UNIVERSIDADE FEDERAL DE PELOTAS
Centro de Desenvolvimento Tecnológico
Bacharelado em Biotecnologia

Trabalho de Conclusão de Curso

ESTUDOS DE BIOINFORMÁTICA APLICADOS À FERTILIDADE EM EQUINOS

Alessandra Neis

Pelotas, 2018

Pelotas, 2018

Alessandra Neis

ESTUDOS DE BIOINFORMÁTICA APLICADOS À FERTILIDADE EM EQUINOS

Trabalho acadêmico apresentado ao Curso de Biotecnologia da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Bacharel em Biotecnologia.

Orientadora: Priscila Marques Moura de Leon

Universidade Federal de Pelotas / Sistema de Bibliotecas Catalogação na Publicação

N416e Neis, Alessandra

Estudos de bioinformática aplicados à fertilidade em equinos / Alessandra Neis ; Priscila Marques Moura de Leon, orientadora. — Pelotas, 2018.

98 f. : il.

Trabalho de Conclusão de Curso (Bacharelado em Biotecnologia) — Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, 2018.

1. Biomarcador. 2. Modelagem de proteínas. 3. Prospecção de ncrnas. 4. Éguas. 5. Reprodução. I. Leon, Priscila Marques Moura de, orient. II. Título.

CDD : 636.10824

Elaborada por Ubirajara Buddin Cruz CRB: 10/901

Alessandra Neis

ESTUDOS DE BIOINFORMÁTICA APLICADOS À FERTILIDADE EM EQUINOS

Trabalho de Conclusão de Curso aprovado, como requisito parcial, para obtenção do grau de Bacharel em Biotecnologia, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas.

Data da Defesa: 20/11/2018

Banca examinadora:

Profª. Drª. Priscila Marques Moura de Leon (Orientadora)
Médica Veterinária, Doutora em Biotecnologia pela Universidade Federal de Pelotas

Profª. Drª. Vanessa Galli
Bióloga, Doutora em Biologia Celular e Molecular pela Universidade Federal do Rio Grande do Sul

Frederico Schmitt Kremer
Biotecnologista, Doutorando em Biotecnologia pela Universidade Federal de Pelotas

Gino Luigi Bonilla Lemos Pizzi
Médico Veterinário, Mestrando em Zootecnia pela Universidade Federal de Pelotas

Dedico este trabalho a meus pais e meu irmão.

Não teria conseguido sem vocês.

Agradecimentos

A meus pais Adelaide e Jandir, e meu irmão Guilherme, pelo amor e apoio incondicional durante estes anos de estudo e incentivo a sempre ir além.

A minha orientadora Priscila, que sempre demonstrou confiança em meu potencial, em nossos projetos e contribuiu enormemente para minha formação, mostrando diariamente competência, caráter e perseverança para me guiar nos diversos desafios da vida acadêmica e pessoal.

Ao Frederico, que contribuiu enormemente com meus projetos durante os últimos anos com sabedoria, paciência e gentileza, fazendo com que eu olhasse com muito carinho para a Bioinformática.

Aos meus queridos amigos Andréia Santos, Ana Luísa Callegari, Elias Barbosa, Josiane Otalakoski, Júlia Terra, Karina Pereira, Laura Del Valle, Marina Medeiros, Pâmela Caballero, Patrícia Oliveira, Pedro Albuquerque e Yasmine Alves. Obrigada por tudo, o apoio de vocês foi fundamental e sou muito grata por ter vocês em minha vida.

Resumo

NEIS, Alessandra. **Estudos de bioinformática aplicados à fertilidade em equinos**. 2018. 98f. Trabalho de Conclusão Curso - Bacharelado em Biotecnologia, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Pelotas, 2018.

A reprodução é essencial à equinocultura, visto o apelo comercial e a valorização individual da espécie equina. Diversas patologias interferem na fertilidade e algumas disfunções de relevância podem ter causas multifatoriais e/ou multigênicas, como endometrite, cistos ovarianos e endometriais, folículos anovulatórios e difunções hormonais. A bioinformática tem se mostrado essencial à análise dos dados e o desenvolvimento de *softwares* e algoritmos de predição têm possibilitado que análises *in silico* delimitassem alvos concretos antes da realização dos ensaios-padrão. Com a escassez de estudos de detecção de candidatos a biomarcadores da fertilidade em éguas, o presente trabalho tem como objetivos utilizar ferramentas bioinformáticas para: 1) modelar pela primeira vez as moléculas da prolactina e seu receptor; 2) realizar uma análise do genoma equino para a identificação de sequências referentes à RNAs não-codificantes, comparadas às que já são presentes nos bancos de dados, visando um potencial para aplicação na fertilidade em éguas. No primeiro artigo, as estruturas foram modeladas *in silico* através do software I-TASSER, sendo então corrigidas e submetidas a diversas etapas de refinamento conformacional e de dinâmica molecular. Nossos resultados demonstraram que os modelos preditos apresentaram 96% de confiabilidade (ePRL) e 92,5% de aminoácidos em posições corretas ou favorecidas (ePRLR) e a ePRL obteve um *score* de -4.09 na avaliação da qualidade global, enquanto ePRLR apresentou -8.45. A partir deste estudo, foi possível obter duas novas estruturas proteicas equinas de qualidade, o que permite posteriores análises funcionais e estruturais sobre a interferência de mutações nas vias de sinalização celular e sua utilização como candidatos a biomarcadores da fertilidade. No segundo artigo, avaliaram-se as sequências referentes à RNAs não-codificantes no genoma equino em relação aos bancos de dados Rfam e miRBase. Após uma extensiva revisão bibliográfica foram elencados 27 miRNAs relacionados aos processos oocitários e desordens reprodutivas associadas à infertilidade nas espécies humana, bovinos, equinos, ovinos, suínos e murinos. Os mais relevantes sugeridos para futuras análises são os mir-17, -18, -21, -23, -34, -133, -181, -378, -503 e let-7. A análise do genoma equino revelou que as classes de ncRNAs mais encontradas foram os miRNAs (65,1%), snRNAs (24,92%), lncRNAs (2,58%) e moléculas com ações regulatórias diversas (7,4%). Os miRNAs mais encontrados foram mir-544, mir-1255 e mir-562, que atuam em vias gênicas responsáveis pela manutenção do ciclo celular. Pela primeira vez relatada, foi possível identificar as diferentes classes, distribuição de elementos não-codificantes no genoma e fornecer uma lista de miRNAs que comprovadamente tem papel na fertilidade e que devem ser considerados para futuros estudos na espécie equina. Dessa forma, com o uso de ferramentas bioinformáticas, espera-se auxiliar na elucidação da genômica envolvida no desenvolvimento de desordens da reprodução e na detecção de candidatos a biomarcadores de fertilidade em éguas.

Palavras-chave: modelagem de proteínas; prospecção de ncRNAs; reprodução; éguas; biomarcador.

Abstract

NEIS, Alessandra. **Bioinformatics studies applied to fertility in the equine species.** 2018. 98p. Graduation in Biotechnology. Bachelor's Degree in Biotechnology, Trabalho de Conclusão Curso - Bacharelado em Biotecnologia, Center for Technological Development, Federal University of Pelotas, Pelotas, 2018.

Reproduction is essential to equine breeding for their commercial appeal and individual appreciation. Several disorders interfere with fertility and some may have a multifactorial and/or multigenic cause, such as endometritis, ovarian and endometrial cysts, anovulatory follicles and hormonal dysregulation. After the horse genome sequencing, the data generated could be applied to horse breeding, mainly in biomarkers identification. In this context, bioinformatics proved essential for data analysis, software development, and prediction algorithms have enabled in silico analyzes to determine concrete targets prior to standard assays. Once there are few reports relating genomics to mare fertility, the aim of this study is to apply bioinformatic tools to: 1) model the first molecules of the prolactin and its receptor is used; 2) perform an equine genome analysis to identify non-coding RNA sequences, already present in the databases, in order to enhance the application of mares fertility. The former were modeled in silico using I-TASSER software, corrected and then submitted to a diverse range of conformational refinement and molecular dynamics tools. The results demonstrated that the models are 96% confidence (ePRL) and 92.5% amino acids in correct or favored positions (ePRLR), according to Rampage. The ePRL obtained a score of -4.09 in the overall quality assessment performed by QMEAN, while ePRLR showed -8,45 in QMEANbrane software. From this study, it was possible to obtain two new, high quality equine molecular protein structures, which allows further analysis on the interference of mutations in cell signaling pathways and their use as candidates for fertility biomarkers. In the second article, an evaluation of non-coding RNAs sequences on the equine genome in relation to Rfam and miRBase databases. An extensive bibliographic review provided 27 miRNAs, related to oocyte processes and reproductive disorders associated with infertility in human, bovine, equine, ovine, swine and murine species were listed. The most promising molecules suggested are mir-17, -18, -21, -23, -34, -133, -181, -378, -503 and let-7, for their presence in both reproductive disorders and steroidogenesis. Analysis of different ncRNAs types showed the most found were miRNAs (65.1%), snRNAs (24.92%), lncRNA (2.58%) and diverse regulating RNAs (7.4%). The miRNAs were mir-544, mir-1255 and mir-562, which act in signaling pathways related to cell cycle maintenance. For the first time reported, it was possible to identify different classes, distribution of non-coding elements in the genome and provide of a list of miRNAs proven fertility-related and should be considered for future studies of the equine species. Thus, using bioinformatics tools, we aim to contribute in the elucidation of the genomics associated to reproduction disorders and detection of biomarkers of fertility in mares.

Keywords: protein modelling; ncRNAs prospection; reproduction; mare.

Lista de figuras

Figuras Revisão de Literatura

Figura 1 Diferentes estágios da oogênese e foliculogênese na espécie equina.....36

Figura 2 Regulação endócrina do desenvolvimento folicular na espécie equina.....38

Figura 3 Estrutura de um precursor de miRNA equino.....44

Figura 4 Biogênese dos miRNAs.....46

Figuras Artigo

Figure 1 *Predicted model for activated equine prolactin receptor, containing two receptor molecules forming the binding site. The three different domains are separated (ECD, TMD and ICD)*……..62

Figure 2 *Protein model of equine prolactin. Indicated in the structure are the C-terminal and N-terminal, as well as the main four alpha helixes*……..62

Figuras Manuscrito

Figura 1 Metodologia utilizada para a obtenção dos miRNAs de relevância a partir da literatura disponível nos bancos de dados. 76

Figura 2 Distribuição de ncRNAs identificados no genoma equino.....83

Figura 3 Distribuição dos miRNAs identificados no genoma equino 84

Figura 4 Distribuição dos ncRNAs por cromossomo.....85

Figura 5 Distribuição dos miRNAs por cromossomo.....86

Lista de quadros

Quadro 1 miRNAs relacionados à esteroidogênese selecionados a partir da revisão de literatura.....78

Quadro 2 miRNAs relacionados a desordens reprodutivas selecionados a partir da revisão de literatura.....80

Lista de abreviaturas e siglas

ACVR1	*Activin A Receptor Type 1* (PT: Receptor de Activina A Tipo 1)
ACVR2A	*Activin Receptor A Type 2A* (PT: Receptor de Activina A tipo 2A)
Ago	Argonauta
AKT1	*AKT Serine/Threonine Kinase 1* (PT: AKT serina/treonina quinase 1)
ARID3B	*AT-Rich Interaction Domain 3B* (PT: Domínio 3B de Interação Rico em A-T)
BCL-2	*B-Cell CLL/Lymphoma 2* (PT:Linfoma de células B 2)
cAMP	AMP cíclico
CASP	Avaliação Crítica da Predição Estrutural
CASP3	Caspase 3
CBFB	Fator Core-Binding Subunidade Beta
CCNB1	Ciclina B1
CCND1	Ciclina D1
CCND2	Ciclina D2
CCNE2	Ciclina E2
CCNE3	Ciclina E3
CDC25A	*Cell Division Cycle 25A*
CDK4	*Cyclin Dependent Kinase 4* (PT: Quinase dependente de Ciclina 4)
CDK6	*Cyclin Dependent Kinase 6* (PT: Quinase dependente de Ciclina 6)

CDKN1A	*Cyclin Dependent Kinase Inhibitor 1A* (PT: Inibidor da quinase dependente de ciclina 1A)
CGP	Células germinativas primordiais
CNV	*Copy Number Variation* (PT: Variação no Número de Cópias)
COX-2	Ciclo-oxigenase 2
CTBP1	*C-Terminal Binding Protein 1* (PT: Proteina ligante de C-terminal 1)
CYP11A1	Citocromo P450 Família 11 Subfamília A Membro 1
DLL1	*Delta Like Canonical Notch Ligand 1* (PT: Ligante NOTCH1 Canônico similar à delta)
DNA	Ácido desoxirribonucleico
E2F3	*E2F Transcription Factor 3*
ECD	*Extracellular Domain* (PT: Domínio extracellular)
EGF	*Epidermal Growth Factor* (PT: Fator de crescimento epidermal)
EGFR	*Epidermal Growth Factor Receptor* (PT: Receptor do fator de crescimento epidermal)
ePRL	*Equine Prolactin (PT: Prolactina Equina)*
ePRLR	*Equine Prolactin Receptor (PT: Receptor de Prolactina Equina)*
ERBB2	*Erb-B2 Receptor Tyrosine Kinase 2* (PT: Receptor de tirosina-quinase 2 de Erb-B2)
ERBB3	*Erb-B2 Receptor Tyrosine Kinase 3* (PT: Receptor de tirosina-quinase 3 de Erb-B2)
ESR1	*Estrogen Receptor 1* (PT: Receptor de Estrógeno 1)
Ezh2	*Enhancer of Zeste 2 Polycomb Repressive Complex 2 Subunit*
FIV	Fertilização In Vitro
FOXL2	*Box Forkhead L2*
FSH	*Follicule Stimulating Hormone* (PT: Hormônio Folículo Estimulante)
G6PDH	Glicose-6-fosfato-desidrogenase
Gb	Gigabase
GnRH	*Gonadotrophin Releasing Hormone* (PT: Hormônio Liberador de Gonadotrofinas)
hCG	*Human Chrionic Gonadotrophin* (PT: Gonadotrofina Coriônica Humana)

HMGA2	High-mobility group AT-hook 2
ICD	*Intracellular Domain*
ICSI	*Intracytoplasmic Sperm Injection* (PT: Injeção Intracitoplasmática de Espermatozóides)
ID2	*Inhibitor Of DNA Binding 2* (PT: Inibidor de Ligação ao DNA 2)
ID3	*Inhibitor Of DNA Binding 3* (PT: Inibidor de ligação ao DNA 3)
IKBKB	*Inhibitor Of Nuclear Factor Kappa B Kinase Subunit Beta* (PT: Quinase Inibidora do fator nuclear Kappa B subunidade beta)
IL-10	Interleucina 10
IL-12B	Interleucina 12 Beta
IL-1B	Interleucina 1 Beta
IL-37	Interleucina 37
IL-8	Interleucina 8
INHBB	*Inhibin Subunit Beta B* (PT: Inibina Subunidade Beta B)
INSR	*Insulin Receptor* (PT: Receptor de insulina)
IPO	Insuficiência Ovariana Prematura
IRS2	*Insulin Receptor Substrate 2* (PT: Receptor de Substrato de Insulina 2)
JAG1	*Jagged 1*
JAK2	Janus-quinase 2
JUN	Jun Proto-Oncogene
Kb	Kilobase
KLF13	*Kruppel Like Factor 13* (PT: Fator similar a Kruppel 13)
KRT7	Keratin 7 (PT: Queratina 7)
LH	*Luteinizing Hormone* (PT: Hormônio Luteinizante)
LIF	*Interleukin 6 Family Cytokine* (PT: interleucina 6 da família de citocinas)
LIN28	Lin-28 homólogo A
lincRNA	RNA não-codificante nuclear longo
lncRNA	RNA longo não-codificante
MAP3K1	*Mitogen-Activated Protein Kinase 1* (PT: Proteína quinase ativada por mitógeno 1)
MAPK14	*Mitogen-Activated Protein Kinase* 14 (PT: Proteína quinase ativada por mitógeno 14

Mb	Megabase
MET	Proto-oncogene MET
miRNA	micro-RNA
mRNA	RNA mensageiro
MYC	Proto-oncogene MYC
NCBI	*National Center for Biotechnology Information*
ncRNA	RNA não-codificante
NMR	Ressonância Magnética Nuclear
NOS2	*Nitric Oxide Synthase 2* (PT: Óxido nítrico sintase 2)
NOTCH1	Notch 1
Nt	Nucleotídeo
NTKR3	*Neurotrophic Receptor Tyrosine Kinase 3* (PT: Receptor de tirosina-quinase Neutrófico 3)
ORF	Fase de leitura aberta
OTXR	*Oxytocin Receptor* (PT: Receptor de Ocitocina)
P53	*Tumor supressor p53* (PT: supressor tumoral p53)
PAZ	Piwi-Argonauta-Zwillie
PCR	Reação em Cadeia da Polimerase
PDCD4	*Programmed Cell Death 4* (PT: Morte celular programada 4)
piRNA	RNA que interage com PIWI
PLA2G4A	*Phospholipase A2 Group IVA* (PT: Fosfolipase A2 Grupo IVA)
pré-miRNA	micro-RNA precursor
pri-miRNA	micro-RNA primário
Prl	Prolactina
Prlr	Receptor de prolactina
PTEN	*Phosphatase and Tensin Homolog* (PT: Homólogo da Fosfatase e Tensina)
qRT-PCR	*Quantitative Real Time PCR* (PT: PCR em tempo real quantitativo)
RAB5B	Família Oncogênica RAS
RBMS1	*RNA Binding Motif Single Stranded Interacting Protein 1*
RHOA	*Ras Homolog Family Member A* (PT: Homólogo da família RAS membro A)

RISC	*RNA Induced Silencing Complex* (PT: Complexo de Indução do Silenciamento de RNA)
RNA	Ácido ribonucleico
RNA-pol	RNA polimerase
RNA-Seq	Sequenciamento de RNA
ROCK1	*Rho Associated Coiled-Coil Containing Protein Kinase 1*
ROCK2	*Rho Associated Coiled-Coil Containing Protein Kinase 2*
rRNA	RNA ribossômico
siRNA	*Small Interference RNA* (PT: Pequeno RNA de interferência)
SMAD4	*Membro da família SMAD 4*
snoRNA	*Small Nucleolar RNA* (PT: Pequeno RNA Nuceololar)
SNP	*Single Nucleotide Polymorphism* (PT: Polimorfismo de Nucleotídeo Único)
snRNA	*Small Nuclear RNA* (PT: Pequeno RNA nuclear)
SOP	Síndrome do Ovário Policístico
SP1	Fator de Transcrição SP1
STAR	*Steroidogenic Acute Regulatory Protein* (PT: Proteína Regulatória Esterodoigênica Aguda)
STAT	*Signal Transducer and Activator of Transcription proteins* (PT: Proteínas Transdutoras de sinal e Ativadoras da Transcrição)
SYT1	Sinaptogamin (PT: Sinaptogamina 1)
TAGLN2	*Transgelin 2* (PT: Transgelina 2)
TGFB1	Fator de Crescimento Transformante Beta 1
TM	*Transmembrane Domain* (PT: Domínio Transmembrana)
TP53	*Tumor Protein P53* (PT: Proteína Tumoral 53)
TRIM71	*Tripartite Motif Containing 71*
tRNA	RNA transportador
UTR	*Untranslated region*, Região não-traduzida
VEGFA	*Vascular Endothelial Growth Factor A* (PT: Fator de Crescimento Vascular Endotelial A)
WNT	*Wingless-Type*
XIAP	X-Linked Inhibitor Of Apoptosis (PT: Inibidor de apoptose ligado ao cromossomo X)
XPO-5	Exportina-5

Sumário

1 Introdução Geral 18
2 Revisão de literatura 21
2.1 Biotecnologia 21
2.2 Equinocultura 22
2.2.1 Fisiologia reprodutiva das éguas 23
2.2.3 Desordens reprodutivas 28
2.3 Genômica 30
2.4 ncRNAs 33
2.5 Bioinformática 39
2.5.1 Modelagem de proteínas 40
2.5.2 Detecção *in silico* de ncRNAs 42
3 Objetivos 44
4 Artigo 45
5 Manuscrito 62
6 Conclusão Geral 87
Referências 89

1 Introdução Geral

A equinocultura refere-se à criação de equinos, e, no Brasil movimenta em torno de 16 bilhões de reais, contando com 5,5 milhões de cabeças (MINISTÉRIO DA AGRICULTURA, 2016). Nesse contexto, a área reprodutiva equina está sempre em evidência, devido ao apelo comercial crescente e a valorização individual destes animais. A reprodução em éguas apresenta um enorme desafio, devido à complexidade fisiológica e genética, assim como a quantidade de disfunções (SNIDER, 2015) e a complexidade genômica relacionada à fertilidade (CHOWDHARY, 2013).

Diversas patologias reprodutivas às quais as éguas são acometidas ainda carecem de estudos relacionados à genética e genômica, embora a fisiopatologia das desordens já seja descrita. Disfunções de relevância que possuem uma causa desconhecida e/ou multifatorial incluem endometrite, folículos anovulatórios, cistos ovarianos e endometriais e disfunções hormonais. Tais fatores podem levar à infertilidade e estão associados com baixo desempenho reprodutivo (SENA et al., 2016).

Após o sequenciamento do genoma equino (WADE et al., 2009), mesmo com uma grande quantidade de dados disponíveis, apenas uma pequena parcela aplica-se no estudo da reprodução das fêmeas, e a extrapolação dos dados associados à fertilidade, em conjunto com a transcriptômica e proteômica, ainda é escassa.

A bioinformática surge, nesse contexto, como uma importante ferramenta ainda pouco explorada na referida área. O desenvolvimento da bioinformática ocorreu em conjunto com o avanço do estudo de proteínas, e posteriormente, com as técnicas moleculares, principalmente o sequenciamento de DNA, técnica do DNA recombinante e o desenvolvimento da PCR (Reação em Cadeia da Polimerase, do inglês, *Polymerase Chain Reaction*). A criação de bancos de dados de sequências e o surgimento de ferramentas que facilitam a manipulação computacional do DNA e permitiram análises mais complexas (GAUTHIER et al., 2018), como análises

estruturais e funcionais *in silico*, que hoje são empregadas antes mesmo da realização de experimentos *in vitro. Softwares* de predição ganharam espaço, a fim de auxiliar na elucidação de sequências não designadas nas bases de dados (NAIR, 2007). Três principais métodos computacionais são empregados nas análises de dados, que incluem a predição *ab initio*, baseada nas características da sequência ou estrutura; predição comparativa, que relaciona-se com a conservação evolucionária da sequência, e a predição integrada do método *ab initio* e comparativo (LESK, 2014).

A modelagem de proteínas fornece uma importante base para estudos de análise estrutural e funcional focados em uma molécula. Um dos métodos de predição de proteínas amplamente utilizado denomina-se *threading,* que avalia dentro de uma biblioteca de estruturas conhecidas, aquelas que possuem maior similaridade conformacional com a sequência de uma proteína de estrutura desconhecida. São construídos diversos modelos aproximados, usando os diferentes alinhamentos possíveis das sequências das proteínas conhecidas com a da proteína desconhecida. Um padrão geral de conformação é atribuído à proteína, com uma precisão de 70 a 80% na análise resíduo-por-resíduo (LESK, 2014).

Na reprodução, as moléculas da prolactina e seu receptor desempenham importantes papéis fisiológicos e ainda não possuem sua estrutura proteica elucidada. A prolactina (PRL) é secretada pela glândula pituitária e entre suas funções reprodutivas estão estímulo ao desenvolvimento das mamas, lactogênese e galactopoiese, regulação da esteroidogênese, secreção de progesterona, receptividade reprodutiva e comportamento materno. Disfunções hormonais na PRL e no seu receptor (PRLR) podem levar ao aparecimento de desordens como galactorreia, amenorreia, infertilidade e ciclos anovulatórios (revisado em FREEMAN et al., 2000). Dessa maneira, análises funcionais e estruturais de candidatos a biomarcadores seriam de grande relevância para a elucidação de diferentes aspectos relacionados à infertilidade em éguas. Entretanto, estas análises acabam sendo dificultadas pela inexistência de modelos estruturais para a espécie equina. Visto isso, uma análise através do método de *threading* se apresenta como uma alternativa viável para a realização de estudos bioinformáticos voltados à reprodução em éguas.

Além da predição de proteínas, é possível empregar a bioinformática no estudo de sequências gênicas relacionadas a elementos não-codificantes do

genoma (HUANG et al., 2017). Vários métodos computacionais foram projetados e aplicados à detecção de genes de micro-RNAs (miRNAs), através da análise da estrutura secundária e a estabilidade termodinâmica das estruturas *hairpin*. Os métodos de mapeamento de genes de miRNAs baseados em homologia utilizam miRNAs disponíveis e validados experimentalmente a fim de encontrar estruturas e sequências similares em espécies relacionadas (YOUSEF; JENS, 2014). Se um miRNA é identificado em um genoma, seus homólogos podem ser encontrados em outras espécies (YOUSEF; JENS, 2014).

A importância da regulação gênica exercida pelos ncRNAs (RNAs não-codificantes, do inglês, *Noncoding RNA*) abre portas para a utilização dos bancos de dados e ferramentas de análise de sequências. O baixo número de estudos em éguas e a similaridade no perfil de ncRNAs regulatórios entre espécies, os fazem muito interessantes de serem aplicados na fertilidade das fêmeas. A partir da bibliografia disponível relacionada a miRNAs, é possível elencar aqueles associados a desordens reprodutivas e ao desenvolvimento folicular e oocitário em diferentes espécies, organizando um estudo base para a transposição deste conhecimento para equinos. Portanto, uma análise no genoma equino através de ferramentas computacionais permite elencar os ncRNAs existentes e prospectar moléculas candidatas a biomarcadores que podem estar associados à fertilidade equina.

2 Revisão de literatura

2.1 Biotecnologia

A biotecnologia refere-se à utilização de microrganismos vivos e suas partes a fim de gerar novos produtos, processos, tecnologias e inovações. A biotecnologia pode ser dividida em duas grandes áreas: clássica e moderna. A biotecnologia clássica emprega microrganismos como agentes fermentadores para a produção de queijos, vinhos, cervejas, pães, iogurtes, entre outros. Já a biotecnologia moderna conta com a evolução dos conhecimentos relacionados principalmente à genética, microbiologia e biologia molecular (VIALTA, 2018), estimulando o desenvolvimento de novos compostos farmacêuticos, combustíveis renováveis, produção de vacinas, testes diagnósticos, organismos transgênicos, terapia gênica, estudos com células-tronco, produção de moléculas recombinantes, entre outros.

Os estudos de biologia molecular alavancaram com o desenvolvimento de técnicas como a Reação em Cadeia da Polimerase (do inglês, *Polimerase Chain Reaction* - PCR) e o sequenciamento de DNA. A partir do momento em que tornou-se possível analisar qualitativa e quantitativamente o material genético, conseguiu-se desenvolver técnicas mais sofisticadas, como a qRT-PCR quantitativa (PCR em tempo real, do inglês, *quantitative real time PCR*), microarranjos e sequenciamento de RNA (RNA-seq). Com isso, surgiram novos dados referentes a regiões gênicas de interesse, associá-las a determinadas características, e então prospectar e definir biomarcadores. Da mesma forma, foi possibilitada a análise de mutações, manipulação de genes e clonagem, análises gênicas estruturais e funcionais, o que permitiu haver uma infinidade de aplicações na pesquisa e diagnóstico.

A partir da análise do DNA, diversos campos se desenvolveram, como a genômica, transcriptômica e proteômica. Uma vez determinada a sequência do genoma, esta pode ser utilizada para identificar as sequências codificantes e não-codificantes, que atuam como molde ou regulador da síntese de RNA – o

transcritoma. Geralmente, técnicas de hibridização como os microarranjos são utilizados para a detecção do perfil de expressão de RNAs tecido-específico. Estas análises também detectam RNAs não-codificantes, aos quais têm funções regulatórias sobre outros RNAs. A detecção dos níveis de expressão fornece uma estimativa aproximada do nível de expressão da proteína codificada, entretanto, uma análise do conjunto completo de proteínas, em determinados tecidos ou condições, representa de fato os produtos gerados pelo código genético – a proteômica. O proteoma, assim como o transcritoma, não é estático e possui um perfil diferente em determinadas condições às quais é submetida a célula (WATSON, 2015). Juntas, estas três áreas contribuem enormemente para a elucidação dos fenômenos que ocorrem tanto em condições fisiológicas quanto patológicas, determinando as moléculas-chave atuantes nestes processos.

As técnicas moleculares foram sendo associadas à criação de equinos, onde ferramentas já foram descritas na literatura e aplicadas à equinocultura (HAAS; TORRES, 2016), entretanto, sua utilização comercial ainda é um pouco limitada. Estudos aplicados à genômica visam elucidar aspectos relacionados a desordens genéticas, encontrar biomarcadores de desempenho físico, relacionados à e à clínica, analisando o papel destas mutações (CHOWDHARY, 2013). Além disso, o uso de técnicas moleculares se faz presente na genealogia e registro em associações, através da exigência de testes de paternidade.

2.2 Equinocultura

A equinocultura refere-se à criação de cavalos, especialmente de raça. Devido ao apelo comercial crescente, ao prestígio por estes animais e pela sua valorização individual, a equinocultura tornou-se uma importante área econômica do agronegócio. De acordo com estimativa feita pela *American Horse Council* (2018), em 2017 a equinocultura movimentou em torno de 50 bilhões de dólares diretamente, apenas nos Estados Unidos. No Brasil, esse montante foi de 16 bilhões de reais, de acordo com estimativa de 2016, contando com um rebanho de quase 5,5 milhões de cabeças (MINISTÉRIO DA AGRICULTURA, 2016). Vista a aplicação da criação de cavalos em atividades de esportes hípicos, enduro, lazer e trabalho, a reprodução torna-se um fator indispensável para os criadores, uma vez que é peça-chave na manutenção da qualidade genética do plantel.

Ao longo da criação de equinos, por muito tempo apenas utilizaram-se observações fenotípicas para a seleção de reprodutores. Atualmente, se dispõe de grande quantidade de informações referentes à fisiologia, anatomia e genética da reprodução, principalmente para machos. O desenvolvimento de novas técnicas aplicadas à reprodução, e, principalmente, as biotécnicas, auxiliaram enormemente a evolução da qualidade genética, além de diminuir os custos relacionados às coberturas em algumas raças. As biotécnicas mais utilizadas em equinos incluem inseminação artificial, transferência de embriões, criopreservação de sêmen e injeção intracitoplasmática de espermatozoides (do inglês, *IntraCytoplamatic Sperm Iinjection* - ICSI) (BERTOZZO et al., 2014).

A fertilidade em éguas representa um fator crucial na criação de equinos, tanto pela complexidade da genética da reprodução em si (CHOWDHARY, 2013) quanto pela quantidade de disfunções que podem afetar as fêmeas (SNIDER, 2015). Além disso, a maioria das disfunções de fertilidade são bem compreendidas em relação a sua fisiopatologia, porém seus mecanismos genéticos e moleculares ainda são escassamente descritos. Neste sentido, uma revisão dos principais fatores relacionados à reprodução é realizada a seguir.

2.2.1 Fisiologia reprodutiva das éguas

As éguas são monovulatórias, o que significa que apenas um folículo alcança o estado necessário para ovulação num período de 21 dias, o chamado ciclo estral. O estro é o período de 5 dias em que ocorre a ovulação, enquanto o diestro corresponde aos períodos anterior e posterior ao estro (revisado em MOREL, 2008). Nas éguas, a reprodução é chamada de poliéstrica, em que o período ovulatório ocorre diversas vezes, e sazonal, em um determinado período do ano - primavera e verão (revisado em PTASZYNSKA, 2007). A estação anovulatória pode ser diferenciada em: fase de transição de atividade cíclica para um anestro profundo, período anovulatório médio e uma segunda fase de transição para atividade cíclica na primavera (revisado em AURICH, 2011). A idade da égua também pode estar relacionada à sua resposta reprodutiva – aquelas acima de 15 anos de idade (revisado em MOREL, 2008), apresentam a duração do ciclo estral mais longa e a ocorrência do pico hormonal é significantemente diminuída (GINTHER et al., 2008).

A regulação endócrina é regida pelo eixo hipotalâmico-pituitário-gonadal, que libera os principais hormônios do ciclo estral (revisado em PTASZYNSKA, 2007). O Hormônio Liberador de Gonadotrofinas (GnRH) induz a glândula pituitária a liberar o Hormônio Folículo Estimulante (FSH) e Luteinizante (LH), que por sua vez estimularão a produção dos hormônios ovarianos, consequentemente o desenvolvimento folicular e a ovulação. O FSH encontra-se em seus maiores níveis entre os dias 9 a 12 e no final do ciclo. Já o LH é liberado através de pulsos de secreção e um pico de secreção no dia 1 do ciclo resulta na extrusão do oócito (revisado em PTASZYNSKA, 2007).

A prolactina (Prl) também é secretada pela glândula pituitária e entre suas funções reprodutivas estão estímulo ao desenvolvimento das mamas, lactogênese e galactopoiese, assim como, a regulação da esteroidogênese, secreção de progesterona, receptividade reprodutiva e o comportamento materno (revisado em FREEMAN et al., 2000). Dentre suas funções não-reprodutivas, estão a regulação da homeostase (revisado em FREEMAN et al., 2000) e mudanças metabólicas nos meses de inverno (revisado em MOREL, 2008). A molécula interage com o receptor de prolactina (PRLR) e inicia sua ativação mediante a formação de um complexo trimérico entre dois receptores e o hormônio, que leva à fosforilação e ativação da Janus-quinase 2 (JAK2) e transdução de sinal ao núcleo por meio das proteínas transdutoras de sinal e ativadoras de transcrição (STATs) (IHLE et al., 1994). Disfunções hormonais e no receptor podem levar ao aparecimento de desordens reprodutivas como galactorreia, amenorreia, infertilidade e ciclos anovulatórios (revisado em FREEMAN et al., 2000).

A primeira molécula de Prl pituitária equina foi descrita em 1988 em dois tipos específicos de lactotrófos no tecido adeno-hipofisário (LEHRMAN et al., 1988). Em relação à sua secreção, é observado um aumento acentuado das concentrações pouco antes do início do estro durante os ciclos estrais de outono em éguas (THOMPSON; OBERHAUS, 2015). Sazonalmente, o padrão das concentrações séricas é maior durante o verão e menor no inverno. Assim, submeter as éguas à iluminação artificial suplementar nos meses de inverno auxilia no aumento das concentrações de Prl (JOHNSON, 1986). Já a indução hormonal com Prl permitiu às éguas ovular no inverno, sem interferência do FSH e antagonistas dopaminérgicos exógenos, indicando que a Prl individualmente é capaz de estimular a ovulação em éguas (BRENDEMUEHL; CROSS, 2000; THOMPSON; HOFFMAN; DEPEW, 1997).

Embora sua aplicação seja promissora, não existe uma fonte em massa de Prl equina disponível e o tratamento com Prl suína recombinante levou à geração de resposta imune (THOMPSON; OBERHAUS, 2015).

Os hormônios secretados pelos ovários são estradiol e progesterona. A secreção de estradiol está associada ao comportamento receptivo da égua e é dependente do FSH. A progesterona está associada ao desenvolvimento folicular e oocitário final e tem papel inibidor da liberação de LH. No caso de ausência do reconhecimento da prenhes, a progesterona automaticamente voltará aos níveis basais, senão, sua produção é mantida pelo folículo remanescente (revisado em MOREL, 2008).

O desenvolvimento dos oócitos começa após a colonização das gônadas primordiais e reorganização genômica pelas células germinativas primordiais (CGP), que passam por intensa divisão mitótica com uma citocinese incompleta originando um grande número de células. As CGPs formam aglomerados contendo em torno de 16 células unidas por pontes citoplasmáticas, sendo então chamadas de oogônias (ROJAS et al., 2015) (Figura 1). A partir disso, o ciclo mitótico para e inicia-se a meiose, com a ativação por meio do ácido retinóico, transformando as oogônias em oócitos primários haploides envoltos em células epiteliais (SÁNCHEZ; SMITZ, 2012). Os oócitos primários entram em quiescência no final da prófase I da meiose I, e esperam o período de indução pelo LH, já sendo bem maiores que as oogônias e contendo um elevado número de organelas citoplasmáticas (SCHOENWOLF et al., 2016).

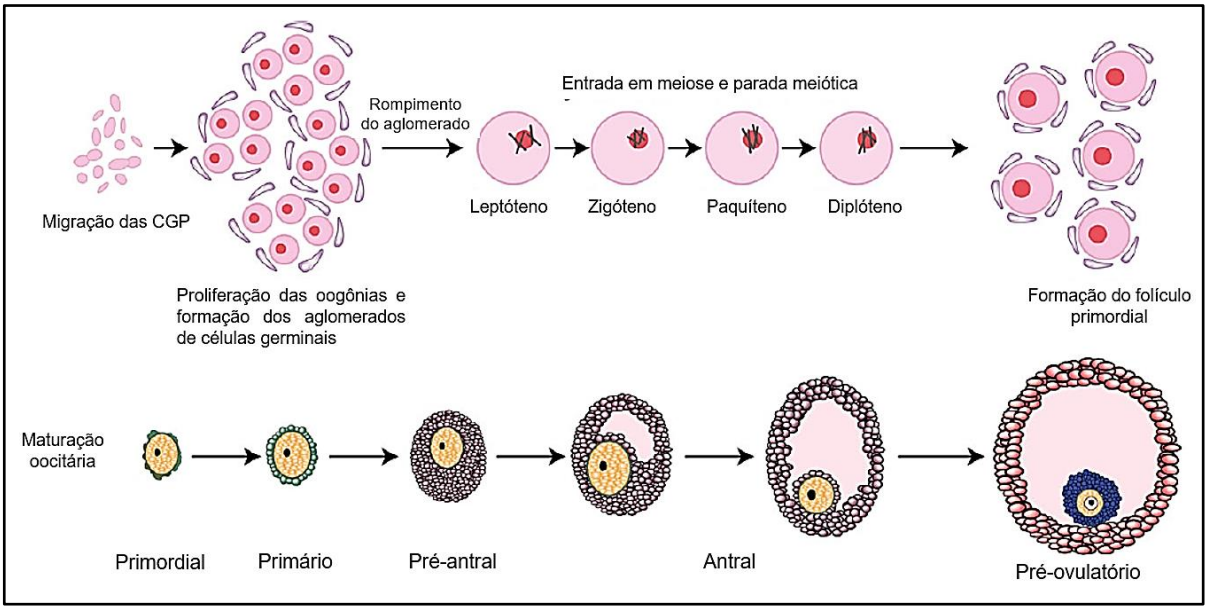


Figura 1. Diferentes estágios da oogênese e foliculogênese. Adaptado de (ANDERSON; TELFER, 2018).

A reativação dos oócitos primários ocorre sob ação dos hormônios FSH e LH e possui duas fases, a folicular e a luteal. Com a diminuição dos níveis de AMP cíclico (cAMP), o oócito primário conclui a meiose I (SÁNCHEZ; SMITZ, 2012), gerando um oócito haploide e um corpo polar (SCHOENWOLF et al., 2016). A estimulação ovariana leva a um aumento gradual dos níveis de estradiol, produzidos pelas células da granulosa. O oócito secundário agora é dependente da ação do FSH e estradiol, está circundado por múltiplas camadas de células da granulosa cuboides juntamente com as camadas interna e externa da teca, que sintetizam os andrógenos posteriormente aromatizados (ROJAS et al., 2015). A segunda divisão meiótica, por sua vez, só será completada mediante presença do DNA paterno (SCHOENWOLF et al., 2016).

A foliculogênese corresponde ao desenvolvimento do folículo. A quantidade de células germinais para toda a vida é determinada anteriormente ao nascimento e, assim como em humanos, a primeira fase do desenvolvimento começa a ocorrer qualquer momento após o nascimento, uma vez que não é dependente dos hormônios gonadotóficos (revisado em MOREL, 2008).

O folículo primordial é composto pelo oócito e uma camada de células foliculares achatadas; quando as células foliculares epiteliais começam a formar uma camada espessa em formato cuboide, o folículo é chamado de primário. Inicia-se a secreção de glicoproteínas e forma-se a zona pelúcida em torno do oócito. O

folículo é denominado em crescimento quando as células foliculares epiteliais encontram-se em multicamadas, podendo haver 6 ou 7. Células do estroma são recrutadas para formar a membrana granulosa e teca interna. O folículo continua a crescer e acumula líquido em sua cavidade interna (antro), sendo denominando agora folículo antral. As células da granulosa projetam o oócito para o antro, formando a estrutura *cumulus oophorus* (SCHOENWOLF et al., 2016). A segunda fase da foliculogênese ocorre continuamente, e ao final de 21 dias estará formado um folículo grande e dilatado, o folículo de Graaf (revisado em MOREL, 2008). O desenvolvimento folicular é exemplificado na Figura 2.

O folículo que aumenta de tamanho tem grandes chances de se tornar dominante e começa a se movimentar para a fossa ovulatória. Essa estrutura é uma particularidade da anatomia ovariana em equinos, representando o único local em que a ovulação ocorre. O oócito é liberado do folículo com o pico de LH, e o folículo remanescente colapsa. As células da teca se agrupam na cavidade folicular, formando o corpo lúteo. O corpo lúteo produz progesterona para manutenção de uma possível prenhês (ROJAS et al., 2015). Em caso negativo, o corpo lúteo regride e é invadido por fibroblastos, tornando-se um tecido cicatrizado de coloração amarela a branca (*corpo albicans*) (revisado em MOREL, 2008).

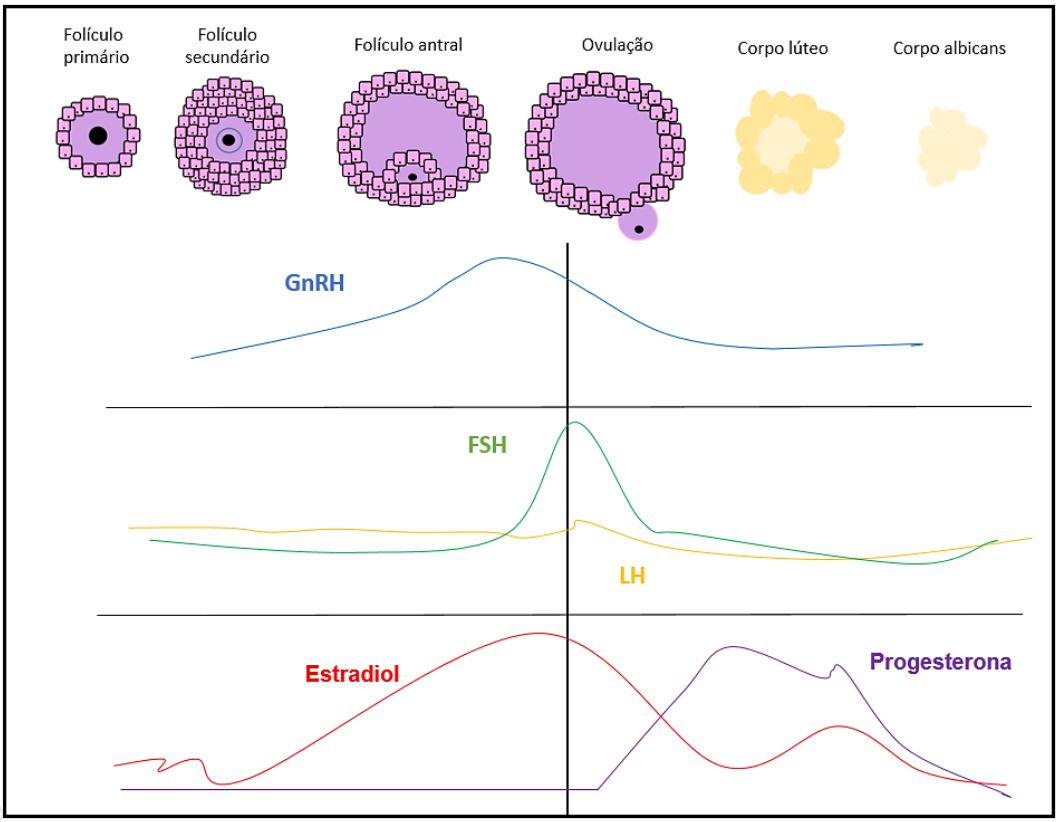


Figura 2. Regulação endócrina do desenvolvimento folicular.

2.2.3 Desordens reprodutivas

A infertilidade pode decorrer de diversas complicações nos mais diversos aspectos da reprodução, vista sua complexidade, e geralmente possuem causas multifatoriais e/ou multigênicas. Algumas das principais desordens reprodutivas em éguas incluem: endometrite, folículos anovulatórios, cistos ovarianos e endometriais e disfunções hormonais. Tais fatores podem levar à infertilidade e estão associados com baixo desempenho reprodutivo e infertilidade (SENA et al., 2016).

Uma das principais causas de distúrbios da fertilidade em éguas é a endometrite, que leva à perda gestacional nos estágios embrionários iniciais e atraso no ciclo reprodutivo (LEBLANC; CAUSEY, 2009). Esta pode ser bacteriana ou não, sendo caracterizada pela inflamação do endométrio. Após a cobertura, é frequente ocorrer uma endometrite transitória, onde há uma resposta inflamatória intensa mesmo em éguas com o trato reprodutivo normal. Na endometrite persistente pós-cobertura, há uma reação inflamatória aos fluidos seminais e pode ser identificada se ocorrer recorrentemente cerca de 30 minutos após a cobertura (CRUZ, 2016). No

caso de uma endometrite persistente, alguns fatores predisponentes como insuficiência na eliminação de *debris* pelo miométrio, (PTASZYNSKA, 2007) e má conformação do períneo (revisado em MOREL, 2008) podem contribuir para seu desenvolvimento. Já a endometriose, por sua vez, é irreversível, e caracteriza-se principalmente pela fibrose periglandular (CRUZ, 2016). Geralmente, estão presentes 2 a 3 camadas de tecido fibrótico ao redor das glândulas endometriais, e quanto maior o número de camadas, maior o comprometimento endometrial (revisado em PTASZYNSKA, 2007) e insensibilização das células estromais fibróticas quanto à expressão de receptores dos hormônios esteroides, que encontram-se reduzidos (CRUZ, 2016).

Os cistos são uma anormalidade não-inflamatória, e podem ser classificados em endometriais e linfáticos/glandulares (STANTON; STEINER; PUGH, 2004). Os cistos endometriais são menores, estão espalhados pela região uterina e propõe-se que estejam sujeitos às mudanças hormonais (STANTON; STEINER; PUGH, 2004). Por outro lado, os cistos linfáticos são maiores e localizam-se em determinados pontos (revisado em SENA et al., 2016) onde há acúmulo da linfa ou obstrução dos vasos linfáticos. Éguas acima dos 10 anos de idade possuem maior incidência de cistos endometriais e linfáticos. Foram detectados cistos persistentes contendo maiores níveis de testosterona, menores de estradiol e maior porcentagem de ligantes IGF-I (Fator de Crescimento Insulínico I, do inglês *Insulin Growth Factor I*), possivelmente similar ao que ocorre em bovinos (BELTMAN et al., 2014). Sua importância na fertilidade se dá pela perda embrionária resultante de: implantação imprópria que resulta em nutrição inadequada, falha no reconhecimento maternal devido à falta de mobilidade do concepto, falha na liberação de prostaglandinas e modificações na placentação (STANTON; STEINER; PUGH, 2004).

Já os folículos anovulatórios ou hemorrágicos surgem a partir da falha na ovulação e persistência de uma estrutura dominante não-rompida, sem a liberação do oócito (revisado em CUERVO-ARANGO; NEWCOMBE, 2010). As características do folículo anovulatório incluem a presença de placas flutuantes, perda do formato circular do folículo, formação de uma protrusão cônica no lugar onde ocorreria a extrusão do oócito e turbidez do fluido folicular. É uma causa distinta de falha ovulatória, e sua relevância clínica reside na falha na ovulação do folículo dominante (revisado em GINTHER et al., 2007). A incidência de folículos anovulatórios foi detectada principalmente no final da temporada reprodutiva, afetando em média

20% das éguas avaliadas. Da mesma forma, ao avaliar a presença de folículos em ciclos estrais subsequentes em éguas velhas, foi detectado que 44% apresentaram a existência dessa estrutura. Nenhuma relação consistente com hormônios ou outras desordens foi encontrada (revisado em GINTHER et al., 2007).

Disfunções hormonais também representam uma preocupante fonte de infertilidade. A hiperprolactinemia leva à galactorreia, amenorreia e infertilidade e (NEWEY et al., 2013; ORMANDY et al., 1997) e observou-se que hormônios LH e FSH sofrem interferência desta desordem nas vias neuronais, levando a ocorrência de ciclos anovulatórios (SONIGO et al., 2012). Camundongos *knockout PRLR* apresentaram baixa taxa de ovulação e esterilização, oogênese aberrante, falha de implantação (BOLE-FEYSOT et al., 1998) e níveis de progesterona insuficientes (ORMANDY et al., 1997). Em cânceres de ovário, endometrial (LEVINA et al., 2009) e de mama (CLEVENGER et al., 2003), a Prl vem sendo estudada como um potencial biomarcador (HACHIM et al., 2016). Mutações do tipo SNP (Polimorfismo de Nucleotídeo Único, do inglês, *Single Nucleotide Polymorphism*) no gene PRLR foram relatadas em associação com esta desordem e infertilidade em humanos (NEWEY et al., 2013) e associada a taxa de fertilidade em bovinos (HUANG et al., 2009). Já em equinos, dois polimorfismos intrônicos na sequência do *PRLR* demonstraram maior taxa de prenhes por estro em garanhões Hanoverianos (GIESECKE et al., 2009). Não existem dados disponíveis sobre a influência desses SNPs em éguas, o que abre portas na identificação de novos biomarcadores da fertilidade associados a *PRL* e *PRLR*.

2.3 Genômica

A partir do sequenciamento do genoma equino realizado por (WADE et al., 2009), revelou que o mesmo possui aproximadamente 2.7 Gb. Foram encontradas 20.322 regiões codificantes e mais de um milhão de SNPs. Este estudo contribuiu enormemente na elucidação de aspectos evolutivos, permitiu detectar os polimorfismos entre as diferentes raças, identificar genes parálogos (originados antes da especiação) e ortólogos (originados por um ancestral comum). Além disso, houve a determinação dos blocos cromossômicos homo e heterozigóticos, variação no número de cópias (do inglês, *Copy Number Variation* - CNV) e que 46% do genoma é formado por sequências repetitivas. Análises transcriptômicas de 8

tecidos confirmaram a expressão de 87% dos 18.039 genes presentes no banco de dados Ensembl (http://www.ensembl.org/index.html). A sintenia dos cromossomos equinos com um cromossomo humano foi de 53%, além de 15.027 genes ortólogos com os humanos. SNPs associados à homozigose na pelagem leopardo, típica da raça Apaloosa, foram identificados como candidatos a biomarcadores, uma vez que estão associados ao desenvolvimento da detecção da cegueira noturna estacionária congênita. A detecção de marcadores moleculares evidencia o potencial do genoma para a detecção e diagnóstico de doenças e características, além de representar um bom parâmetro na escolha dos reprodutores.

A reprodução, como fator fundamental na criação equina, ainda apresenta dificuldades de aplicação da genômica, tendo literatura limitada, devido à complexidade da fertilidade equina (CHOWDHARY, 2013). Estudos relacionados à fertilidade de garanhões são mais abundantes. No estudo de Giesecke; Sieme; Distl (2010), revisa-se os achados de candidatos biomarcadores da reprodução em machos, relacionados à espermatogênese, ligação ao óvulo, maturação espermática e esteroidogênese. Dois polimorfismos intrônicos na sequência do *PRLR* demonstraram maior taxa de prenhes por estro em garanhões Hanoverianos (GIESECKE et al., 2009). Porém, não existem dados disponíveis sobre a influência desses SNPs em éguas. Pouco ainda se sabe sobre a influência da genômica sobre a reprodução das éguas, e estudos relacionados aos folículos e oócitos ainda são escassos comparado à outras espécies, como bovinos e humanos.

A expressão de genes em diferentes estágios do desenvolvimento oocitário foi descrita pela primeira vez por Leon et al., (2013). Utilizando oócitos imaturos (em metáfase I) e cultivados *in vitro* (em metáfase II), foi avaliada a expressão dos genes *Survivin* e *BCL-2*, anti-apoptóticos, *BAX* e *P53*, pró-apoptóticos. Não houve diferença de expressão relacionado ao estágio de maturação, mas em comparação às células do *cumulus*, a taxa de *Survivin* mostrou-se maior em oócitos, enquanto que a p53 mostrou-se maior em células da granulosa provenientes de oócitos com diferentes classificações morfológicas.

Em estudo anterior realizado pelo mesmo grupo, foi avaliada a associação entre SNPs no gene da *P53* e a ocorrência de abortos em éguas Puro Sangue Inglês. A distribuição do SNP no códon 72, que codifica para o genótipo mutado Arg/Pro, foi avaliado em éguas em diferentes condições fisiológicas e apresentando alguma desordem reprodutiva, como endometrite, abortos e ciclos anovulatórios. A

ocorrência de abortos demonstrou maior relação com o fenótipo mutado avaliado, indicando um possível biomarcador para esta desordem reprodutiva (LEON et al., 2012).

Ao avaliar a atividade da glicose-6-fosfato-desidrogenase (G6PDH), que tem atividade em oócitos em crescimento por seu papel no metabolismo, é possível considera-la um candidato biomarcador de viabilidade oocitária. O grupo de Mohammadi-sangcheshmeh et al. (2011) detectou que os oócitos equinos que possuíam maior atividade de G6PDH tinham maior proporção de complexos *cumulus-oócito* expandidos, maior tamanho e volume que oócitos sem a atividade enzimática. Além disso, o grupo ativo atingiu um grau maior de maturação (59.0% vs. 28.7%) e maior taxa de clivagem e embriões no estágio de blastocisto, o que corrobora com a expressão dos genes avaliados. Oito genes apresentaram-se em altos níveis de expressão em oócitos equinos imaturos, em comparação com o grupo que não continha essa atividade enzimática, atuando durante a replicação do DNA mitocondrial (*ATPV6E, IF-3* e *TFAM*) e regulação do ciclo celular durante a maturação oocitária (*DNMT1, STAT3, AURORA-A, ODC1* e *CKS2*). Outro estudo do mesmo grupo detectou que a atividade da G6PDH em células do *cumulus* expandidas em comparação com células do *cumulus* compactadas interferia em oito genes, nomeadamente *TFAM, STAT3* e *CKS2*, com expressão aumentada, e *COX1, ATPV6E* e *DNMT1*, com baixos níveis de expressão no *cumulus* expandido em éguas (MOHAMMADI-SANGCHESHMEH et al., 2014).

O transcriptôma equino foi analisado e anotado por diferentes grupos, sendo Coleman et al. (2010) um dos pioneiros a realizar o sequenciamento de RNA (RNA-seq) em equinos. Seis tecidos, cartilagem da articulação, membrana sinovial, vilosidades placentárias, testículos, cerebelo e tecidos embrionários de 34 dias foram utilizados na análise. A cobertura das sequências foi suficiente para refinar a anotação estrutural de 11.356 dos genes previstos e identificou 456 transcritos adicionais não listados nos bancos de dados Ensembl (http://www.ensembl.org/index.html) ou NCBI (https://www.ncbi.nlm.nih.gov/).

A fim de sair do campo das predições *in silico* e análise interespécies, em 2015 realizou-se o maior estudo de transcriptômica equina. A partir deste estudo, detectou-se que 780 variantes não estavam anotadas como ORFs (Fase de Leitura Aberta, do inglês, *Open Reading Frame*) no projeto genoma equino. Ao total, 68.594

transcritos provenientes de RNA-seq foram inseridos nos bancos de dados, utilizando 43 tecidos (HESTAND et al., 2015).

Da mesma forma, Mansour et al. (2017) utilizaram 59 indivíduos e oito tecidos para o refinamento do transcritoma, detectando 7.494 novos *loci* transcricionais candidatos e 36.876 genes. Os novos transcritos foram selecionados com base na sobreposição com genes da anotação equina atual e sua capacidade de codificação de proteínas, assim como a sobreposição com genes de outras espécies alinhados ou previsões de genes *ab initio*. Conseguiu-se produzir um transcritoma que incluía as regiões UTR (regiões não-traduzidas, do inglês, *Untranslated Region*), detecção de diferentes isoformas gênicas e novos genes.

As análises genômicas e transcriptômicas ainda estão em processo de aprimoramento no genoma, sendo que apenas recentemente foi completada a terceira versão do genoma equino (EquCab3.0). Com uma maior cobertura, foram reportados mais genes ortólogos aos humanos e também uma maior quantidade de genes similares anotados em espécies relacionadas (KALBFLEISCH et al., 2018). Com a evolução dos estudos genômicos, uma tendência promissora é avaliação de regiões não-codificantes, que têm importantes funções na regulação e manutenção e ainda permanecem sem um total descrição em equinos.

2.4 ncRNAs

Os ncRNAs são uma classe de moléculas que não resultarão em uma proteína. Estes são representados pelos RNAs ribossômicos (rRNA), que correspondem a 90% do total de RNAs em uma célula, seguido dos RNAs transportadores (tRNA). Acordou-se que os RNAS com mais de 200 nucleotídeos (nt) e que contenham ou não função conhecida são denominados lcnRNAs (RNAs longos não codificantes, do inglês, *long noncoding RNA*) (PALAZZO; LEE, 2015). Dentre os lncRNAs, existem os lincRNAs (ncRNAs Intergênicos Longos, do inglês, *long intergenic noncoding RNA*), que agem como *scaffold* dos complexos DNA-cromatina (HOTTIP, HOTAIR), inativando o cromossomo X (XIST), regulando a ação da telomerase e em mecanismos de *imprinting* (ESTELLER, 2011).

RNAs de tamanho intermediário incluem os snRNAs (Pequenos RNAs Nucleares, do inglês, *Small Nuclear RNA*), que apresentam entre 100 e 200 nt e podem ainda ser subdivididos em snoRNAs (Pequenos RNAs Nucleolares, do

inglês, *Small Nucleolar RNA*), que contêm até 100 nt. Sua função é associada principalmente com a modificação dos ribossomos e formação do complexo, tendo como exemplo os SNORDs. Além disso, podem interferir na formação de outros ncRNAs, como U6, que liga-se ao *spliceossomo* (ESTELLER, 2011).

Os pequenos RNAs dividem-se em pequenos RNAs de interferência (do inglês, *small interference RNA* - siRNA), micro-RNAs e RNAs que interagem com PIWI (*do inglês, PIWI interacting RNA* - piRNA), que geralmente apresentam-se em grande quantidade, o que tem relação direta com o tipo celular (PALAZZO; LEE, 2015). Dentre estes, os siRNAs agem principalmente no silenciamento de transposons, genes repetitivos e vírus (KIM, 2005); os miRNAs tem como função principal a regulação da transcrição de mRNAs, e os piRNAs estão envolvidos com a manutenção da integridade do genoma nas linhagens germinativas (ESTELLER, 2011).

Os miRNAs são moléculas não-codificantes amplamente descritas na literatura devido ao seu papel essencial na regulação pós-transcricional. Estes foram descritos pela primeira vez por Lee; Feinbaum; Ambros (1993) em *Caenorhabditis elegans*. Sendo em seguida descoberta a família let-7, por Reinhart et al. (2000). A estrutura de um precursor de miRNA típico equino é mostrada na Figura 3.

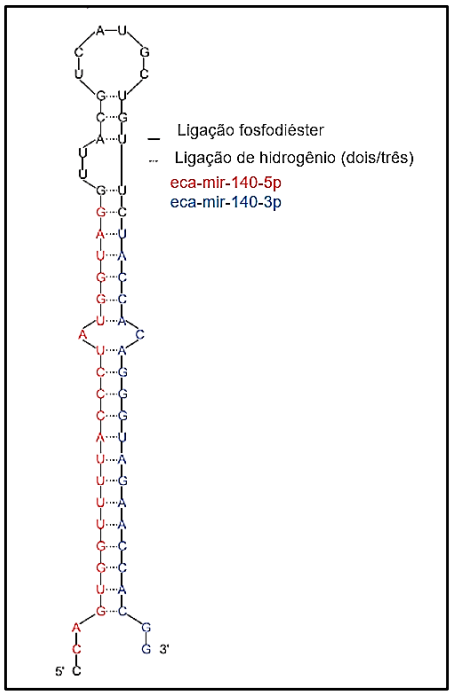


Figura 3. Estrutura de um precursor de miRNA equino. Adaptado de VAN DER KOLKA, PACHOLEWSKA, GERBER (2015).

Os miRNAs são RNAs endógenos de fita simples contendo em torno de 25 nt, codificados no genoma. Sua origem pode ser canônica, quando encontra-se em regiões intergênicas e é clivado pela proteína Drosha, ou não-canônica, através dos *mirtrons* – pré-miRNAs clivados a partir de íntrons através de *splicing* alternativo (ESTELLER, 2011). Em plantas, os genes que codificam para miRNAs (genes *MIR*) estão presentes em regiões intergênicas e geralmente provém de uma duplicação inversa do seu gene-alvo (VOINETT, 2009). Já em animais, os genes de miRNAs costumam se encontrar em íntrons de regiões codificantes ou não e agrupam-se em unidades de transcrição policistrônica, formando *clusters* de miRNAs (VOINETT, 2009).

A via canônica inicia com a transcrição das regiões gênicas correspondentes ao miRNA pela RNA polimerase (RNA Pol) II ou III, gerando um transcrito primário (pri-miRNAs) (Figura 4). Os pri-miRNAs tem uma conformação em cadeia dupla e são mais longos, geralmente com mais de 1 kb e contém em sua estrutura um ou mais *loops*. No núcleo, sofrem processamento substancial pelo complexo microprocessador, formado pela proteína Drosha e seu cofator DGCR8, clivam a estrutura em suas partes finais, resultando na geração de um miRNA de aproximadamente 65 nt (pré-miRNA). Drosha possui dois domínios de RNAse III em tandem que se dimerizam intramolecularmente e cortam as fitas 3’ e 5’, e DGCR8 possui dois domínios de ligação à fita dupla. O transporte para o citoplasma ocorre por meio da exportina-5 (XPO-5), associada a GTPAse Ran (revisado em HA; KIM, 2014; KIM, 2005)

Uma vez no citoplasma, o pré-miRNA é submetido ao processamento pela Dicer, outra RNAse III formada por um centro catalítico de repetições em tandem e o domínio PAZ (piwi-argonauta-zwili). O corte é de aproximadamente 2 nt na extremidade 3’ gerada pela Drosha e a distância entre os domínios PAZ e a RNAse III é usada como régua para determinar o local de clivagem. Geralmente, esse corte ocorre a 25 nt de distância, em posições variáveis que podem gerar diferentes miRNAs maduros (revisado em HA; KIM, 2014).

O pequeno duplex de miRNA clivado pela Dicer é inserido no complexo RISC (Complexo de Silenciamento Induzido por RNA, do inglês, *RNA Interfering Silencing Complex*), formado com a ligação da proteína Argonauta (Ago) à Dicer. No duplex de miRNA, um deles será definido como RNA-guia, enquanto que o outro é o RNA-passageiro e geralmente é degradado. A fita passageira, geralmente degradada, é

designada com a adição de um asterisco ao código do miRNAs. O domínio PIWI é similar a uma RNAse H e contém um sítio ativo de clivagem do mRNAs (RNA mensageiro) alvos entre as posições nucleotídicas 10 e 11. Apenas AGO2 cliva mRNAs perfeitamente encaixados no complexo, (revisado em HA; KIM, 2014). O complexo resultante é dirigido ao mRNA-alvo que contenha sequências complementares ao miRNA-guia, sendo que a inibição da tradução ou degradação do mRNA dependerá da complementaridade de ligação entre as duas moléculas. Os miRNAs altamente correspondentes influenciarão a degradação do mRNA pela Ago, enquanto que uma menor complementaridade levará à inibição da tradução (KIM, 2005).

Existem isoformas dos miRNAs, causadas por uma variabilidade na região 3' ou substituições nucleotídicas fora da sequência-guia e hipotetiza-se que são resultantes das ações de enzimas envolvidas na clivagem da sequência-guia. Embora os efeitos funcionais destas isoformas não sejam totalmente conhecidos, a eficiência de ligação ao alvo, incorporação à Ago e estabilidade podem estar sendo comprometidas (NEILSEN; GOODALL; BRACKEN, 2012).

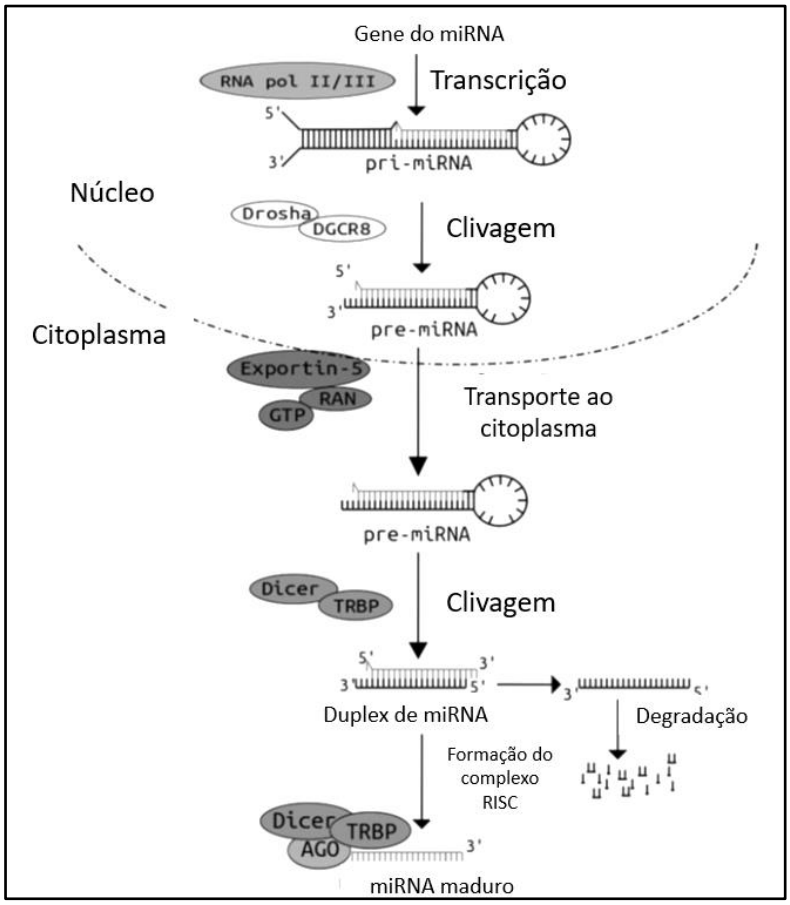


Figura 4. Biogênese dos miRNAs. Adaptado de (YOUSEF; JENS, 2014).

Existem diversos mecanismos pelos quais ocorre a regulação pós-transcricional, sendo esta geralmente negativa: quanto maior a expressão do miRNA, menor a do gene, e vice-versa. Após o início da tradução, os miRNAs podem agir degrandando os mRNAs alvos, reprimindo sua tradução, bloqueando o alongamento da tradução ou promovendo a dissociação prematura dos ribossomos. Uma degradação proteica co-traducional também pode ocorrer, onde a tradução não seria inibida, mas haveria a degradação da cadeia polipeptídica nascente concomitantemente. Os miRNAs podem interferir no alongamento da cadeia polipeptídica através da competição das proteínas Ago pela ligação à estrutura *cap*, assim como impedir que as subunidades ribossomais se unam (revisado em EULALIO; HUNTZINGER; IZAURRALDE, 2008).

A identificação do perfil de expressão de lncRNAs equinos foi realizada apenas em 2017. Após o estudo de Mansour et al. (2017), as sequências curtas, de baixa expressão, baixa conservação entre espécies e menor diversidade de éxons foram analisadas por Scott et al. (2017). Estas sequências geralmente são descartadas durante os processos de refinamento por não terem nenhum gene similar. O estudo detectou sequências de 20.800 transcritos, com uma distribuição similar àquela dos genes anotados (SCOTT et al., 2017).

Já a descrição do perfil de miRNAs em tecidos equinos foi relatada por Kim et al. (2014), que analisou o cólon, musculo esquelético e fígado. Foram descritos 292 miRNAs já conhecidos e 329 novos, em comparação com estudos anteriores *in silico.* Pacholewska et al. (2016) identificaram 698 novos miRNAs em nove tecidos através de RNA-seq incluindo 3 diferentes origens de músculo esquelético, coração, fígado, cartilagem, ossos e sangue.

O perfil de miRNAs associados aos tecidos reprodutivos ou a desordens reprodutivas em fêmeas ou machos não se encontram disponíveis na literatura até o momento. Alguns estudos pontuais associando os miRNAs ao desenvolvimento folicular e oocitário foram realizados e serão descritos a seguir.

O primeiro estudo relacionando miRNAs à fertilidade foi realizado por da Silveira et al. (2012), que detectaram a presença de microvesículas e exosomos no fluido folicular, que possuíam mRNA e miRNAs, sendo os maiores números obtidos nas microvesículas. 25 miRNAs foram detectados nas microvesículas e exosomos, sendo relacionados a via *WNT*, envolvida na esteroidogênese, luteogênese e

crescimento folicular. Supõe-se que estas estruturas participem da comunicação celular, entregando mRNAs e miRNAs como mensageiros, embora especule-se que sua origem seja intra-ovariana. Em uma avaliação entre o fluido folicular de éguas jovens e velhas, obteve-se que os miRNAs mir-181a, mir-375 e mir-513a-3p estão diferencialmente expressos no fluido folicular de éguas velhas em comparação às jovens. Estes participam da via de sinalização do *TGFB*, crucial para o correto desenvolvimento folicular.

Os miRNAs envolvidos no desenvolvimento folicular foram obtidos através do estudo de Schauer et al. (2013), que encontrou expressão diferencial relacionada a sobrevivência das células da granulosa (mir-21, mir-23a), proliferação (mir-145, mir-503, mir-224) e/ou produção de estradiol (mir-224, mir-383, mir-378). Neste estudo também foi avaliada a expressão em folículos dominantes e subordinados tendo um aumento nos níveis de mir-145, mir-378 e mir-132 no fluido folicular de folículos subordinados e um aumento nos níveis foliculares de mir-21, mir-132, mir-212 e mir-224 após a administração de uma dose ovulatória de hCG (Gonadotrofina Coriônica Humana, do inglês, *Human Chorionic Gonadotrophin*).

A análise da expressão diferencial de folículos ovulatórios e anovulatórios foi realizada por Donadeu; Schauer, (2013), conforme repertório encontrado em outras espécies. Os folículos foram aspirados e foi realizado qRT-PCR para a quantificação da expressão. Os folículos anovulatórios apresentaram altos níveis de mir-21, mir-23b, mir-378 e mir-202 no fluido folicular em comparação com folículos ovulatórios. O mir-224 e mir-383 não puderam ser detectados. Este é o único estudo que relaciona aspectos da infertilidade aos miRNAs (DONADEU; SCHAUER, 2013).

A atividade dos genes *ACVR1* e *ID2*, cruciais para o desenvolvimento oocitário, foi avaliada nos exossomos do fluido folicular no diestro e em folículos pré-ovulatórios. Os miRNAs mir-27b, mir-372 e mir-382 estavam presentes nos exossomos; os dois primeiros, em altos níveis de expressão no diestro, e o último, em baixos níveis, no folículo pré-ovulatório. Esse dado corrobora com o fato do gene *ACVR1* estar 3x mais expresso no folículo pré-ovulatório (DA SILVEIRA et al., 2014).

Recentemente, foi demonstrado que mais de 60% dos miRNAs equinos conhecidos têm correspondências perfeitas com miRNAs associados a doenças humanas (BUZA et al., 2014). Aliado a isso, em mulheres e camundongos, existe extensa literatura experimental relacionando o papel dos miRNAs à reprodução e

viabilidade folicular (LI et al., 2015), sendo possível fazer uma transposição dos conhecimentos já adquiridos de outras espécies para a fertilidade em éguas.

Na reprodução, os miRNAs foram associados a casos de infertilidade relacionados a síndrome do ovário policístico (SOP) (ROTH et al., 2014). Mulheres em procedimento de FIV, férteis ou portadoras de SOP, foram analisadas quanto ao perfil de expressão de miRNAs por *microarray*. Vinte e nove miRNAs estavam diferencialmente expressos e cinco se mostraram elevados no grupo com SOP, sendo estes mir-32, mir-34c, mir-135a, mir-18b e mir-9. Uma análise quantitativa de diversos potenciais genes-alvo revelou que *IL8, SYT1* e *IRS2* estavam significantemente diminuídos no grupo portador de SOP. Estes abrangem aspectos da fisiopatologia da doença, como participação no metabolismo de carboidratos e função das células beta, comunicação intercelular e síntese de esteroides (ROTH et al., 2014).

Mulheres com endometriose também foram avaliadas no estudo de Burney et al. (2009). Em comparação com mulheres normais, as famílias mir-9 e mir-34 se mostraram diferencialmente expressas no ensaio de *microarray*, com uma redução de 25x no primeiro. Os genes-alvo detectados para cada mirRNA foram *CCNE2, CDK4, CDK6, CDC25A* e *BCL2*, envolvidos na manutenção do ciclo celular.

A insuficiência prematura ovariana (IOP) foi avaliada através de qRT-PCR e descreveu níveis aumentados de mir-23a no plasma de mulheres diagnosticadas com IOP, em comparação com mulheres com ciclos normais. Este miRNA demonstrou ter como gene-alvo XIAP e a cascata de sinalização das caspases, que tem relação íntima com a regulação da apoptose. Assim, foi possível relacionar o mir-23 com a apoptose das células da granulosa, ao diminuir os níveis de expressão de XIAP, inibidor endógeno das caspases (YANG et al., 2012).

Mesmo em casos sem causa aparente, como múltiplas falhas em tentativas de fertilização *in vitro,* foi relatado um perfil de expressão desregulado. O estudo de Karakaya et al. (2015) coletou oócitos de mulheres em procedimentos de FIV e detectou que o mir-21 estava com expressão aumentada em 5x nas mulheres pouco responsivas.

2.5 Bioinformática

O desenvolvimento da bioinformática ocorreu em conjunto com o avanço do estudo de proteínas, e posteriormente, com as técnicas moleculares. O sequenciamento de DNA, que surgiu na década de 1970, assim como a utilização do DNA recombinante e o desenvolvimento da PCR, gerou uma quantidade cada vez maior de arquivos e dados. A criação de bancos de dados de sequências e surgimento de ferramentas que facilitassem a manipulação computacional do DNA, a partir dos anos 90, permitiu análises cada vez mais complexas, como as dos genomas completos de organismos procariotos e eucariotos (GAUTHIER et al., 2018).

Com a evolução dos estudos genômicos, transcriptômicos e proteômicos, e a partir do momento em que se tornou possível analisar qualitativa e quantitativamente o material genético, técnicas mais sofisticadas foram desenvolvidas, como a qRT-PCR, microarranjos e RNA-seq. *Softwares* de predição ganharam espaço, a fim de auxiliar na elucidação de sequências não designadas nas bases de dados geradas por estes ensaios, tendo em vista diversas funções: encontrar sequências codificantes, descrever sua função e comparar sua similaridade com outras sequências (tanto para o DNA quanto RNA e proteínas), encontrar sítios ativos em proteínas, descrever a interação com outras moléculas e sua possível conformação (NAIR, 2007). As análises estruturais e funcionais *in silico* foram crescendo e hoje são empregadas antes mesmo da realização de experimentos *in vitro.* Existem três principais métodos computacionais empregados nas análises de dados, que incluem a predição *ab initio*, baseada nas características da sequência ou estrutura, comparativa, que leva em conta a conservação evolucionária e a predição integrada destes dois últimos métodos (LESK, 2014).

2.5.1 Modelagem de proteínas

Os métodos experimentais de determinação de estruturas fornecem informações de qualidade sobre a estrutura de subconjunto de proteínas (BAKER; SALI, 2001). As técnicas padrão para o estabelecimento de uma estrutura proteica incluem a cristalografia de raios X e NMR (Ressonância Nucelar Magnética, do inglês, *Nuclear Magnetic Ressonance*) (revisado em LESK, 2014). Embora efetivas, são experimentos laboriosos e custosos. Assim, os métodos de predição

computacional fornecem informações valiosas para uma grande parte de sequências cujas estruturas não serão determinadas experimentalmente (BAKER; SALI, 2001).

Para a avaliação de proteínas preditas, existe um evento bienal que realiza testes cegos de avaliação das predições de estrutura, o CASP (Avaliação Crítica da Predição Estrutural, do inglês, *Critical Assesment of Structure Prediction*). As predições de estrutura tradicionalmente foram categorizadas em modelagem comparativa, predição de reconhecimento conformacional e predição de nova conformação. Alguns critérios separaram a predição de estruturas de proteínas em três categorias, como revisado em Lesk, (2014):

(1) Homologia: se uma proteína de estrutura desconhecida tiver homólogos de estrutura conhecida, com 40% ou mais resíduos idênticos em um alinhamento de qualidade, os métodos de modelagem de homologia auxiliam na obtenção de um modelo estrutural quase completo e de qualidade boa o suficiente para interpretar a função da proteína. Quanto maior a similaridade da sequência, mais preciso é o modelo.

(2) *Threading*: caso nenhum homólogo estrutural conhecido apresentar similaridade com 40% ou mais de identidade por resíduo, ainda será possível atribuir um padrão geral de conformação à proteína de estrutura desconhecida. Deve ser possível prever sua estrutura secundária com uma precisão de 70 a 80% numa base resíduo-por-resíduo.

(3) Se nenhum homólogo de estrutura conhecida for reconhecível a partir das sequências, o último recurso é usar um método de previsão geral o suficiente para lidar com novas dobras. Tais métodos incluem abordagens *a priori* e utilização de *softwares* baseados no conhecimento estrutural de proteínas.

Threading é um valioso método para o reconhecimento da conformação. Dada uma biblioteca de estruturas conhecidas e a sequência de uma proteína de estrutura desconhecida, constroem-se diversos modelos aproximados usando diferentes alinhamentos possíveis destas sequências (BAKER; SALI, 2001). Ambos os modelos de *threading* e homologia lidam com a estrutura tridimensional induzida por um alinhamento da sequência de consulta com estruturas conhecidas de homólogos. A modelagem de homologia se concentra em um conjunto de alinhamentos, tendo como objetivo um modelo muito detalhado; já a modelagem por *threading* explora muitos alinhamentos possíveis e lida com modelos aproximados

(LESK, 2014); e podem às vezes revelar relações mais distantes do que métodos puramente baseados em sequência (BAKER; SALI, 2001).

Nesse contexto, ainda diversas proteínas de relevância biológica permanecem sem elucidação estrutural. Na reprodução, as moléculas da prolactina e seu receptor desempenham importantes papéis fisiológicos. Em humanos (NEWEY et al., 2013) e bovinos (HUANG et al., 2009) já é conhecida a interferência de mutações na prolactina e/ou seu receptor na fertilidade, gerando distúrbios como hipo e hiperprolactinemia, galactorreia, ciclos anovulatórios e infertilidade. Porém, em éguas, ainda é ausente uma descrição entre polimorfismos genéticos e modificações na fertilidade, embora a tentativa de fornecer dados estruturais e funcionais sobre as proteínas equinas já tenha aberto caminho para possíveis estudos de impacto dessas mutações (NEIS et al., 2016). Entretanto, análises funcionais e estruturais e sua avaliação como um candidatos a biomarcadores são dificultadas pela inexistência de modelos estruturais na espécie equina. Dessa maneira, a análise através do método de *threading* se apresenta como uma alternativa viável para a realização de estudos bioinformáticos voltados à reprodução em éguas.

2.5.2 Detecção *in silico* de ncRNAs

Estudos *in vitro* voltados à detecção de ncRNAs empregam ensaios de amplificação, como qRT-PCR, testes de hibridização, como os microarranjos, e sequenciamento de nova geração, os quais são os métodos mais utilizados para sua detecção e validação como biomarcadores (LESK, 2014). Muitas vezes, as abordagens experimentais podem ser inadequadas, tanto pelas limitações de cada técnica, quanto pelo preparo e tipo de amostra. Além disso, muitos possuem expressão tecido-específica, baixos níveis de expressão ou condição específica que dificulta sua detecção experimental (YOUSEF; JENS, 2014).

A fim de solucionar alguns destes problemas, vários métodos computacionais foram projetados e aplicados à detecção de genes de miRNAs, sendo que inúmeras abordagens para a predição *in silico* foram criadas até agora. Esses programas geralmente consideram a estrutura secundária do miRNA precursor como a característica mais importante de um gene de miRNA. Os algoritmos de predição geralmente são empregados na análise da estrutura secundária e a estabilidade

termodinâmica das estruturas *hairpin*, e os métodos existentes consistem de uma estimativa genômica ampla destas estruturas, filtragem ou classificação com base em sua similaridade aos miRNAs conhecidos, e confirmação experimental de candidatos putativos. Para extrair possíveis miRNAs de um genoma, métodos de baseados em homologia e *ab initio* são usados (HUANG et al., 2017).

Os métodos de mapeamento de genes de miRNAs baseados em homologia ou *learning-based* utilizam miRNAs disponíveis e validados experimentalmente a fim de encontrar estruturas e sequências similares em espécies relacionadas. Se um miRNA é identificado em um genoma, seus homólogos podem ser encontrados em outras espécies. Porém, caso a finalidade for encontrar novos miRNAs, a abordagem por homologia representa uma limitação; incluído a isso, muitos miRNAs são espécie-específicos. Neste sentido, a predição *ab initio* oferece uma boa alternativa, pois requer a utilização de parâmetros manualmente selecionados, necessitando apenas a sequência primária. A principal dificuldade deste método é a escolha adequada destes parâmetros, para que permitam determinar se uma sequência é realmente um miRNA baseado em suas propriedades. A estrutura do *hairpin* e a energia livre mínima são muito empregados para a identificação de miRNAs (YOUSEF; JENS, 2014).

Levando em conta a importância da regulação gênica através de ncRNAs, o baixo número de estudos relacionados em éguas e a similaridade no perfil de ncRNAs regulatórios entre espécies, se propõe utilizar a bibliografia relacionada a desordens reprodutivas e uma análise no genoma equino para selecionar miRNAs candidatos a biomarcadores que podem relacionar-se a disfunções reprodutivas em éguas.

3 Objetivos

Objetivo geral

Utilizar ferramentas bioinformáticas aplicadas ao estudo de fertilidade em éguas, através do estabelecimento da estrutura do hormônio prolactina e seu receptor e da identificação de RNAs não-codificantes no genoma equino.

Objetivos específicos

-Estabelecer a estrutura *in silico* da proteína prolactina equina, através de ferramentas bioinformáticas de predição de estrutura e refinamento;

-Estabelecer a estrutura *in silico* do receptor da prolactina equina através de ferramentas bioinformáticas de predição de estrutura e refinamento;

-Identificar RNAs não-codificantes no genoma equino possivelmente impactantes na fertilidade de éguas;

-Relacionar os RNAs não-codificantes encontrados com distúrbios no desenvolvimento oocitário, folicular e esteroidogênese;

-Fornecer RNAs não-codificantes candidatos a biomarcadores para a realização de estudos experimentais *in vitro* em éguas.

4 Artigo

Artigo submetido à revista Molecular Diversity (fator de impacto 2.229)

Title:

***In silico* prediction of prolactin molecules as a tool for equine genomics reproduction**

Short title: Prediction of PRL and PRLR equine structures'

Neis A [1]; Kremer F S[2]; Pinto L S[2]; Leon P M M[1#]

[1]Grupo de Pesquisa em Genômica de Equinos - GenE, Núcleo de Biotecnologia, Centro de Desenvolvimento Tecnológico, Campus Universitário, Universidade Federal de Pelotas, Pelotas, Rio Grande do Sul, Brazil, Caixa Postal 96010-900.

[2]Laboratório de Bioinformática e Proteômica, Núcleo de Biotecnologia, Centro de Desenvolvimento Tecnológico, Campus Universitário, Universidade Federal de Pelotas, Pelotas, Rio Grande do Sul, Brazil, Caixa Postal 96010-900.

corresponding author:

E-mail: primleon@gmail.com

Phone number: +55532757350

Acknowledgements: The authors declare there was no conflict of interest. This project was developed by a Biotechnology undergraduate student through supporting scholarship from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). FSK is student of the Graduate Program in Biotechnology at Universidade Federal de Pelotas also supported by CNPQ.

Abstract

The prolactin hormone is involved in several biological functions, although its main role resides on reproduction. As it interferes on fertility changes, studies focused on human health have stablished a linkage of this hormone to fertility losses. Regarding animal research, there is still a lack of information about the structure of prolactin. In case of horse breeding, prolactin has a particular influence, once there is an individualization of these animals and equines are known for presenting several reproductive disorders. As there is no molecular structure available for the prolactin hormone and receptor we performed several bioinformatics analyses through prediction and refinement softwares, as well as manual modifications. Aiming to elucidate the first computational structure of both molecules and analyse structural and functional aspects related to these proteins, here we provide the first known equine model for prolactin and prolactin receptor, which obtained high global quality scores in diverse software's for quality assessment. QMEAN overall score obtained for ePrl was (-4.09) and QMEANbrane for ePrlr was (-8.45), which proves the structures' reliability. This study will implement another tool in equine genomics in order to give light to interactions of these molecules, structural and functional alterations and therefore help diagnosing fertility problems, contributing in selection of a high genetic herd.

Keywords: mares; infertility; PRL; PRLR; bioinformatics.

Full Text

Introduction

The prolactin hormone (Prl) is a molecule synthesized and secreted by lactotrophs in the pituitary gland. PRL molecule interacts with the prolactin receptor (Prlr) and its activation starts when a trimeric complex of the two receptors and the hormone is formed (BOLE-FEYSOT et al., 1998; GOFFIN et al., 1996). The Prl-Prlr interaction stimulates several biological processes related to reproductive regulation acting mainly in mammogenesis, lactogenesis, galactopoiesis, reproductive receptiveness and maternal behaviour (FREEMAN et al., 2000).

The major dysfunction is hyperprolactinemia, causing galactorrhea, amenorrhea and infertility, as observed in humans and rodents (NEWEY et al., 2013; ORMANDY et al., 1997). It also interferes in neuronal pathways related to luteinizing hormone (LH) and follicule stimulating hormone (FSH), thus leading to anovulatory cycles (SONIGO et al., 2012). In knockout *PRLR* mice, infertility related to decreased ovulation rate, aberrant oogenesis, implantation failure (BOLE-FEYSOT et al., 1998) and insufficient progesterone levels (ORMANDY et al., 1997) confirm the luteal dependence of Prl. Single-nucleotide polymorphisms (SNPs) in PRLR causing genetic hyperprolactinemia and infertility in humans were described (NEWEY et al., 2013) as well as in stallions, where they could be implied as biomarkers for fertility (GIESECKE et al., 2010). Prl can be used to stimulate ovulation in mares, but since there is no mass source of this molecule its use has been restraint (THOMPSON; OBERHAUS, 2015).

For its relevance in fertility, we modelled the equine prolactin (ePrl) and entire mature structure of equine prolactin receptor (ePrlr) long isoform including a probable intracellular

domain (ICD). Once there are only human solved models available structures in databases, we used bioinformatics tools for prediction, refinement and quality assessment to obtain the first equine molecules. This can provide a better understanding of the molecular features', protein-protein interactions and potential functional and structural impacts of mutations in coding sequences. Therefore, the number of animals and individual enrichment can be affected, decreasing economic loss with reproductive failures and promoting novel discoveries in mare reproduction and fertility.

Materials and Methods

Sequence accession

Information about ePrl was obtained from UniProt (APWEILER et al., 2004) (Uniprot:P12420). The signal-peptide was removed once the final protein structures does not include this sequence. In order to determine the molecular structure of ePrlr, the protein sequence available at Uniprot (Uniprot: F6Z164) provided information about its domains: extracellular domain (ECD), transmembrane (TM), ICD and signal-peptide. The absence of a whole model and physic-chemical characteristics of TM proteins made necessary to run each part of ePrlr separately (ECD, TM and ICD). In this way, the sequence was split using PyMOL Molecular Graphics System (Version 1.8 Schrödinger, LLC).

Structure modelling

I-TASSER (ROY; KUCUKURAL; ZHANG, 2010) software was used to generate the 3D predicted model from amino acid sequences, once it does not use solved structures as model to predict proteins. For ePrl, the generated 3D model was run straight through refinement tools. After ePrlr parts were modelled, AIDA (XU et al., 2014) software enabled their attachment.

Structural refinement

The atomic-level protein structure refinement was made through Fragment Guided Molecular Dynamics (FG-MD) (XU et al., 2011), re-shaping the molecular arrangements energy landscape. ModLoop (FISER; SALI, 2003) was implied to predict loop conformations without relying on a database of known protein structures. However, in case of ePrlr, a manually adjustment of the molecular arrangement by correcting wrong positioned atoms and chains was made before, using PyMOL Molecular Graphics System (Version 1.8 Schrödinger, LLC).

Quality assessment

To assess model reliability ModFold6 (MAGHRABI; MCGUFFIN, 2017) was used, measuring global model quality. QMEAN (BENKERT; BIASINI; SCHWEDE, 2011) and QMEANbrane (STUDER; BIASINI; SCHWEDE, 2014) evaluate global and local quality of database models and predicted transmembrane structures. RAMPAGE (BATUT; GINGERAS, 2013) verifies correctly positioned amino acids and model reliability. DFIRE (ZHOU; ZHOU, 2002) and dDFIRE (YANG; ZHOU, 2008) measure conformational free energy. SolvX (HOLM; SANDER, 1992) determines solvent availability. ProQ2 (RAY; LINDAHL; ORN WALLNER, 2012) was used for residue quality and PsiPred (MCGUFFIN; BRYSON; JONES, 2000) for secondary structure prediction. Comparative analysis with human molecules were made, using PDB solved structures for Prl (PDB: 1RW5), although Prlr did not present any complete ECD and TM model. After quality assessment and proving of their reliability, both molecules were submitted to ModelArchive database of predicted structures (www.modelarchive.org).

Results and Discussion

ePrl has 229 amino acids and is composed by four main alpha helix. The first predicted ePrl molecule can be seen in Figure 1. It was modelled using I-TASSER and had few sequence adjustments, relative to removal of signal-peptide. The hormone did not require manual modifications. ePRLR structure is formed by three domains within 625 amino acids. The first ePrlr 3D molecule predicted was constructed respecting these domains, that were then gathered as seen in Figure 2. ECD is related to hormone binding, TMD cross the cell membrane bilayer and ICD is intrinsically flexible and structurally disordered that interacts with intracellular signalling molecules. A JAK2 (Janus-kinase 2) constitutively bind to the structure is located in BOX1, relatively conserved proline-rich motif region necessary for consensus folding of the molecule recognized by the transduction molecules. For JAK2 activation, both receptor dimerization and box 1 presence are required. Manual correction evaluating incorrectly positioned amino acids and secondary structure was used to adjust global conformation and membrane orientation. It is important to notice that this model of the long isoform ePrlr is not dimerized, what occurs via hormone binding and activation of a trimeric complex.

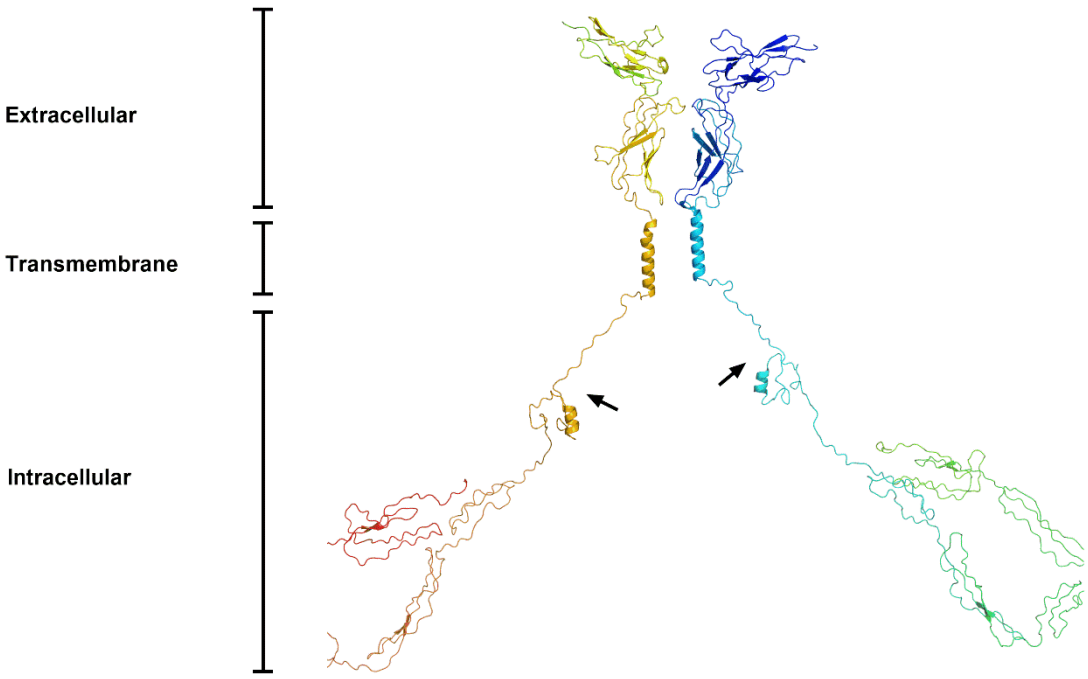


Fig. 1 Predicted model for activated equine prolactin receptor, containing two receptor molecules forming the binding site. The three different domains are separated (ECD, TMD and ICD). The arrows indicate BOX1, where JAK2 constitutively bind to the structure is located.

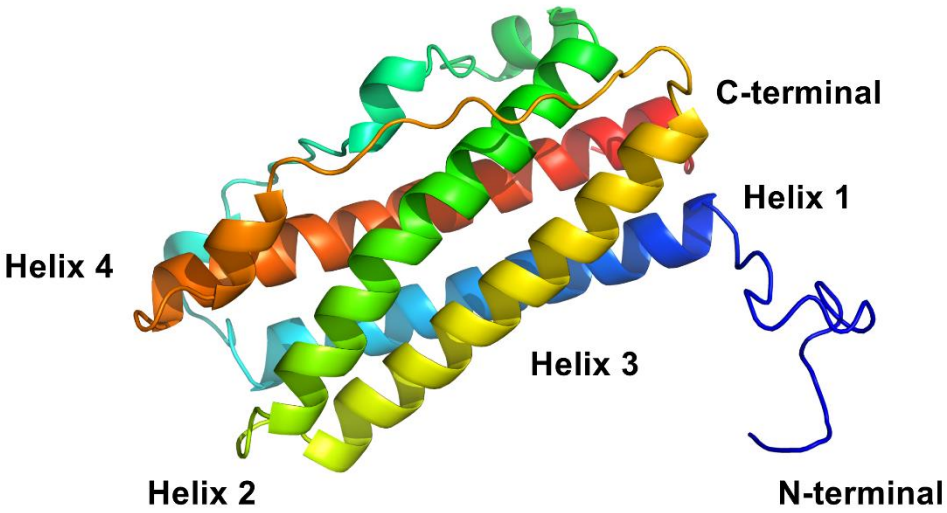


Fig. 2 Protein model of equine prolactin. Indicated in the structure are the C-terminal and N-terminal, as well as the main four alpha helixes

In this prediction, threading modelling was performed, once the equine structure is not available in public databases, comparing the target sequence directly to solved template proteins even in the absence of an evolutionary relationship (ZHANG, 2009). I-TASSER structures' quality resembles RMSD ranges from 1-2 Å from solved experimental structures in diverse organisms (READ; CHAVALI, 2007) and has approximately 70% reliability on designing (DU et al., 2015). There are available tools for modelling TM proteins, but I-TASSER was used once it was a threading modelling and other tools does not support this kind of prediction. Even not being available a 'transmembrane mode', constantly made Critical Assessment of protein Structure Prediction (CASP) experiments prove its reliability and usage over time, with predictions similar to native models (YANG et al., 2016). ePRL and ePRLR generated structures were refined at the atomic level by FG-MD software, and a new structure was generated for both after refinement. On CASP refinement experiments, FG-MD demonstrated significant potential (ZHANG; LIANG; ZHANG, 2011).

To assess the quality of 3D models, scores greater than 0.4 indicate more complete and confident models in ModFold6. In both designed molecules, values of 0.6996 for ePrl and 0.3950 for ePrlr were achieved. Both structures were evaluated for loop conformations and molecular arrangement. The receptor molecule had a decreased score due to present different characteristics combined in a single model, increasing residues' error. The software provides reliable information about predicted models, even though there is no specific analysis on TM quality, ModFold6 series of methods proved to be the best in estimating absolute scores of models in CASP12 (KRYSHTAFOVYCH et al., 2018)

QMEAN Z-score provides an estimate of absolute quality of an individual model in relation to reference structures solved by crystallography. ePrl score (-4.09) shown a good model reliability (Online Resource 1), once it is measured by the resulting scoring function independent of the size of the protein. Generally, the more negative index, the better result

(BENKERT; BIASINI; SCHWEDE, 2011). ePrlr scored -8.45 in QMEANbrane, a specific TM tool, which is within the expected range for TM proteins (Online Resource 2). QMEAN is still in the best performing methods for model quality (KRYSHTAFOVYCH et al., 2018) and QMEANbrane also showed superiority among other TM quality assessment softwares (STUDER; BIASINI; SCHWEDE, 2014).

ePrl structure demonstrates high reliability by DFIRE (-363.43) and dDFIRE (-514.85), once the lower index the less instable the model is. After refinement analysis and considering the complexity of ICD in ePrlr, the model presents a considerable reliability (-612.36 on DIFRE and -732.95 on dDFIRE), once it was designed by threading modelling and the molecule has specific physic-chemical properties, which interfered in the obtained scores.

ProQ2 analysis demonstrates ePrl high quality model, closer to score 1 (Online Resource 3A). ePrlr obtained a lower residue score due to its multiple characteristics and disordered conformation of ICD (Online Resource 3B). In recent analyses of ProQ2, it was between the single-model methods that performed best in local and global model accuracy (KRYSHTAFOVYCH et al., 2018; UZIELA; WALLNER, 2016). Quality assessment revealed that ePRL reliability resides in 96% and ePrlr long isoform obtained 92.5% of favoured or allowed amino acids in right position, according to RAMPAGE.

As SolvX assess the solvation profile and a well-packed structure should have an overall solvation preference below zero (HOLM; SANDER, 1992), it can be inferred that our ePrl model is trustworthy (score -17.2). However, ePrlr structure (258.3) presented high levels of solvent accessibility, probably due its membrane protein nature and different environments where it is included (KOEHLER LEMAN; ULMSCHNEIDER; GRAY, 2015).

PsiPred results for single residue quality and topology (MCGUFFIN; BRYSON; JONES, 2000) show our structures having an optimal residue-per-residue quality and high trustworthy range (Online Resources 4 and 5). Secondary structure prediction of the four main

alpha helix in ePrl and the TM region in ePrlr are accurate. Although generally implied for soluble proteins (ALMEIDA et al., 2017), its score provide accurate usefulness also for TM proteins (KOEHLER LEMAN; ULMSCHNEIDER; GRAY, 2015) being the most described tool for secondary structure refinement (KASHANI-AMIN et al., 2018).

Conclusions

ePrl has shown to be a highly trustworthy model, that can be used for further structural and functional analyses regarding interactions with the receptor or other molecules. ePrlr, a TM protein, had to pass through more rounds of refinement tools and the quality assessment carefully analysed regarding its properties. Considering that we modelled a TM protein that had no previous structure in the databases, our model presents usability and biological relevance. We also defined a hypothetic ICD region, domain that still does not have a solved structure. In this study, we present the first ePrl and ePrlr molecules to date obtained and analysed by bioinformatics tools for structure prediction, loop corrections and molecular dynamics refinements. Our predicted models could contribute to novel discoveries related to mare reproduction, stimulating *in vitro* studies to analyse mutations and other genetic related structure alterations, understanding molecular dynamics of ligands, protein-receptor interaction, as well as docking with other molecules and impact of structural changes. From this new molecules it will be possible to analyse and predict interactions with analogs of prolactin, qhat can be possibly implied as an alternative inductor of ovulation.

The predicted structures are available in ModelArchive database and can be accessed through: ePrl (DOI: 10.5452/ma-ahb5e) and ePrlr (DOI: 10.5452/ma-acf71). The respective PDB files are available in Online Resource 6 and 7.

References

1. Bole-Feysot C, Goffin V, Edery M, et al (1998) Prolactin (PRL) and Its Receptor: Actions, Signal Transduction Pathways and Phenotypes Observed in PRL Receptor Knockout Mice. Endocr Rev 19:225–268. doi: 10.1210/edrv.19.3.0334

2. Goffin V, Kinet S, Ferrag F, et al (1996) Antagonistic Properties of Human Prolactin Analogs That Show Paradoxical Agonistic Activity in the Nb2 Bioassay. J Biol Chem 271:16573–16579.

3. Freeman ME, Kanyicska B, Lerant A, Nagy G (2000) Prolactin: Structure, Function, and Regulation of Secretion. Physiol Rev 80:1523–1631.

4. Ormandy CJ, Camus A, Barra J, et al (1997) Null mutation of the prolactin receptor gene produces multiple reproductive defects in the mouse. Genes Dev 11:167–78.

5. Newey PJ, Phil D, Gorvin CM, et al (2013) Mutant Prolactin Receptor and Familial Hyperprolactinemia. N Engl J Med 369:2012–2020. doi: 10.1056/NEJMoa1307557

6. Sonigo C, Bouilly J, Carré N, et al (2012) Hyperprolactinemia-induced ovarian acyclicity is reversed by kisspeptin administration. J Clin Invest 122:3791–5. doi: 10.1172/JCI63937

7. Giesecke K, Hamann H, Sieme H, Distl O (2010) Evaluation of prolactin receptor (prlr) as candidate gene for male fertility in hanoverian warmblood horses. Reprod Domest Anim 45:124–130. doi: 10.1111/j.1439-0531.2009.01533.x

8. Apweiler R, Bairoch A, Wu CH, et al (2004) UniProt: the Universal Protein knowledgebase. Nucleic Acids Res 32:D115-9. doi: 10.1093/nar/gkh131

9. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5:725–38. doi: 10.1038/nprot.2010.5

10. Xu D, Jaroszewski L, Li Z, Godzik A (2014) AIDA: ab initio domain assembly server. Nucleic Acids Res 42:W308–W313. doi: 10.1093/nar/gku369

11. Xu D, Zhang J, Roy A, Zhang Y (2011) Automated protein structure modeling in CASP9 by I-TASSER pipeline combined with QUARK-based ab initio folding and FG-MD-based structure refinement. Proteins 147–60. doi: 10.1002/prot.23111

12. Fiser A, Sali A (2003) ModLoop: automated modeling of loops in protein structures. Bioinforma Appl NOTE 19:2500–2501. doi: 10.1093/bioinformatics/btg362

13. Maghrabi AHA, McGuffin LJ (2017) ModFOLD6: an accurate web server for the global and local quality estimation of 3D protein models. Nucleic Acids Res 45:W416–W421. doi: 10.1093/nar/gkx332

14. Benkert P, Biasini M, Schwede T (2011) Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27:343–50. doi: 10.1093/bioinformatics/btq662

15. Studer G, Biasini M, Schwede T (2014) Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane). Bioinformatics 30:i505-11. doi: 10.1093/bioinformatics/btu457

16. Batut P, Gingeras TR (2013) RAMPAGE: promoter activity profiling by paired-end

sequencing of 5'-complete cDNAs. Curr Protoc Mol Biol 104:Unit 25B.11. doi: 10.1002/0471142727.mb25b11s104

17. Zhou H, Zhou Y (2002) Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. Protein Sci 11:2714–2726. doi: 10.1110/ps.0217002

18. Yang Y, Zhou Y (2008) Specific interactions for ab initio folding of protein terminal regions with secondary structures. Proteins 72:793–803. doi: 10.1002/prot.21968

19. Holm L, Sander C (1992) Evaluation of protein models by atomic solvation preference. J Mol Biol 225:93–105.

20. Ray A, Lindahl E, Orn Wallner B (2012) Improved model quality assessment using ProQ2. BMC Bioinformatics 13:1–12. doi: 10.1186/1471-2105-13-224

21. McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16:404–5.

22. Zhang Y (2009) Protein structure prediction: when is it useful? Curr Opin Struct Biol 19:145–55. doi: 10.1016/j.sbi.2009.02.005

23. Read RJ, Chavali G (2007) Assessment of CASP7 predictions in the high accuracy template-based modeling category. Proteins 69 Suppl 8:27–37. doi: 10.1002/prot.21662

24. Du H, Brender JR, Zhang J, Zhang Y (2015) Protein structure prediction provides comparable performance to crystallographic structures in docking-based virtual screening. Methods 71:77–84. doi: 10.1016/j.ymeth.2014.08.017

25. Yang J, Zhang W, He B, et al (2016) Template-based protein structure prediction in CASP11 and retrospect of I-TASSER in the last decade. Proteins 84 Suppl 1:233–46. doi: 10.1002/prot.24918

26. Zhang J, Liang Y, Zhang Y (2011) Atomic-level protein structure refinement using fragment-guided molecular dynamics conformation sampling. Structure 19:1784–95. doi: 10.1016/j.str.2011.09.022

27. Kryshtafovych A, Monastyrskyy B, Fidelis K, et al (2018) Assessment of model accuracy estimations in CASP12. Proteins Struct Funct Bioinforma 86:345–360. doi: 10.1002/prot.25371

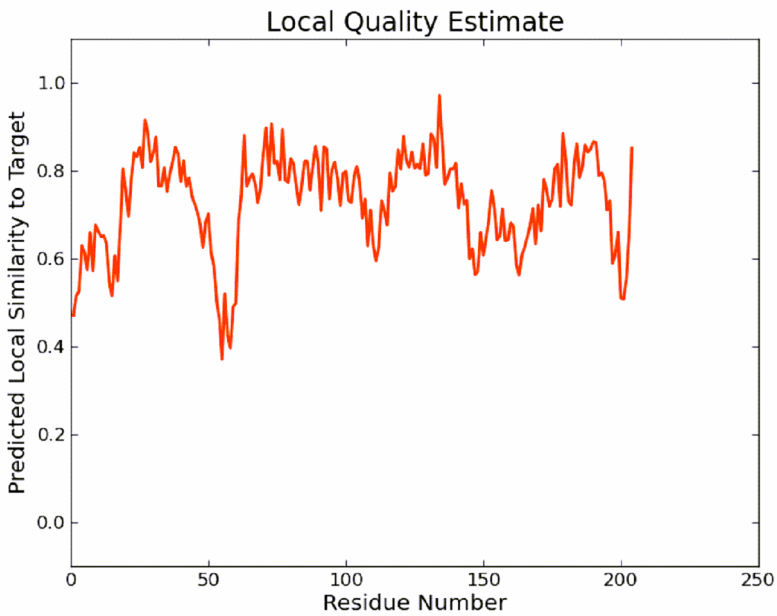
28. Uziela K, Wallner B (2016) ProQ2: estimation of model accuracy implemented in Rosetta. Bioinformatics 32:1411–1413. doi: 10.1093/bioinformatics/btv767

29. Koehler Leman J, Ulmschneider MB, Gray JJ (2015) Computational modeling of membrane proteins. Proteins 83:1–24. doi: 10.1002/prot.24703

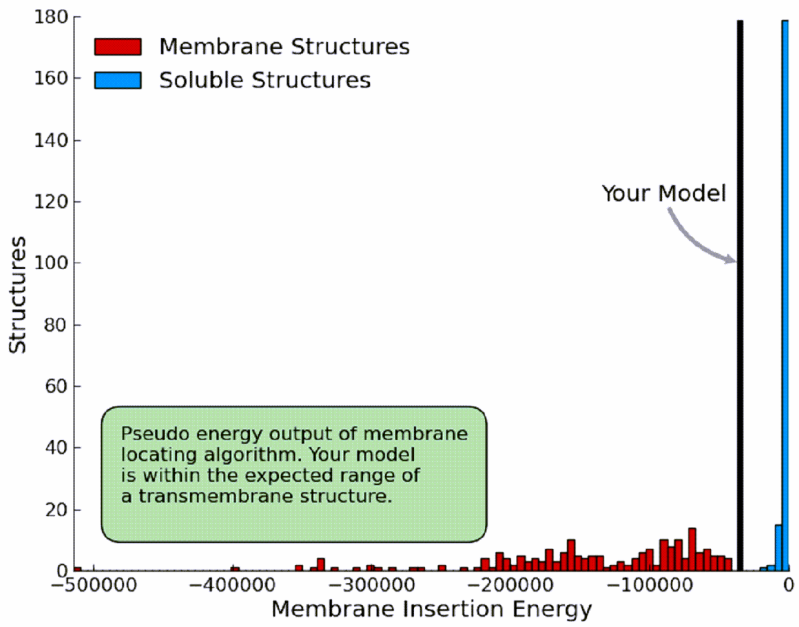
30. Almeida JG, Preto AJ, Koukos PI, et al (2017) Membrane proteins structures: A review on computational modeling tools. Biochim Biophys Acta - Biomembr 1859:2021–2039. doi: 10.1016/j.bbamem.2017.07.008

31. Kashani-Amin E, Tabatabaei-Malazy O, Sakhteman A, et al (2018) A systematic review on popularity, application and characteristics of protein secondary structure prediction tools. Curr Drug Discov Technol. doi: 10.2174/1570163815666180227162157

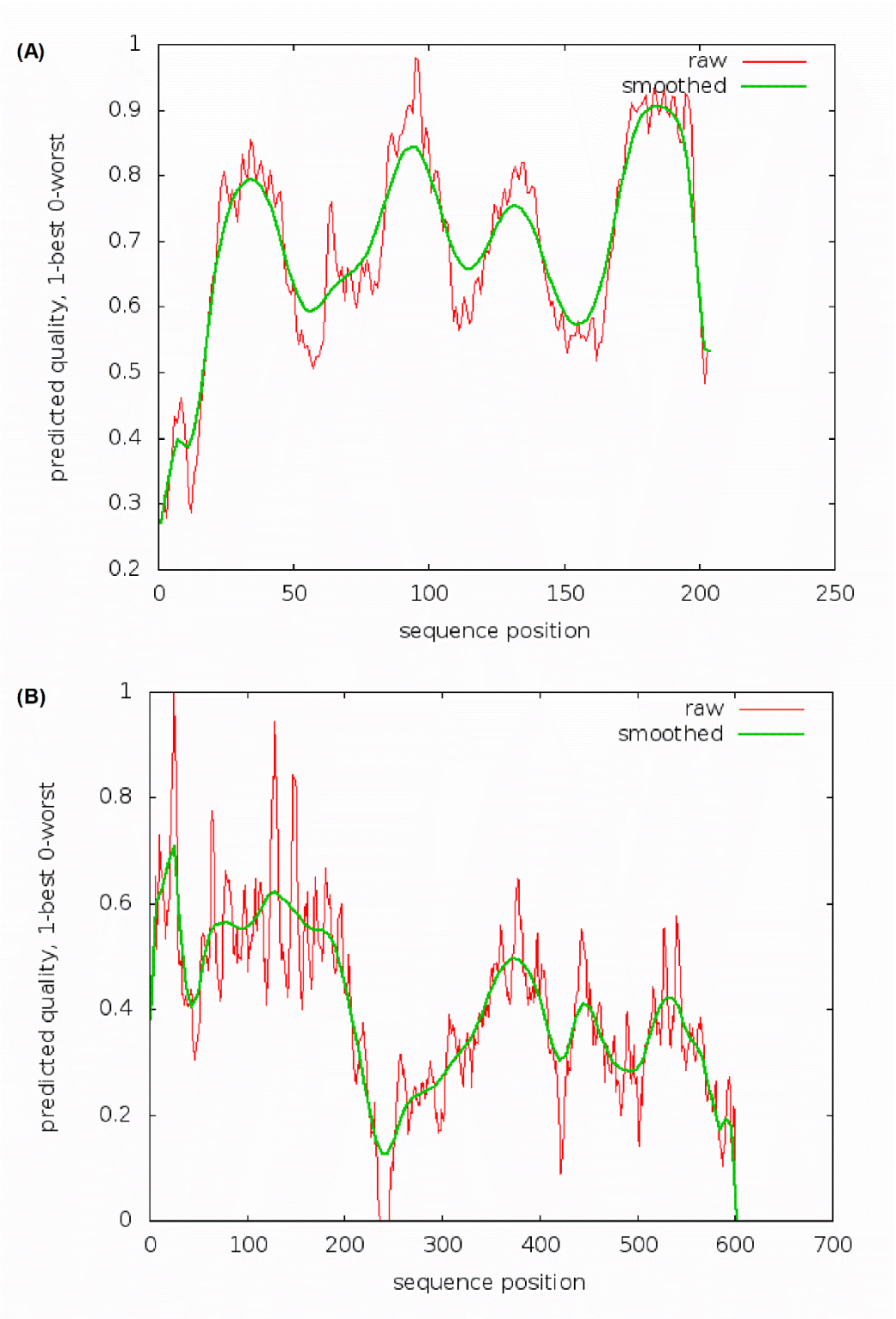
Supplementary Material



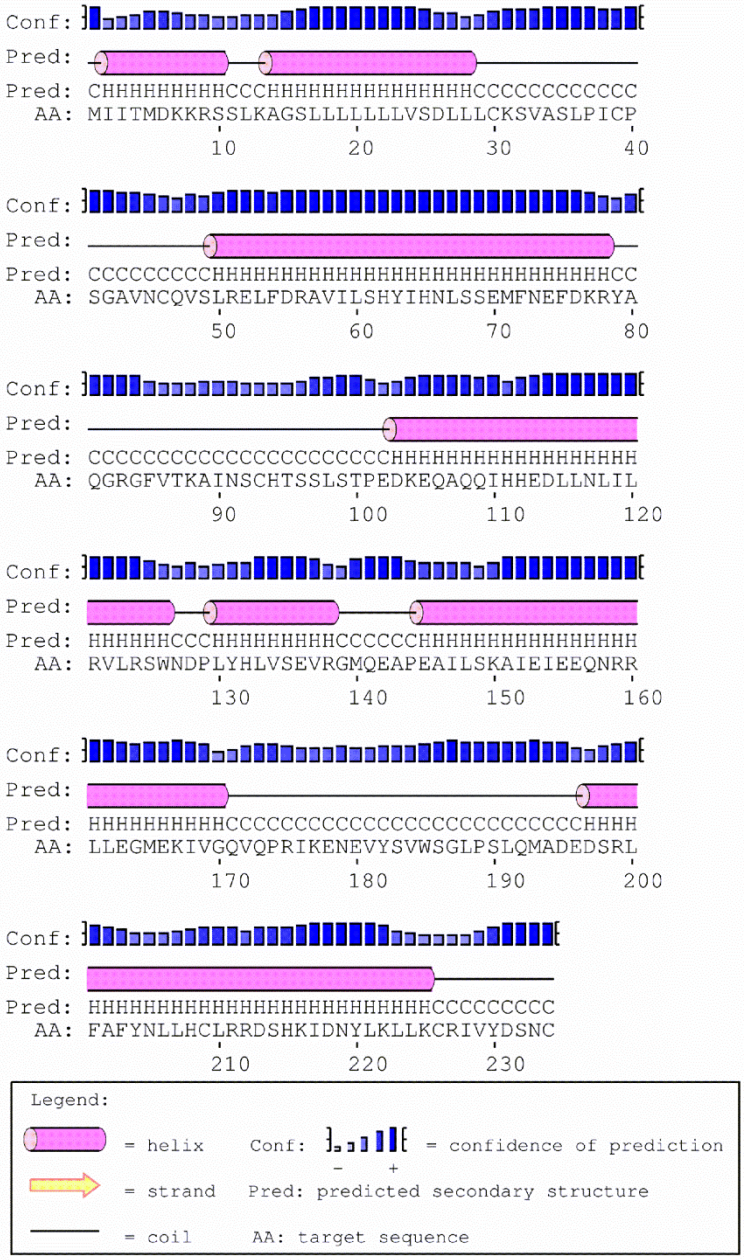
Online Resource 1 QMEAN graphical result for ePrl predicted model. The more negative, more reliable is the predicted structure



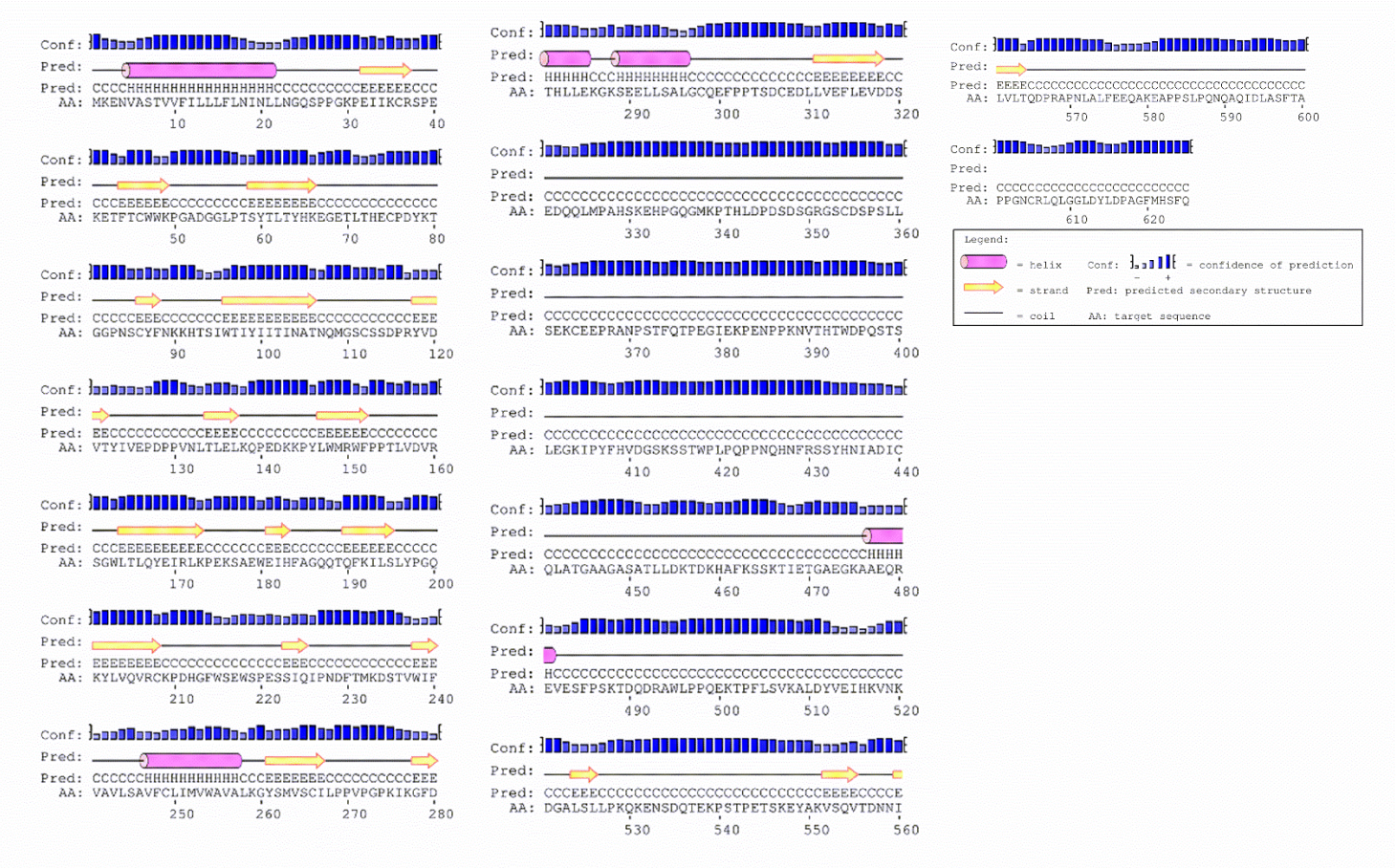
Online Resource 2 QMEANbrane result for ePrlr



Online Resource 3 Graphical results for quality estimate of the predicted models obtained in ProQ2. (A) Equine prolactin (ePrl) global quality. (B) Equine prolactin receptor (ePrlr) global quality



Online Resource 4 PsiPred specific residue quality for ePrl



Online Resource 5 PsiPred confidence of secondary structure prediction for ePrlr

5 Manuscrito

Prospecção *in silico* de ncRNAs na espécie equina aplicados a estudos de fertilidade

Resumo

A fertilidade em éguas é um tema complexo, envolvendo um grande número de disfunções reprodutivas, muitas vezes de causa desconhecida e/ou multifatorial. A elucidação de determinadas desordens é dificultada pela quantidade de disfunções sem compreensão total, a complexidade genética da fertilidade e a escassez de estudos genômicos voltados às fêmeas. Algumas regiões aparentemente "sem função" foram detectadas no sequenciamento do genoma equino e algumas foram determinadas como RNAs não-codificantes (ncRNAs). Dentre estes, os miRNAs são os RNAs regulatórios mais abundantes e desempenham um papel crucial na regulação pós-transcricional. Em mulheres, já é conhecida a influência dos miRNAs em desordens reprodutivas, como síndrome do ovário policístico, insuficiência ovariana prematura e endometriose. Sabe-se que 60% dos miRNAs conhecidos de equinos têm correspondências perfeitas com miRNAs associados a doenças humanas, o que torna possível transpor o conhecimento entre espécies. O objetivo do estudo foi elencar miRNAs encontrados em disfunções reprodutivas e esteroidogênese em outras espécies e analisar o genoma equino para a identificação dos ncRNAs presentes e sua distribuição. Através da análise da literatura, foram elencados 27 miRNAs que apresentaram relevância biológica para a fertilidade, sendo que sete não foram encontrados na análise do genoma equino. A análise dos ncRNAs se deu através da ferramenta INFERNAL, que procura sequências complementares nos bancos de dados de proteínas e miRNAs. Os miRNAs foram as moléculas mais abundantes, representando 65,1% do total analisado, seguido de snRNAs (24,92%), lncRNAs (2,58%) e moléculas com outras ações regulatórias (7,4%). Os miRNAs que obtiveram mais *hits* no genoma foram mir-544, mir-1255, mir-562 e mir-1302, envolvidos com a regulação do ciclo celular. A partir deste trabalho, foi possível elencar miRNAs que já comprovadamente interferem na fertilidade em outras espécies, sendo sugeridos os mir-17, -18, -21, -23, -34, -133, -181, -378, -503 e let-7, por seu papel diversificado, atuando tanto em desordens reprodutivas quanto hormonais, fatores chave na fertilidade. Assim, tornam-se bons candidatos a estudos futuros *in silico* e *in vitro*. Os ncRNAs identificados no genoma representam papel essencial para a elucidação de suas funções e genes-alvo, quantificação em condições fisiológicas ou desordens.

Palavras-chave: biomarcador; bioinformática; reprodução; éguas.

Introdução

A fertilidade em éguas é um tema complexo, envolvendo um grande número de disfunções reprodutivas, muitas vezes de causa desconhecida e/ou multifatorial. Algumas destas disfunções estão relacionadas a endometrite, folículos anovulatórios e cistos ovarianos/endometriais. A elucidação da etiologia e fisiopatologia de desordens é dificultada pela quantidade de disfunções sem compreensão total que estão descritas em éguas (SNIDER, 2015). Incluído a isso, está a complexidade genética relacionada a fertilidade (CHOWDHARY, 2013) e a escassez de estudos genômicos voltados a fêmeas. Assim, a relação entre genômica e fertilidade ainda é pouco explorada.

A partir do sequenciamento do genoma equino (WADE et al., 2009) obteve-se grande quantidade de dados genômicos, os quais possibilitaram a determinação de biomarcadores para doenças, análise da interferência de mutações e também a descoberta de regiões aparentemente “sem função”. Através da transcriptômica, descobriu-se que eram geradas moléculas de RNA, porém não eram codificantes para uma proteína, sendo designados RNAs não-codificantes (ncRNAs). Os ncRNAs são uma classe de moléculas que não são traduzidos em uma proteína, sendo divididos por sua função e tamanho, podendo exercer funções regulatórias pós-transcricionais (HA; KIM, 2014).

Os miRNAs maduros desempenham um papel crucial na regulação pós-transcricional e são os RNAs regulatórios mais abundantes (HA; KIM, 2014). Ocorre a ligação da fita guia às regiões 3' não traduzidas (3'-UTRs) de seus mRNAs-alvo, o que pode levar à degradação do mRNA alvo, desadenilação da cauda poli(A), competição do complexo de silenciamento com fatores de iniciação da tradução e estrutura *cap*, bloqueio do alongamento da tradução, promoção da dissociação prematura dos ribossomos ou degradação da cadeia polipeptídica nascente (EULALIO; HUNTZINGER; IZAURRALDE, 2008).

Na reprodução, em mulheres, a influência dos miRNAs é descrita em casos de infertilidade relacionados à síndrome do ovário policístico (SOP) (ROTH et al., 2014), endometriose (BURNEY et al., 2009), insuficiência ovariana primária (IOP) (GUO; SUN; LAI, 2017), e mesmo em casos sem causa aparente de falhas recidivas da fertilização *in vitro* (FIV) (KARAKAYA et al., 2015). Já em éguas, é possível obter dados referentes aos aspectos ovulatórios e de maturação folicular, como descrito

por SCHAUER et al. (2013), que descreveu os miRNAs detectados em folículos dominantes e subordinados. Donadeu; Schauer (2013) avaliaram a expressão de alguns miRNAs já descritos para outras espécies em folículos ovulatórios e anovulatórios.

Considerando que mais de 60% dos miRNAs conhecidos de equinos têm homologia perfeita com miRNAs associados a doenças humanas (BUZA et al., 2014). O objetivo do estudo foi elencar miRNAs associados a desordens reprodutivas e hormonais, com foco no desenvolvimento folicular e oocitário de diferentes espécies. A fim de ampliar o conhecimento e determinação efetiva do papel dos miRNAs na fertilidade das fêmeas, uma análise computacional aprofundada do genoma equino foi realizada, procurando identificar os ncRNAs existentes, avaliar sua distribuição e prospectar biomarcadores que podem estar associados à fertilidade em éguas.

Metodologia

Análise de miRNAs candidatos a biomarcadores de fertilidade na literatura

Os ncRNAs de interesse em outras espécies foram obtidos através de pesquisa na literatura disponível no banco de dados Pubmed NCBI *(National Center for Biotechnology Information) (https://ncbi.nlm.nih.gov/pubmed)*, relacionando aos dados obtidos *in silico* e *in vitro* das seguintes espécies: equinos, humanos, camundongos, bovinos, suínos e ovinos. A seleção dos miRNAs foi baseada no seu envolvimento na oogênese, foliculogênese e em desordens reprodutivas, que incluíram: cistos ovarianos, endometriais e folículos anovulatórios. Para isso, foram utilizadas diferentes combinações das palavras-chave: *"ncRNA", "miRNA", "anovulatory", "infertility", "fertility", "cyst", "foliculle", "equine", "mare", "bovine", "human" "women", "ovine", "murine", "reproduction", "disorder".* O gene-alvo dos ncRNAs e sua função foram buscados no banco de dados GeneCards (SAFRAN et al., 2010) e MirPath (VLACHOS et al., 2015). Os ncRNAs que apresentaram relação ou possível relação com os problemas de fertilidade foram elencados por sua função, gene-alvo relatado e espécie. Um fluxograma da metodologia utilizada pode ser visto na Figura 1.

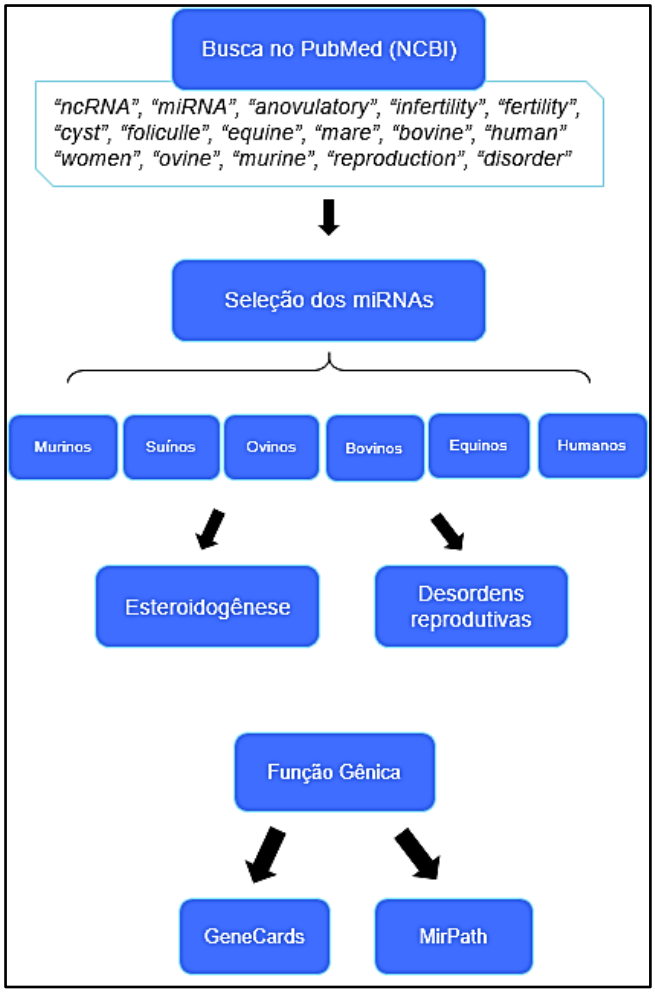


Figura 1. Metodologia utilizada para a obtenção dos miRNAs de relevância a partir da literatura disponível nos bancos de dados. Foi realizada uma busca na literatura disponível no PubMed utilizando as palavras-chave descritas em diferentes combinações. Os miRNAs encontrados englobaram murinos, suínos, ovinos, bovinos, equinos e humanos, sendo relacionados à esteroidogênese e desordens reprodutivas associadas a infertilidade. A função gênica descrita foi obtida através do banco de informações sobre genes GeneCards e MirPath.

Prospecção de ncRNAs no genoma equino

Para analisar os ncRNAs descritos no genoma, foram obtidos os 32 cromossomos equinos através do banco de dados Ensembl (https://www.ensembl.org/Equus_caballus/Info/Index), na versão EquCab2.0 (Ensembl:GCA_000002305.1). As sequências foram analisadas através da ferramenta INFERNAL (http://eddylab.org/infernal/), que prediz RNAs não-codificantes em relação ao banco de dados de famílias de RNA, o Rfam (KALVARI et al., 2018) e miRBase v22.0 (GRIFFITHS-JONES, 2006). Foram elencados os ncRNAs reconhecidamente transcritos no genoma através dos bancos de dados e segregados por abundância e função dos genes-alvo. Os parâmetros utilizados em

nossa busca foram o *E-value " -- E 1e-200 --rfam"* (valor de significância estatística dos *hits* na base de dados), sendo obrigatoriamente menor que $1x10^{-1}$. Os resultados foram filtrados através de *scripts* em Python para obter apenas as sequências com mais de 99% de similiaridade ao alinhamento local, retirada das sequências alinhadas com procariotos, árqueas, plantas e quanto aos tipos de RNAs em maior quantidade nas células (mRNA, tRNA e rRNA).

Resultados

Análise de miRNAs candidatos a biomarcadores da fertilidade na literatura

Foram elencados 27 miRNAs que já possuem evidências experimentais de associação a fatores de importância para a reprodução, incluindo sua influência na esteroidogênese e no desenvolvimento folicular e oocitário. Os resultados são listados nos quadros 1 e 2, respectivamente. De acordo com a literatura, os miRNAs que potencialmente se apresentam como biomarcadores foram segregados conforme genes-alvo descritos, espécie (murinos, ovinos, suínos, bovinos, equinos e humanos), função e desordem às quais podem estar relacionados. Posteriormente, estes miRNAs foram buscados dentre os resultados obtidos nos dados genômicos de miRNAs equinos. Sete miRNAs não foram encontrados no genoma equino, sendo estes mir-18, mir-125, mir-202, mir-222, mir-320, mir-483 e mir-520. Destes, apenas três (mir-18, mir-320 e mir- 520) não foram encontrados no banco de dados mirBase. Dos miRNAs analisados na literatura, todos ficaram abaixo do *threshold* de 10 *hits* estabelecido, sendo que cada hit representa uma cópia do gene. Os mais encontrados foram let-7, com 10 *hits*, mir-17, com 6 *hits*, mir-133 e mir-29, com 5 *hits*. Foram definidos os miRNAs obtidos a partir da literatura mais interessantes a serem posteriormente estudados, sendo estes let-7, mir-17, -34, -181, -503.

Quadro 1 - miRNAs relacionados à esteroidogênese selecionados a partir da literatura.

miRNA	Gene-alvo	Ação	Espécie	Referência
mir-17-5p; mir-17-3p	STAR, CYP-19A1 COX-2, IL-1b, ERb	Proliferação, diferenciação e síntese de esteroides.	Humanos	TOLOUBEYDOKHTI et al., 2007; SIROTKIN et al., 2009 KASIMANICKAM; KASTELIC, 2016
# mir-18, mir-18b	IL-8, SYT1, IRS2	Estímulo à liberação de progesterona, estradiol e testosterona nas CGs.	Humanos	SIROTKIN et al., 2009
mir-24	RAB5B, EGFR, TGFB1	Regulação da liberação de progesterona.	Humanos	SANG et al., 2013
mir-29a; mir-29b	PLA2G4A, OTXR	(A) estímulo à síntese de prostaglandinas; (B) estímulo à secreção de progesterona.	(A) Ratos (B) Bovinos	KUANG et al., 2014; XU et al., 2018
mir-32	IL-8, SYT1, IRS2	Estímulo à liberação de progesterona nas CGs humanas, inibição da liberação de testosterona e estradiol.	Humanos	SIROTKIN et al., 2009
mir-34	CDK6, VEGFA, CCNE3, CDC25A, BCL-2, MYC, WNT, JAG1, INHBB, P53, DLL1, NOTCH1, MET, EZH2, MYC, E2F3,CCND1	Leva à apoptose das CGs suínas. Mediação da supressão da proliferação através da p53 em camundongos. Inibição da liberação de progesterona e estradiol nas CGs humanas. Participa da transição folicular/luteal em ovinos.	Suínos Camundongo Humanos Ovinos	TU et al., 2014; CORNEY et al., 2007 BOMMER et al., 2007; SIROTKIN et al., 2009; MCBRIDE et al., 2012
# mir-125a e mir-125b	ARID3B, BAK1, ERBB2/3, LIF, NTRK3, TP53, KRT7, CDKN1A, CBFB, MYC, HMGA2, LIN28A, TRIM71, AKT1, KLF13	Estímulo à liberação de progesterona nas CGs humanas. É abundante durante todo desenvolvimento folicular ovino.	Humano Ovino	SIROTKIN et al., 2013; MCBRIDE et al., 2012
mir-129-2-3p	MAP3K1	Envolvidos no aumento na produção de progesterona e inibição da síntese de estradiol	Bovinos	ZIELAK-STECIWKO et al., 2014
mir-133a, mir-133b	TAGLN2, FOXL2, CYP11A1	Quando superexpresso, reduz os níveis de FoxL2, que reprime produção de progesterona. Inibição da liberação de progesterona nas CGs humanas.	Camundongo Humano	XIAO et al., 2014; DAI et al., 2013; SIROTKIN et al., 2009
mir-143	CDK4, CDK6, CCNB1, CCND2, CCNE2	Inibição da formação folicular primordial em camundongos e da liberação de progesterona nas CGs em humanos.	Camundongo Humano	ZHANG et al., 2013; SIROTKIN et al., 2009
mir-181a - mir-181d	ACVR2A; CCND2; MAPK, TGFB	Expressão aumentada em folículos murinos primários. Presente nas microvesículas foliculares de éguas velhas. Inibe a liberação de progesterona nas CGs	Camundongo Equinos Humanos	ZHANG et al., 2013b; DA SILVEIRA et al., 2012; SIROTKIN et al., 2009

		humanas.		
mir-193b	ESR1	Regulação da liberação de progesterona	Humanos	SANG et al., 2013
# mir-222	ESR1	Envolvido com a liberação de estrógeno	Ratos	HOSSAIN et al., 2013
mir-224	TGFB1, SMAD4, CYP19A1	Encontrado em folículos equinos em luteinização. Promove a liberação de estradiol em camundongos.	Equinos Camundongo	SCHAUER et al., 2013; YAO et al., 2010
# mir-320	IL-1A, IL-10, IL-12B, IL-37, RAB5B	Regulação da liberação de estrógeno	Humanos	SANG et al., 2013
mir-378	CYP19A1, IGF-1	Síntese de estradiol	Suínos Equinos	SUI et al., 2014; DONADEU & SCHAUER, 2013
mir-383	RBMS1, C-MYC	Síntese e liberação de estradiol	Humanos Bovinos	YIN et al., 2012; ZIELAK-STECIWKO et al., 2014
# mir-483-5p	MAPK3, NOTCH3, IL-1A, IL-10, IL-12B, IL-37	Regulação da liberação de progesterona	Humanos	XU et al., 2015; SANG et al., 2013
# miRNA-520c-3p	ESR1	Regulação da liberação de estrógeno	Humanos	SANG et al., 2013
Let-7	PTEN, CASP3 e p53	Abundante no ovário de humanos, suínos e ovinos; Inibição da liberação de estrógeno e progesterona em CGs humanas	Humanos Ovinos Suínos	LI et al., 2015; MCBRIDE et al., 2012; SIROTKIN et al., (2009).

“ # “ refere-se aos miRNAs não encontrados na análise genômica em equinos.

CG = células da granulosa

Quadro 2 - miRNAs relacionados à desordens reprodutivas obtidos a partir da literatura.

miRNA	Gene-alvo	Função	Espécies	Desordem	Referências
mir-17-5p; mir-17-3p	STAR, CYP-19A1 COX-2, IL-1b, ERb	Níveis elevados em endometriose humana e metrite bovina.	Humanos Bovinos	EH, M	TOLOUBEYDOKHTI et al., 2007; KASIMANICKAM & KASTELIC, 2016
# mir-18, mir-18b	IL8, SYT1, IRS2	Níveis elevados no fluido folicular de mulheres com SOP	Humanos	SOP	ROTH et al., 2014; ZIELAK-STECIWKO et al., (2014)
mir-21 (mir-21-3p/5p)	PDCD4, PTEN	Níveis elevados em FA e folículos subordinados. Mulheres não responsivas à FIV. 2x mais expresso em EBS. Altos níveis levam à atresia folicular, em baixos, Ap e dificuldade de expansão do *cumulus*	Bovinos Equinos Humanos	EBS, FA,FIV	DONADEU & SCHAUER, 2013; KARAKAYA et al., 2014; TSCHERNER et al., 2018; SCHAUER et al., 2013
mir-23a e mir-23b	STAR, CYP-19A1 COX-2, IL-1b, ERb, TGFB, WNT, MAPK, AKT, XIAP, CAS3	Induz apoptose das células da granulosa. Expressão aumentada em casos de IOP. Em altos níveis na IOP causa apoptose. Expressão descrita em exossomos/microvesículas no fluido folicular em éguas jovens.	Humanos Equinos	IOP	(A E B) DONADEU & SCHAUER, 2013; TOLOUBEYDOKHTI et al., 2008; DA SILVEIRA et al., 2012; YANG et al., 2012; ZHOU et al., 2012
mir-24	RAB5B, EGFR, TGFB1, IKIBK, JUN, SP1, INRS, MAPK14, NOS2, ID3	Interfere na liberação de progesterona	Humanos Bovinos	SOP, ESB	SANG et al., 2013; HAILEMARIAM et al., 2014
mir-32	IL-8, SYT1, IRS2	Níveis elevados no fluido folicular de mulheres não-responsivas apresentando SOP	Humanos	SOP	ROTH et al., 2014;
família mir-34	CCNE, CCNE2, CDK4, CDK6, CDC25A, BCL-2, IL-8, SYT1, IRS2	Níveis diminuídos em EH.	Humanos	EH	ROTH et al., 2014; BURNEY et al., 2009;
mir-132	HMGA2, CTBP1	Envolvido com a liberação de estrógeno	Humanos	SOP	SANG et al., 2013
mir-193b	ESR1	Regulação da liberação de progesterona	Humanos	SOP	SANG et al.,2013
# mir-202	TGFB; IL-8, SYT1, IRS2	Expressão aumentada em POI e folículos anovulatórios	Humanos Equinos	IOP, FA	DONADEU & SCHAUER, 2013; ZHOU et al., 2012
# mir-222	ESR1	Envolvido com a liberação de estrógeno	Ratos Humanos	SOP	HOSSAIN et al., 2013; SANG et al., 2013
mir-224	TGFB1, SMAD4, CYP19A1	Diferencialmente expresso em casos de pré-eclampsia; encontrado em folículos equinos na fase luteal	Humanos Equinos	PE	YAO et al., 2010; SCHAUER et al., 2013
# mir-320	RAB5B	Regulação da liberação de estrógeno	Humanos	SOP	SANG et al., 2013
# mir-483-5p	MAPK3 NOTCH3, IL-1A, IL-10, IL-12B, IL-37	Níveis aumentados em SOP	Humanos	SOP	XU et al., 2015; SANG et al., 2013

mir-503	MAPK14, IKBKB, JUN, CCND1, BCL-2, RHO A, ROCK1, ROCK2	Inibe a proliferação de células endometrióticas. 5x mais expresso em EBS.	Humanos Bovinos	EH, EBS	HIRAKAWA et al., 2016; HAILEMARIAM et al., 2014; ZIELAK-STECIWKO et al., 2014
# mirna-520c-3p	ESR1	Regulação a liberação de estradiol	Humanos	SOP	SANG et al., 2013
Let-7	PTEN, CASP3 e p53	Níveis de expressão diminuídos em IOP	Humanos	IOP	YANG et al., (2012); ZHOU et al., (2011).

“ # “ refere-se aos miRNAs não encontrados na análise genômica em equinos.

Abreviações:

CG = células da granulosa

CT = células da teca

EH = endometriose humanos

ESB = endometrite subclínica bovina

FA = folículo anovultório

FIV = mulheres não responsivas à FIV sem causa aparente

IOP = insuficiência ovariana prematura

M = metrite

PE = pré-eclampsia

SOP = síndrome do ovário policístico

Prospecção de ncRNAs no genoma equino

No presente trabalho, utilizando ferramentas bionformáticas, pela primeira vez é apresentada uma análise de ncRNAs em todo genoma equino. A partir da utilização do software INFERNAL, foi possível filtrar sequências não-codificantes no genoma já descritas nos bancos de dados de famílias de proteínas e miRNA por sua estrutura primária, secundária e homólogos. Através dos parâmetros estabelecidos pela busca no *software* INFERNAL e *scripts* de filtragem, existe apenas 1% de chance de haver um *mismatch* ou gap na sequência dos ncRNAs descritos, demonstrando uma flexibilidade mínima em relação a falsos-negativos.

Foram encontrados 4.482 ncRNAs no genoma equino (Figura 2). Dentre as classes de ncRNAs mais encontradas estão os miRNAs, representando 65,1% do total. Em seguida, snRNAs (24,92%), lncRNA (2,58%) e moléculas com outras ações regulatórias, descritas aqui como “Outros” (7,4%).

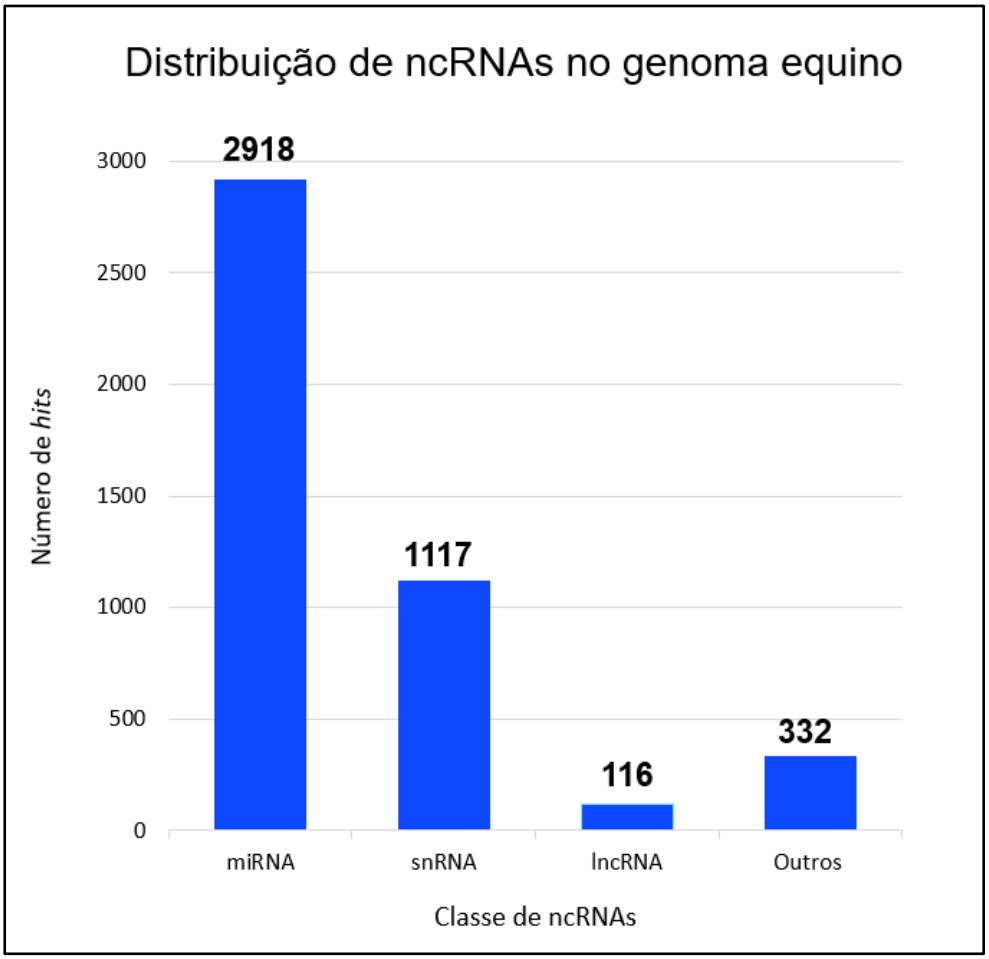


Figura 2. Distribuição de ncRNAs identificados no genoma equino. Os valores indicam a quantidade de *hits* no genoma. miRNA = micro-RNAs; snRNA = pequenos RNAs nucleares; lcnRNA = longos RNAs não codificantes; outros = RNAs não-codificantes de funções diversas.

Dentro da família de snRNAs foram encontrados as sequências referentes a: pequenos RNAs nucleolares (snoRNA), que guiam as modificações em tRNAs, rRNAs e RNAs nucleares, *motif hammerhead* (RNA que catalisa quebras e ligações em outras moléculas de RNA), elementos específicos de ligação à 5' UTR dos mRNAs - os sítios internos de entrada no ribossomo (do inglês, *Internal Ribosome Entry Site* - IRES) e snRNAs que compõe o spliceossomo (por exemplo, U6). Foram encontrados também 116 lncRNA, associados a regulação de genes específicos. Na categoria "Outros" foram incluídos os ncRNAs de ligação a sítios do rRNA, RNAs que fazem parte da formação de complexos proteicos do tipo ribonucleoprotease, ribozimas e moléculas que promovem modificações no *frameshift* ribossomal.

Os miRNAs representaram a maior quantidade de ncRNAs regulatórios. Os miRNAs que obtiveram mais *hits* no genoma são apresentados na Figura 3, sendo

que mir-544, mir-1255, mir-562, mir-1302 e mir-584 foram os mais encontrados pelo *software*. Os números indicam quantas cópias do gene foram encontradas durante a análise genômica. Aqueles que não atingiram o *threshold* de 10 detecções foram incluídos em “Outros”.

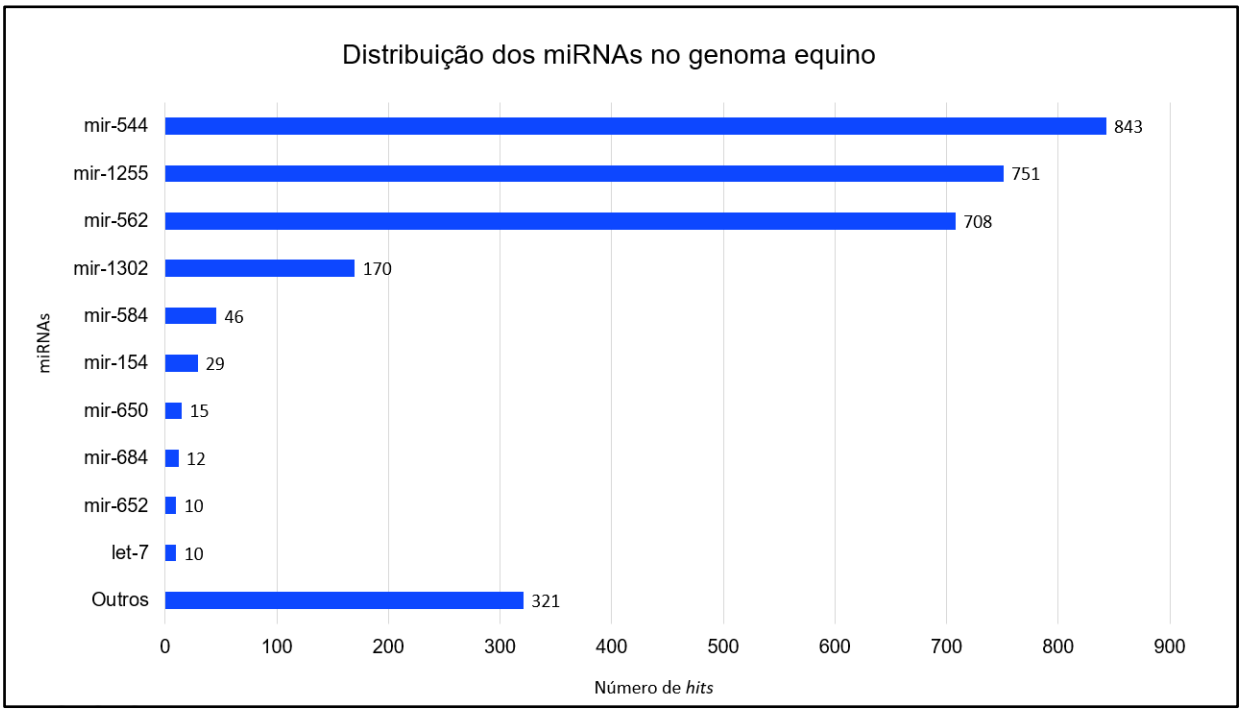


Figura 3.Distribuição dos miRNAs identificados no genoma equino. Os valores indicam a quantidade de *hits* no genoma. Outros = miRNAs que atingiram 10 *hits* ou menos dentro do genoma.

Os miRNAs que obtiveram a maior quantidade de *hits* no genoma foram analisados em virtude da sua função e vias de sinalização que interferem. Detectou-se que os mir-544, mir-1255, mir-562 e mir-584 estão intimamente relacionados à manutenção do ciclo celular, processos metabólicos básicos, mitose, apoptose, regulação de mRNAs e biossíntese de aminoacetil-tRNA. Da mesma forma, estes miRNAs estavam relacionados às vias de sinalização de receptores andrógenos.

A distribuição dos ncRNAs por cromossomo pode ser vista na Figura 4. A análise revelou que a maior quantidade de ncRNAs estava concentrada nos cromossomos 1, X e 4, enquanto que os menores número de sequências identificadas foram os cromossomos 29, 30 e 31.

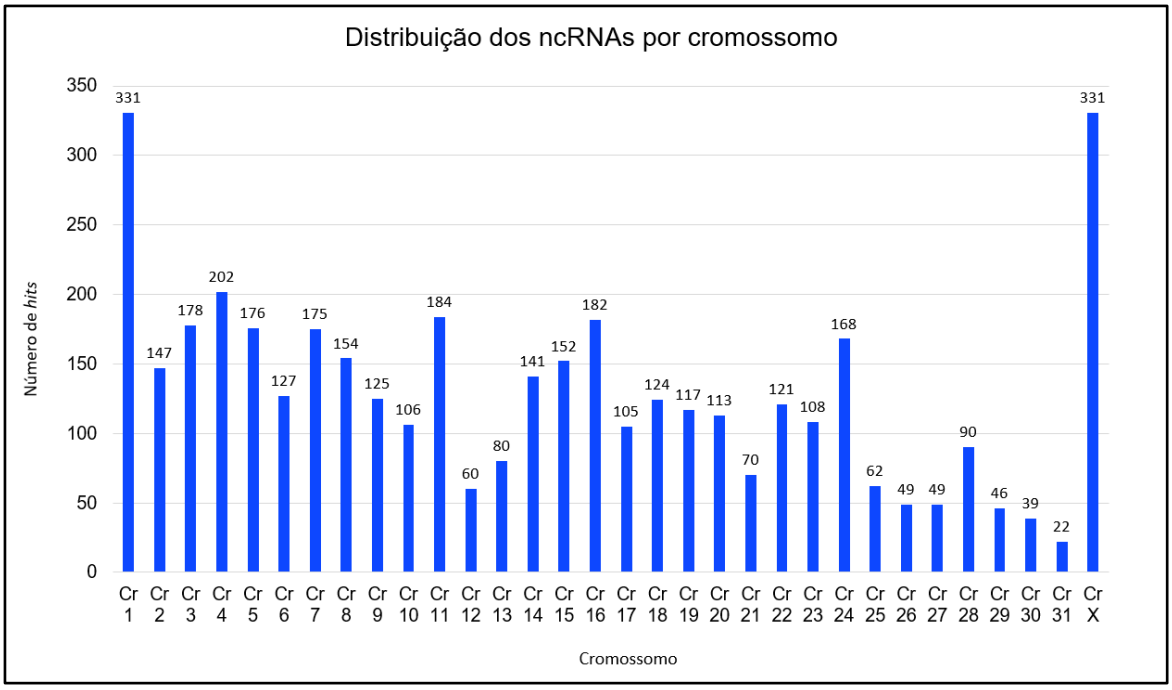


Figura 4 Distribuição dos ncRNAs por cromossomo. Cr = cromossomo. Os valores indicam o número de *hits* observados para cada cromossomo.

Avaliamos a distribuição dos miRNAs por cromossomo, como visto na Figura 5. Os miRNAs que identificamos tiveram mais *hits* nos cromossomos X (267), 1 (223) e 16 (144), enquanto os que apresentaram menos miRNAs foram 29, 30 e 31.

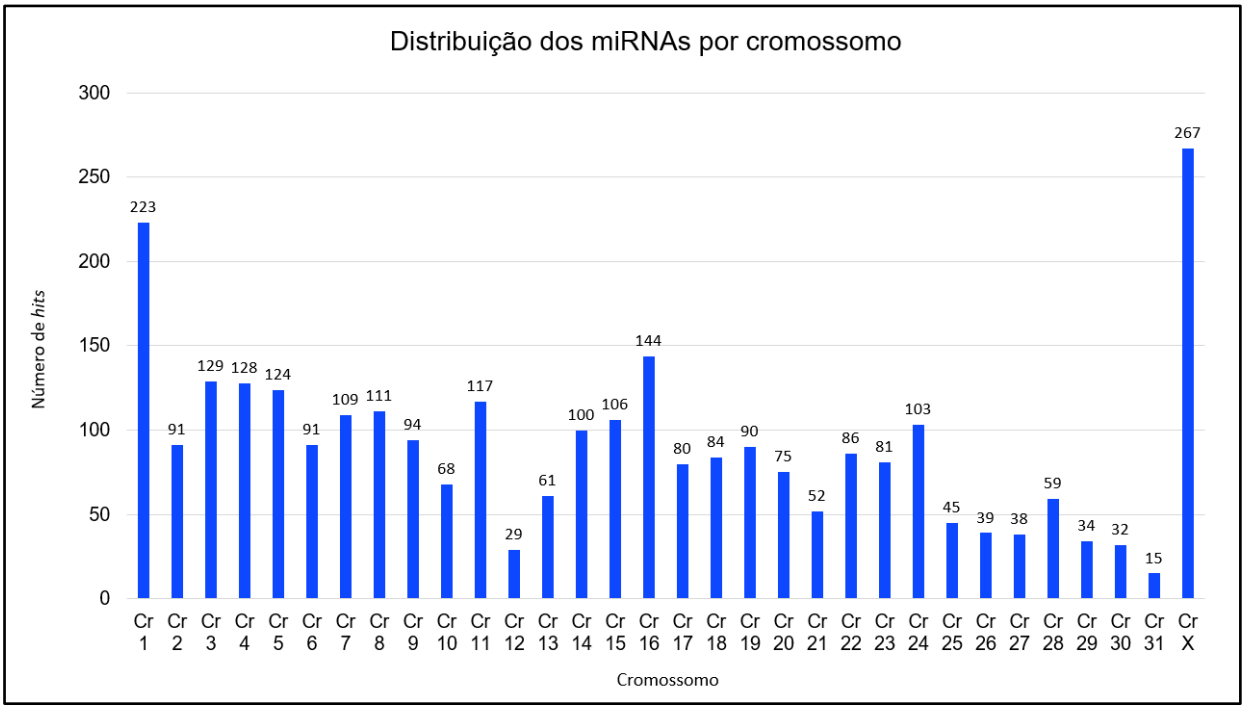


Figura 5. Distribuição do miRNAs por cromossomo. Cr = cromossomo. Os valores indicam o número de *hits* observados para cada cromossomo.

Discussão

Análise de miRNAs candidatos a biomarcadores da fertilidade na literatura

Devido à maior disponibilidade e similaridade genômica, foi utilizada principalmente a literatura experimental relacionada às espécies humana e bovina, mas também dados referentes a murinos, suínos e ovinos foram considerados. A busca por estes miRNAs no genoma equino não discriminou as nomenclaturas específicas dos miRNAs. Por exemplo, no genoma, o mir-133 foi detectado, mas sua relevância biológica para a fertilidade foi descrita em relação à diferentes posições nucleotidícas - mir-133a e mir-133b. A nomenclatura dos miRNAs leva em conta diferenças em um ou dois nucleotídeos na estrutura madura, adicionando-se uma letra minúscula logo após sua numeração quando possuem sequências quase idênticas. Já a adição de números após a nomenclatura significa que são idênticos. Quando duas sequências são geradas em fitas opostas, adiciona-se 5p (para a fita 5') ou 3p (para a fita 3') (VAN DER KOLK; PACHOLEWSKA; GERBER, 2015). Entretanto, os miRNAs encontrados durante a revisão de literatura incluem espécies

que já possuem uma ampla descrição de seu perfil transcriptômico e, portanto, o motivo destes miRNAs não terem sido encontrados em nossa análise pode ser por ainda não estarem reportadas nos bancos de dados tais diferenças nas sequências maduras. Outra possível causa é a existência de isoformas destes miRNAs, causadas por variabilidade na região 3' ou substituições nucleotídicas fora da sequência guia. Embora os efeitos funcionais destas isoformas não sejam totalmente conhecidos, a eficiência de ligação ao alvo, incorporação à Ago e estabilidade podem estar sendo comprometidas (NEILSEN; GOODALL; BRACKEN, 2012). Como o miRBase leva em consideração a fita-guia, estas isoformas não estão compreendidas no banco de dados (NEILSEN; GOODALL; BRACKEN, 2012).

Em nossa análise em relação a literatura, let-7 foi o miRNA mais encontrado, apresentado 10 *hits* no genoma. A família de miRNAs let-7 é a mais abundante encontrada nos ovários em suínos, humanos (LI et al., 2015) e ovinos (MCBRIDE et al., 2012) e atua na manutenção e maturação oocitária (LI et al., 2015). Por ter sido observado um grande número de vezes no genoma, é possível afirmar que nosso resultado corrobora com essa afirmação. É sugerido seu papel regulatório ao inibir a liberação de estrógeno e progesterona em células humanas (SIROTKIN et al., 2009), estar em baixos níveis de expressão em casos de IOP sendo os genes-alvo *PTEN, CASP3* e *p53*, apontando para a possível associação com a ocorrência da menopausa precoce e infertilidade (YANG et al., 2012; ZHOU et al., 2011). Em equinos, a família let-7 foi descrita no cólon, músculo esquelético e fígado (KIM et al., 2014), porém não há relatos até o momento em tecidos reprodutivos.

Em nossa análise do genoma, o mir-17 obteve 6 *hits*, sendo um dos mais encontrados. Em casos de metrite bovina, o mir-17-3p foi encontrado em níveis elevados (KASIMANICKAM; KASTELIC, 2016). Em casos de endometriose foi descrita expressão diminuída do gene-alvo *STAR*, regulador da aromatase, além do mir-17-5p, que tem como genes-alvo *COX-2, IL-1b, STAR, CYP-19A1* e *ESR*, relacionados à síntese de esteroides (TOLOUBEYDOKHTI et al., 2007). Sirotkin et al., (2009) detectou que mir-17-3p inibe a liberação de progesterona e testosterona nas células da granulosa. O mir-34 apresentou três *hits* em nossa análise genômica. Os miRNAs 34b e 34c foram descritos como mediadores da supressão pela p53 em células de câncer de ovário, evidenciando seu papel na repressão do crescimento em níveis de expressão normais (CORNEY et al., 2007). Em mulheres com

endometriose estavam em níveis de expressão diminuídos, o que corrobora com a manutenção do perfil proliferativo destas células (BURNEY et al., 2009). Tu et al., (2014) reportou apoptose das células da granulosa em folículos suínos e *INHBB* como gene-alvo, associado à secreção de FSH. A liberação de testosterona e estradiol foram inibidas por estes miRNA (SIROTKIN et al., 2009) e foram detectados níveis aumentados em folículos ovinos (MCBRIDE et al., 2012). O mir-34c também apresentou-se elevado em casos de SOP (ROTH et al., 2014).

A análise *in silico* relevou que o mir-181 teve dois *hits* no genoma equino. No fluido folicular de éguas velhas em comparação às jovens, maiores níveis de expressão deste miRNA foram encontrados. Um de seus genes-alvo é *TGFB*, crucial para o correto desenvolvimento folicular (DA SILVEIRA et al., 2012). Foi descrito em níveis aumentados de expressão em pacientes com IOP, podendo estar envolvido na patofisiologia da doença por suprimir o gene *ACVR1* (ZHANG et al., 2013b). Por fim, demonstrou um papel inibitório na liberação de progesterona (SIROTKIN et al., 2009) e descrição *in silico* em equinos (ZHOU et al., 2009a). O mir-503 teve apenas um *hit* no genoma equino. Hirakawa et al., (2016) descreveram que mir-503 está associado a repressão da proliferação das células endometrióticas em mulheres, ao interferir nos genes relacionados ao controle do ciclo celular, proliferação e apoptose. Em camundongos e ovelhas este miRNA também já foi relatado interferindo nas funções oocitárias e foliculares (revisado em DONADEU; SCHAUER; SONTAKKE, 2012). Como foi encontrado em folículos subordinados bovinos, sugere-se sua relação com a atresia (ZIELAK-STECIWKO et al., 2014). Na endometrite subclínica bovina este miRNA também foi reportado relacionado às vias de NF-kB, crucial na regulação da inflamação (HAILEMARIAM et al., 2014). Em equinos foi descrito em análises *in silico* (ZHOU et al., 2009a).

Uma das propostas do trabalho foi sumarizar os miRNAs relacionados à esteroidogênese e desordens de fertilidade de relevância descritos em outras espécies. Um estudo similar elencou miRNAs de importância para o desenvolvimento folicular e luteal descritos até a data, porém não incluiu equinos. Estes foram: mir-224, -383, -378, -23a, -26b, -132/mir-212, -21, 125b, -145, -199a-3p, -503, let-7b/mir-17-5p (DONADEU; SCHAUER; SONTAKKE, 2012). Isso mostra que nosso estudo está em concordância com o sugerido pelos autores, já que os miRNAs citados apresentam-se relacionados com os aspectos que analisamos

(Quadros 1 e 2), sendo potenciais candidatos a testes mais aprofundados *in silico* e *in vitro*. Além dos miRNAs que descrevemos como melhores candidatos (let-7, mir-17, -34, -181, -503), outros que também se destacaram foram mir-18, -21, -125, -133, -378, -202 e -224. Entretanto, priorizamos aqueles ainda não descritos em equinos e que possuem significativo relato na literatura.

Nossa análise não detectou *hits* referentes a sete miRNAs. Além disso, apenas dois não têm descrição para a espécie equina no miRBase (mir-202 e mir-320), o que significa que ainda não foram identificados experimental ou computacionalmente. A partir de todos os miRNAs candidatos que elencamos, observamos que muitos possuem papel de importância nos processos da esteroidogênese e desordens associadas a quadros de infertilidade. Aqueles que mais se destacaram em nossa análise e são sugeridos para futuros estudos são let-7, mir-17, -34, -181, -503.

Prospecção de ncRNAs no genoma equino

Nossa análise identificou que os ncRNAs tiveram mais *hits* no genoma nos cromossomos 1, X (331 *hits*) e 4, (202 *hits*). Os cromossomos com menor número de sequências identificadas foram 29, 30 e 31. Nossos resultados corroboram em parte com o estudo de Zhou et al. (2009), que observaram que os miRNAs em maior densidade por Mb estavam nos cromossomos 11, 24 e X, e ausentes nos cromossomos 29 e 31. Já os miRNAs que identificamos tiveram mais *hits* nos cromossomos X (267 *hits*), 1 (223 *hits*) e 16 (144 *hits*), enquanto os que apresentaram menos *hits* foram 29 (34 *hits*), 30 (32 *hits*) e 31 (15 *hits*).

Os miRNAs que foram mais detectados pelo *software* foram mir-544, mir-1255, mir-562 e mir-1302, que participam da manutenção do ciclo celular, processos metabólicos básicos, mitose, apoptose, regulação de mRNAs e biossíntese de aminoacetil-tRNA. Da mesma forma, todos os miRNAs estavam relacionados às vias de sinalização de receptores andrógenos. Esse resultado já era esperado, visto a abrangência da análise. Como foram as moléculas mais encontradas na análise genômica, os miRNAs comprovam-se úteis na busca por biomarcadores. Isso foi realizado por KIM et al. (2014), que avaliou possíveis biomarcadores para aplicação na clínica equina analisando-se o perfil normal de expressão de miRNAs. Através do

software predição de novos miRNAs Mireap, foram analisados tecidos do cólon, fígado e músculo esquelético, frequentemente associados a patologias. PACHOLEWSKA et al. (2016) identificaram 698 novos miRNAs através das ferramentas de predição de novos miRNAs, miRDeep e miRdentify, em 71 pequenas bibliotecas de RNA-seq. Algoritmos e *softwares* foram utilizados para detectar potenciais biomarcadores para diferenciação inter-racial entre pôneis e cavalos, revelando miR-122, miR-200, miR-483 e miR-328 como candidatos. Nossa análise do genoma identificou os ncRNAs por similaridade e, portanto, novas moléculas não são reportadas. Como existem apenas estudos transcriptômicos nesta área, nossa análise é importante como um estudo base de verificação da presença destas moléculas e avaliação das sequências que estão disponibilizadas nos bancos de dados.

Com a realização do nosso estudo, é possível afirmar que os ncRNAs descritos por nós estão presentes no genoma e provavelmente são transcritos, mas os níveis de transcrição dependem da condição fisiológica e do tipo celular. Dessa maneira, os miRNAs obtidos em nossa análise não representam o transcritoma, mas o genoma, podendo haver modificações na transcrição destas sequências. Sendo assim, é muito importante a detecção de ncRNAs *in vitro* após as análises *in silico* para a validação da sua função. Foi sugerido por SCHRIMPF et al., (2014) a associação do *locus* da Fosfolipase C com a fertilidade em garanhões, em que um SNP estaria potencialmente interferindo em sítios de ligação de fatores de transcrição e nas vias regulatórias dos miRNAs. Em éguas, não existe relato deste tipo de associação até o presente momento, mas o fato de os miRNAs estarem presentes em vesículas (DA SILVEIRA et al., 2012) e na corrente sanguínea (PACHOLEWSKA et al.,2016) estimula seu uso como biomarcador. A expressão de miRNAs durante a gestação foi avaliada em éguas , o que pode ser de relevância clínica para a detecção de complicações relacionadas à gestação (KLOHONATZ et al., 2016; LOUX et al., 2017), como já ocorre em humanos (MOUILLET et al., 2015).

Nosso estudo levou em conta as sequências de DNA que possuem complementaridade com os bancos de dados de proteínas (Rfam) e miRNAs (miRBase), sendo, portanto, detectados aqueles já obtidos experimental ou computacionalmente. É importante ressaltar que a análise *in silico* do genoma foi realizada em nível de identificação, porém a quantidade produzida destes miRNAs a

partir dos genes não são conhecidas. O estudo de ZHOU et al. (2009) foi pioneiro na detecção *in silico* de miRNAs, utilizando a ferramenta BLAST na comparação do genoma equino com sequências já anotadas de animais. Diferentemente de nosso estudo, que teve como objetivo identificar e relatar as sequências existentes no genoma com base naquelas que se alinham diretamente com os bancos de dados de proteínas e miRNAs. ZHOU et al. (2009) procuraram identificar novos miRNAs utilizando as sequências maduras dos mesmos em relação ao genoma. Além disso, diversos filtros relacionados à energia livre conformacional e estrutura secundária do miRNA foram aplicados, a fim de diferenciar pseudo-genes. Através do software INFERNAL, obtivemos um perfil a partir de um alinhamento de múltiplas sequências de uma família de RNA, com um sistema de pontuação específico para substituições, inserções e deleções. A pontuação gerada leva em conta a estrutura primária da sequência e a estrutura secundária, permitindo identificar homólogos de RNA que tem sua estrutura secundária mais conservada do a sequência primária, o que não é realizado pelo BLAST.

Conclusão

A partir do nosso estudo, foi possível obter uma lista de miRNAs que possuem um papel de relevância na fertilidade em diferentes espécies, visando empregá-los em posteriores estudos de avaliação em oócitos equinos. Os mais relevantes sugeridos para futuras análises são let-7, mir-17, -34, -181, -503. por sua presença tanto em desordens reprodutivas quanto na esteroidogênese. Incluído a isso, estes miRNAs elencados podem fornecer importantes informações sobre sua função em éguas, auxiliando na elucidação de fatores que podem interferir na fertilidade. Através de nosso estudo, é possível estabelecer um rumo concreto para pesquisas de avaliação do perfil de expressão em condições normais, em associação com disfunções, como candidatos a biomarcadores.

A análise do genoma equino através de ferramentas bioinformáticas demonstrou a presença de 21 dos 27 miRNAs elencados na revisão de literatura, demonstrando que já foram obtidos em outras análises por estarem presentes nos bancos de dados e que futuras análises são favorecidas, uma vez que já foram descritos anteriormente. Foi descrita pela primeira vez as diferentes classes de ncRNAs presentes no sequenciamento do genoma e sua distribuição, em que a maior prevalência foi de miRNAs. Os mais encontrados pelo *software* foram mir-544, mir-1255, mir-562 e mir-1302, que participam da regulação do ciclo celular e vias de sinalização de receptores andrógenos. A presença de genes codificantes para os ncRNAs encontrados aumenta a confiabilidade da busca por biomarcadores da fertilidade, uma vez se estivessem ausentes no genoma seriam fracos candidatos. Nossa análise dos ncRNAs presentes no genoma é um passo-chave para futuros estudos relacionados a avaliação da função e gene-alvo destas moléculas, quantificação através de análises como microarranjos, em condições fisiológicas ou desordens. A partir disso, podem ser considerados genes candidatos a biomarcadores e partirem para etapas de validação molecular.

Referências

BOMMER, G. T. et al. p53-Mediated Activation of miRNA34 Candidate Tumor-Suppressor Genes. **Current Biology**, v. 17, n. 15, p. 1298–1307, 7 ago. 2007.

BURNEY, R. O. et al. MicroRNA expression profiling of eutopic secretory endometrium in women with versus without endometriosis. **Molecular Human Reproduction**, v. 15, n. 10, p. 625–631, 2009.

BUZA, T. et al. Computational prediction of disease microRNAs in domestic animals. **BMC Research Notes**, v. 7, n. 1, p. 1–13, 2014.

CHOWDHARY, B. P. **Equine Genomics**. Texas, USA: John Wiley & Sons, Inc., 2013.

CORNEY, D. C. et al. MicroRNA-34b and MicroRNA-34c Are Targets of p53 and Cooperate in Control of Cell Proliferation and Adhesion-Independent Growth. **Cancer Research**, v. 67, n. 18, p. 8433–8438, 15 set. 2007.

DA SILVEIRA, J. C. et al. Cell-Secreted Vesicles in Equine Ovarian Follicular Fluid Contain miRNAs and Proteins: A Possible New Form of Cell Communication Within the Ovarian Follicle1. **Biology of Reproduction**, v. 86, n. 3, p. 1–10, 2012.

DAI, A. et al. MicroRNA-133b stimulates ovarian estradiol synthesis by targeting Foxl2. **FEBS Letters**, v. 587, n. 15, p. 2474–2482, 2013.

DONADEU, F. X.; SCHAUER, S. N. Differential miRNA expression between equine ovulatory and anovulatory follicles. **Domestic Animal Endocrinology**, v. 45, n. 3, p. 122–125, 2013.

DONADEU, F. X.; SCHAUER, S. N.; SONTAKKE, S. D. Involvement of miRNAs in ovarian follicular and luteal development. **Journal of Endocrinology**, v. 215, n. 3, p. 323–334, 2012.

ESTELLER, M. Non-coding RNAs in human disease. **Nature Reviews Genetics**, v. 12, n. 12, p. 861–874, 2011.

EULALIO, A.; HUNTZINGER, E.; IZAURRALDE, E. Getting to the root of miRNA-mediated gene silencing. **Cell**, v. 132, n. 1, p. 9–14, 11 jan. 2008.

GRIFFITHS-JONES, S. miRBase: The MicroRNA Sequence Database. In: **MicroRNA Protocols**. New Jersey: Humana Press, 2006. v. 342p. 129–138.

GUO, Y.; SUN, J.; LAI, D. Role of microRNAs in premature ovarian insufficiency. **Reproductive Biology and Endocrinology**, v. 15, n. 1, p. 1–7, 2017.

HA, M.; KIM, V. N. Regulation of microRNA biogenesis. **Nature Reviews Molecular Cell Biology**, v. 15, n. 8, p. 509–524, 2014.

HAILEMARIAM, D. et al. MicroRNA-regulated molecular mechanism underlying bovine subclinical endometritis. **Reproduction, Fertility and Development**, v. 26, n. 6, p. 898–913, 2014.

HIRAKAWA, T. et al. MiR-503, a microRNA epigenetically repressed in endometriosis, induces apoptosis and cell-cycle arrest and inhibits cell proliferation, angiogenesis, and contractility of human ovarian endometriotic stromal cells. **Human Reproduction**, v. 31, n. 11, p. 2587–2597, 2016.

HOSSAIN, M. M. et al. Altered expression of miRNAs in a dihydrotestosterone-induced rat PCOS model. **Journal of ovarian research**, v. 6, n. 1, p. 36, 15 maio 2013.

KALVARI, I. et al. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. **Nucleic Acids Research**, v. 46, n. D1, p. D335–D342, 4 jan. 2018.

KARAKAYA, C. et al. Poor ovarian response in women undergoing in vitro fertilization is associated with altered microRNA expression in cumulus cells. **Fertility and Sterility**, v. 103, n. 6, p. 1469–1476.e3, 2015.

KASIMANICKAM, V.; KASTELIC, J. Circulating cell-free mature microRNAs and their target gene prediction in bovine metritis. **Scientific Reports**, v. 6, n. July, p. 1–8, 2016.

KIM, M. C. et al. Identification and Characterization of microRNAs in normal equine tissues by next generation sequencing. **PLoS ONE**, v. 9, n. 4, p. 1–10, 2014.

KLOHONATZ, K. M. et al. Circulating miRNAs as Potential Alternative Cell Signaling Associated with Maternal Recognition of Pregnancy in the Mare. **Biology of Reproduction**, v. 95, n. 6, p. 124–124, 16 dez. 2016.

KUANG, H. et al. Profiling of differentially expressed microRNAs in premature ovarian failure in an animal model. **Gynecological Endocrinology**, v. 30, n. 1, p. 57–61, 4 jan. 2014.

LOUX, S. C. et al. Evaluation of circulating miRNAs during late pregnancy in the mare. **PLoS ONE**, v. 12, n. 4, p. 1–12, 2017.

MCBRIDE, D. et al. Identification of miRNAs associated with the follicular–luteal transition in the ruminant ovary. **REPRODUCTION**, v. 144, n. 2, p. 221–233, ago. 2012.

MOUILLET, J.-F. et al. MicroRNAs in placental health and disease. **American Journal of Obstetrics and Gynecology**, v. 213, n. 4, p. S163–S172, out. 2015.

NEILSEN, C. T.; GOODALL, G. J.; BRACKEN, C. P. IsomiRs – the overlooked repertoire in the dynamic microRNAome. **Trends in Genetics**, v. 28, n. 11, p. 544–549, 2012.

PACHOLEWSKA, A. et al. Novel equine tissue miRNAs and breed-related miRNA expressed in serum. **BMC Genomics**, v. 17, n. 1, p. 1–15, 2016.

ROTH, L. W. et al. Altered microRNA and gene expression in the follicular fluid of women with polycystic ovary syndrome. **Journal of Assisted Reproduction and Genetics**, v. 31, n. 3, p. 355–362, 2014.

SAFRAN, M. et al. GeneCards Version 3: the human gene integrator. **Database : the journal of biological databases and curation**, v. 2010, p. baq020, 5 ago. 2010.

SANG, Q. et al. Identification of MicroRNAs in Human Follicular Fluid: Characterization of MicroRNAs That Govern Steroidogenesis in Vitro and Are Associated With Polycystic Ovary Syndrome in Vivo. **The Journal of Clinical Endocrinology & Metabolism**, v. 98, n. 7, p. 3068–3079, jul. 2013.

SCHAUER, S. N. et al. Involvement of miRNAs in equine follicle development. **Reproduction**, v. 146, n. 3, p. 273–282, 2013.

SCHRIMPF, R. et al. Genome-Wide Association Study Identifies Phospholipase C zeta 1 (PLCz1) as a Stallion Fertility Locus in Hanoverian Warmblood Horses. **PLoS ONE**, v. 9, n. 10, 29 out. 2014.

SIROTKIN, A. V. et al. Identification of MicroRNAs controlling human ovarian cell steroidogenesis via a genome-scale screen. **Journal of Cellular Physiology**, v. 219, n. 2, p. 415–420, maio 2009.

SNIDER, T. A. Reproductive Disorders in Horses. **Veterinary Clinics of North America - Equine Practice**, v. 31, n. 2, p. 389–405, 2015.

SUI, S. et al. Maternal Low-protein Diet Alters Ovarian Expression of Folliculogenic and Steroidogenic Genes and Their Regulatory MicroRNAs in Neonatal Piglets. **Asian-Australasian journal of animal sciences**, v. 27, n. 12, p. 1695–704, dez. 2014.

TOLOUBEYDOKHTI, T. et al. The expression of microRNA (miRNA), mir-17, mir-211 and mir-542 and their target genes, StAR, IL-1b and Cox2 in follicular cells derived from women undergoing ART. **Fertility and Sterility**, v. 88, p. S165–S166, 1 set. 2007.

TSCHERNER, A. et al. STAT3 signaling stimulates miR-21 expression in bovine cumulus cells during in vitro oocyte maturation. **Scientific Reports**, v. 8, n. 1, p. 11527, 2018.

TU, F. et al. miR-34a targets the inhibin beta B gene, promoting granulosa cell apoptosis in the porcine ovary. **Genetics and Molecular Research**, v. 13, n. 2, p. 2504–2512, 14 jan. 2014.

VAN DER KOLK, J. H.; PACHOLEWSKA, A.; GERBER, V. The role of microRNAs in equine medicine: a review. **Veterinary Quarterly**, v. 35, n. 2, p. 88–96, 2015.

VLACHOS, I. S. et al. DIANA-miRPath v3.0: deciphering microRNA function with experimental support. **Nucleic acids research**, v. 43, n. W1, p. W460-6, 1 jul. 2015.

WADE, C. M. et al. Genome Sequence, Comparative Analysis, and Population Genetics of the Domestic Horse. **Science**, v. 326, n. 5954, p. 865–867, 2009.
XIAO, G. et al. MiR-133b Regulates the Expression of the Actin Protein TAGLN2 during Oocyte Growth and Maturation: A Potential Target for Infertility Therapy. **PLoS ONE**, v. 9, n. 6, p. e100751, 24 jun. 2014.

XU, X. J. et al. Methylation status of the FHIT gene in the transformed human mesenchymal F6 stem cell line. **Oncology Letters**, v. 9, n. 6, p. 2661–2666, 2015.

XU, Y. et al. TGF-β1 resulting in differential microRNA expression in bovine granulosa cells. **Gene**, v. 663, n. 2017, p. 88–100, 2018.

YANG, X. et al. Differentially expressed plasma microRNAs in premature ovarian failure patients and the potential regulatory function of mir-23a in granulosa cell apoptosis. **REPRODUCTION**, v. 144, n. 2, p. 235–244, ago. 2012.

YAO, G. et al. MicroRNA-224 Is Involved in Transforming Growth Factor-β-Mediated Mouse Granulosa Cell Proliferation and Granulosa Cell Function by Targeting Smad4. **Molecular Endocrinology**, v. 24, n. 3, p. 540–551, mar. 2010.

YIN, M. et al. Transactivation of microRNA-383 by Steroidogenic Factor-1 Promotes Estradiol Release from Mouse Ovarian Granulosa Cells by Targeting RBMS1. **Molecular Endocrinology**, v. 26, n. 7, p. 1129–1143, 2012.

ZHANG, J. et al. miR-143 is critical for the formation of primordial follicles in mice. **Frontiers in bioscience (Landmark edition)**, v. 18, p. 588–97, 1 jan. 2013a.

ZHANG, Q. et al. MicroRNA-181a suppresses mouse granulosa cell proliferation by targeting activin receptor IIA. **PloS one**, v. 8, n. 3, p. e59667, 2013b.

ZHOU, M. et al. In silico detection and characteristics of novel microRNA genes in the Equus caballus genome using an integrated ab initio and comparative genomic approach. **Genomics**, v. 94, n. 2, p. 125–131, 1 ago. 2009.

ZHOU, Y. et al. MicroRNA expression profiles in premature ovarian failure patients and its potential regulate functions. **Chinese Journal of Birth Health & Heredity**, 2011.

ZIELAK-STECIWKO, A. E. et al. Expression of microRNAs and their target genes and pathways associated with ovarian follicle development in cattle. **Physiological Genomics**, v. 46, n. 19, p. 735–745, 2014.

6 Conclusão Geral

As pesquisas realizadas neste Trabalho de Conclusão de curso empregando a bioinformática no estudo da fertilidade representam uma importante etapa de avanço na elucidação da genômica envolvida em desordens que afetam éguas. Além disso, a bioinformática aplicada a fêmeas demonstra grande potencial de utilização no diagnóstico molecular, com muito ainda a ser explorado.

Ao proporcionar as estruturas referentes à prolactina e seu receptor, estudos de avaliação estrutural e funcional auxiliam a desvendar características associadas à fisiopatologia de doenças reprodutivas, nível de expressão gênica em diferentes desordens. Ainda, a avaliação *in silico* e *in vitro* prospecta o impacto estrutural e funcional de SNPs nas vias de sinalização, assim como da capacidade de ligação das moléculas, as quais podem acarretar em perda de função.

A análise dos RNAs não-codificantes no genoma abre novas portas para a associação a desordens reprodutivas, pois fornece um conhecimento básico para o estudo de seu impacto nas funções foliculares e oocitárias. A partir disso, é possível detectar no transcritoma o nível de expressão em condições fisiológicas, comparando-as a diferentes condições.. As análises moleculares também podem ser realizadas em tecidos reprodutivos e fluidos biológicos de animais que apresentam desordens reprodutivas. Com isso, testes de validação para marcadores moleculares da fertilidade em éguas podem determinar aqueles que possuem maior relevância biológica para utilização no diagnóstico clínico.

Refletindo sobre o presente trabalho, observei que a área da genômica equina é altamente promissora e merece maior atenção, principalmente para fêmeas. Ainda existe uma infinidade de informações a serem descobertas relacionadas à infertilidade, o que gera uma grande oportunidade devido às diferentes causas que inviabilizam a prenhes e suprimem os índices reprodutivos da espécie equina. Durante estes três anos, nossos projetos sempre se mostraram

desafiadores, por focarmos em uma área onde a maioria dos estudos ainda só havia sido realizada em outras espécies. Ao mesmo tempo, se mostraram inovadores e acredito que contribuíram muito para a formação de meu perfil profissional, por demonstrar todas as fases de execução de um estudo e o que é necessário ao se pesquisar em um campo novo. Espero ter contribuído de alguma forma para outros grupos de pesquisa em genômica da reprodução, e, acima de tudo, gerar melhorias que se apliquem a equinocultura.

Referências

ALMEIDA, J. G. et al. Membrane proteins structures: A review on computational modeling tools. **Biochimica et Biophysica Acta - Biomembranes**, v. 1859, n. 10, p. 2021–2039, 2017.

AMERICAN HORSE COUNCIL. **Economic Impact of the United States Horse Industry**. Disponível em: <https://www.horsecouncil.org/>. Acesso em: 19 set. 2018.

ANDERSON, R. A.; TELFER, E. E. Being a good egg in the 21st century. **British Medical Bulletin**, v. 127, n. 1, p. 83–89, 2018.

APWEILER, R. et al. UniProt: the Universal Protein knowledgebase. **Nucleic acids research**, v. 32, n. Database issue, p. D115-9, 1 jan. 2004.

AURICH, C. Reproductive cycles of horses. **Animal Reproduction Science**, v. 124, n. 3–4, p. 220–228, 2011.

BAKER, D.; SALI, A. Protein structure prediction and structural genomics. **Science**, v. 294, n. 5540, p. 93–96, 2001.

BATUT, P.; GINGERAS, T. R. RAMPAGE: promoter activity profiling by paired-end sequencing of 5’-complete cDNAs. **Current protocols in molecular biology**, v. 104, p. Unit 25B.11, 11 nov. 2013.

BELTMAN, M. E. et al. Hormonal composition of follicular fluid from abnormal follicular structures in mares. **Research in Veterinary Science**, v. 97, n. 3, p. 488–490, 1 dez. 2014.

BENKERT, P.; BIASINI, M.; SCHWEDE, T. Toward the estimation of the absolute quality of individual protein structure models. **Bioinformatics (Oxford, England)**, v. 27, n. 3, p. 343–50, 1 fev. 2011.

BERTOZZO, B. R. et al. Vantagens e desafios das biotécnicas avançadas utilizadas na reprodução equina assistida. **Boletim de Indústria Animal**, v. 71, n. 1, p. 84–93, 2014.

BOLE-FEYSOT, C. et al. Prolactin (PRL) and Its Receptor: Actions, Signal Transduction Pathways and Phenotypes Observed in PRL Receptor Knockout Mice. **Endocrine Reviews**, v. 19, n. 3, p. 225–268, jun. 1998.

BOMMER, G. T. et al. p53-Mediated Activation of miRNA34 Candidate Tumor-Suppressor Genes. **Current Biology**, v. 17, n. 15, p. 1298–1307, 7 ago. 2007.

BRENDEMUEHL, J. P.; CROSS, D. L. Influence of the dopamine antagonist domperidone on the vernal transition in seasonally anoestrous mares. **Journal of reproduction and fertility. Supplement**, n. 56, p. 185–93, 2000.

BURNEY, R. O. et al. MicroRNA expression profiling of eutopic secretory endometrium in women with versus without endometriosis. **Molecular Human Reproduction**, v. 15, n. 10, p. 625–631, 2009.

BUZA, T. et al. Computational prediction of disease microRNAs in domestic animals. **BMC Research Notes**, v. 7, n. 1, p. 1–13, 2014.

CHOWDHARY, B. P. **Equine Genomics**. Texas, USA: John Wiley & Sons, Inc., 2013.

CLEVENGER, C. V. et al. The Role of Prolactin in Mammary Carcinoma. **Endocrine Reviews**, v. 24, n. 1, p. 1–27, fev. 2003.

COLEMAN, S. J. et al. Structural annotation of equine protein-coding genes determined by mRNA sequencing. **Animal Genetics**, v. 41, n. SUPPL. 2, p. 121–130, 2010.

CORNEY, D. C. et al. MicroRNA-34b and MicroRNA-34c Are Targets of p53 and Cooperate in Control of Cell Proliferation and Adhesion-Independent Growth. **Cancer Research**, v. 67, n. 18, p. 8433–8438, 15 set. 2007.

CRUZ, J. A. DA. **Processo inflamatório no útero de éguas: Endometrite**. [s.l.] UNIVERSIDADE FEDERAL DE CAMPINA GRANDE - UFCG, 2016.

CUERVO-ARANGO, J.; NEWCOMBE, J. R. Risk factors for the development of haemorrhagic anovulatory follicles in the mare. **Reproduction in Domestic Animals**, v. 45, n. 3, p. 473–480, 2010.

DA SILVEIRA, J. C. et al. Cell-Secreted Vesicles in Equine Ovarian Follicular Fluid Contain miRNAs and Proteins: A Possible New Form of Cell Communication Within the Ovarian Follicle1. **Biology of Reproduction**, v. 86, n. 3, p. 1–10, 2012.

DA SILVEIRA, J. C. et al. Regulation of ACVR1 and ID2 by cell-secreted exosomes during follicle maturation in the mare. **Reproductive Biology and Endocrinology**, v. 12, n. 1, p. 1–9, 2014.

DAI, A. et al. MicroRNA-133b stimulates ovarian estradiol synthesis by targeting Foxl2. **FEBS Letters**, v. 587, n. 15, p. 2474–2482, 2013.

DONADEU, F. X.; SCHAUER, S. N. Differential miRNA expression between equine ovulatory and anovulatory follicles. **Domestic Animal Endocrinology**, v. 45, n. 3, p. 122–125, 2013.

DONADEU, F. X.; SCHAUER, S. N.; SONTAKKE, S. D. Involvement of miRNAs in ovarian follicular and luteal development. **Journal of Endocrinology**, v. 215, n. 3, p.

323–334, 2012.

DU, H. et al. Protein structure prediction provides comparable performance to crystallographic structures in docking-based virtual screening. **Methods (San Diego, Calif.)**, v. 71, p. 77–84, jan. 2015.

ESTELLER, M. Non-coding RNAs in human disease. **Nature Reviews Genetics**, v. 12, n. 12, p. 861–874, 2011.
EULALIO, A.; HUNTZINGER, E.; IZAURRALDE, E. Getting to the root of miRNA-mediated gene silencing. **Cell**, v. 132, n. 1, p. 9–14, 11 jan. 2008.

FISER, A.; SALI, A. ModLoop: automated modeling of loops in protein structures. **BIOINFORMATICS APPLICATIONS NOTE**, v. 19, n. 18, p. 2500–2501, 2003.

FREEMAN, M. E. et al. Prolactin: Structure, Function, and Regulation of Secretion. **Physiological Reviews**, v. 80, n. 4, p. 1523–1631, 2000.

GAUTHIER, J. et al. A brief history of bioinformatics. **Briefings in Bioinformatics**, n. February, p. 1–16, 2018.

GIESECKE, K. et al. Evaluation of prolactin receptor (PRLR) as candidate gene for male fertility in Hanoverian warmblood horses. **Reproduction in Domestic Animals**, v. 45, n. 5, p. e124-30, out. 2009.

GIESECKE, K. et al. Evaluation of prolactin receptor (prlr) as candidate gene for male fertility in hanoverian warmblood horses. **Reproduction in Domestic Animals**, v. 45, n. 5, p. 124–130, 2010.

GIESECKE, K.; SIEME, H.; DISTL, O. Infertility and candidate gene markers for fertility in stallions: A review. **The Veterinary Journal**, v. 185, n. 3, p. 265–271, set. 2010.

GINTHER, O. J. et al. Incidence, Endocrinology, Vascularity, and Morphology of Hemorrhagic Anovulatory Follicles in Mares. **Journal of Equine Veterinary Science**, v. 27, n. 3, p. 130–139, mar. 2007.

GINTHER, O. J. et al. Effects of age on follicle and hormone dynamics during the oestrous cycle in mares. **Reproduction, Fertility and Development**, v. 20, n. 8, p. 955–963, 2008.

GOFFIN, V. et al. Antagonistic Properties of Human Prolactin Analogs That Show Paradoxical Agonistic Activity in the Nb2 Bioassay. **THE JOURNAL OF BIOLOGICAL CHEMISTRY**, v. 271, n. 28, p. 16573–16579, 1996.

GRIFFITHS-JONES, S. miRBase: The MicroRNA Sequence Database. In: **MicroRNA Protocols**. New Jersey: Humana Press, 2006. v. 342p. 129–138.

GUO, Y.; SUN, J.; LAI, D. Role of microRNAs in premature ovarian insufficiency. **Reproductive Biology and Endocrinology**, v. 15, n. 1, p. 1–7, 2017.

HA, M.; KIM, V. N. Regulation of microRNA biogenesis. **Nature Reviews Molecular Cell Biology**, v. 15, n. 8, p. 509–524, 2014.

HAAS, D. J.; TORRES, A. C. D. Aplicações Das Técnicas De PCR No Diagnóstico De Doenças Infecciosas Dos Animais. **Revista Científica De Medicina Veterinária**, v. 26, p. 15, 2016.

HACHIM, I. Y. et al. Prolactin Receptor Expression is an Independent Favorable

Prognostic Marker in Human Breast Cancer. **Applied Immunohistochemistry & Molecular Morphology**, v. 24, n. 4, p. 238–245, abr. 2016.

HAILEMARIAM, D. et al. MicroRNA-regulated molecular mechanism underlying bovine subclinical endometritis. **Reproduction, Fertility and Development**, v. 26, n. 6, p. 898–913, 2014.

HESTAND, M. S. et al. Annotation of the Protein Coding Regions of the Equine Genome. **PLoS ONE**, v. 10, n. 6, p. 13, 2015.

HIRAKAWA, T. et al. MiR-503, a microRNA epigenetically repressed in endometriosis, induces apoptosis and cell-cycle arrest and inhibits cell proliferation, angiogenesis, and contractility of human ovarian endometriotic stromal cells. **Human Reproduction**, v. 31, n. 11, p. 2587–2597, 2016.

HOLM, L.; SANDER, C. Evaluation of protein models by atomic solvation preference. **Journal of molecular biology**, v. 225, n. 1, p. 93–105, 5 maio 1992.

HOSSAIN, M. M. et al. Altered expression of miRNAs in a dihydrotestosterone-induced rat PCOS model. **Journal of ovarian research**, v. 6, n. 1, p. 36, 15 maio 2013.

HUANG, J. et al. **Bioinformatics in MicroRNA Research**. Methods in ed. Nova York: Springer Science, 2017. v. 1617

HUANG, W. et al. Interactions of the bovine placental lactogen and prolactin receptor genes are associated with fertility traits in cattle. **Animal The Animal Consortium**, v. 3, n. 12, p. 1743–1745, 2009.

IHLE, J. N. et al. Signaling by the cytokine receptor superfamily: JAKs and STATs. **Trends in biochemical sciences**, v. 19, n. 5, p. 222–7, maio 1994.

JOHNSON, A. L. Serum concentrations of prolactin, thyroxine and triiodothyronine relative to season and the estrous cycle in the mare. **Journal of animal science**, v. 62, n. 4, p. 1012–20, abr. 1986.

KALBFLEISCH, T. S. et al. EquCab3, an Updated Reference Genome for the Domestic Horse. **bioRxiv**, p. 306928, 25 abr. 2018.

KALVARI, I. et al. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. **Nucleic Acids Research**, v. 46, n. D1, p. D335–D342, 4 jan. 2018.

KARAKAYA, C. et al. Poor ovarian response in women undergoing in vitro fertilization is associated with altered microRNA expression in cumulus cells. **Fertility and Sterility**, v. 103, n. 6, p. 1469–1476.e3, 2015.

KASHANI-AMIN, E. et al. A systematic review on popularity, application and characteristics of protein secondary structure prediction tools. **Current Drug Discovery Technologies**, v. 15, 27 fev. 2018.

KASIMANICKAM, V.; KASTELIC, J. Circulating cell-free mature microRNAs and their target gene prediction in bovine metritis. **Scientific Reports**, v. 6, n. July, p. 1–8, 2016.

KIM, M. C. et al. Identification and Characterization of microRNAs in normal equine tissues by next generation sequencing. **PLoS ONE**, v. 9, n. 4, p. 1–10, 2014.

KIM, V. N. Small RNAs: classification, biogenesis, and function. **Molecules and cells**, v. 19, n. 1, p. 1–15, 2005.

KLOHONATZ, K. M. et al. Circulating miRNAs as Potential Alternative Cell Signaling Associated with Maternal Recognition of Pregnancy in the Mare. **Biology of Reproduction**, v. 95, n. 6, p. 124–124, 16 dez. 2016.

KOEHLER LEMAN, J.; ULMSCHNEIDER, M. B.; GRAY, J. J. Computational modeling of membrane proteins. **Proteins**, v. 83, n. 1, p. 1–24, 2015.

KRYSHTAFOVYCH, A. et al. Assessment of model accuracy estimations in CASP12. **Proteins: Structure, Function and Bioinformatics**, v. 86, n. July 2017, p. 345–360, 2018.

KUANG, H. et al. Profiling of differentially expressed microRNAs in premature ovarian failure in an animal model. **Gynecological Endocrinology**, v. 30, n. 1, p. 57–61, 4 jan. 2014.

LEBLANC, M.; CAUSEY, R. Clinical and subclinical endometritis in the mare: Both threats to fertility. **Reproduction in Domestic Animals**, v. 44, n. SUPPL. 3, p. 10–22, 2009.

LEE, R. C.; FEINBAUM, R. L.; AMBROS, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. **Cell**, v. 75, n. 5, p. 843–54, 3 dez. 1993.

LEHRMAN, S. R. et al. Primary structure of equine pituitary prolactin. **International journal of peptide and protein research**, v. 31, n. 6, p. 544–54, jun. 1988.

LEON, P. M. M. DE et al. Association between single nucleotide polymorphisms in p53 and abortion in Thoroughbred mares. **The Veterinary Journal**, v. 193, n. 2, p. 573–575, ago. 2012.

LEON, P. M. M. et al. Expression of apoptotic genes in immature and in vitro matured equine oocytes and cumulus cells. **Zygote**, v. 21, n. 03, p. 279–285, 21 ago. 2013.

LESK, A. M. **Introduction to Bioinformatics**. 4. ed. Oxford, UK: Oxford University Press, 2014.

LEVINA, V. V et al. Biological Significance of Prolactin in Gynecologic Cancers. **Cancer Res**, v. 69, n. 12, p. 5226–33, 2009.

LI, Y. et al. MicroRNAs in ovarian function and disorders. **Journal of Ovarian**

Research, 2015.

LOUX, S. C. et al. Evaluation of circulating miRNAs during late pregnancy in the mare. **PLoS ONE**, v. 12, n. 4, p. 1–12, 2017.

MAGHRABI, A. H. A.; MCGUFFIN, L. J. ModFOLD6: an accurate web server for the global and local quality estimation of 3D protein models. **Nucleic Acids Research**, v. 45, n. W1, p. W416–W421, 3 jul. 2017.

MANSOUR, T. A. et al. Tissue resolved, gene structure refined equine transcriptome. **BMC Genomics**, v. 18, n. 103, p. 12, 2017.

MCBRIDE, D. et al. Identification of miRNAs associated with the follicular–luteal transition in the ruminant ovary. **REPRODUCTION**, v. 144, n. 2, p. 221–233, ago. 2012.

MCGUFFIN, L. J.; BRYSON, K.; JONES, D. T. The PSIPRED protein structure prediction server. **Bioinformatics (Oxford, England)**, v. 16, n. 4, p. 404–5, abr. 2000.

MINISTÉRIO DA AGRICULTURA, P. E A. (MAPA). **Estudo do Complexo do Agronegócio do Cavalo**. Brasília: [s.n.]. Disponível em: <http://www.agricultura.gov.br/assuntos/camaras-setoriais-tematicas/documentos/camaras-setoriais/equideocultura/anos-anteriores/revisao-do-estudo-do-complexo-do-agronegocio-do-cavalo>. Acesso em: 15 ago. 2018.

MOHAMMADI-SANGCHESHMEH, A. et al. G6PDH-activity in equine oocytes correlates with morphology, expression of candidate genes for viability, and preimplantative in vitro development. **Theriogenology**, v. 76, n. 7, p. 1215–1226, out. 2011.

MOHAMMADI-SANGCHESHMEH, A. et al. Developmental competence of equine oocytes: impacts of zona pellucida birefringence and maternally derived transcript expression. **Reproduction, Fertility and Development**, v. 26, n. 3, p. 441, 2014.

MOREL, M. C. G. D. **Equine Reproductive Physiology, Breeding and Stud Management**. 3. ed. Aberystwyth: Cambridge University Press, 2008.

MOUILLET, J.-F. et al. MicroRNAs in placental health and disease. **American Journal of Obstetrics and Gynecology**, v. 213, n. 4, p. S163–S172, out. 2015.

NAIR, A. S. Computational Biology & Bioinformatics: A Gentle Overview. **Communications of the Computer Society of India**, p. 1–13, jan. 2007.

NEIS, A. et al. **ABORDAGEM IN SILICO : IMPACTO DE SNPs SOBRE A ESTRUTURA E FUNÇÃO BIOLÓGICA DO RECEPTOR DA PROLACTINA EQUINA**. (L. H. et Al., Ed.)Anais do I Congresso de Biotecnologia da Região Sul: Cenário Atual e Perspectivas Futuras. **Anais**...Lajeado, RS: Editora Univates, 2016Disponível em: <https://www.univates.br/editora-univates/media/publicacoes/171/pdf_171.pdf>

NEWEY, P. J. et al. Mutant Prolactin Receptor and Familial Hyperprolactinemia. **New England Journal of Medicine**, v. 369, n. 21, p. 2012–2020, 2013.

NEILSEN, C. T.; GOODALL, G. J.; BRACKEN, C. P. IsomiRs – the overlooked repertoire in the dynamic microRNAome. **Trends in Genetics**, v. 28, n. 11, p. 544–549, 2012.

ORMANDY, C. J. et al. Null mutation of the prolactin receptor gene produces multiple reproductive defects in the mouse. **Genes & development**, v. 11, n. 2, p. 167–78, 15 jan. 1997.

PACHOLEWSKA, A. et al. Novel equine tissue miRNAs and breed-related miRNA expressed in serum. **BMC Genomics**, v. 17, n. 1, p. 1–15, 2016.

PALAZZO, A. F.; LEE, E. S. Non-coding RNA: What is functional and what is junk? **Frontiers in Genetics**, v. 5, n. JAN, p. 1–11, 2015.

PTASZYNSKA, M. **Compêndio de Reprodução Animal**. [s.l: s.n.].

RAY, A.; LINDAHL, E.; ORN WALLNER, B. Improved model quality assessment using ProQ2. **BMC Bioinformatics**, v. 13, p. 1–12, 2012.

READ, R. J.; CHAVALI, G. Assessment of CASP7 predictions in the high accuracy template-based modeling category. **Proteins**, v. 69 Suppl 8, p. 27–37, 2007.

REINHART, B. J. et al. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. **Nature**, v. 403, n. 6772, p. 901–906, 24 fev. 2000.

ROJAS, J. et al. Physiologic Course of Female Reproductive Function: A Molecular Look into the Prologue of Life. **Journal of Pregnancy**, v. 2015, 2015.

ROTH, L. W. et al. Altered microRNA and gene expression in the follicular fluid of women with polycystic ovary syndrome. **Journal of Assisted Reproduction and Genetics**, v. 31, n. 3, p. 355–362, 2014.

ROY, A.; KUCUKURAL, A.; ZHANG, Y. I-TASSER: a unified platform for automated protein structure and function prediction. **Nature protocols**, v. 5, n. 4, p. 725–38, abr. 2010.

SAFRAN, M. et al. GeneCards Version 3: the human gene integrator. **Database : the journal of biological databases and curation**, v. 2010, p. baq020, 5 ago. 2010.

SÁNCHEZ, F.; SMITZ, J. Molecular control of oogenesis. **Biochimica et Biophysica Acta - Molecular Basis of Disease**, v. 1822, n. 12, p. 1896–1912, 2012.

SANG, Q. et al. Identification of MicroRNAs in Human Follicular Fluid: Characterization of MicroRNAs That Govern Steroidogenesis in Vitro and Are Associated With Polycystic Ovary Syndrome in Vivo. **The Journal of Clinical Endocrinology & Metabolism**, v. 98, n. 7, p. 3068–3079, jul. 2013.

SCHAUER, S. N. et al. Involvement of miRNAs in equine follicle development. **Reproduction**, v. 146, n. 3, p. 273–282, 2013.

SCHOENWOLF, G. et al. **Larsen Embriologia Humana**. 5. ed. Rio de Janeiro: Elsevier, 2016.

SCHRIMPF, R. et al. Genome-Wide Association Study Identifies Phospholipase C zeta 1 (PLCz1) as a Stallion Fertility Locus in Hanoverian Warmblood Horses. **PLoS ONE**, v. 9, n. 10, 29 out. 2014.

SCOTT, E. Y. et al. Identification of long non-coding RNA in the horse transcriptome. **BMC Genomics**, v. 18, n. 1, p. 1–11, 2017.

SENA, L. M. et al. Principais causas de perdas gestacionais na espécie equina: Revisão. **Pubvet**, v. 10, n. 12, p. 933–945, 2016.

SIROTKIN, A. V. et al. Identification of MicroRNAs controlling human ovarian cell steroidogenesis via a genome-scale screen. **Journal of Cellular Physiology**, v. 219, n. 2, p. 415–420, maio 2009.

SNIDER, T. A. Reproductive Disorders in Horses. **Veterinary Clinics of North America - Equine Practice**, v. 31, n. 2, p. 389–405, 2015.

SONIGO, C. et al. Hyperprolactinemia-induced ovarian acyclicity is reversed by kisspeptin administration. **The Journal of clinical investigation**, v. 122, n. 10, p. 3791–5, out. 2012.

STANTON, M. B.; STEINER, J. V; PUGH, D. . Endometrial cysts in the mare. **Journal of Equine Veterinary Science**, v. 24, n. 1, p. 14–19, 1 jan. 2004.

STUDER, G.; BIASINI, M.; SCHWEDE, T. Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane). **Bioinformatics (Oxford, England)**, v. 30, n. 17, p. i505-11, 1 set. 2014.

SUI, S. et al. Maternal Low-protein Diet Alters Ovarian Expression of Folliculogenic and Steroidogenic Genes and Their Regulatory MicroRNAs in Neonatal Piglets. **Asian-Australasian journal of animal sciences**, v. 27, n. 12, p. 1695–704, dez. 2014.

THOMPSON, D. L.; HOFFMAN, R.; DEPEW, C. L. Prolactin administration to seasonally anestrous mares: reproductive, metabolic, and hair-shedding responses. **Journal of animal science**, v. 75, n. 4, p. 1092–9, abr. 1997.

THOMPSON, D. L.; OBERHAUS, E. L. Prolactin in the Horse: Historical Perspective, Actions and Reactions, and Its Role in Reproduction. **Journal of Equine Veterinary Science**, v. 35, n. 5, p. 343–353, 2015.

TOLOUBEYDOKHTI, T. et al. The expression of microRNA (miRNA), mir-17, mir-211 and mir-542 and their target genes, StAR, IL-1b and Cox2 in follicular cells derived from women undergoing ART. **Fertility and Sterility**, v. 88, p. S165–S166, 1 set.

2007.

TSCHERNER, A. et al. STAT3 signaling stimulates miR-21 expression in bovine cumulus cells during in vitro oocyte maturation. **Scientific Reports**, v. 8, n. 1, p. 11527, 2018.

TU, F. et al. miR-34a targets the inhibin beta B gene, promoting granulosa cell apoptosis in the porcine ovary. **Genetics and Molecular Research**, v. 13, n. 2, p. 2504–2512, 14 jan. 2014.

UZIELA, K.; WALLNER, B. ProQ2: estimation of model accuracy implemented in Rosetta. **Bioinformatics**, v. 32, n. 9, p. 1411–1413, 1 maio 2016.

VAN DER KOLK, J. H.; PACHOLEWSKA, A.; GERBER, V. The role of microRNAs in equine medicine: a review. **Veterinary Quarterly**, v. 35, n. 2, p. 88–96, 2015.

VIALTA, A. **O que é biotecnologia?** Disponível em: <https://cib.org.br/faq/o-que-e-biotecnologia/>. Acesso em: 28 out. 2018.

VLACHOS, I. S. et al. DIANA-miRPath v3.0: deciphering microRNA function with experimental support. **Nucleic acids research**, v. 43, n. W1, p. W460-6, 1 jul. 2015.

VOINNET, O. Review Origin , Biogenesis , and Activity of Plant MicroRNAs. **Cell**, v. 136, n. 4, p. 669–687, 2009.

WADE, C. M. et al. Genome Sequence, Comparative Analysis, and Population Genetics of the Domestic Horse. **Science**, v. 326, n. 5954, p. 865–867, 2009.

WATSON, J. D. Técnicas de Biologia Molecular. In: TECHBOOKS (Ed.). . **Biologia Molecular do Gene**. 7. ed. São Paulo: Artmed Editora LTDA, 2015. p. 147–190.

XIAO, G. et al. MiR-133b Regulates the Expression of the Actin Protein TAGLN2 during Oocyte Growth and Maturation: A Potential Target for Infertility Therapy. **PLoS ONE**, v. 9, n. 6, p. e100751, 24 jun. 2014.

XU, D. et al. Automated protein structure modeling in CASP9 by I-TASSER pipeline combined with QUARK-based ab initio folding and FG-MD-based structure refinement. **Proteins**, n. Suppl 10, p. 147–60, 2011.

XU, D. et al. AIDA: ab initio domain assembly server. **Nucleic Acids Research**, v. 42, n. W1, p. W308–W313, 1 jul. 2014.

XU, X. J. et al. Methylation status of the FHIT gene in the transformed human mesenchymal F6 stem cell line. **Oncology Letters**, v. 9, n. 6, p. 2661–2666, 2015.

XU, Y. et al. TGF-β1 resulting in differential microRNA expression in bovine granulosa cells. **Gene**, v. 663, n. 2017, p. 88–100, 2018.

YANG, J. et al. Template-based protein structure prediction in CASP11 and retrospect of I-TASSER in the last decade. **Proteins**, v. 84 Suppl 1, n. Suppl 1, p.

233–46, 2016.

YANG, X. et al. Differentially expressed plasma microRNAs in premature ovarian failure patients and the potential regulatory function of mir-23a in granulosa cell apoptosis. **REPRODUCTION**, v. 144, n. 2, p. 235–244, ago. 2012.

YANG, Y.; ZHOU, Y. Specific interactions for ab initio folding of protein terminal regions with secondary structures. **Proteins**, v. 72, n. 2, p. 793–803, ago. 2008.

YAO, G. et al. MicroRNA-224 Is Involved in Transforming Growth Factor-β-Mediated Mouse Granulosa Cell Proliferation and Granulosa Cell Function by Targeting Smad4. **Molecular Endocrinology**, v. 24, n. 3, p. 540–551, mar. 2010.

YIN, M. et al. Transactivation of microRNA-383 by Steroidogenic Factor-1 Promotes Estradiol Release from Mouse Ovarian Granulosa Cells by Targeting RBMS1. **Molecular Endocrinology**, v. 26, n. 7, p. 1129–1143, 2012.

YOUSEF, M.; JENS, A. **miRNomics - MicroRNA Biology and Computational Analysis**. Methods in ed. Nova Iorque: Springer Science+Business Media, 2014.

ZHANG, J. et al. miR-143 is critical for the formation of primordial follicles in mice. **Frontiers in bioscience (Landmark edition)**, v. 18, p. 588–97, 1 jan. 2013a.

ZHANG, J.; LIANG, Y.; ZHANG, Y. Atomic-level protein structure refinement using fragment-guided molecular dynamics conformation sampling. **Structure (London, England : 1993)**, v. 19, n. 12, p. 1784–95, 7 dez. 2011.

ZHANG, Q. et al. MicroRNA-181a suppresses mouse granulosa cell proliferation by targeting activin receptor IIA. **PloS one**, v. 8, n. 3, p. e59667, 2013b.

ZHANG, Y. Protein structure prediction: when is it useful? **Current opinion in structural biology**, v. 19, n. 2, p. 145–55, abr. 2009.

ZHOU, H.; ZHOU, Y. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. **Protein Science**, v. 11, n. 11, p. 2714–2726, 2002.

ZHOU, M. et al. In silico detection and characteristics of novel microRNA genes in the Equus caballus genome using an integrated ab initio and comparative genomic approach. **Genomics**, v. 94, n. 2, p. 125–131, 1 ago. 2009a.

ZHOU, Y. et al. MicroRNA expression profiles in premature ovarian failure patients and its potential regulate functions. **Chinese Journal of Birth Health & Heredity**, 2011.

ZIELAK-STECIWKO, A. E. et al. Expression of microRNAs and their target genes and pathways associated with ovarian follicle development in cattle. **Physiological Genomics**, v. 46, n. 19, p. 735–745, 2014.