam vários outros ncRNAs, com outras características e funções diferentes ainda não descobertas. Dessa maneira, imagina-se um novo ciclo para o dogma central da biologia, conforme apresentado na Figura 3 (LIMA, 2006).

Figura 3A- Dogma central da biologia B - novo fluxo com as novas descobertas.

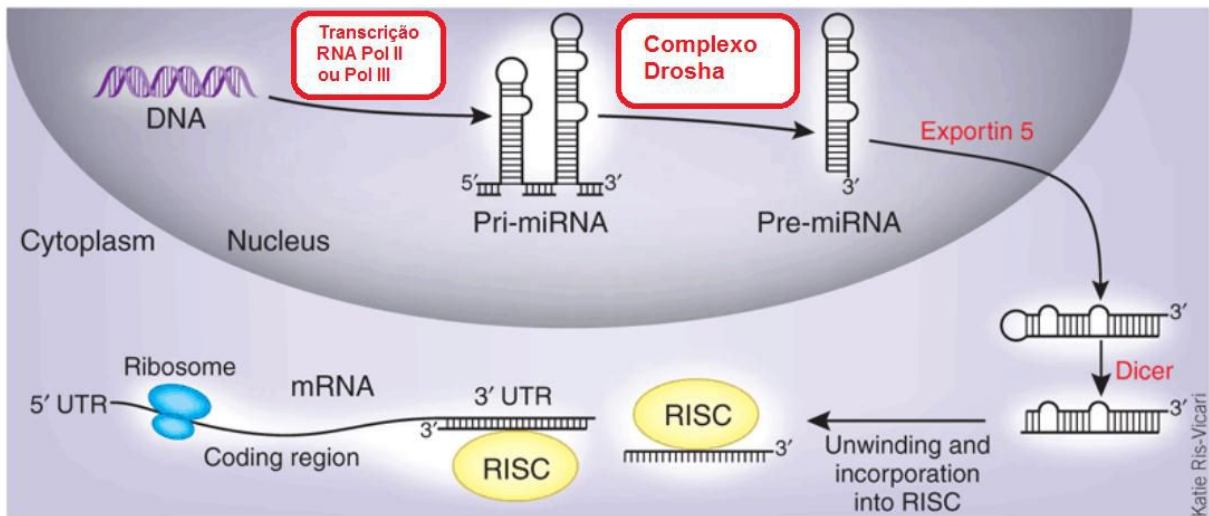


Fonte: LIMA (2006).

Mais recentemente os mecanismos epigenéticos envolvendo a participação de RNAs não codificantes (ncRNA), como os microRNAs (miRNAs), têm recebido atenção da comunidade científica. Os miRNAs constituem um novo campo de intensa pesquisa científica associada ao câncer, na qual se procura entender seus papéis em diversos processos biológicos como o crescimento, a diferenciação celular e a apoptose (SATO, TSUCHIYA; MELTZER, 2011).

A Figura 4 apresenta o processo para o surgimento dos microRNAs: os miRNAs primários transcritos são gerados pela RNA polimerase II ou polimerase III, seja como unidades de transcrição separadas ou incorporados dentro dos introns de genes codificadores de proteínas. Dentro do núcleo, o processo que contém a enzima Rnase III Drosha separa o transcrito primário para liberar o “hairpin” pré miRNA que será exportado para o citoplasma pela exportina 5 (XPO5). No citoplasma, um complexo de proteínas, incluindo DICER e TRBP, engloba o pré-miRNA para produzir o miRNA maduro de cadeia simples, o qual irá incorporar na sequência o RNA de silenciamento, induzindo o complexo (RISC) (MARTIN; ANDERS, 2012).

Figura 4 - Ciclo do microRNA



Fonte: Adaptado JEFFREY (2008)

Os miRNAs regulam a expressão gênica por meio de hibridação com RNAs mensageiros (RNAm), impedindo a tradução ou disparando o processo de degradação do RNAm. A relação hipotética é a de que a baixa expressão de determinados miRNAs está relacionada à expressão dos RNAm correspondentes, enquanto uma alta expressão dos miRNAs inibiria a expressão das atividades biológicas relacionadas àquele RNAm (HUANG; JIA, 2011). Por sua vez, a regulação da expressão de miRNAs pode também ocorrer por mecanismos epigenéticos envolvendo a metilação do DNA em regiões de ICGs (HAN; WITMER, 2007).

O miRNA miR-506 foi o primeiro identificado em primatas no cromossomo X, sendo que tal miRNA serviu de *cluster* para identificação de mais dez miRNAs, o que formou sete diferentes famílias e padrões de identificação, gerando mais 7 *clusters* diferentes (miR-506, miR-507, miR-508, miR-509, miR-510, miR-513 e miR-514) (JINGFANG; JIAN, 2016). Uma das técnicas utilizadas pela bioinformática é a utilização de *cluster* para identificação de novos miRNAs.

2.4 Bioinformática

A bioinformática ou biocomputação integra conhecimentos de diversas áreas com a finalidade de decifrar processos biológicos por meio de modelos lógico-matemáticos e estatísticos. Dentre as diversas aplicações, podemos dar atenção à busca pela identificação de

proteínas diferencialmente expressas em condições patológicas, apresentando possíveis caminhos para a produção de fármacos antitumorais (ARAÚJO; FARIAS, 2008).

A bioinformática surgiu juntamente com a tecnologia de sequenciamento automatizado, e começou a ganhar espaço na área de biologia, mas foi limitada pela capacidade computacional das máquinas da época e a disponibilidade de dados. A partir da década de 90 vem ocorrendo um desenvolvimento de novas tecnologias de sequenciamento de DNA, contribuindo para o aumento do volume de dados biológicos a serem analisados com a utilização de ferramentas de bioinformática.

A bioinformática atua em diversas áreas para interpretação de dados biológicos, algumas como: 1) tipos de informação biológica e base de dados, 2) análises de sequências e modelagem molecular, 3) análises genômicas e 4) sistemas biológicos. As novas descobertas auxiliadas pela bioinformática geraram um volume grande de informações e exigem novas técnicas e ferramentas para análise, armazenamento e interpretação dos dados (DINIZ; CANDURI, 2017).

A melhor forma de armazenar esses novos dados é por meio de bancos de dados públicos, dentre os existentes pode-se destacar: *GenBank at the National Center for Biotechnology Information* (NCBI), criado em 1982, *DNA Database of Japan* (DDBJ), criado em 1986, e o *European Molecular Biology Laboratory* (EMBL), criado em 1997 (PEVSNER, 2015).

A colaboração entre esses três bancos de dados citados deu origem ao *International Nucleotide Sequence Database Collaboration* (INSDC) (<<http://www.insdc.org>>), o qual estabeleceu uma padronização para o formato dos dados e protocolos para a validação dos dados recebidos.

Os três bancos de dados que compõem o INSDC recebem, armazenam e compartilham várias sequências de nucleotídeos recebidos com informações associadas; com a queda nos custos de sequenciamento, o INSDC lida com um grande volume de dados (2400 trilhões de pares de bases em 2015) e desenvolve serviços e novas tecnologias para atender a demanda (COCHRANE; KARSCH-MIZRACHI, 2016).

Esses bancos de dados citados armazenam diversos tipos de informações, como genomas completos, genes específicos, proteínas, miRNAs etc. Porém, quando se trata somente de miRNAs, existe um banco de dados específico (miRBase) para várias espécies diferentes que recebem novas informações e validam novos miRNAs.

O banco de dados miRBase reúne informações sobre 28645 microRNAs descritos para 223 espécies. Estão catalogados 1883 miRNAs de humanos e 502 miRNAs de cães

(www.mirbase.org). Estes miRNAs podem estar ligados a várias doenças, inclusive o câncer (NADINE; HELLMUTH-ALEXANDER, 2013).

Além dos bancos de dados há, também, o trabalho para a identificação dos processos regulatórios de RNAs/miRNAs. O método frequentemente utilizado para detecção e identificação de novos miRNAs computacionalmente ainda é um desafio, sendo baseado em genômica comparativa com organismos, *loci* genéticos, tamanho da transcrição, estrutura do RNA primário e a energia livre da estrutura. Algumas ferramentas que utilizam esse método são o MiRscan57 e miRseeker, os quais utilizam regiões conservadas com estruturas de *stem-loop* (ANAND; PANDEY, 2017).

Foram identificados 359 doenças humanas como câncer de estômago, câncer de próstata, câncer de pulmão, doença de Parkinson dentre outras relacionadas a 287 miRNAs humanos, os quais estão também associados com outros animais domésticos que possuem homologia com a espécie humana, 50% destes são homólogos aos cães (BUZA; ARICK, 2014).

Além dos miRNAs, as ilhas de CpG também estão associados à metilação do DNA. Pronina (2017) teve como objetivo de estudo avaliar computacionalmente o papel da metilação em regiões promotoras de ilhas de CpG na desregulação de 13 miRNAs associados ao câncer. Concluiu-se, por meio de estudos de PCR quantitativo, juntamente com 58 amostras de câncer de mama e análises de bioinformática, que todos os miRNAs que possuíam ilhas de CpG nas regiões promotoras estavam associados de alguma maneira à metilação do DNA e a algum tipo de câncer.

As análises *in silico* estimam a presença de 37729 ICG no genoma humano, sendo 12917 associadas à regiões promotoras e à porção 5' dos genes, e no genoma canino existem 58327 ICG e 9.825 associadas à regiões promotoras (HAN; ZHAO, 2009). É possível detectar ICG no genoma utilizando técnicas computacionais e matemáticas que, aplicadas na área de estudo da biologia, são conhecidas como bioinformática.

Em 1987, Gardiner-Garden e Frommer estabeleceram um algoritmo para a identificação das ilhas CpG como sendo uma sequência de DNA com um comprimento superior a 200 pares de base, uma quantidade de CG em que região de mais do que 50% e uma razão entre o observado/esperado (O/E) superior a 0,6 (tamanho>200pb, CG>50% e O/E>0.6) (GARDINER-GARDEN; FROMMER, 1987)(KOZOMARA; GRIFFITHS-JONES, 2010). Em 2002, Takai e Jones reavaliaram a definição de Gardiner-Garden e Frommer e propuseram um novo critério onde a quantidade de CG seria maior que 55%, razão O/E maior que 0,65 e o tamanho maior que 500 pares de base (tamanho>500pb, CG>55%, O/E>0.65)

(TAKAI; JONES, 2002). Esse algoritmo seria capaz de excluir os falsos positivos das pequenas sequências repetitivas (CHUANG; YANG, 2012).

Baseados nos algoritmos de Gardiner-Garden e Frommer e Takai e Jones, surgiram várias ferramentas computacionais para descoberta de ilhas de CpG como: CpGPlot, CpGProD, CpGIS e CpGcluster. Cada *software* apresenta suas características e diferenças como apresentação das ICG, se a plataforma é para *web* ou *desktop*, dentre outras apresentadas na Tabela 3.

Os *softwares* utilizados para a detecção de ICG são baseados em funções matemáticas de combinação como o *Sliding window*, em combinações com sequências já conhecidas como os *clusters* e algoritmos de complementação como CGA e CPSO. Todos esses algoritmos são baseados em funções estatísticas e matemáticas já utilizadas em soluções de problemas computacionais.

Tabela 3 - Comparação das funções dos softwares de ICG

Ano	2011	2006	2003	2002	2000
<i>Software</i>	CpGPAP	CpGcluster	CpGIS	CpGProD	CpG Plot
Mudança de Parâmetros	✓	✓	✓	✓	✓
Resultado ICG	✓	✓	✓	✓	✓
CpG Dinucleotídeo	✓		✓		
<i>Length</i>				✓	
Dados de O/E	✓				✓
Dados de CG%	✓				
<i>Upload</i> de sequências	✓			✓	✓
Integração com outras análises					
<i>Data integrator</i>	✓				

Fonte: adaptado de CHUANG; YANG (2012).

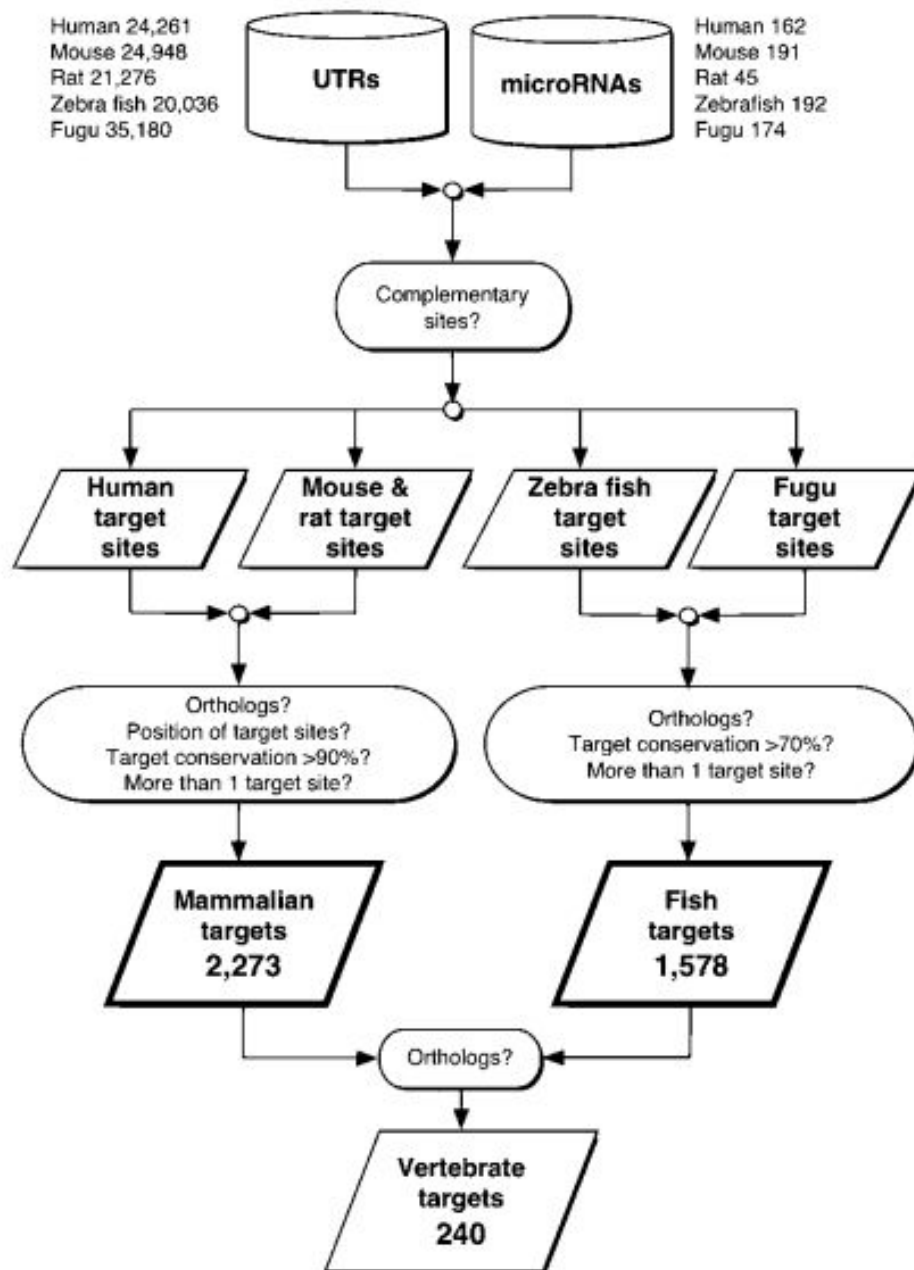
Além dos *softwares* de ICG, existem também algoritmos que são utilizados para detectar microRNAs através da técnica de *clusters*, conforme *pipeline* apresentado na Figura 5. Essa técnica se baseia em encontrar relações estatísticas e complementariedade entre as sequências confirmadas como miRNAs e as novas sequências, levando em consideração a relação entre as espécies pesquisadas, como taxa de conservação do DNA e homologia da região pesquisada. O miRBase é o principal repositório de dados de microRNAs, onde é possível buscar microRNAs por sequência, predição dos genes alvos e por nome (GRIFFITHS-JONES; SAINI, 2008).

Além dos genes alvos dos miRNAs apresentados na Figura 5, o sistema de nomenclatura também é importante para os miRNAs. Enquanto os pré-miRNAs tem a notação como “miR-X”, os miRNAs “maduros” possui a mesma notação porém o “X” é um valor numérico. A espécie é representada por um código de 3 letras acrescentado anteriormente ao “miR”, e ainda segundo LAGES e IPAS (2012) pode ser diferenciado por mais uma letra para poucos pares de bases diferentes:

- miR-181a: aacauucaacgcugucggugagu
- miR-181b: aacauucauugcugucggugggu

Um sufixo numeral é adicionado aos nomes de todos os miRNAs *multi loci*, por exemplo, miR-181a-1 e miR-181a-2 (miR-181a-1 originário de um local no cromossomo 1 e miR-181a-2 de um local no cromossomo 9). Entretanto, temos alguns miRNAs que não seguem o mesmo padrão, como hsa-let-7 e cel-lin-4 (LAGES; IPAS, 2012). Neste raciocínio, uma pesquisa de miRNAs, de cães e humanos, no miRBase, pode apresentar relações interessantes entre as ilhas de CpG e o câncer.

Figura 5 - Pipeline de targets em organismos vertebrados



Fonte: BINO; ENRIGHT (2004)

3 Motivação

Com o aumento da população e da expectativa de vida canina, vários casos de câncer estão sendo diagnosticados e, diferentemente do que ocorre nos casos de câncer em humanos, não existe um empenho tão grande para os estudos desta patologia em cães.

Neste sentido, uma análise computacional e comparativa de predição de microRNAs e Ilhas de CpG e busca de similaridades com o genoma humano apresenta caráter inovador e um passo inicial para o desenvolvimento de novas drogas.

4 Objetivos

4.1 Objetivo geral

Realizar uma análise completa sobre microRNAs e ilhas de CpG, nos genomas caninos e humanos, por meio de ferramentas computacionais que permitam armazenar, pesquisar, comparar e visualizar esses dados relacionados ao câncer.

4.2 Objetivos específicos

- 1) Catalogar os microRNAs caninos e os seus potenciais genes alvos.
- 2) Identificar ICGs na região genômica englobando os microRNAs caninos e seus potenciais genes alvos.
- 3) Analisar as ICGs para a presença de elementos de regulação da expressão gênica.
- 4) Apresentar graficamente a localização genômica das ICGs, microRNAs e potenciais genes alvos.
- 5) Identificar e comparar as associações obtidas com as potencialmente similares em seres humanos.
- 6) Identificação e análise da distribuição de sítios de restrição de enzimas sensíveis à metilação nas ICGs.
- 7) Construção de ferramenta para visualização e comparação da estrutura das ICGs caninas e humanas.

5 Materiais e métodos

As análises foram realizadas em um *notebook* com processador Intel Core i3, com 8 GB de memória RAM, sistema operacional Linux Mint 17.1 e linguagem de programação Python. Os dados a serem analisados foram obtidos na base de dados do *National Center for Biotechnology Information* (NCBI) e no MiRBase (base de dados de microRNAs).

5.1 Linguagem de Programação

Cada vez mais são necessárias aplicações que tenham alto desempenho, seja pelo grande volume de dados ou pela complexidade da lógica; neste caso é importante a escolha de uma linguagem de programação que seja robusta, rápida e eficiente.

Dentre as linguagens científicas destaca-se o Python, que é uma linguagem de programação dinâmica de alto nível e orientada a objetos, além de possuir uma sintaxe simples e clara, permitindo grandes ganhos de produtividade e, ainda, integração com outras linguagens e ferramentas (COELHO, 2007).

O Python possui suporte para vários formatos de dados para bioinformática como: Clustalw, FASTA, GenBank, PubMed and Medline, ExPASy files, like Enzyme and Prosite, SCOP, including ‘dom’ and ‘lin’ files, UniGene e SwissProt. Isto acontece devido à tipagem dinâmica do Python, não sendo necessário declarar os tipos de dados.

Além disso, grandes empresas e aplicativos utilizam a linguagem Python, como a Globo.com, Netflix, Google, Amazon Web Service etc., pois é uma linguagem completamente *open source*, ou seja, podemos alterar e utilizar suas funções e classes de acordo com a necessidade do usuário, além de ser gratuita e multi-plataforma.

Juntamente com o Python utilizou-se um *software* chamado Glade, para a implementação de uma interface gráfica, para facilitar o manuseio para qualquer usuário, tornando possível a variação dos parâmetros e realização de novas análises de maneira intuitiva e simples. O Glade é um *software* gratuito e *open source* que interage com várias linguagens de programação como: C, C++, C#, Java, Perl, Python e outras (GLADE, 2014).

O arquivo gerado pela interface é em um formato XML, podendo funcionar em qualquer sistema operacional e se adequando a qualquer resolução de sistema e tela. O arquivo contendo a interface estará disponível a sessão de anexos deste trabalho.

5.2 Aquisição dos dados

O genoma canino foi obtido da base de dados do *National Center for Biotechnology Information* (NCBI). O genoma obtido foi a última versão (CanFam 3.1), atualizada em 24/09/2013 e encontra-se disponível para *download* em: <ftp://ftp.ncbi.nlm.nih.gov/genomes/Canis_lupus_familiaris/>

O genoma humano obtido foi atualizado em 15/03/2015, sendo a última versão e disponível em: <ftp://ftp.ncbi.nlm.nih.gov/genomes/Homo_sapiens>.

Foi desenvolvido um *script* apresentado na Figura 6, nativo ao sistema operacional Linux, para realização do *download* de modo automático, dentre os arquivos de interesse do NCBI.

Figura 6 - Script para aquisição dos genomas

```

1  #!/bin/bash
2
3  for(( i=1; i<23; i++ ))
4  do
5  if [ $i -le 9 ]
6  then
7      wget ftp://ftp.ncbi.nlm.nih.gov/genomes/Homo_sapiens/CHR_0$i/hs_alt_CHM1_1.1_chr$i.fa.gz
8  else
9      wget ftp://ftp.ncbi.nlm.nih.gov/genomes/Homo_sapiens/CHR_$i/hs_alt_CHM1_1.1_chr$i.fa.gz
10 fi
11 done
12 wget ftp://ftp.ncbi.nlm.nih.gov/genomes/Homo_sapiens/CHR_X/hs_alt_CHM1_1.1_chrX.fa.gz
13 wget ftp://ftp.ncbi.nlm.nih.gov/genomes/Homo_sapiens/CHR_X/hs_ref_GRCh38.p2_chrY.fa.gz
14

```

Fonte: Autor

O miRBase (<http://www.mirbase.org>) é um repositório público de sequência de microRNAs criado em 2002, atualizado e utilizado constantemente pela comunidade científica (KOZOMARA; GRIFFITHS-JONES, 2014), as sequências de microRNAs caninas e humanas foram obtidas e organizadas em 26/11/2014. Os dados obtidos do site miRBase são organizados de acordo com o nome do miRNA, um número de registro “*accession*”, o cromossomo, as posições no cromossomo e o “*strand*”, conforme Figura 7.

Figura 7 - microRNAs *Canis familiaris* miRBase

Canis familiaris miRNAs (502 sequences) [CanFam3.1]

ID	Accession	Chromosome	Start	End	Strand
cfa-let-7a-1	MI0007989	chr10	20034319	20034387	-
cfa-let-7a-2	MI0010328	chr5	12052786	12052881	+
cfa-let-7b	MI0010329	chr10	20033388	20033472	-
cfa-let-7c	MI0008076	chr31	13259522	13259588	+
cfa-let-7d	MI0027994	chr1	97902012	97902136	-

Fonte: Autor

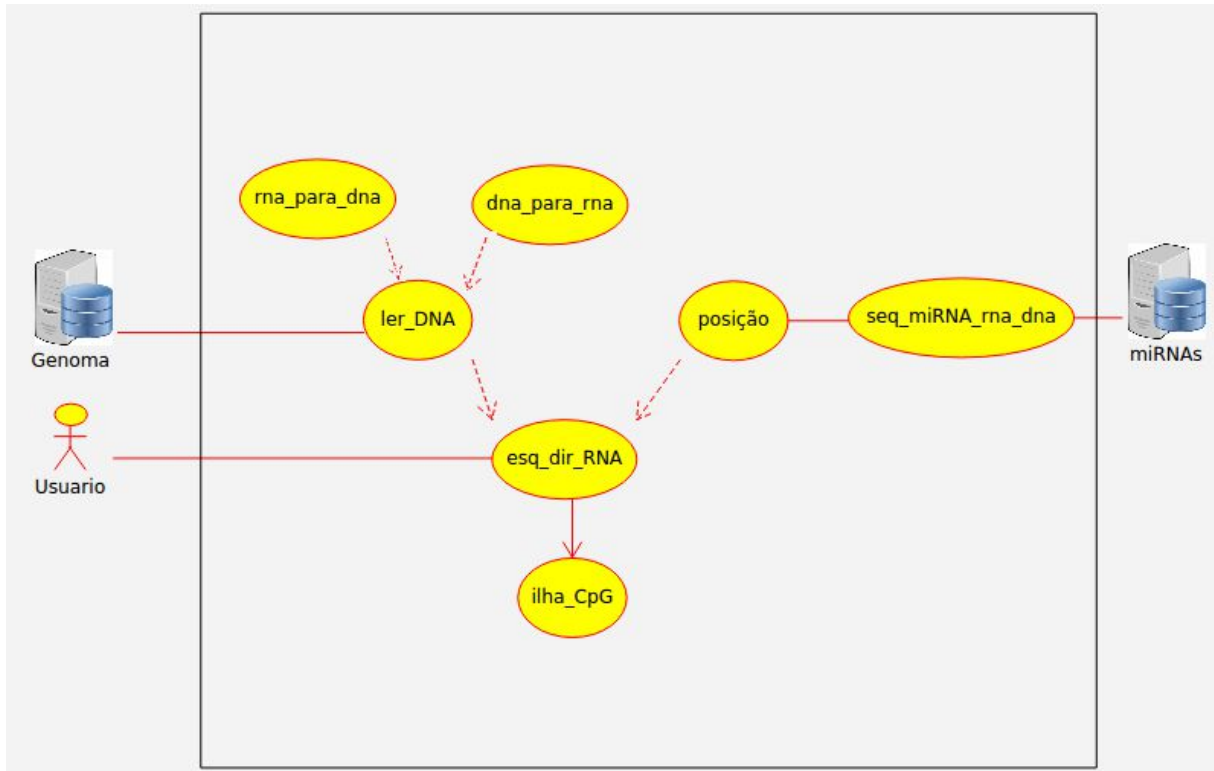
Os dados com a sequência estão dispostos em outro arquivo diferente, sendo necessária a criação de um algoritmo em Python para organizar e catalogar todos os dados dos miRNAs em um único arquivo, pois as consultas serão baseadas neste arquivo (Figura 8).

Figura 8 - Arquivo de miRNAs com sequências

ID	Chromosome	Start	End	Strand	Sequence
cfa-let-7a-1	chr10	20034319	20034387	-	UGAGGUAGUAGGUUGUAUAG
cfa-let-7a-2	chr5	12052786	12052881	+	CUGCAUGUUCCCAGGUUGAGC
cfa-let-7b	chr10	20033388	20033472	-	CGGGGUGAGGUAGUAGGUUG
cfa-let-7c	chr31	13259522	13259588	+	UGAGGUAGUAGGUUGUAUGG
cfa-let-7d	chr1	97902012	97902136	-	AAAUGGGUCCUAGGAAGAG

Fonte: Autor

As sequências dos nucleotídeos são de RNAs e devem ser convertidas para DNA, de forma a serem localizadas nos arquivos do genoma canino, obtido no site NCBI. Os valores de “*strand*”, positivo ou negativo, informa em qual fita e sentido a sequência se encontra e a necessidade de se converter para a mesma informação do genoma, na Figura 9 é apresentado o fluxograma de funcionamento.

Figura 9 – Fluxograma funcionamento do *software* miRNA_CG Finder

Fonte: Autor

O algoritmo foi desenvolvido na linguagem Python e possui sete funções para obter a lista de sequências com ilhas de CpG próximos. Inicialmente, a função “ler_DNA” realiza a indexação de cada cromossomo, para que seja realizada as buscas; posteriormente, duas funções de conversão “rna_para_dna” e “dna_para_rna” são utilizadas para facilitar as conversões nas buscas, se necessário. A interface do *software* desenvolvido é apresentada na Figura 10.

O campo “Search ICGs near miRNAs” é utilizado para o usuário limitar a quantidade de pares de bases, antes e depois, das sequências dos miRNAs que irão busca ilhas de CpG, e os demais campos – “CG content”, “Ratio” e “Lenght” – são os parâmetros definidos para determinar se um trecho de nucleotídeos possui ou não ilhas de CpG.

Os resultados são exibidos e também é criado uma tabela no formato .CVS, com todos os resultados, que poderá ser aberta em qualquer *software* de pacote Office.

Figura 10 - Interface do Software

The image shows a software window titled "miRNA_CG finder" with a standard Windows-style title bar (minimize, maximize, close buttons). The main content area has a heading "Analysis of miRNAs and CpG islands". Below this, there are several input fields and buttons:

- "DNA Folder:" followed by a dropdown menu showing "(Nenhum)".
- "miRNAs File:" followed by a file selection button showing "(Nenhum)".
- "Search ICGs near miRNAs (Base Pairs):" followed by a text input field containing "2000".
- "CG Content:" followed by a text input field containing "55.0".
- "Ratio:" followed by a text input field containing "0.65".
- "Length" followed by a text input field containing "500".
- Two buttons: "Clear" and "RUN".
- A section at the bottom labeled "miRNAs with ICG" with an empty list area below it.

Fonte: Autor

As sequências dos miRNAs estão nos dois sentidos da fita, sendo importante realizar a conversão de acordo com o “*strand*” contido no autor “miRNAs”, que nada mais é que um arquivo com os dados das sequências de miRNAs, descrito na Figura 8. A função “seq_mirna_rna_dna(seq_rna)” receberá como entrada a sequência de RNA e substituirá o nucleotídeo “U” (Uracila) por “T” (Timina); quando a sequência estiver na mesma fita que o DNA do Genoma, a segunda função “rna_para_dna(rna)”, conforme código apresentado na Figura 11, cujo objetivo é obter a sequência complementar de DNA, realizará as trocas de nucleotídeos necessárias.

Figura 11- Funções em Python para converter nucleotídeos

```

def seq_mirna_rna_dna(seq_rna):
    """Retorna a sequencia trocando U por T"""
    seq_dna = seq_rna.replace("U", "T")
    return seq_dna

def rna_para_dna(rna):
    """ Retorna o DNA """
    dna = rna.replace("A", "T")
    dna = dna.replace("U", "A")
    dna = dna.replace("C", "X")
    dna = dna.replace("G", "C")
    dna = dna.replace("X", "G")
    return dna

```

Fonte: Autor

Após converter a sequência de microRNA para DNA de acordo com o “*strand*” e com o arquivo do cromossomo disponível pela função “*ler_DNA*”, é possível realizar a busca dentro do cromossomo correspondente.

Depois de localizar a sequência do microRNA no cromossomo do DNA, obtém-se uma sequência de tamanho determinado pelo autor “*usuario*”, antes e depois do miRNA, pela função “*esq_dir_RNA*”, e depois essas sequências, antes e após o miRNA, são enviadas para a função “*ilha_CpG*”.

A função “*ilhaCpG(qtidadePb)*” recebe como entrada um valor (variável *qtidadePb*), fornecido pelo autor “*usuario*”, que será a quantidade de pares de bases a ser analisada antes e após o miRNA. Após a função ser iniciada, é gerado um arquivo com cabeçalho informando qual miRNA terá sua região próxima analisada (*ID_antes*), em qual cromossomo (*Cromossome*), posição final e inicial (*Start* e *End*), qual a porcentagem de Citosina e Guanina (%CG) na região, qual a taxa de Citosina e Guanina que aparecem juntas (*Ratio*) e a sequência analisada (*Sequência*). Os critérios de Gardiner-Garden e Frommer e Takai e Jones são detalhados na página 23.

Este processo é realizado de maneira automatizada para todos os 492 microRNAs caninos, e ao final é apresentado um arquivo com todos os microRNAs que possuem ilhas de CpG ao seu entorno (Figura 12).

Desta maneira, no final, obtém-se todas as sequências que possuem Ilhas de CpG e miRNAs próximos.

Figura 12 - Algoritmo para detecção de Ilhas de CpG

```

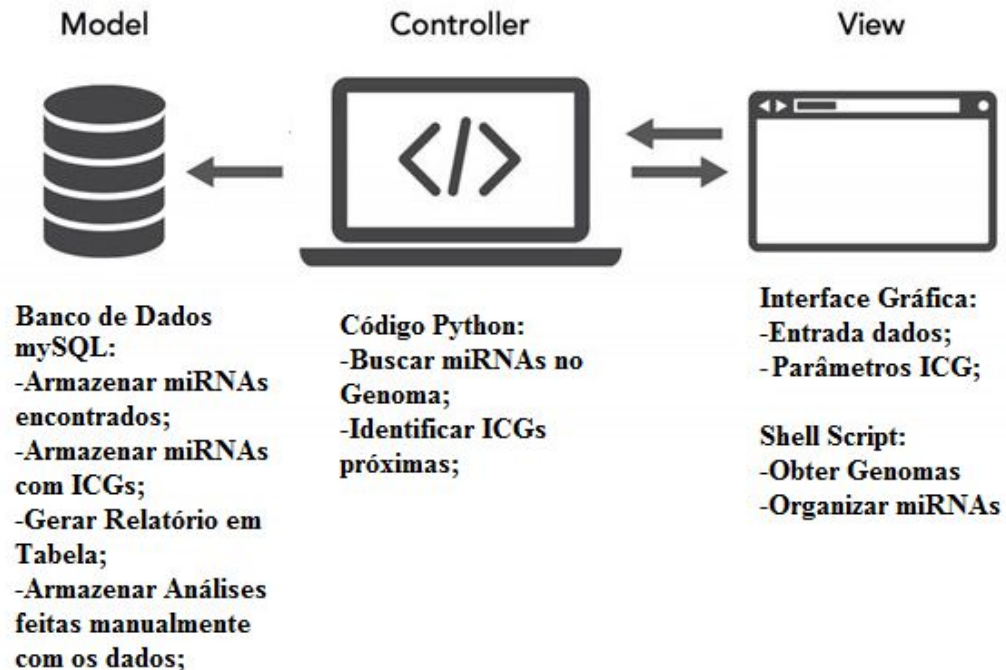
93 def ilhaCpG(qtidadePb):
94     x=0
95     y=0
96     cr = csv.reader(open("Resultado"+str(qtidadePb)+".csv", "rb"))
97     listaTemporariaAntes=[]
98     listaTemporariaAntes.append("ID_Antes; Chromosome; Start; End; %CG; Ratio; Sequencia\n")
99     listaTemporariaDepois=[]
100    listaTemporariaDepois.append("ID_Depois; Chromosome; Start; End; %CG; Ratio; Sequencia\n")
101    for seq_dna in cr:
102        seq_dna = seq_dna[0].split(";")
103        if seq_dna[3] != " SEQ depois":
104            tamanho = len(seq_dna[3])
105            num_c = seq_dna[3].count("C")#numero de citosina
106            num_g = seq_dna[3].count("G")#numero de guanina
107            gc_content = 100.0 * (float(num_c + num_g)) / tamanho
108            #print "The GC content of your sequence is {0:.1f}%".format(gc_content)
109            expected = float(num_c * num_g) / tamanho
110            if expected != 0:
111                ratio = float(seq_dna[3].count("CG")) / expected
112            else:
113                ratio = 0
114            if gc_content > 55.0 and tamanho >= 500 and ratio > 0.65: #parametros
115                linha = str(seq_dna[0])+";"+str(seq_dna[1])+";"+str(seq_dna[5])+";"+str(int(seq_c
116                listaTemporariaDepois.append(linha)
117                x += 1

```

Fonte: Autor

O software desenvolvido *miRNA_CG Finder* atende ao padrão de desenvolvimento *Model, View e Controller* (MVC), onde a camada *View* é responsável pela comunicação com o usuário, há as entradas dos dados para realização das análises, conforme apresentado na Figura 13. A camada *Controller* é onde fica toda a lógica de programação e serão fornecidas as instruções computacionais para realização das buscas e análises. E a camada *Model* é o local em que fica o armazenamento dos dados, ou seja, onde os resultados serão salvos em um Banco de Dados e gerada uma tabela no formato *Comma Separated Values* (CSV), para facilitar o entendimento e análises de maneira manual posteriormente.

Figura 13 - Padrão de desenvolvimento do software



Fonte: Autor

Com os resultados obtidos pelo *software* desenvolvido, os miRNAs com ICGs foram submetidos à análise dos *motifs*, com a finalidade de verificar os padrões genômicos nas sequências obtidas. Esses padrões genômicos incluem *tandem repeats*, trechos de sequências ricas em GC ou AT (ilhas de CpG) e locais de ligações de proteínas associadas ao DNA (sítios de fatores de transcrição) (BOEVA, 2016).

Para realizar essa análise, utilizamos o *software HOMER (Hypergeometric Optimization of Motif EnRichment)* (<<http://homer.salk.edu/homer/>>), de código aberto e desenvolvido para sistemas UNIX, que possui uma coletânea de ferramentas computacionais para análises e algoritmos de descobertas de *motifs* e enriquecimento de sequências (HEINZ, 2010).

Para a descoberta dos *motifs* acontecem os seguintes processos:

- 1) Extração das Sequências: verifica se os dados de entrada estão corretos, se foram extraídos do DNA de maneira apropriada.
- 2) Seleção do *Background*: será definido quais promotores, de acordo com a espécie ou definido automaticamente e quais os tipos de análises serão realizadas.
- 3) Normalização do GC: realiza a verificação e busca de ilhas de CpG e o alvo das sequências.

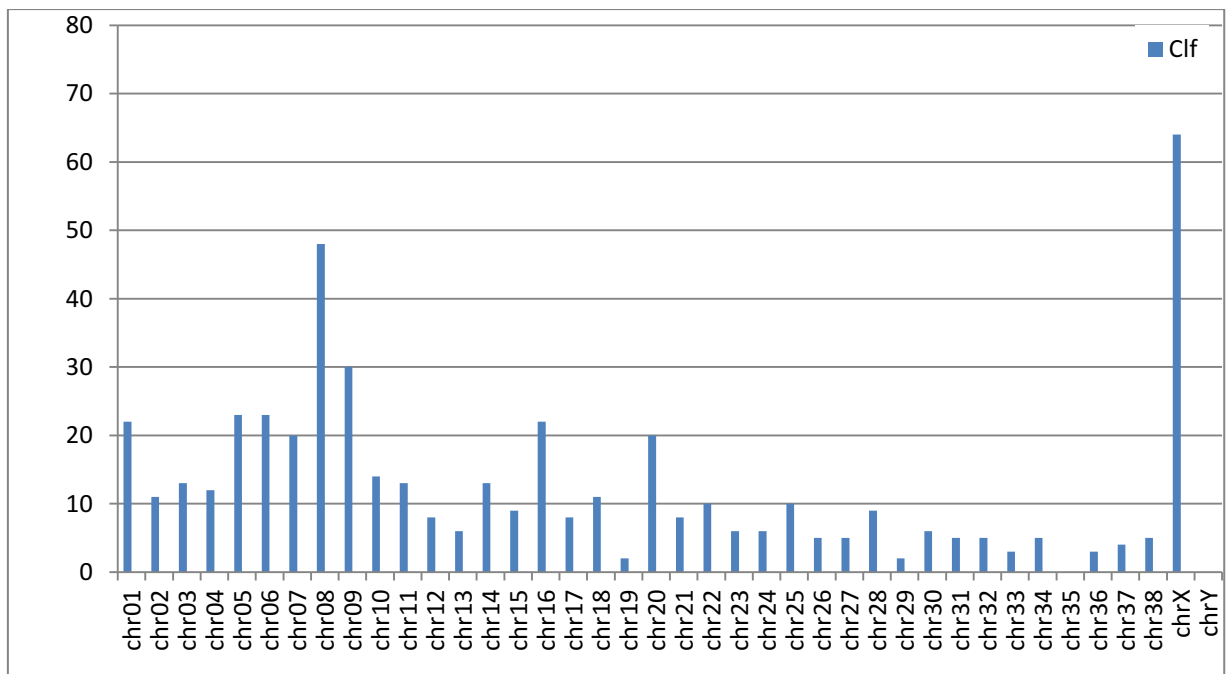
6 Resultados e Discussão

A partir da organização dos dados de miRNA, foi verificado graficamente a quantidade de miRNAs em cada cromossomo para o *Canis lupus familiaris* (clf), apresentada na Figura 14, e para o *Homo sapiens* (hs), na Figura 15. É importante visualizar graficamente a distribuição dos miRNAs ao longo de todo o genoma, pois isso determinará em qual cromossomo nosso *software* realizará mais buscas e provavelmente obterá mais resultados.

Podemos destacar um número elevado de miRNAs em alguns cromossomos, como por exemplo, os cromossomos 8 e X para o genoma canino e o cromossomo 1, 19 e X para o genoma humano. Segundo McKierman (2017) o cromossomo X das espécies concentram cerca de 7% de todos os miRNAs, isto porque é um local onde ocorre muita expressão e regulação de genes (MOLLOY; MCKIERNAN, 2017).

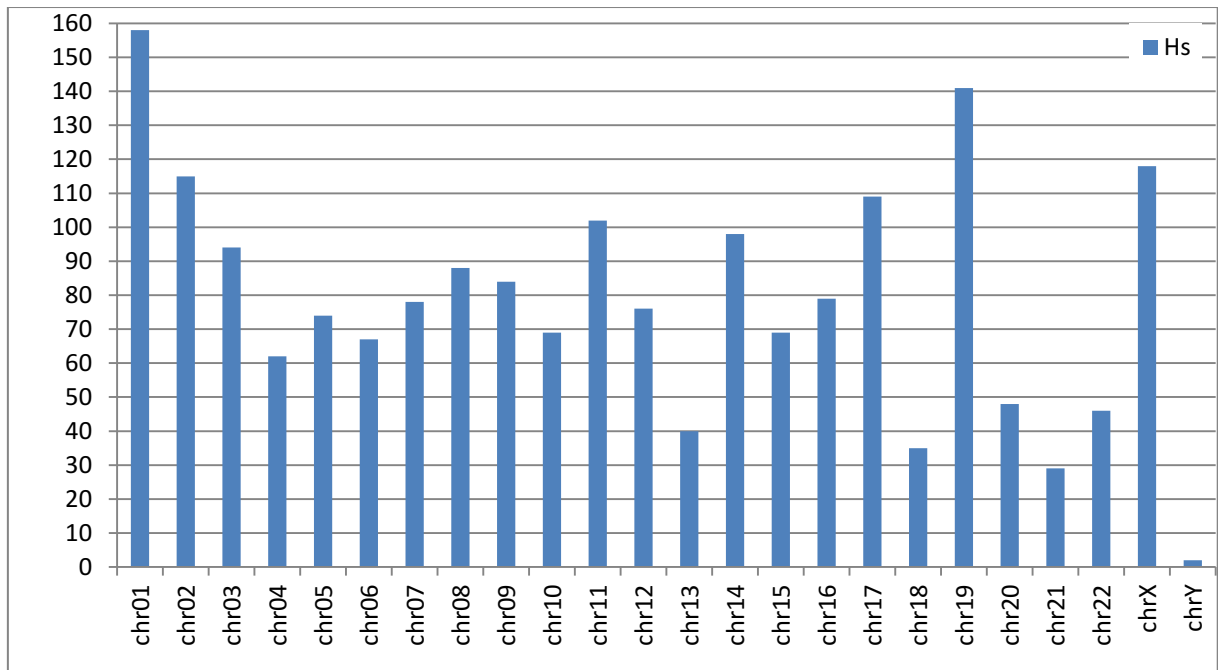
Porém nosso intuito é buscar quais cromossomos possuem ilhas de CpG próximas aos miRNAs e, por isso, é necessária a realização de análise em buscas destas ilhas.

Figura 14 - Quantidade de microRNAs por cromossomo canino.



Fonte: Autor

Figura 15 - Quantidade de miRNAs por cromossomo Humano.



Fonte: Autor

Para a análise da identificação e organização de ilhas de CpG nos genomas humano e canino foi desenvolvido um *software* denominado *miRNA_ICG Finder*. O *miRNA_ICG Finder* apresenta algumas características diferentes em relação a outros sistemas para buscar ilhas de CpG existentes como: integração com outras análises como os miRNAs, mudança de parâmetros, ser genérico para qualquer espécie e outras diferenças que podem ser verificadas na Tabela 4.

Tabela 4 - Comparação entre os softwares de ICG

Ano	2011	2006	2003	2002	2000	2017
<i>Software</i>	CpGPAP	CpGcluster	CpGIS	CpGProD	CpG plot	miRNA _CG Finder
Mudança de Parâmetros	✓	✓	✓	✓	✓	✓
Resultado ICG	✓	✓	✓	✓	✓	✓
CpG Dinucleotídeo	✓		✓			✓
<i>Length</i>				✓		✓
Dados de O/E	✓				✓	✓
Dados de CG%	✓					✓
<i>Upload</i> de sequências	✓			✓	✓	✓
Integração com outras análises						✓
Data integrator	✓					✓

Fonte: adaptado de CHUANG; YANG (2012).

Algumas pesquisas apontam uma relação entre os miRNAs e as ICG, como por exemplo, os miRNAs da família miR-200 que estão associados com a regulação da migração/invasão celular, sendo definida que as regiões de “cortes” desses miRNAs predizem a doença e, por sua vez, revelam que sua expressão pode estar frequentemente associada a alteração na metilação de CpG (LONG; SMIRAGLIA, 2017).

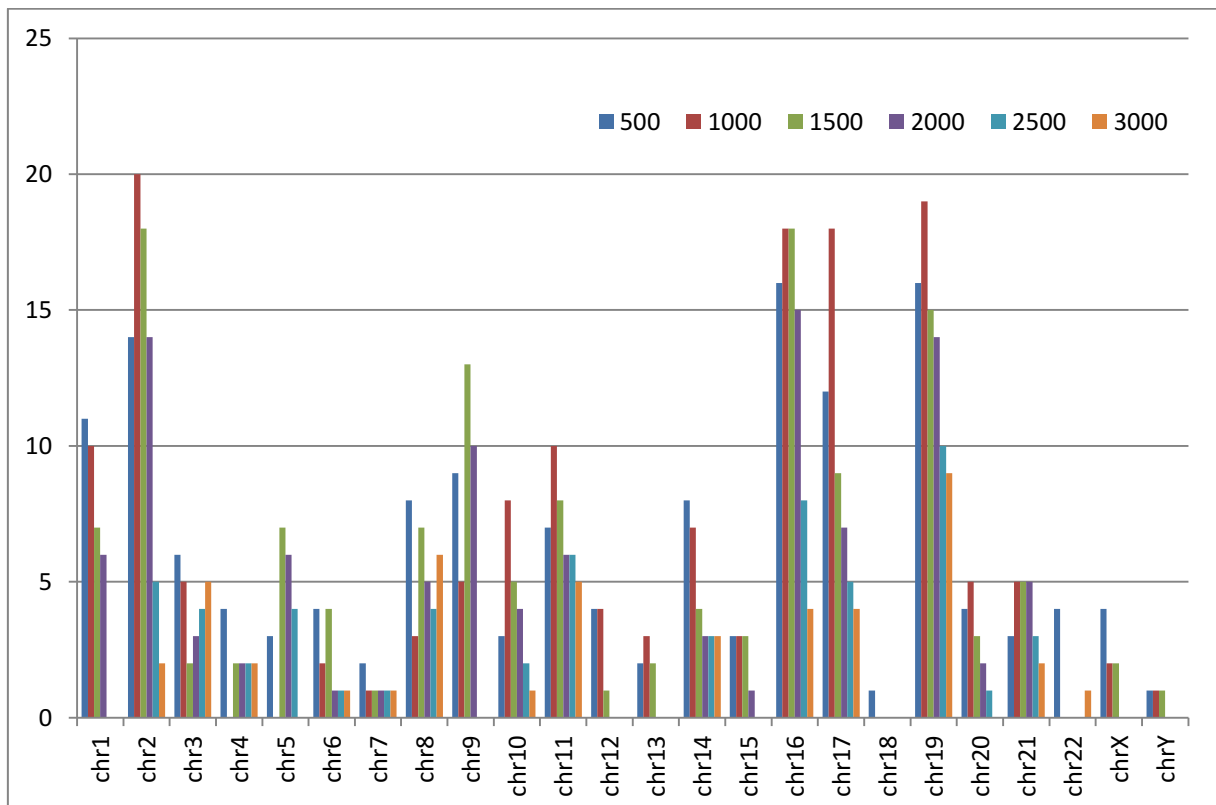
6.1 Ilhas de CpG

O CpG Finder segue as sugestões dos algoritmos propostos por Gardiner-Garden e Frommer e Takai e Jones, os quais mencionam apenas o tamanho mínimo da sequência de 500 pares de base, mas não limitam o tamanho máximo da sequência; desta maneira buscou-se ilhas de CpG em sequências, antes e após os miRNAs, de tamanho variado: 500 pb, 1000 pb, 1500 pb, 2000 pb, 2500 pb e 3000 pb.

A partir desses dados obteve-se os resultados que são apresentados na Figura 16 para o genoma humano e na Figura 17 para o genoma canino. Observando os dois gráficos é possível notar que, de uma maneira geral, quanto maior é o tamanho da sequência ao entorno do miRNA, menor será o número de ilhas de CpG que o miRNA_CG Finder localiza. Embora as pesquisas sobre o genoma não-codificante ainda estejam incompletas, existem evidências de que as ilhas de CpG, que não eram associadas a um gene conhecido, podem estar associadas a ncRNA, mais especificamente aos microRNAs (LONG; SMIRAGLIA, 2017). Dessa maneira, buscar uma relação entre microRNAs e ilhas de CpG próximas podem revelar uma associação importante.

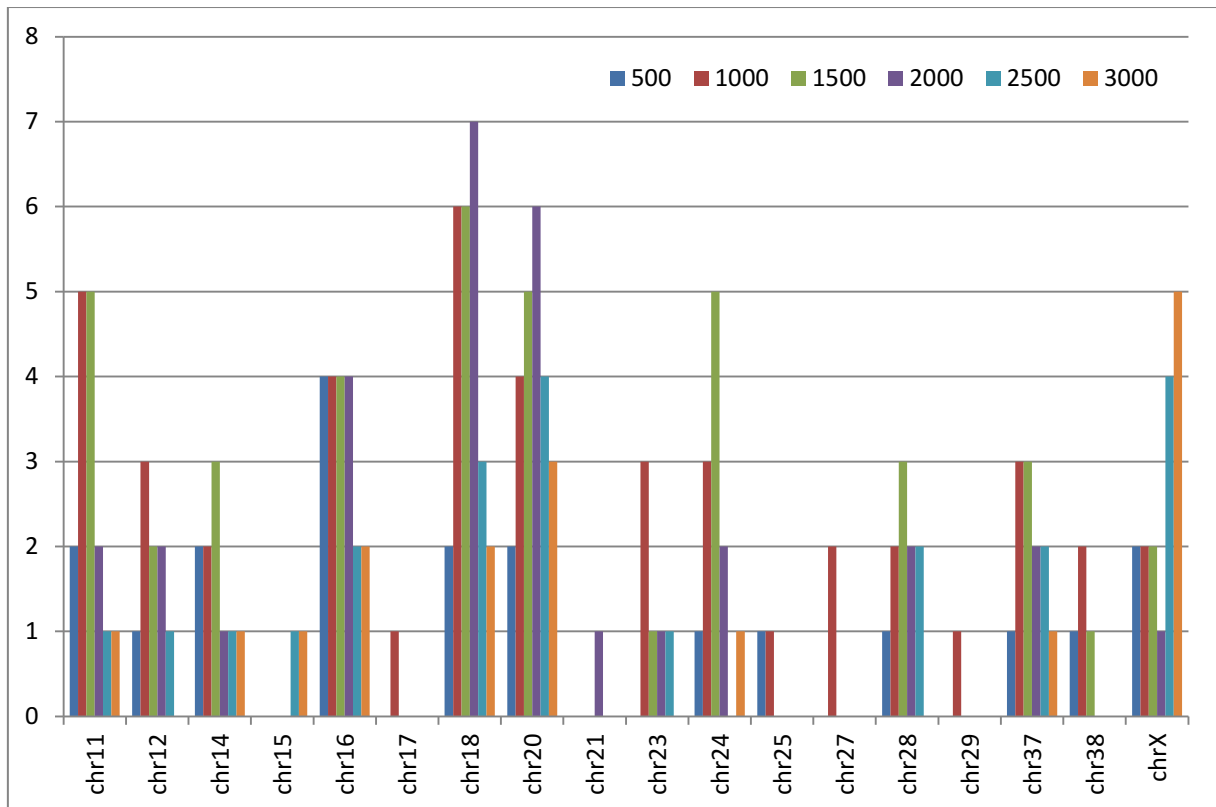
Sendo assim, foram consideradas todas as regiões de miRNAs com ilhas de CpG próximas, análise que levou em consideração diferentes distâncias em termos de pares de bases sendo a variação de 500, 1000, 1500, 2000, 2500 e 3000 conforme apresentado na Figura 16 para o genoma humano e na Figura 17 para o genoma canino. A variação do número de pares de bases ao redor do miRNA apresentava apenas algumas ilhas de CpG a mais em cada caso, sendo que todas apresentaram as mesmas ilhas de CpG independente da variação.

Figura 16 - Quantidade de regiões de ICG por cromossomo em Humanos



Fonte: Autor

Figura 17 - Quantidade de regiões de ICG por cromossomo em Cães.



*Os cromossomos não apresentados não obtiveram nenhuma ilha CpG

Fonte: Autor

A Tabela 5 apresenta os miRNAs com ilhas de CpG próximas, as posições das regiões, os valores da razão de CG observado e esperado (*ratio*) e a quantidade de CG (%CG) para o genoma canino. Foram obtidas 67 sequências de regiões a serem analisadas.

Tabela 5 - Regiões de miRNAs com ICG próximas em cães

miRNAs	Cromossomo	Start	End	%CG	Ratio
cfa-mir-8891	chr01	104782434	104784434	56,8	0,87
cfa-mir-574	chr03	73492019	73494019	60,73	0,89
cfa-mir-145	chr04	59530027	59532027	57,2	0,79
cfa-mir-378	chr04	59275306	59277306	73,30	0,84
cfa-mir-33b	chr05	41685158	41687158	68,87	0,65
cfa-mir-8872	chr05	41262560	41264560	60,00	0,83
cfa-mir-8875	chr06	21077100	21079100	61,53	1,08
cfa-mir-8877	chr06	39745933	39747933	67,4	0,69
cfa-mir-187	chr07	54258842	54260842	56	0,87
cfa-mir-8865	chr07	80970911	80972911	69,87	1,18
cfa-mir-92b	chr07	42335379	42337379	57,63	0,73
cfa-mir-127	chr08	69103897	69105897	59,47	0,70
cfa-mir-203	chr08	71771146	71773146	59,13	0,66

cfa-mir-369	chr08	69290972	69292972	55,83	0,79
cfa-mir-376a-2	chr08	69269456	69271456	56,33	0,79
cfa-mir-376b	chr08	69269856	69271856	57,27	0,68
cfa-mir-376c	chr08	69269085	69271085	55,33	0,82
cfa-mir-381	chr08	69271882	69273882	55,27	0,67
cfa-mir-409	chr08	69290681	69292681	55,90	0,86
cfa-mir-410	chr08	69291276	69293276	56,07	0,71
cfa-mir-411	chr08	69251993	69253993	57,2	0,66
cfa-mir-433	chr08	69102811	69104811	57,90	0,72
cfa-mir-487b	chr08	69274453	69276453	58,07	0,65
cfa-mir-8884	chr08	53822710	53824710	57,80	0,76
cfa-mir-132	chr09	46150490	46152490	70,73	0,84
cfa-mir-193a	chr09	41311380	41313380	71	0,86
cfa-mir-196a-1	chr09	24892656	24894656	64,53	0,73
cfa-mir-212	chr09	46150874	46152874	68,13	0,74
cfa-mir-219-2	chr09	55133006	55135006	73,8	0,69
cfa-mir-632	chr09	40607485	40609485	58,4	0,72
cfa-mir-6516	chr09	3774662	3776662	55,07	0,84
cfa-mir-8830-1	chr11	14016415	14018415	63,67	0,76
cfa-mir-8830-2	chr11	14016059	14018059	56,33	0,71
cfa-mir-8830-3	chr11	14016656	14018656	58,93	0,71
cfa-mir-219-1	chr12	2670890	2672890	65,73	0,73
cfa-mir-148a	chr14	39301806	39303806	59,67	0,71
cfa-mir-183	chr14	7066441	7068441	66,2	0,68
cfa-mir-196b	chr14	40343962	40345962	55,60	0,89
cfa-mir-8840	chr15	32469020	32471020	55,47	0,67
cfa-mir-8852	chr16	1146069	1148069	58,17	0,85
cfa-mir-8853	chr16	1275368	1277368	55,53	0,76
cfa-mir-8862	chr16	1099248	1101248	61,77	0,75
cfa-mir-8906	chr16	13916184	13918184	70,33	0,68
cfa-mir-129-2	chr18	26153496	26155496	58,00	0,73
cfa-mir-210	chr18	25674446	25676446	65,80	0,82
cfa-mir-8859a	chr18	46689398	46691398	71,00	0,82
cfa-mir-8859b	chr18	46689579	46691579	68,13	0,75
cfa-mir-1199	chr20	48437915	48439915	55,90	0,71
cfa-mir-191	chr20	40143098	40145098	60,07	0,67
cfa-mir-425	chr20	40143575	40145575	58,60	0,88
cfa-mir-8804	chr20	55096919	55098919	60,20	0,68
cfa-mir-8806	chr23	12234964	12236964	70,67	0,86
cfa-mir-124-3	chr24	46916689	46918689	72,67	0,72
cfa-mir-133c	chr24	46490486	46492486	70,33	0,66
cfa-mir-8809	chr24	47405814	47407814	66,17	0,68
cfa-mir-320	chr25	35013575	35015575	74,6	0,77
cfa-mir-202	chr28	40840046	40842046	74,73	0,69
cfa-mir-1840	chr37	25734749	25736749	58,47	0,78
cfa-mir-375	chr37	25565951	25567951	62,10	0,66
cfa-mir-8798	chr38	17579190	17581190	55,13	0,78
cfa-mir-106a	chrX	104895574	104897574	57,63	0,91

cfa-mir-18b	chrX	104895415	104897415	57,83	0,91
cfa-mir-19b-2	chrX	104895025	104897025	57,37	0,89
cfa-mir-20b	chrX	104895158	104897158	58,10	0,90
cfa-mir-503	chrX	105180138	105182138	56,8	0,80
cfa-mir-718	chrX	121861688	121863688	72,13	0,67
cfa-mir-92a-2	chrX	104894886	104896886	56,13	0,85

Fonte: Autor

Para o genoma humano foi realizada a mesma análise e o resultado é apresentado na Tabela 6; entretanto foi encontrado uma quantidade maior, com 178 regiões. Acredita-se que isso acontece pela quantidade maior de miRNAs catalogados devido ao maior número de estudos dedicados aos humanos.

Tabela 6 - Regiões de miRNAs com ICG próximas em humanos

miRNAs	Cromossomo	Start	End	%CG	Ratio
hsa-mir-1302-2	chr1	137	2137	59,65	0,73
hsa-mir-2682	chr1	98216538	98218538	57,267	0,79
hsa-mir-4632	chr1	60	2060	59,20	0,72
hsa-mir-4742	chr1	84	2084	59,20	0,72
hsa-mir-5095	chr1	87	2087	59,30	0,72
hsa-mir-6084	chr1	20811080	20813080	57,60	0,79
hsa-mir-6733	chr1	43343847	43345847	62,40	0,68
hsa-mir-6808	chr1	1152116	1154116	67,20	0,68
hsa-mir-760	chr1	94017463	94019463	55,30	0,93
hsa-mir-9-1	chr1	136077205	136079205	62,45	0,70
hsa-mir-1915	chr10	21726397	21728397	64,30	0,79
hsa-mir-2110	chr10	112557834	112559834	55,30	0,67
hsa-mir-3663	chr10	115551525	115553525	56,65	0,69
hsa-mir-4482	chr10	102652049	102654049	58,00	0,76
hsa-mir-4678	chr10	85886083	85888083	57,85	0,78
hsa-mir-4683	chr10	35870140	35872140	63,00	0,80
hsa-mir-1237	chr11	60752191	60754191	64,77	0,67
hsa-mir-129-2	chr11	43540617	43542617	55,10	0,73
hsa-mir-1908	chr11	58198927	58200927	60,43	0,70
hsa-mir-210	chr11	507120	509120	64,77	0,69
hsa-mir-3656	chr11	115359006	115361006	56,40	0,74
hsa-mir-4488	chr11	57892131	57894131	64,17	0,68
hsa-mir-5691	chr11	9050839	9052839	60,96	0,66
hsa-mir-6090	chr11	124861054	124863054	65,00	0,79
hsa-mir-611	chr11	58176248	58178248	55,533	0,67
hsa-mir-4497	chr12	107079049	107081049	65,00	0,78
hsa-mir-6125	chr12	59512145	59514145	58,20	0,96
hsa-mir-615	chr12	51284517	51286517	63,20	0,72
hsa-mir-8072	chr12	120605554	120607554	57,533	0,70
hsa-mir-3613	chr13	31518592	31520592	57,933	0,82

hsa-mir-3665	chr13	59220204	59222204	56,667	0,75
hsa-mir-1247	chr14	82965240	82967240	68,10	0,81
hsa-mir-127	chr14	82285503	82287503	56,20	0,65
hsa-mir-203b	chr14	85522298	85524298	74,40	0,75
hsa-mir-409	chr14	82468658	82470658	55,80	0,66
hsa-mir-412	chr14	82470896	82472896	57,20	0,79
hsa-mir-433	chr14	82286503	82288503	56,867	0,69
hsa-mir-4505	chr14	55165198	55167198	62,30	0,73
hsa-mir-1302-10	chr15	82101024	82103024	58,45	0,80
hsa-mir-1469	chr15	76466578	76468578	58,467	0,73
hsa-mir-7706	chr15	65515344	65517344	56,067	0,99
hsa-mir-9-3	chr15	69502381	69504381	67,20	0,69
hsa-mir-1225	chr16	2078110	2080110	63,63	0,67
hsa-mir-1538	chr16	59797720	59799720	57,28	0,73
hsa-mir-3176	chr16	531213	533213	59,52	0,66
hsa-mir-3177	chr16	1725040	1727040	59,53	1,07
hsa-mir-3178	chr16	2521955	2523955	58,15	0,66
hsa-mir-3180-1	chr16	16428165	16430165	68,65	0,74
hsa-mir-3180-2	chr16	16428163	16430163	68,65	0,74
hsa-mir-3180-4	chr16	152	2152	57,800	0,80
hsa-mir-3181	chr16	40973814	40975814	63,60	0,77
hsa-mir-4519	chr16	57	2057	57,85	0,67
hsa-mir-4719	chr16	83	2083	58,05	0,67
hsa-mir-5189	chr16	113	2113	57,90	0,67
hsa-mir-6511a-1	chr16	16442853	16444853	68,20	0,69
hsa-mir-6511a-2	chr16	16442853	16444853	68,20	0,69
hsa-mir-6511a-3	chr16	16442853	16444853	68,20	0,69
hsa-mir-6770-3	chr16	15197083	15199083	55,667	0,80
hsa-mir-6862-2	chr16	69	2069	58,05	0,67
hsa-mir-762	chr16	32162392	32164392	56,12	0,67
hsa-mir-7854	chr16	64	2064	58,10	0,68
hsa-mir-1268b	chr17	74949008	74951008	57,33	0,69
hsa-mir-132	chr17	1852069	1854069	70,17	0,65
hsa-mir-152	chr17	42969602	42971602	61,200	0,67
hsa-mir-193a	chr17	26739724	26741724	64,80	0,67
hsa-mir-196a-1	chr17	43565915	43567915	57,00	0,66
hsa-mir-212	chr17	1850331	1852331	56,48	0,88
hsa-mir-3615	chr17	69597571	69599571	56,32	0,68
hsa-mir-4522	chr17	22473817	22475817	60,70	0,76
hsa-mir-4733	chr17	26274664	26276664	66,10	0,69
hsa-mir-4734	chr17	33883890	33885890	61,23	0,86
hsa-mir-4738	chr17	70635763	70637763	57,600	0,67
hsa-mir-6080	chr17	59630809	59632809	70,60	0,70
hsa-mir-6781	chr17	38001573	38003573	60,60	0,68
hsa-mir-6787	chr17	77020334	77022334	68,40	0,67
hsa-mir-4741	chr18	17430572	17432572	64,60	0,92
hsa-mir-1181	chr19	10354946	10356946	65,800	0,73
hsa-mir-1199	chr19	14024862	14026862	61,96	0,69

hsa-mir-1470	chr19	15400226	15402226	60,50	0,66
hsa-mir-371a	chr19	51032773	51034773	57,00	0,65
hsa-mir-372	chr19	51032988	51034988	60,20	0,76
hsa-mir-4321	chr19	2190220	2192220	71,73	0,80
hsa-mir-4322	chr19	10179356	10181356	55,05	0,81
hsa-mir-4530	chr19	36641021	36643021	68,20	0,82
hsa-mir-4745	chr19	744597	746597	61,67	0,72
hsa-mir-4746	chr19	4383560	4385560	62,10	0,67
hsa-mir-4750	chr19	47133323	47135323	64,90	0,66
hsa-mir-4999	chr19	8343720	8345720	58,95	0,71
hsa-mir-639	chr19	14480412	14482412	59,60	0,72
hsa-mir-6789	chr19	2175429	2177429	57,70	0,73
hsa-mir-6790	chr19	6332618	6334618	58,00	0,67
hsa-mir-6791	chr19	6676639	6678639	64,90	0,68
hsa-mir-7108	chr19	2374497	2376497	61,40	0,69
hsa-mir-935	chr19	51227434	51229434	59,36	0,70
hsa-mir-1258	chr2	176054579	176056579	55,50	0,80
hsa-mir-1302-3	chr2	109768265	109770265	58,15	0,76
hsa-mir-149	chr2	88	2088	57,75	0,76
hsa-mir-153-1	chr2	215487766	215489766	65,80	0,65
hsa-mir-3131	chr2	62	2062	57,60	0,77
hsa-mir-375	chr2	215194819	215196819	61,08	0,66
hsa-mir-4444-1	chr2	173406375	173408375	61,10	0,96
hsa-mir-4784	chr2	127675430	127677430	63,83	0,73
hsa-mir-4785	chr2	156591648	156593648	56,400	0,67
hsa-mir-5001	chr2	228744381	228746381	56,600	0,66
hsa-mir-5703	chr2	223665948	223667948	56,40	0,87
hsa-mir-663b	chr2	128443205	128445205	59,52	0,65
hsa-mir-6810	chr2	69	2069	57,60	0,77
hsa-mir-7157	chr2	59	2059	57,55	0,77
hsa-mir-7704	chr2	172382503	172384503	60,600	0,80
hsa-mir-7845	chr2	98	2098	57,70	0,76
hsa-mir-933	chr2	171361412	171363412	62,50	0,69
hsa-mir-124-3	chr20	58451098	58453098	71,467	0,67
hsa-mir-1289-1	chr20	30733291	30735291	56,60	0,74
hsa-mir-1292	chr20	2573895	2575895	61,40	0,77
hsa-mir-663a	chr20	26129090	26131090	56,60	0,93
hsa-mir-6869	chr20	1313606	1315606	55,20	0,77
hsa-mir-3197	chr21	29789894	29791894	57,08	0,88
hsa-mir-3687-1	chr21	315008	317008	56,27	0,97
hsa-mir-3687-2	chr21	315008	317008	56,27	0,97
hsa-mir-1281	chr22	25148293	25150293	65,60	0,87
hsa-mir-1306	chr22	3873351	3875351	55,20	0,68
hsa-mir-3618	chr22	3873042	3875042	58,00	0,74
hsa-mir-658	chr22	21899131	21901131	56,00	0,90
hsa-mir-6816	chr22	3899892	3901892	62,33	0,68
hsa-mir-1226	chr3	47832365	47834365	66,20	0,66
hsa-mir-191	chr3	49000703	49002703	63,27	0,69

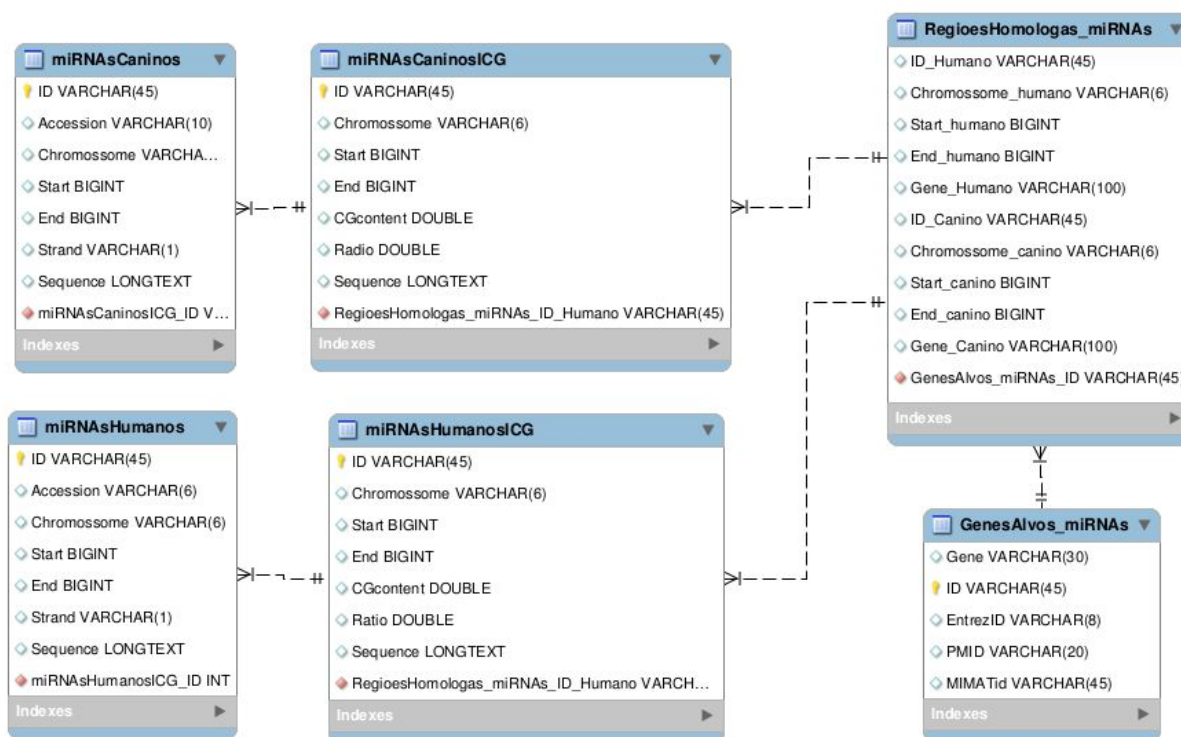
hsa-mir-425	chr3	49000228	49002228	65,40	0,65
hsa-mir-4442	chr3	25646421	25648421	59,40	0,66
hsa-mir-4787	chr3	50652703	50654703	65,40	0,66
hsa-mir-564	chr3	44843485	44845485	59,70	0,75
hsa-mir-6872	chr3	50250948	50252948	56,33	0,78
hsa-mir-4274	chr4	90	2090	63,23	0,95
hsa-mir-4449	chr4	50357837	50359837	56,80	0,80
hsa-mir-4454	chr4	54	2054	63,70	0,94
hsa-mir-572	chr4	11284123	11286123	61,00	1,01
hsa-mir-2277	chr5	89329203	89331203	55,200	0,86
hsa-mir-3661	chr5	129934117	129936117	56,24	0,69
hsa-mir-3912	chr5	167086288	167088288	55,56	0,80
hsa-mir-4281	chr5	172328925	172330925	59,48	0,68
hsa-mir-4635	chr5	1050351	1052351	60,75	0,70
hsa-mir-4638	chr5	176920575	176922575	60,35	0,65
hsa-mir-6831	chr5	136266584	136268584	60,400	0,70
hsa-mir-219a-1	chr6	33117599	33119599	60,200	0,69
hsa-mir-3939	chr6	164463940	164465940	60,17	0,88
hsa-mir-4466	chr6	154153097	154155097	55,80	0,88
hsa-mir-6720	chr6	1332751	1334751	63,467	0,67
hsa-mir-148a	chr7	25981208	25983208	57,37	0,70
hsa-mir-196b	chr7	27198899	27200899	63,50	0,66
hsa-mir-4651	chr7	72374678	72376678	68,20	0,71
hsa-mir-124-1	chr8	9766513	9768513	59,00	0,66
hsa-mir-3150b	chr8	92860369	92862369	55,80	0,73
hsa-mir-320a	chr8	22194219	22196219	57,200	0,65
hsa-mir-378d-2	chr8	91703530	91705530	57,47	0,78
hsa-mir-4469	chr8	42688324	42690324	58,47	1,02
hsa-mir-4664	chr8	141588185	141590185	60,27	0,66
hsa-mir-6850	chr8	142688791	142690791	56,57	0,71
hsa-mir-937	chr8	141668268	141670268	64,90	0,65
hsa-mir-939	chr8	142292624	142294624	64,40	0,68
hsa-mir-219a-2	chr9	111166910	111168910	56,75	0,70
hsa-mir-219b	chr9	111166913	111168913	56,75	0,70
hsa-mir-27b	chr9	77857290	77859290	56,40	0,66
hsa-mir-2861	chr9	110562824	110564824	57,533	0,70
hsa-mir-3152	chr9	73	2073	56,90	0,68
hsa-mir-3621	chr9	119825400	119827400	70,667	0,68
hsa-mir-3689c	chr9	71	2071	56,95	0,68
hsa-mir-3689d-1	chr9	73	2073	56,90	0,68
hsa-mir-3689f	chr9	65	2065	57,00	0,68
hsa-mir-3960	chr9	110562740	110564740	58,533	0,73
hsa-mir-4477b	chr9	80	2080	56,95	0,68
hsa-mir-4665	chr9	5998609	6000609	64,10	0,71
hsa-mir-4674	chr9	119202435	119204435	68,20	0,71
hsa-mir-548a	chr9	64	2064	57,00	0,68
hsa-mir-873	chr9	76	2076	56,95	0,68
hsa-mir-4767	chrX	6737038	6739038	56,533	0,75

hsa-mir-503	chrX	129882181	129884181	59,80	0,70
hsa-mir-6089-1	chrX	2196987	2198987	56,600	0,65
hsa-mir-718	chrX	149349846	149351846	72,20	0,69
hsa-mir-6089-2	chrY	2415605	2417605	56,600	0,65

Fonte: Autor

Os dados foram organizados em uma base de dados relacional, de forma que as tabelas podem ser inter-relacionadas e catalogadas de maneira mais otimizada, conforme apresentado na Figura 18, foi utilizado o *software MySQL Workbench 6.0* para elaboração das tabelas.

Figura 18- Modelo Entidade-Relacionamento(MER) do banco de dados mySQL




















Fonte: Autor

Após obter as posições dos microRNAs com ilhas de CpG próximas, no genoma canino e humano, procuramos por *motifs* e homologias dessas regiões em buscas de miRNAs e genes de interesse em comum.

Foram realizadas as análises dos *motifs* para as sequências caninas e humanas que possuem miRNAs e ICG, ou seja, as sequências apresentadas nas Tabela 5 e Tabela 6. Os resultados estão apresentados na Tabela 7 para as sequências humana e na Tabela 8 para as sequências canina geradas pelo *software Homer*.

Nessas tabelas pode-se evidenciar que as sequências consenso para os domínios comuns aos microRNAs. Estas regiões podem ser analisadas, posteriormente, com possíveis sítios de ligação de elementos de interação com microRNAs, porém, há necessidade de experimentos adicionais para testar estas possibilidades.

Tabela 7 - Domínios de Sequências miRNAs Humano geradas pelo *Homer*

Rank	Motif	P-value	log P-value	% de Alvos	% de Background	STD(Bg STD)	Best Match/Details
1		1e-24	-5.750e+01	57.14%	1.30%	588.2bp (379.3bp)	SeqBias: CA-repeat(0.907) More Information Similar Motifs Found
2		1e-24	-5.639e+01	60.71%	2.45%	538.1bp (311.7bp)	SeqBias: polyA-repeat(0.887) More Information Similar Motifs Found
3		1e-23	-5.300e+01	80.36%	12.56%	573.0bp (681.5bp)	Znf263(Zf)/K562-Znf263-ChIP-Seq(GSE31477)/Homer(0.762) More Information Similar Motifs Found
4		1e-22	-5.122e+01	87.50%	18.76%	594.4bp (712.0bp)	SeqBias: GCW-triplet(0.790) More Information Similar Motifs Found
5		1e-17	-4.101e+01	58.93%	6.19%	519.9bp (563.1bp)	PHD1/PHD1_BUT90/8-SUT1(Harison)/Yeast(0.758) More Information Similar Motifs Found
6		1e-16	-3.813e+01	44.64%	2.21%	528.4bp (705.5bp)	SKN7(MacIsaac)/Yeast(0.641) More Information Similar Motifs Found
7		1e-15	-3.535e+01	37.50%	1.00%	545.7bp (98.0bp)	STAT3/MA0144.2/Jaspar(0.759) More Information Similar Motifs Found
8		1e-15	-3.534e+01	64.29%	11.35%	557.5bp (446.2bp)	Trl(Zf)/S2-GAGA factor-ChIP-Seq(GSE40646)/Homer(0.901) More Information Similar Motifs Found
9		1e-15	-3.497e+01	51.79%	5.26%	529.9bp (726.7bp)	PB0097.1_Zfp281_1/Jaspar(0.852) More Information Similar Motifs Found
10		1e-14	-3.350e+01	69.64%	15.83%	528.8bp (606.1bp)	SKN7(MacIsaac)/Yeast(0.716) More Information Similar Motifs Found
11		1e-13	-3.204e+01	64.29%	12.98%	606.5bp (772.7bp)	ASH1/Literature(Harison)/Yeast(0.683) More Information Similar Motifs Found
12		1e-13	-3.177e+01	58.93%	10.22%	553.8bp (838.8bp)	Bcl6(Zf)/Liver-Bcl6-ChIP-Seq(GSE31578)/Homer(0.723) More Information Similar Motifs Found
13		1e-12	-2.820e+01	69.64%	19.50%	474.8bp (705.3bp)	che-1/MA0260.1/Jaspar(0.713) More Information Similar Motifs Found
14 *		1e-11	-2.541e+01	28.57%	1.02%	530.8bp (711.6bp)	pan/dmmpmm(SeSiMCMC)/fly(0.794) More Information Similar Motifs Found
15 *		1e-10	-2.522e+01	53.57%	11.04%	553.6bp (615.7bp)	SUT1?/SacCer-Promoters/Homer(0.589) More Information Similar Motifs Found
16 *		1e-10	-2.452e+01	57.14%	13.50%	544.0bp (515.7bp)	TDA9/MA0431.1/Jaspar(0.708) More Information Similar Motifs Found
17 *		1e-10	-2.404e+01	30.36%	1.68%	587.5bp (572.9bp)	Foxo1(Forkhead)/RAW-Foxo1-ChIP-Seq(Fan_et_al.)/Homer(0.763)

Total de sequencias alvos = 56

Total de sequencias de background = 270

* - possíveis falsos positivos

Fonte: Autor

Tabela 7 - Domínios de Sequências miRNAs Humano (Continuação)

18 *		1e-10	-2.363e+01	64.29%	19.22%	576.1bp (676.7bp)	PB0110.1_Bcl6b_2/Jaspar(0.781) More Information Similar Motifs Found
19 *		1e-9	-2.185e+01	26.79%	1.17%	515.4bp (0.0bp)	Zfp809(Zf)/ES-Zfp809-ChIP-Seq(GSE70799)/Homer(0.644) More Information Similar Motifs Found
20 *		1e-9	-2.183e+01	46.43%	8.91%	559.5bp (466.9bp)	CCA1/MA0972.1/Jaspar(0.828) More Information Similar Motifs Found
21 *		1e-9	-2.101e+01	67.86%	24.20%	544.1bp (769.6bp)	eve/dmmpmm(Bigfoot)/fly(0.678) More Information Similar Motifs Found
22 *		1e-8	-1.932e+01	57.14%	17.61%	560.0bp (718.4bp)	eor-1/MA0543.1/Jaspar(0.684) More Information Similar Motifs Found
23 *		1e-6	-1.601e+01	28.57%	4.07%	463.8bp (590.7bp)	Atf2(bZIP)/3T3L1-Atf2-ChIP-Seq(GSE56872)/Homer(0.727) More Information Similar Motifs Found
24 *		1e-4	-9.809e+00	23.21%	5.06%	676.6bp (512.2bp)	MAC1/MA0326.1/Jaspar(0.747) More Information Similar Motifs Found
25 *		1e-3	-8.998e+00	10.71%	0.70%	188.2bp (226.9bp)	RAP1/MA0359.1/Jaspar(0.755) More Information Similar Motifs Found
26 *		1e-3	-8.824e+00	48.21%	22.87%	464.4bp (732.1bp)	GCR2(MacIsaac)/Yeast(0.696) More Information Similar Motifs Found
27 *		1e-2	-6.494e+00	50.00%	28.47%	601.5bp (782.3bp)	ERF094/MA1049.1/Jaspar(0.888) More Information Similar Motifs Found
28 *		1e-2	-5.330e+00	5.36%	0.25%	338.5bp (0.0bp)	SKN7/SKN7_H2O2Lo[] (Harbison)/Yeast(0.658) More Information Similar Motifs Found

Fonte: Autor

Tabela 8 - Domínios de Sequências miRNAs Canina geradas pelo *Homer*

Rank	Motif	P-value	log P-value	% of Alvos	% of Background	STD(Bg STD)	Best Match/Details
1		1e-20	-4.709e+01	79.55%	9.13%	534.0bp (538.3bp)	Erra(NR)/HepG2-Erra-ChIP-Seq(GSE31477)/Homer(0.685) More Information Similar Motifs Found
2		1e-19	-4.527e+01	84.09%	12.87%	529.8bp (665.0bp)	SeqBias: polyA-repeat(0.882) More Information Similar Motifs Found
3		1e-17	-4.070e+01	59.09%	2.83%	539.1bp (949.8bp)	SeqBias: GA-repeat(0.844) More Information Similar Motifs Found
4		1e-16	-3.852e+01	75.00%	10.95%	471.9bp (536.3bp)	SeqBias: CA-repeat(0.962) More Information Similar Motifs Found
5		1e-15	-3.538e+01	59.09%	5.07%	488.6bp (761.3bp)	E2A(bHLH),near_PU.1/Bcell-PU.1-ChIP-Seq(GSE21512)/Homer(0.814) More Information Similar Motifs Found
6		1e-15	-3.531e+01	61.36%	6.03%	537.3bp (588.3bp)	E2F6/MA0471.1/Jaspar(0.784) More Information Similar Motifs Found
7		1e-14	-3.401e+01	70.45%	11.25%	606.5bp (658.9bp)	ABI4(1)/AP2/EREBP/Zeamays/AthaMap(0.751) More Information Similar Motifs Found
8		1e-14	-3.231e+01	40.91%	0.78%	514.9bp (502.5bp)	Eip74EF/dmmpmm(Bigfoot)/fly(0.797) More Information Similar Motifs Found
9		1e-12	-2.916e+01	52.27%	4.96%	571.0bp (338.9bp)	STAT3/MA0144.2/Jaspar(0.799) More Information Similar Motifs Found
10		1e-12	-2.870e+01	43.18%	1.97%	518.0bp (249.3bp)	cad/dmmpmm(Bergman)/fly(0.821) More Information Similar Motifs Found
11 *		1e-10	-2.482e+01	72.73%	19.34%	607.0bp (715.0bp)	SeqBias: GCW-triplet(0.819) More Information Similar Motifs Found
12 *		1e-10	-2.471e+01	43.18%	3.68%	564.7bp (569.8bp)	LIN54/MA0619.1/Jaspar(0.828) More Information Similar Motifs Found
13 *		1e-10	-2.429e+01	36.36%	1.70%	542.4bp (475.0bp)	ZNF416(Zf)/HEK293-ZNF416.GFP-ChIP-Seq(GSE58341)/Homer(0.663) More Information Similar Motifs Found
14 *		1e-10	-2.403e+01	31.82%	0.47%	512.8bp (0.0bp)	Smad4(MAD)/ESC-SMAD4-ChIP-Seq(GSE29422)/Homer(0.649) More Information Similar Motifs Found
15 *		1e-9	-2.258e+01	43.18%	4.30%	548.5bp (798.7bp)	EBF1(EBF)/Near-E2A-ChIP-Seq(GSE21512)/Homer(0.811) More Information Similar Motifs Found
16 *		1e-9	-2.224e+01	52.27%	8.84%	625.5bp (453.5bp)	SKN7(MacIsaac)/Yeast(0.735) More Information Similar Motifs Found
17 *		1e-9	-2.122e+01	50.00%	8.15%	650.8bp (466.7bp)	RDS1/MA0361.1/Jaspar(0.677) More Information Similar Motifs Found

Total de sequências alvos = 44

Total de sequências de background = 212

* - possível falso positivo

Fonte: Autor

Tabela 8 - Domínios de Sequências miRNAs Canina (Continuação)

18 *		1e-8	-2.023e+01	52.27%	10.34%	568.5bp (494.0bp)	Unknown4/Drosophila-Promoters/Homer(0.729) More Information Similar Motifs Found
19 *		1e-8	-1.902e+01	52.27%	10.87%	610.6bp (524.5bp)	vv1/dmmpmm(Down)/fly(0.759) More Information Similar Motifs Found
20 *		1e-7	-1.728e+01	29.55%	2.23%	393.4bp (248.6bp)	RIM101/MA0368.1/Jaspar(0.731) More Information Similar Motifs Found
21 *		1e-5	-1.182e+01	27.27%	3.78%	489.5bp (778.4bp)	HoxA2(Homeobox)/mES-Hoxa2-ChIP-Seq(Donaldson_et_al.)/Homer(0.743) More Information Similar Motifs Found
22 *		1e-3	-8.344e+00	45.45%	18.76%	369.6bp (704.7bp)	SKN7/SKN7_H2O2Lo[1](Harbison)/Yeast(0.698) More Information Similar Motifs Found
23 *		1e-2	-6.770e+00	15.91%	2.55%	255.4bp (592.1bp)	ttk/dmmpmm(Papatsenko)/fly(0.824) More Information Similar Motifs Found
24 *		1e-2	-6.234e+00	11.36%	1.16%	286.9bp (0.0bp)	che-1/MA0260.1/Jaspar(0.683) More Information Similar Motifs Found
25 *		1e-2	-4.724e+00	9.09%	1.24%	386.7bp (475.8bp)	SKN7/MA0381.1/Jaspar(0.694) More Information Similar Motifs Found
26 *		1e0	-1.914e+00	38.64%	29.50%	509.2bp (783.2bp)	Maz(Zf)/HepG2-Maz-ChIP-Seq(GSE31477)/Homer(0.974) More Information Similar Motifs Found

Fonte: Autor

As análises dos *motifs* para as duas espécies apresentaram alguns falsos positivos, porém a partir da coluna *best match/Details* podemos verificar a relação das regiões específicas associadas aos possíveis promotores, e também visualizar de forma ilustrada, através da coluna *logos*, a incidência de cada base na sequência.

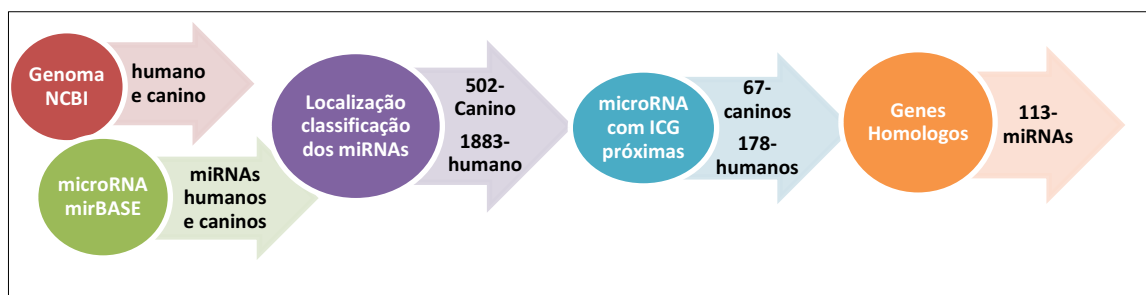
Após a análise dos *motifs*, foi realizada também a homologia dessas regiões entre as espécies, ou seja, através das posições em cada cromossomo das sequências de humanos descritos na Tabela 6, buscou-se a homologia na espécie canina.

Para isto utilizamos o LiftOver tools, no UCSC *Genome Bioinformatics Site* (<<https://genome.ucsc.edu/cgi-bin/hgLiftOver>>). Dessa maneira, foram submetidas as regiões caninas e obteve-se as posições dos genes homólogos em humanos.

Seguindo o fluxo das análises dos dados apresentados na Figura 19. Foram obtidos 67 microRNAs com ilhas de CpG próximas, para o genoma canino e 178 microRNAs com ilhas de CpG próximas, para o genoma humano.

Dentre os 113 miRNAs homólogos as duas espécies foi realizado uma análise das regiões no banco de dados do NCBI (<<https://www.ncbi.nlm.nih.gov/guide/genomes-maps/>>) afim de encontrar genes alvos, outros miRNAs e similaridades entre as regiões, conforme apresentado na Tabela 10.

Figura 19 - Fluxograma dos dados após análises



Fonte: Autor

Ao analisar os microRNAs com ilhas de CpG caninos e humanos, encontramos 8 microRNAs comuns às duas espécies: mir-210, mir-212, mir-1199, mir-425, mir-375, mir-132, mir-148a e mir-191. Esses miRNAs foram pesquisados na literatura e associados a alguns tipos de câncer, conforme descritos na Tabela 9 (BUZA; ARICK, 2014).

Tabela 9 - miRNAs associados aos tipos de câncer.

miRNAs	Tipo de Câncer
Mir-210	Câncer de esôfago, câncer de pulmão e câncer de mama.
Mir-212	Câncer de pâncreas e câncer de pulmão.
Mir-425	Câncer de mama, câncer de pâncreas e câncer de esôfago.
Mir-375	Câncer de ovário, câncer de laringe, câncer de pulmão, câncer de mama, câncer de próstata, câncer do colo retal, câncer de esôfago, câncer de boca, câncer cervical uterino, câncer de fígado e câncer da glândula salivar.
Mir-132	Câncer da bexiga urinária, câncer de estômago, câncer de pâncreas e câncer de pulmão.
Mir-148a	Câncer gástrico, câncer de colorretal, câncer de mama, câncer de estômago, câncer de ovário, câncer de pâncreas, câncer gastrointestinal, câncer de esôfago, câncer de pulmão e câncer de próstata.
Mir-132	Câncer de mama, câncer de estômago, câncer de tireoide, câncer de
Mir-191	próstata, câncer de ovário, câncer de pulmão e câncer de pâncreas.

Fonte: Adaptado de BUZA; ARICK (2014).

Os miRNAs mir-219 e mir-210 estão relacionados com o câncer de colo retal, sendo que o mir-219 possui como principal gene alvo o PDGFR α (XIONG; ZHANG, 2015). O mir-212 e o mir-132 tem como gene alvo o TMEM106B (BUZA; ARICK, 2014). O mir-212 também desempenha um papel importante na regulação do CREB/CRTC1, responsável por mecanismos ligados espinha dorsal (HOU; CUI, 2016).

Estudos apontam que o microRNA mir-191 está ligado ao câncer de ovário. Por meio da relação do microRNA e o TNF- α com DAPK1 constatou-se que pode contribuir para a transformação maligna da endometriose.

Na Tabela 10 apresentamos a análise realizada nas regiões de todos os miRNAs com ICGs humanos em comparação as regiões homólogas em cães.

Tabela 10- Análise dos genes e regiões dos miRNAs com ICG

Homo Sapiens					Canis Lupus				
chr	start	end	miRNA	GENE	chr	start	end	miRNA	GENE
chr11	505010	507010	hsa-mir-210	PHRF1	chr18	25613808	25614430	cfa-mir-210	PHRF1 LOC100684026/
chr14	82963104	82965104	hsa-mir-1247	DIO3/DIO30/LOC105370674	chr8	55079484	55080013	cfa-mir-8884	LOC102154023
chr16	32160309	32162309	hsa-mir-762	CTF1/MIR4519	chr3	31842247	31843872		NIPA1
chr16	2078110	2080110	hsa-mir-1225	TSC2/NTHL1/PKD1	chr6	38873763	38875815	cfa-mir-1842	TSC2/NTHL1/PKD1
chr16	1722958	1724958	hsa-mir-3177	MRPS34/NME3/HN1L	chr6	39164056	39164993		MRPS34/NME3/HN1L
chr17	1850331	1852331	hsa-mir-212	DPH1/HIC1/SMG6/MIR132	chr9	45985197	45986602	cfa-mir-212	DPH1/HIC1/SMG6/MIR132
chr17	69597571	69599571	hsa-mir-3615	SLC9A3R1/NAT9/RAB37	chr9	15998368	16000348		KCNJ2/MAP2K6
chr19	14022743	14024743	hsa-mir-1199	PALM3/SAMD1	chr20	48479380	48480180	cfa-mir-1199	PALM3/SAMD1
chr19	2188140	2190140	hsa-mir-4321	AMH/JSRP1/SF3A2	chr20	56840381	56841916	cfa-mir-8804	MYDGF/TNFAIP8L1/SEM
chr19	10179356	10181356	hsa-mir-4322	S1PR2/MRPL4/ICAM1/DNMT1	chr20	50889390	50890218	*possível mirna	S1PR2/MRPL4/ICAM1/DN
chr19	742535	744535	hsa-mir-4745	PTBP1/PLPR3/MIR3187	chr20	57850101	57851573	cfa-mir-8804	MYDGF/TNFAIP8L1/SEM
chr19	47131267	47133267	hsa-mir-4750	TBC1D17/AKT1S1/IL4I1/ATF5	chr1	108834918	108835298	*possível mirna	TBC1D17/AKT1S1/IL4I1/
chr2	128441090	128443090	hsa-mir-663b	ANKRD30BL/CDC27P1	chr19	22217095	22218918		LIMS2/
chr2	127673353	127675353	hsa-mir-4784	RHOQP2/TUBA3D/ MDVFAF3 /WDR6/P4HTM/DALRD3/	chr19	22937217	22939426		LIMS2/GPR17/MYO7B
chr3	48998141	49000141	hsa-mir-425	MIR191	chr20	40164997	40167064	cfa-mir-425	MDVFAF3/WDR6/P4HTM
chr5	172326863	172328863	hsa-mir-4281	CDHR2	chr4	39955643	39957614	*sem mirna	CDHR2
chr5	1050351	1052351	hsa-mir-4635	NKD2/SLC12A7/	chr34	11440842	11442203	cfa-mir-8790	NKD2/SLC12A7/
chr8	141588185	141590185	hsa-mir-4664	MAPk15/FAM83H/CCDC166	chr13	36014480	36016692	cfa-mir-151	PTK2/AGO2/CHRAC1
chr1	94017463	94019463	hsa-mir-760	DNTTIP2	chr6	55158262	55159993	*possível mirna	DNTTIP2
chr10	115551525	115553525	hsa-mir-3663	ENO4/SHTN1/VAX1	chr28	26197388	26199613		ATRNL1/GFRA1/CCDC17

chr10	85886083	85888083	hsa-mir-4678	CTSLP1/MINPP1	chr4	33742354	33744365		GRID1/CDHR1
chr11	60752191	60754191	hsa-mir-1237	CCDC88B	chr18	55497703	55499107	cfa-mir-192	CDC42BPG
chr11	58198927	58200927	hsa-mir-1908	FADS3/FADS1	chr18	38145261	38147837	cfa-mir-130a	CLP1/YPEL4
chr11	9050839	9052839	hsa-mir-5691	SCUBE2/ KRT8P41	chr21	32398838	32400503	cfa-mir-8810	RNF141/ LYVE1
chr14	82965240	82967240	hsa-mir-1247	LOC105370675/LOC105370674	chr8	55080014	55081469	cfa-mir-8884	LOC102153790/LOC102153791
chr15	82101024	82103024	hsa-mir-1302-10	FAM138E/WASH3P	chr3	55858592	55859228	cfa-mir-8898	HOMER2/ FSD2
chr16	2080200	2082200	hsa-mir-1225	PKD1/TSC2/NTHL1	chr6	38871890	38873672	cfa-mir-1842	PKD1/TSC2/NTHL1
chr16	59797720	59799720	hsa-mir-1538	NFAT5	chr5	88197668	88198112	*possível mirna	NFAT5
chr16	1725040	1727040	hsa-mir-3177	MAPK8IP3	chr6	39163473	39163980	cfa-mir-8877	MAPK3
chr16	2521955	2523955	hsa-mir-3178	CEMP1/AMDHD2/ mir-3179-1 /mir-3670-1/mir-3511-a1	chr6	38516980	38517516	cfa-mir-1842	BRICD5/MLST8
chr16	16428165	16430165	hsa-mir-3180-1	/mir-6770-1/ mir-3670-2 /mir-3179-2/mir-6511-a2	chr6	27686909	27688188		mir-365-1/mir-193b
chr16	16428163	16430163	hsa-mir-3180-2	/mir-6770-2	chr6	27686911	27688190		mir-365-1/mir-193b
chr17	1852069	1854069	hsa-mir-132	DPH1/OVCA2/HIC1/SMG6/mir-212	chr9	45986281	45989049	cfa-mir-132	DPH1/OVCA2/HIC1/SMG6
chr17	43565915	43567915	hsa-mir-196a-1	HOXB7/HOXB8/RPL9P28	chr9	19696147	19697251		MEOX1
chr17	1852441	1854441	hsa-mir-212	DPH1/OVCA2/HIC1/SMG6/mir-132	chr9	45986707	45989049	cfa-mir-212	DPH1/OVCA2/HIC1/SMG6
chr17	71587442	71589442	hsa-mir-636	MFSD11/JMJD6/SRSF2	chr9	17587495	17589109		NMT1/DCAKD/C1QL1
chr17	33883890	33885890	hsa-mir-4734	MLLT6/CISD3/mir-4726	chr9	39306022	39308054		CCL2/ASIC2
chr19	2190220	2192220	hsa-mir-4321	AMH/SF3A2/mir-6789	chr20	56839357	56840325	cfa-8802	AMH/SF3A2
chr19	744597	746597	hsa-mir-4745	PTBP1	chr20	57849308	57849763	cfa-8805	PTBP1
chr19	6332618	6334618	hsa-mir-6790	KHSRP/GTF2F1	chr20	53819064	53819979		DNMT1/STXBP2
chr2	215196883	215198883	hsa-mir-375	CRYBA2/FEV	chr37	22348034	22349297	cfa-mir-375	CRYBA2/FEV
chr2	128443205	128445205	hsa-mir-663b	RNA5-8SP5/CDC27P1	chr19	22215812	22216992		HS6ST1/UGGT1
chr2	176054579	176056579	hsa-mir-1258	CWC22/ZNF385B	chr36	19866113	19867346	cfa-mir-10b	HOXD3/HOXD8

chr2	109768265	109770265	hsa-mir-1302-3	FAM138B/WASH2P	chr10	42593615	42595787		RPL31/NPAS2
chr2	127675430	127677430	hsa-mir-4784	RHOQP2/TUBA3D	chr19	22934957	22937084		LIMS2/GPR17/WDR33
chr2	223665948	223667948	hsa-mir-5703	AGFG1/STIP1P2	chr37	29494742	29496966		SCG2/AP1S3
chr20	26129090	26131090	hsa-mir-663a	NCOR1P1	chr5	39691554	39692728		NCOR1
chr21	29789894	29791894	hsa-mir-3197	PLAC4/BACE2/ MDVFAF3 /WDR6/P4HTM/DALRD3/	chr31	24697158	24698539		CCT8/USP16
chr3	49000703	49002703	hsa-mir-191	MIR425/ MDVFAF3 /WDR6/P4HTM/DALRD3	chr20	40162571	40164606	cfa-mir-191	MDVFAF3/WDR6/P4HTM
chr3	49000228	49002228	hsa-mir-425	/MIR191	chr20	40162979	40164921	cfa-mir-425	MDVFAF3/WDR6/P4HTM
chr5	129934117	129936117	hsa-mir-3661	PPP2CA/RPS13P6	chr11	18646291	18649337		SLC27A6/ISOC1
chr5	167086288	167088288	hsa-mir-3912	NPM1/RPSAP71	chr4	44558791	44560590		TENM2/LOC102152300
chr5	172328925	172330925	hsa-mir-4281	NPM1/RPL10P8/FGF18	chr4	39954264	39955587	*possível mirna	NPM1/RPL26L1/FGF18
chr5	176920575	176922575	hsa-mir-4638	TRIM41/TRV-CAC1-5	chr4	36449288	36450291	cfa-mir-1271	ARL10/NOP16/HIGD2A
chr7	25981208	25983208	hsa-mir-148a	NFE2L3/LOC105375199	chr14	39331080	39332696	cfa-mir-148a	NFE2L3/LOC102154152
chr8	91703530	91705530	hsa-mir-378d-2	PDP1/LOC105375646	chr29	36976320	36977004		SLC26A7/LOC100685367
chr8	42688324	42690324	hsa-mir-4469	HOOK3/RNF170/THAP1	chr16	23056145	23056525		IKBKB/PLAT/AP3M2

Fonte: Autor

Considerando que cada gene possui vários miRNAs associados, buscou-se a interação entre esses genes e os miRNAs existentes para humanos e todos os genes humanos encontrados na Tabela 10 foram analisados através do algoritmo de predição de gene alvo de miRNA miRWalk (<<http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/index.html>>) (PRONINA & LOGINOV, 2016).

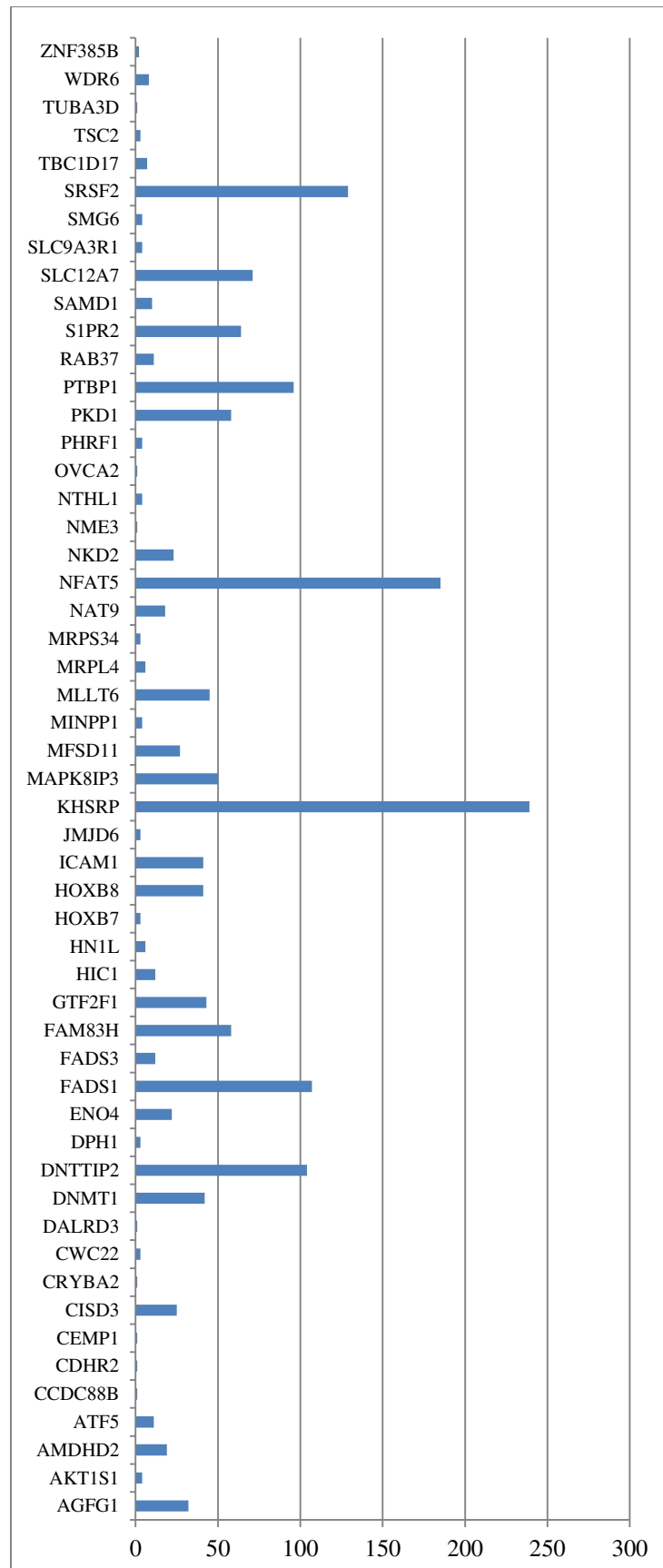
Foram encontrados ainda 5 regiões homólogas entre os genomas humano e canino, que possuem os mesmos genes, conforme apresentado na Tabela 11. Essas regiões apresentam uma grande similaridade, com exceção da presença de miRNAs no genoma canino. Essa grande similaridade pode indicar a existência de microRNAs caninos não catalogados. Essas regiões indicam a presença de microRNAs apenas no genoma humano, sendo estes: hsa-mir-4322, hsa-mir-4750, hsa-mir-760, hsa-mir-1538 e hsa-mir-4281.

Além do miRWalk, onde obtiveram-se os miRNAs associados a possíveis genes alvos, pesquisou-se os mesmos miRNAs no *Catalogue of Somatic Mutations in Cancer* (COSMIC) (<<http://cancer.sanger.ac.uk/cosmic>>), um banco de dados que associa os principais genes às mutações e também aos tipos de câncer.

Dessa forma, foram destacados os miRNAs com seus genes alvos e quais desses genes estão diretamente ligados ao câncer. Dentre os principais genes descritos na Figura 20, foram encontrados 3 genes no COSMIC, que estão diretamente ligados a mutações e ao câncer: SRSF2, TSC2 e MLLT6.

Segundo Luo e Cheng (2017), o gene SRSF2 regula as variantes da célula no controle das expressões associadas ao câncer e confirmou a abundante existência de *alternative splicing*, sendo um dos principais genes que contribuem para o desenvolvimento de cânceres. Já o gene *tuberous sclerosis complex* (TSC2), juntamente com o (TSC1), está associado ao desenvolvimento de tumores benignos e alguns tipos de lesões e inflamações e, em um pequeno número de casos, apresentam também associação com tumores malignos e câncer (MARTIN; ZHOU, 2017). Dados os efeitos seletivos sobre o crescimento oncogênico da proteína MLLT6, especula-se seu complexo modificador de histonas para um gene alvo de câncer e, mesmo não compreendendo detalhadamente seus mecanismos de atuação, o MLLT6 está associado ao câncer de alguma maneira (BERONJA; JANKI, 2013).

Figura 20 - Quantidade de miRNAs que interagem com genes alvos



Fonte: Autor

Tabela 11 - Possíveis miRNAs Caninos não catalogados

Humano		Canino
hsa-mir-4322	S1PR2/MRPL4/ICAM1/DNMT1	S1PR2/MRPL4/ICAM1/DNMT1
hsa-mir-4750	TBC1D17/AKT1S1/IL4I1/ATF5	TBC1D17/AKT1S1/IL4I1/ATF5
hsa-mir-760	DNTTIP2	DNTTIP2
hsa-mir-1538	NFAT5	NFAT5
hsa-mir-4281	NPM1/RPL10P8/FGF18	NPM1/RPL26L1/FGF18

Fonte: Autor

Utilizando o miRBase foram pesquisadas as regiões caninas onde se suspeita a existência de miRNAs não catalogados. O miRBase possui um sistema de pesquisa de alinhamento com sequências catalogadas e foi evidenciada a associação dessas sequências com miRNAs de outras espécies, conforme apresentado pela Tabela 12. O *score* quanto mais próximo de 100, maior a relação com o miRNA; como os *scores* ficaram todos acima de 75, acredita-se que possam existir miRNAs não catalogados.

Tabela 12 - Resultado cluster de regiões no mirBase

Região pesquisadas	Associação com miRNA	Score
S1PR2/MRPL4/ICAM1/DNMT1	dps-mir-2527 (<i>Drosophila</i>)	77
TBC1D17/AKT1S1/IL4I1/ATF5	hsa-mir-6738 (<i>Homo sapiens</i>)	78
DNTTIP2	eca-mir-1302 (<i>Equus caballus</i>)	91
NFAT5	hhi-mir-723 (<i>Hippoglossus</i>)	76
NPM1/RPL26L1/FGF18	rno-mir-346 (<i>Rattus norvegicus</i>)	78

Fonte: Autor

Os microRNAs e seus alvos genéticos são mais estudados em humanos, porém encontramos muitas semelhanças entre as espécies caninas e humanas. Desta forma, analisamos os genes alvos ortólogos em humanos para microRNAs e de outras espécies. A ferramenta computacional *target human scan* (<<http://www.targetscan.org/>>) lista os principais genes alvos para cada microRNA de outras espécies em humanos, conforme Tabela 13.

Nem todos os miRNAs possuem genes alvos associados. A Tabela 13 apresenta todos os miRNAs caninos que possuem genes alvos ortólogos em humanos.

Tabela 13 – Genes alvos dos miRNAs

miRNA	<i>chromosome</i>	<i>Human ortholog of target gene</i>	<i>Representative transcript</i>	<i>chromosome</i>
cfa-mir-574	chr03	SNRK	NM_001100594	chr03
cfa-mir-378	chr04	TOB2	NM_016272	chr22
cfa-mir-92b	chr07	RBFOX2	NM_001031695	chr22
cfa-mir-92b	chr07	RBFOX2	NM_001031695	chr22
cfa-mir-127	chr08	MAPK4	NM_002747	chr18
cfa-mir-203	chr08	ZNF281	NM_012482	chr01
cfa-mir-369	chr08	DENND4A	NM_001144823	chr15
cfa-mir-381	chr08	SENP6	NM_001100409	chr06
cfa-mir-409	chr08	KCTD9	NM_017634	chr08
cfa-mir-410	chr08	PTX3	NM_002852	chr03
cfa-mir-376a-2	chr08	SETD7	NM_030648	chr04
cfa-mir-376c	chr08	RNGTT	NM_003800	chr06
cfa-mir-381	chr08	SENP6	NM_001100409	chr06
cfa-mir-409	chr08	KCTD9	NM_017634	chr08
cfa-mir-433	chr08	ACTC1	NM_005159	chr15
cfa-mir-494	chr08	RTKN2	NM_145307	chr10
cfa-mir-132	chr09	CDK19	NM_015076	chr06
cfa-mir-196 ^a -1	chr09	HOXC8	NM_022658	chr12
cfa-mir-212	chr09	MGA	NM_001080541	chr15
cfa-mir-219-1	chr12	ZNF238	NM_006352	chr01
cfa-mir-219-1	chr12	ZNF238	NM_006352	chr01
cfa-mir-148a	chr14	SOS2	NM_006939	chr14
cfa-mir-129-2	chr18	FMR1	NM_001185075	chrX
cfa-mir-210	chr18	ATG7	NM_001136031	chr03
cfa-mir-129-2	chr18	FMR1	NM_001185075	chrX
cfa-mir-425	chr20	LDOC1L	NM_032287	chr22
cfa-mir-191	chr20	CHMP5	NM_001195536	chr09
cfa-mir-425	chr20	LDOC1L	NM_032287	chr22

cfa-mir-202	chr28	NAA15	NM_057175	chr04
cfa-mir-375	chr37	SLC16A2	NM_006517	chrX
cfa-mir-718	chrX	PURB	NM_033224	chr07

Fonte: Autor

Utilizando os dados obtidos com os miRNAs caninos e os genes alvos ortólogos em humanos, podemos observar uma grande semelhança entre as espécies e uma relação desses miRNAs com a patologia do câncer.

Os miRNAs 127, 132, 148a, 191 e 375 estão ligados ao câncer, como por exemplo o de mama. Tais miRNAs estão ligados as ilhas de CpG e ocorre a metilação nos promotores dessa região. No estudo de Pronina e Loginov (2016) todos os genes dos miRNAs possuem ilhas de CpG próximas e estavam ligados a algum tipo de câncer.

7 Conclusão

A pesquisa realizada até o presente momento revela possíveis miRNAs caninos não catalogados ainda, miRNAs comuns às duas espécies ligados ao câncer e a classificação de possíveis genes alvos desses miRNAs. Falta a realização de uma análise sobre os sítios de restrição de enzimas sensíveis à metilação nas ICGs dessas regiões identificadas e o desenvolvimento de interfaces gráficas, para auxiliar o usuário a realizar a mesma análise para qualquer espécie.

O trabalho realizado pode contribuir para o desenvolvimento de novas terapias de tratamento de doenças caninas, incluindo o câncer.

Referências Bibliográficas

- ABRINPET. **População de pets cresce 5% ao ano e Brasil é quarto no ranking mundial**, 2012. Disponível em: <<http://www.2pro.com.br/site/populacao-de-pets-cresce-5-ao-ano-e-brasil-e-quarto-no-ranking-mundial/>>
Acesso em: 06 abr 2015
- AMARAL, B. A; NONAKA, C. W. **MicroRNAs – Biogenesis, functions and its potential role in oral carcinogenesis**. *Odontol. Clín.-Cient*, p. 105-109, 2010.
- ANAND, D.; PANDEY, P. **Bioinformatics Challenges and Advances in RNA interference**. *Bioinformatics Review*, p. 1-3, 2017.
- ARAÚJO, N. D.; FARIAS, R. P. **A era da bioinformática: seu potencial e suas implicações para as ciências da saúde**. *Estud Biol.*, p. 143-148, 2008.
- BABENKO, V.; CHADAEVA, I. V. **Genomic landscape of CpG rich elements in human**. *BMC Evolutionary Biology*. 2017.
- Ballestar, E. (2011). **An introduction to epigenetics**. *Advances in Experimental Medicine and Biology*, p. 711, 2011.
- BERONJA, S.; JANKI, P. **RNAi screens in mice identify physiological regulators of oncogenic growth**. *Nature*, 185-190, 2013.
- BINO, J.; ENRIGHT, A. J. **Human MicroRNA Targets**. *PLoS Biol*, 2004.
- BOEVA, V. **Analysis of Genomic Sequence Motifs for Deciphering Transcription Factor Binding and Transcriptional Regulation in Eukaryotic Cells**. *Frontiers in Genetics*, 2016.
- BUZA, T.; ARICK, M. **Computational prediction of disease microRNAs in domestic animals**. *BMC Research Notes*, 2014.
- CAIAFA, P.; ZAMPIERI, M. **DNA methylation and chromatin structure: the puzzling CpG islands**. *Journal of Cellular Biochemistry*, p. 257-265, 2005.
- CHAMORRO, A. C., AMBROSIO-ALBUQUERQUE, E. P.; ROSA, V. A. **MICRO-RNAs E CÂNCER: ABORDAGENS E PERSPECTIVAS**. *Rev Pesq Saúde*, p. 119-124, 2015.
- CHUANG, L. Y.; YANG, C. H. **CpGPAP: CpG island predictor analysis platform**. *BMC Genetics*, 2015.
- COCHRANE, G.; KARSCH-MIZRACHI, I. **The International Nucleotide Sequence Database Collaboration**. *Nucleic Acids Research*, p. 48-50, 2016.
- COELHO, F. C. **Computação Científica com Python - Uma introdução à programação para cientistas**. Petropolis, RJ: Edição do Autor, 2007.

COSTA, E. B.; PACHECO, C. **Epigenética: regulação da expressão gênica em nível transcricional e suas implicações.** *Semina: Ciências Biológicas e da Saúde*, pp. 125-136, 2013.

CUI, W.; WANG, L. H. **Expression and regulation mechanisms of Sonic Hedgehog in breast cancer.** *Cancer Science*, pp. 927-933, 2010.

D'HAESELEER, P. **What are DNA sequence motifs?** *Nature Biotechnology*, 423-425, 2006.

DINIZ, W. J.; CANDURI, F. **Bioinformatics: an overview and its applications.** *Genetics and Molecular Research*, 2017.

EIDHAMMER, I.; JONASSEN, I. **Structure Comparison and Structure Patterns.** *Journal of Computational Biology*, p. 685-716, 2004.

GALM O.; HERMAN G.; BAYLIN B. **The fundamental role of epigenetics in hematopoietic malignancies,** *Blood Reviews*, 1-13, 2006.

GARDINER-GARDEN M.; FROMMER M. **CpG Islands in vertebrate genomes.** *Journal of Molecular Biology*, 261-282, 1987

GLADE. **A User Interface Designer**, 2014, disponível em: <<https://glade.gnome.org/>>
Acesso em: 24 Jul 2016

GRIFFITHS-JONES, S.; SAINI, H. K. **miRBase: tools for microRNA genomics.** *Nucleic Acids Research*, p. 154–158, 2008.

HAN, L.; WITMER, D. **DNA methylation regulates microRNA expression.** *Cancer Biology; Therapy*, pp. 1290-1294, 2007.

HAN, L.; ZHAO, Z. **Contrast features of CpG islands in the promoter and other regions in the dog genome.** *Genomics*, pp. 117-124, 2009.

HEINZ, S. **Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities,** *Molecular Cell*, 2010.

HERMAN, S. H.; BAYLIN, S. B. **Gene silencing in cancer is association with promoter hypermethylation.** *The New England Journal of Medicine*, 2042-2054, 2003.

HORTAI, R. S.; COSTA, M. P. **Fatores prognósticos e preditivos dos tumores caninos definidos com auxílio da imuno-histoquímica.** *Ciência Rural*, 1033 - 1039, 2012.

HOU, B.; CUI, X. **Positive feedback regulation between microRNA-132 and CREB in spinal cord contributes to bone cancer pain in mice.** *European Journal of Pain*, 2016.

HUANG, Y.; JIA, X. **Biological functions of microRNAs: a review.** *J Physiol Biochem*, pp. 129–139, 2011.

IBGE. **Pesquisa Nacional de Saúde**. Instituto Brasileiro de Geografia e Estatística , pp. 25-26, 2015.

ILLINGWORTH, R. S.; BIRD, A. S. **CpG islands – ‘A rough guide’**. *FEBS Letters*, pp. 1713-1720, 2009.

INCA. **Estimativa 2016: Incidência de Câncer no Brasil**. Instituto Nacional de Câncer José Alencar Gomes da Silva, Rio de Janeiro, 2015.

JANE, M. D. **Breed-Predispositions to Cancer in Pedigree Dogs**. *Hindawi Publishing Corporation*, 23, 2013.

JEFFREY, S. S. **Cancer biomarker profiling with microRNAs**. *Nature Biotechnology*, 400-401, 2008.

JING , F.; ZHANG , J. **Hypermethylation of tumor suppressor genes BRCA1, p16 and 14-3-3sigma in serum of sporadic breast cancer patients**. *Onkologie*, pp. 14-19, 2007.

JINGFANG, J.; JIAN, L. **The emerging role of miR-506 in cancer**. *Impact Journals*, 2016.

KAMINSKAS, E.; FARRELL, T. **FDA Drug Approval Summary: Azacitidine (5-azacytidine, Vidaza™) for Injectable Suspension**. *The Oncologist*, pp. 176-182, 2005.

KIEFER, J. C. **Epigenetics in development**. *Developmental Dynamics*, 1144-1156, 2007.

Kinney, S. R.; Pradhan, S. **Regulation of expression and activity of DNA (cytosine-5) methyltransferases in mammalian cells**. *Prog Mol Biol Transl Sci*, pp. 311-333, 2011.

KOZOMARA, A.; GRIFFITHS-JONES, S. **miRBase: integrating microRNA annotation and deep-sequencing data**. *Nucleic Acid Research*, 152-157, 2010.

KOZOMARA, A.; GRIFFITHS-JONES, S. **miRBase: annotating high confidence microRNAs using deep sequencing data**. *Nucleic Acids Research*, D68-D73, 2014.

LAGES, E.; IPAS, H. **MicroRNAs: molecular features and role in cancer**. *HAL Arquivos Ouvertes - France*, 2012.

LIANGFENG, H.; WITMER, D. P. **DNA Methylation Regulates MicroRNA Expression**. *Cancer Biology; Therapy*, 1290-1294, 2007.

LIDDLE, R. A.; JIRTLE, R. L. **Epigenetic silencing of genes in human colon cancer**, *Gastroenterology*, 960-962, 2006.

LIMA, A. M. Tese programa de bioinformática USP. **Predição de RNAs não codificantes e sua aplicação na busca do componente RNA da telomerase**, São Paulo, 2006.

LONG, M. D.; SMIRAGLIA, D. J. **The Genomic Impact of DNA CpG Methylation on Gene Expression Relationships in Prostate Cancer**. *biomolecules*, 2017.

LOPES, K. R.; SILVA, A. **Considerações sobre a importância do cão doméstico (*Canis lupus familiaris*) dentro da sociedade humana.** *Acta Veterinaria Brasilica*, pp. 177 - 185, 2012.

LUO, C.; CHENG, Y. **SRSF2 regulates alternative splicing to drive hepatocellular carcinoma development.** *Cancer Research*, 1168-1178, 2017.

MANN, B. S.; JOHNSON, J. R. **FDA Approval Summary: Vorinostat for Treatment of Advanced.** *The Oncologist*, pp. 1247-1252, 2007.

MARTIN, J. D.; ANDERS, L. H. **MicroRNA and cancer.** *Molecular Oncology* 6, 590-610, 2012.

MARTIN, K.; ZHOU, W. **The genomic landscape of tuberous sclerosis complex,** *Nature Research Journal*, 2017.

MINÁROVITS, J. **Microbe-induced epigenetic alterations.** *Acta Microbiologica et Immunologica Hungarica*, pp. 1-19, 2009.

MOLLOY, K.; MCKIERNAN, P. **X chromosome encoded microRNAs are functionally increased in cystic fibrosis monocytes,** *ATS Journal*, 2017.

NADINE, R.; HELLMUTH-ALEXANDER, M. **miRNA Profiling Identifies Candidate miRNAs for Bladder Cancer Diagnosis and Clinical Outcome,** *The Journal of Molecular Diagnostics*, 695-705, 2013.

PAOLONI, M.; KHANNA, C. **Translation of new cancer treatments from pet dogs to humans,** *Nature Reviews Cancer*, 147-156, 2008.

PAULSEN, M.; FERGUSON-SMITH, A. C. **DNA methylation in genomic imprinting, development, and disease,** *The Journal of Pathology*, 2001.

PAVICIC, W.; PERKIÖ, E. **Altered Methylation at MicroRNA-Associated CpG Islands in Hereditary and Sporadic Carcinomas: A Methylation-Specific Multiplex Ligation-Dependent Probe Amplification (MS-MLPA)-Based Approach,** *Molecular Medicine*, 726-735, 2011.

PEVSNER, M. **Bioinformatics and functional genomics.** Chichester: 3rd ed. John Wiley, 2015.

PRONINA, I. V. **DNA methylation contributes to deregulation of 12 cancer-associated microRNAs and breast cancer progression,** *Elsevier Gene*, p. 1-8, 2017.

RICHARDS, E. J. **Inherited epigenetic variation: revisiting soft inheritance,** *Nature Reviews:Genetics*, p. 395-401, 2006.

SATO, F., TSUCHIYA, S.; MELTZER, S. J. **MicroRNAs and epigenetics,** *FEBS Journal*, 2011.

TALBERT, P. B.; HENIKOFF, S. **Spreading of silent chromatin: inaction at a distance**, *Nature Reviews: Genetics*, 793 - 803, 2006.

VINCENT, K.; PICHLER, M. **MicroRNAs, Genomic Instability and Cancer**, *Molecular Sciences*, 14475-14491, 2014.

WAGGONER, D. **Mechanisms of disease: epigenesis**, *Seminars in Pediatric Neurology*, 7-14, 2007.

WEIDMAN, J. R.; DOLINOY, D. C. **Cancer susceptibility: epigenetic manifestation of environmental exposures**. *The Cancer Journal*, 9-16, 2007.

WILSON, A. S.; POWER, B. E. **DNA hypomethylation and human diseases**. *Biochimica et Biophysica Acta*, 138-162, 2007.

Wolffe, A.; Guschin, D. **Review: Chromatin Structural Features and Targets That Regulate Transcription**. *Journal of Structural Biology*, 102-122, 2000.

XIONG, G. B.; ZHANG, G. N. **MicroRNA-219-5p functions as a tumor suppressor partially by targeting platelet-derived growth factor receptor alpha in colorectal cancer**. *Neoplasma*, 2015.

Apêndice

Apêndice A – CÓDIGO-FONTE DO SOFTWARE DESENVOLVIDO (miRNA_CG Finder)

```

1  #!/usr/bin/env python
2  # -*- coding:utf-8 -*-
3
4  import csv
5  import os
6  from gi.repository import Gtk, Gdk
7  import hashlib
8  import base64
9
10 class Software(Gtk.Window):
11
12     def __init__(self):
13         builder = Gtk.Builder()
14         builder.add_from_file(os.getcwd() + "/gui.glade")
15         window = builder.get_object("window")
16         window.set_title("CpG Island and miRNA finder")
17
18         #objetos de entrada/saida de dados
19         self.dna = builder.get_object("localDna")
20         self.mirna = builder.get_object("localMirna")
21         self.distancia = builder.get_object("distICG")
22         self.cgcontent = builder.get_object("CGcontent")
23         self.ratio = builder.get_object("ratio")
24         self.tamanho = builder.get_object("tamanho")
25         self.treeview = builder.get_object("treeview")
26         self.liststore = builder.get_object("liststore")
27
28         #Chama as funções
29         builder.connect_signals({
30             "on_btn_procurar_clicked":self.calculaTudo
31         })
32
33         #Monta a Interface
34         window.show()
35         window.connect("delete-event", Gtk.main_quit)
36
37
38     def calculaTudo(self, widget):
39         z=0
40         listaTemporaria=[]
41         listaTemporaria.append("ID; Chromossome; SEQ antes; SEQ depois; Start;      ↵
42         End\n")
43         cr = csv.reader(open(self.mirna.get_filename(),"rb"))
44         for row in cr:
45             x = row[0].split(";")
46             if x[2] != "Chromosome":
47                 arquivo = self.dna.get_filename()+"/cfa_ref_CanFam3.1_"+str(x[2])+      ↵
48                 ".fa"
49                 dnaL, dnaM = self.ler_dna(arquivo)
50                 p1 = self.seq_mirna_rna_dna(x[7])
51                 p4 = self.rna_para_dna(x[7])[::-1]
52                 if x[5] == "+":
53                     se, sd, pos = self.esq_dir_dna(dnaL, p1, int(self.distancia.      ↵
54                     get_text()))
55                     linha = str(x[0])+";"+str(x[2])+";"+se+";"+sd+";"+str(pos)+";"+      ↵
56                     str(pos+len(p1))+"\n"
57                     print pos

```

```

54         else:
55             se, sd, pos = self.esq_dir_dna(dnaL, p4, int(self.distancia.
56                 get_text()))
57             linha = str(x[0])+";" +str(x[2])+";" +se+";" +sd+";" +str(pos)+";" +
58                 str(pos+len(p4))+"\n"
59             print pos
60             listaTemporaria.append(linha)
61             z+=1
62             print str(x[2])
63         fid = open("Resultado"+str(self.distancia.get_text())+".csv", 'w')
64         fid.writelines(listaTemporaria)
65         fid.close()
66         self.ilhaCpG(self.distancia.get_text())
67         return 1
68
69 def posicao(self, seq):
70     """Gera 4 posições para pesquisa"""
71     pos1 = seq_mirna_rna_dna(seq)
72     pos4 = rna_para_dna(seq)[::-1]
73     return pos1, pos4
74
75 def ler_dna(self, nome_arq):
76     """ Retorna o DNA do cromossomo """
77     dna_M = []
78     arq = open(nome_arq, "r")
79     for linha in arq:
80         if linha[0] != ">":
81             dna_M.append(linha[::-1])
82     dna_L = "".join(dna_M)
83     return dna_L, dna_M
84
85 def seq_mirna_rna_dna(self, seq_rna):
86     """Retorna a sequencia trocando U por T"""
87     seq_dna = seq_rna.replace("U", "T")
88     return seq_dna
89
90 def rna_para_dna(self, rna):
91     """ Retorna o DNA """
92     dna = rna.replace("A", "T")
93     dna = dna.replace("U", "A")
94     dna = dna.replace("C", "X")
95     dna = dna.replace("G", "C")
96     dna = dna.replace("X", "G")
97     return dna
98
99 def dna_para_rna(self, dna):
100     """ Retorna o RNA """
101     rna = dna.replace("A", "U")
102     rna = rna.replace("T", "A")
103
104     rna = rna.replace("C", "X")
105     rna = rna.replace("G", "C")
106     rna = rna.replace("X", "G")
107     return rna
108
109 def esq_dir_dna(self, cromossomo, seq_dna, delta):
110     """ Retorna o número delta de pb antes e depois do mirna """

```



```

109         pos = cromossomo.find(seq_dna)
110         return self.__esq_dir(cromossomo, pos, len(seq_dna), delta)
111
112
113     def esq_dir_rna(self, cromossomo, seq_rna, delta):
114         """Retorna o número delta de pb antes e depois do mirna """
115         dna = rna_para_dna(seq_rna)
116         pos = cromossomo.find(dna[::-1])
117         return self.__esq_dir(cromossomo, pos, len(seq_rna), delta)
118
119
120     def __esq_dir(self, fita, pos, tam_seq, delta):
121         """ """
122         seq_esq = fita[pos-delta: pos]
123         seq_dir = fita[pos+tam_seq: pos+tam_seq+delta]
124         return seq_esq, seq_dir, pos
125
126     def ilhaCpG(self, qtidadePb):
127         x=0
128         y=0
129         cr = csv.reader(open("Resultado"+str(self.distancia.get_text())+".csv", "rb"))
130         listaTemporariaAntes=[]
131         listaTemporariaAntes.append("ID_Antes; Chromosome; Start; End; %CG;
Ratio; Sequencia\n")
132         listaTemporariaDepois=[]
133         listaTemporariaDepois.append("ID_Depois; Chromosome; Start; End; %CG;
Ratio; Sequencia\n")
134         for seq_dna in cr:
135             seq_dna = seq_dna[0].split(";")
136             if seq_dna[3] != " SEQ depois":
137                 tamanho = len(seq_dna[3])
138                 num_c = seq_dna[3].count("C")#numero de citosina
139                 num_g = seq_dna[3].count("G")#numero de guanina
140                 gc_content = 100.0 * (float(num_c + num_g)) / tamanho
141                 #print "The GC content of your sequence is
{0:.1f}%".format(gc_content)
142                 expected = float(num_c * num_g) / tamanho
143                 if expected != 0:
144                     ratio = float(seq_dna[3].count("CG")) / expected
145                 else:
146                     ratio = 0
147                 if gc_content > float(self.cgcontent.get_text()) and tamanho >= int
(self.tamanho.get_text()) and ratio > float(self.ratio.get_text()):
148                     linha = str(seq_dna[0])+";"+str(seq_dna[1])+";"+str(seq_dna[5
])+";"+str(int(seq_dna[5])+2000)+";"+str(gc_content)+";"+str(
ratio)+";"+str(seq_dna[3])+"\n"
149                     self.treeview = self.liststore.append(linha)
150                     listaTemporariaDepois.append(linha)
151                     x += 1
152                 if seq_dna[2] != " SEQ antes":
153                     tamanho = len(seq_dna[2])
154                     num_c = seq_dna[2].count("C")#numero de citosina
155                     num_g = seq_dna[2].count("G")#numero de guanina
156                     gc_content = 100.0 * (float(num_c + num_g)) / tamanho
157                     #print "The GC content of your sequence is
{0:.1f}%".format(gc_content)
158                     expected = float(num_c * num_g) / tamanho

```



```

159         ratio = float(seq_dna[3].count("CG")) / expected
160         if gc_content > float(self.cgcontent.get_text()) and tamanho >= int
            (self.tamanho.get_text()) and ratio > float(self.ratio.get_text()):
            #parametros
161             linha = str(seq_dna[0])+";"+str(seq_dna[1])+";"+str(int(seq_dna
                [4])-2000)+";"+str(seq_dna[4])+";"+str(gc_content)+";"+str(
                ratio)+";"+str(seq_dna[2])+"\n"
162             self.treeview = self.liststore.append(linha)
163             listaTemporariaAntes.append(linha)
164             y += 1
165         #else:
166             #print "This is not a CpG island.""
167         self.treeview.show()
168         fid = open("ilhasCpG"+self.distancia.get_text()+".csv", 'w')
169         fid.writelines(listaTemporariaAntes)
170         fid.writelines(listaTemporariaDepois)
171         fid.close()
172
173     if __name__ == "__main__":
174         app = Software()
175         Gtk.main()
176

```

Apêndice B – CÓDIGO-FONTE DA INTERFACE GRÁFICA DO SOFTWARE
DESENVOLVIDO (miRNA_CG Finder)

```

1  <?xml version="1.0" encoding="UTF-8"?>
2  <!-- Generated with glade 3.16.1 -->
3  <interface>
4    <requires lib="gtk+" version="3.10"/>
5    <object class="GtkListStore" id="liststore">
6      <columns>
7        <!-- column-name ID -->
8        <column type="&lt; defina uma nova coluna &gt;"/>
9        <!-- column-name Chromosome -->
10       <column type="&lt; defina uma nova coluna &gt;"/>
11       <!-- column-name Start -->
12       <column type="&lt; defina uma nova coluna &gt;"/>
13       <!-- column-name End -->
14       <column type="&lt; defina uma nova coluna &gt;"/>
15       <!-- column-name %CG -->
16       <column type="&lt; defina uma nova coluna &gt;"/>
17       <!-- column-name Ratio -->
18       <column type="&lt; defina uma nova coluna &gt;"/>
19       <!-- column-name Sequence -->
20       <column type="&lt; defina uma nova coluna &gt;"/>
21     </columns>
22     <data>
23       <row/>
24       <row/>
25     </data>
26   </object>
27   <object class="GtkWindow" id="window">
28     <property name="can_focus">False</property>
29     <property name="margin_top">3</property>
30     <property name="window_position">center</property>
31     <property name="gravity">center</property>
32     <child>
33       <object class="GtkBox" id="box1">
34         <property name="visible">True</property>
35         <property name="can_focus">False</property>
36         <property name="halign">center</property>
37         <property name="orientation">vertical</property>
38         <child>
39           <object class="GtkLabel" id="label1">
40             <property name="visible">True</property>
41             <property name="can_focus">False</property>
42             <property name="label" translatable="yes">
43 Analysis of miRNAs and CpG islands
44 </property>
45             <attributes>
46               <attribute name="font-desc" value="Aroania Bold 12"/>
47               <attribute name="style" value="normal"/>
48               <attribute name="underline" value="True"/>
49               <attribute name="foreground" value="#000000000000"/>
50             </attributes>
51           </object>
52           <packing>
53             <property name="expand">False</property>
54             <property name="fill">True</property>
55             <property name="position">0</property>
56           </packing>
57         </child>

```

```

58     <child>
59         <object class="GtkBox" id="box2">
60             <property name="visible">True</property>
61             <property name="can_focus">False</property>
62             <property name="homogeneous">True</property>
63             <child>
64                 <object class="GtkBox" id="box3">
65                     <property name="visible">True</property>
66                     <property name="can_focus">False</property>
67                     <property name="orientation">vertical</property>
68                     <property name="homogeneous">True</property>
69                     <child>
70                         <object class="GtkLabel" id="label4">
71                             <property name="visible">True</property>
72                             <property name="can_focus">False</property>
73                             <property name="margin_top">13</property>
74                             <property name="label" translatable="yes">DNA Folder:
75     </property>
76         <attributes>
77             <attribute name="font-desc" value="&lt;Inserir valor&gt; 12"/>
78         </attributes>
79     </object>
80     <packing>
81         <property name="expand">False</property>
82         <property name="fill">True</property>
83         <property name="position">0</property>
84     </packing>
85 </child>
86 <child>
87     <object class="GtkLabel" id="label5">
88         <property name="visible">True</property>
89         <property name="can_focus">False</property>
90         <property name="label" translatable="yes">miRNAs File:
91     </property>
92     <attributes>
93         <attribute name="font-desc" value="&lt;Inserir valor&gt; 12"/>
94     </attributes>
95 </object>
96 <packing>
97     <property name="expand">False</property>
98     <property name="fill">True</property>
99     <property name="position">1</property>
100 </packing>
101 </child>
102 </object>
103 <packing>
104     <property name="expand">False</property>
105     <property name="fill">True</property>
106     <property name="position">0</property>
107 </packing>
108 </child>
109 <child>
110     <object class="GtkBox" id="box4">
111         <property name="visible">True</property>
112         <property name="can_focus">False</property>
113         <property name="orientation">vertical</property>
114         <property name="homogeneous">True</property>

```



```

114     <child>
115         <object class="GtkFileChooserButton" id="localDna">
116             <property name="visible">True</property>
117             <property name="can_focus">False</property>
118             <property name="action">select-folder</property>
119         </object>
120         <packing>
121             <property name="expand">False</property>
122             <property name="fill">True</property>
123             <property name="position">0</property>
124         </packing>
125     </child>
126     <child>
127         <object class="GtkFileChooserButton" id="localMirna">
128             <property name="visible">True</property>
129             <property name="can_focus">False</property>
130         </object>
131         <packing>
132             <property name="expand">False</property>
133             <property name="fill">True</property>
134             <property name="position">1</property>
135         </packing>
136     </child>
137 </object>
138 <packing>
139     <property name="expand">False</property>
140     <property name="fill">True</property>
141     <property name="position">1</property>
142 </packing>
143 </child>
144 </object>
145 <packing>
146     <property name="expand">False</property>
147     <property name="fill">True</property>
148     <property name="position">1</property>
149 </packing>
150 </child>
151 <child>
152     <object class="GtkBox" id="box5">
153         <property name="visible">True</property>
154         <property name="can_focus">False</property>
155         <property name="orientation">vertical</property>
156         <property name="homogeneous">True</property>
157         <child>
158             <object class="GtkBox" id="box7">
159                 <property name="visible">True</property>
160                 <property name="can_focus">False</property>
161                 <property name="orientation">vertical</property>
162                 <child>
163                     <object class="GtkLabel" id="label6">
164                         <property name="visible">True</property>
165                         <property name="can_focus">False</property>
166                         <property name="label" translatable="yes">Search ICGs near
167                             miRNAs (Base Pairs):</property>
168                         <attributes>
169                             <attribute name="font-desc" value="&lt;Inserir valor&gt; 12"/>

```

```

170         </object>
171     <packing>
172         <property name="expand">False</property>
173         <property name="fill">True</property>
174         <property name="position">0</property>
175     </packing>
176 </child>
177 <child>
178     <object class="GtkEntry" id="distICG">
179         <property name="width_request">50</property>
180         <property name="visible">True</property>
181         <property name="can_focus">True</property>
182         <property name="hexpand">False</property>
183         <property name="text" translatable="yes">2000</property>
184     </object>
185     <packing>
186         <property name="expand">False</property>
187         <property name="fill">True</property>
188         <property name="position">1</property>
189     </packing>
190 </child>
191 </object>
192 <packing>
193     <property name="expand">False</property>
194     <property name="fill">True</property>
195     <property name="position">0</property>
196 </packing>
197 </child>
198 <child>
199     <object class="GtkBox" id="box8">
200         <property name="visible">True</property>
201         <property name="can_focus">False</property>
202         <property name="orientation">vertical</property>
203     <child>
204         <object class="GtkLabel" id="label7">
205             <property name="visible">True</property>
206             <property name="can_focus">False</property>
207             <property name="label" translatable="yes">CG Content: </property>
208         </object>
209         <packing>
210             <property name="expand">False</property>
211             <property name="fill">True</property>
212             <property name="position">0</property>
213         </packing>
214     </child>
215     <child>
216         <object class="GtkEntry" id="CGcontent">
217             <property name="width_request">50</property>
218             <property name="visible">True</property>
219             <property name="can_focus">True</property>
220             <property name="text" translatable="yes">55.0</property>
221         </object>
222         <packing>
223             <property name="expand">False</property>
224             <property name="fill">True</property>
225             <property name="position">1</property>
226         </packing>

```

```

227     </child>
228 </object>
229 <packing>
230   <property name="expand">False</property>
231   <property name="fill">True</property>
232   <property name="position">1</property>
233 </packing>
234 </child>
235 <child>
236   <object class="GtkBox" id="box9">
237     <property name="visible">True</property>
238     <property name="can_focus">False</property>
239     <property name="orientation">vertical</property>
240     <child>
241       <object class="GtkLabel" id="label8">
242         <property name="visible">True</property>
243         <property name="can_focus">False</property>
244         <property name="label" translatable="yes">Ratio: </property>
245       </object>
246       <packing>
247         <property name="expand">False</property>
248         <property name="fill">True</property>
249         <property name="position">0</property>
250       </packing>
251     </child>
252     <child>
253       <object class="GtkEntry" id="ratio">
254         <property name="width_request">50</property>
255         <property name="visible">True</property>
256         <property name="can_focus">True</property>
257         <property name="text" translatable="yes">0.65</property>
258       </object>
259       <packing>
260         <property name="expand">False</property>
261         <property name="fill">True</property>
262         <property name="position">1</property>
263       </packing>
264     </child>
265   </object>
266   <packing>
267     <property name="expand">False</property>
268     <property name="fill">True</property>
269     <property name="position">2</property>
270   </packing>
271 </child>
272 <child>
273   <object class="GtkBox" id="box10">
274     <property name="visible">True</property>
275     <property name="can_focus">False</property>
276     <property name="orientation">vertical</property>
277     <child>
278       <object class="GtkLabel" id="label9">
279         <property name="visible">True</property>
280         <property name="can_focus">False</property>
281         <property name="label" translatable="yes">Lenght</property>
282       </object>
283     <packing>

```



```

284         <property name="expand">False</property>
285         <property name="fill">True</property>
286         <property name="position">0</property>
287     </packing>
288 </child>
289 <child>
290     <object class="GtkEntry" id="tamanho">
291         <property name="width_request">50</property>
292         <property name="visible">True</property>
293         <property name="can_focus">True</property>
294         <property name="text" translatable="yes">500</property>
295     </object>
296     <packing>
297         <property name="expand">False</property>
298         <property name="fill">True</property>
299         <property name="position">1</property>
300     </packing>
301 </child>
302 </object>
303 <packing>
304     <property name="expand">False</property>
305     <property name="fill">True</property>
306     <property name="position">3</property>
307 </packing>
308 </child>
309 </object>
310 <packing>
311     <property name="expand">False</property>
312     <property name="fill">True</property>
313     <property name="position">2</property>
314 </packing>
315 </child>
316 <child>
317     <object class="GtkBox" id="box6">
318         <property name="visible">True</property>
319         <property name="can_focus">False</property>
320         <property name="homogeneous">True</property>
321     <child>
322         <object class="GtkButton" id="button2">
323             <property name="label" translatable="yes">Clear</property>
324             <property name="visible">True</property>
325             <property name="can_focus">True</property>
326             <property name="receives_default">True</property>
327         </object>
328         <packing>
329             <property name="expand">False</property>
330             <property name="fill">True</property>
331             <property name="position">1</property>
332         </packing>
333     </child>
334     <child>
335         <object class="GtkButton" id="btn_procurar">
336             <property name="label" translatable="yes">RUN</property>
337             <property name="visible">True</property>
338             <property name="can_focus">True</property>
339             <property name="receives_default">True</property>
340             <property name="has_tooltip">True</property>

```



```

341         <property name="image_position">right</property>
342         <signal name="clicked" handler="on_btn_procurar_clicked" swapped=
"no"/>
343     </object>
344     <packing>
345         <property name="expand">False</property>
346         <property name="fill">True</property>
347         <property name="position">2</property>
348     </packing>
349 </child>
350 </object>
351 <packing>
352     <property name="expand">False</property>
353     <property name="fill">True</property>
354     <property name="position">3</property>
355 </packing>
356 </child>
357 <child>
358     <object class="GtkTreeView" id="treeview">
359         <property name="visible">True</property>
360         <property name="can_focus">True</property>
361         <property name="hscroll_policy">natural</property>
362         <property name="model">liststore</property>
363         <property name="enable_grid_lines">both</property>
364         <child internal-child="selection">
365             <object class="GtkTreeSelection" id="treeview-selection">
366                 <property name="mode">none</property>
367             </object>
368         </child>
369         <child>
370             <object class="GtkTreeViewColumn" id="treeviewcolumn1">
371                 <property name="title" translatable="yes">miRNAs com ICG </property>
372             </object>
373         </child>
374     </object>
375     <packing>
376         <property name="expand">True</property>
377         <property name="fill">True</property>
378         <property name="position">4</property>
379     </packing>
380 </child>
381 </object>
382 </child>
383 </object>
384 </interface>
385

```