Universidade Federal do Pará
Núcleo de Ciências Agrárias e Desenvolvimento Rural
Empresa Brasileira de Pesquisa Agropecuária – Embrapa Amazônia Oriental
Universidade Federal Rural da Amazônia
Programa de Pós-Graduação em Ciência Animal

Priscila di Paula Bessa Santana

Perfil transcriptômico de oócitos maturados e blastocistos bubalinos (*Bubalus bubalis*) produzidos *in vitro*

Belém-Pará
2017

Priscila di Paula Bessa Santana

Perfil transcriptômico de oócitos maturados e blastocistos bubalinos (*Bubalus bubalis*) produzidos *in vitro*

Tese apresentada para obtenção do título de Doutora em Ciência Animal. Núcleo de Ciências Agrárias e Desenvolvimento Rural. Universidade Federal do Pará. Empresa Brasileira de Pesquisa Agropecuária – Amazônia Oriental. Universidade Federal Rural da Amazônia
Área de concentração: Produção Animal
Orientador: Prof. Dr. Moysés dos Santos Miranda.
Co-Orientadora: Profª. Drª. Simone do Socorro Damasceno Santos

Belém-Pará
2017

Priscila di Paula Bessa Santana

Perfil transcriptômico de oócitos maturados e blastocistos bubalinos (*Bubalus bubalis*) produzidos *in vitro*

Tese apresentada para obtenção do título de Doutora em Ciência Animal. Programa de Pós-Graduação em Ciência Animal. Núcleo de Ciências Agrárias e Desenvolvimento Rural. Universidade Federal do Pará. Empresa Brasileira de Pesquisa Agropecuária – Amazônia Oriental. Universidade Federal Rural da Amazônia
Área de concentração: Produção Animal

Belém - PA: 01/09/2017

Banca Examinadora:

Moysés dos Santos Miranda

Prof. Dr. Moysés dos Santos Miranda (Orientador)
Instituto de Ciências Biológicas – UFPA

Marcela da Silva Cordeiro

Profª. Drª. Marcela da Silva Cordeiro
Instituto Federal do Pará– IFPA

Nathalia N. da Costa de Almeida

Profª. Drª. Nathália Nogueira da Costa de Almeida
Instituto de Ciências Biológicas – UFPA

Prof. Dr. Otávio Mitio Ohashi
Instituto de Ciências Biológicas – UFPA

Prof. Dr. Rommel Thiago Jucá Ramos
Centro de Genômica e Biologia de Sistemas – UFPA

Dados Internacionais de Catalogação-na-Publicação (CIP)
Sistema de Bibliotecas da UFPA

Santana, Priscila di Paula Bessa, 1988-
Perfil transcriptômico de oócitos maturados e blastocistos bubalinos (*Bubalus bubalis*) produzidos *in vitro* / Priscila di Paula Bessa Santana. – 2017.

Orientador: Moysés dos Santos Miranda.
Tese (Doutorado) -- Universidade Federal do Pará, Núcleo de Ciências Agrárias e Desenvolvimento Rural, Programa de Pós-graduação em Ciência Animal, Belém, 2017.

1. Búfalo – Embrião. 2. Óvulo. 3. Blastocito. 4. Transcriptoma. 5. Fertilização *in vitro*.

CDD 22. ed. 636.293

Dedico

A minha mãe, Marlene Bessa
pelo apoio e carinho.

AGRADECIMENTOS

Primeiramente, citarei as instituições apoiadoras do projeto. A Universidade Federal do Pará, o Instituto de Ciências Biológicas, o Laboratório de Fertilização *In Vitro*, o Laboratório de Genômica e Bioinformática, a PósGraduação de Ciência Animal e o abatedouro frigorífico SOCIPE. Além da empresa Bubrás e o Laboratório de Genômica e Bioinformática que financiaram a PIVE e o sequencimento.

Aprendi, contudo, que o doutorado não se restinge a projetos, experimentos e artigos também consiste em uma experiência de vida que se bem aproveitada pode proporcionar aprendizados pessoais duradouros. Com este estado de espírito, as pessoas que citarei não foram importantes somente para a execusão desta tese, também são amigos e parceiros da jornada de aprendizado, chamada doutorado.

A família é parte essencial de toda esta história, por isso agradeço a minha mãe Marlene, aos meus irmãos Alex e Wylliam, a Marcela e ao nosso pequeno Guilherme. Agradecimento especial ao meu amado esposo Mariozito, que também viveu este doutorado intensamente. Obrigada pela compreensão, apoio e carinho!

Agradecimento especial e transbordante aos meus orientadores e colegas, porém, sem a formalidade destes títulos, essas pessoas são a minha segunda família: Moysés, Ohashi, Simone, Nath, Priscilla, Thiago, Marcela, Alessandra, Julianne, Moysészinho, Mayra, Eduardo, Miyuka, Karynne, Vanessa, Carol, Marcyelle, Arnaldo, Cynthia, Carla, Paulo, Hamilton, Juliana, Andrey, Ana Júlia. Agradeço imensamente os ensinamentos, a dedicação, a compreensão e o apoio!

Agradecimento ao professor Artur Luiz por literalmente descer dos céus com a idéia deste projeto. Agradeço ao professor Rommel por orientar todas as análises de bioinformática. Agradeço também aos amigos do laboratório de genômica e bioinformática, especialmente Soraya, Flávia, Kenny e Lino que foram parceiros super maravilhosos! Agradeço aos amigos de Macapá, a família Favacho, que nos recebeu com carinho proporcionando as coletas dos oócitos e embriões de búfalo.

Encerro com os sentimentos sinceros de gratidão, amizade e carinho pelas pessoas citadas sem as quais a realização deste trabalho não teria sido possível, e sem as quais esta jornada de aprendizado pessoal não teria sido tão doce.

RESUMO

Os búfalos tem potencial econômico e industrial que está, no entanto, sub-utilizado devido, entre outros fatores, a baixa eficiência de biotecnologias reprodutivas como a Produção *In Vitro* de Embriões (PIVE). A eficiência da PIVE, por sua vez, relaciona-se ao conhecimento sobre o funcionamento de oócitos e embriões. Nesse sentido, o uso de Sequenciamento De Nova Geração (NGS) de RNA (RNA-seq) têm proporcionado avanços para o entendimento dos mecanismos de maturação oocitária e desenvolvimento embrionário em diversas espécies domésticas, e foram revisados no **capítulo 1**. Este trabalho é dedicado ao estudo dos aspectos moleculares da maturação oocitária e desenvolvimento embrionário em bubalinos. No **capítulo 2**, foi feita a descrição e análise dos transcriptomas de oócitos maturados e blastocistos produzidos *in vitro* de búfalos. Para isso, mRNA foi extraído para a construção de bibliotecas *barcoded*, que foram sequenciadas na plataforma Ion Proton™. As reads foram alinhadas ao genoma de referência bovino (*Bos taurus* UMD3.1), sendo utilizados o programa Cufflinks para calcular a abundância relativa dos transcritos e o pacote DESeq2 para analisar a expressão. Observou-se a expressão de 13.976 genes, dos quais houve genes compartilhados (62%) e exclusivos com funções especializadas em oócitos (1,6%) e blastocistos (15,7%). Houve 4.153 genes diferencialmente expressos em blastocistos em relação aos oócitos, dos quais alguns genes com maior variação de expressão foram relacionados ao metabolismo de lipídios, implantação e maturação oocitária. Foi elaborado um painel transcriptômico de oócitos e blastocistos *in vitro* de búfalos, sendo discutida a importância de genes alvo promissores em futuras estratégias de aprimoramento da PIVE na espécie bubalina. No **capítulo 3**, os transcriptomas de oócitos e blastocistos de búfalos foram comparados aos de bovinos relatados na literatura. Dado que os protocolos de PIVE em búfalos são em grande parte inspirados em bovinos, esta comparação foi utilizada para indicar as simailaridades dos transcriptomas dessas espécies. Foi utilizada a análise de redes de co-expressão e preservação de módulos do pacote WGCNA do programa R. Como resultado, foi observada grande similaridade dos perfis transcriptômicos de búfalos e bovinos. Dos 7 módulos de co-expressão identificados em búfalos, 4 foram fortemente preservados (*Z-summary*>10) em bovinos, sendo suas ontologias gênicas relacionadas ao programa de desenvolvimento

embrionário. Porém, o perfil de expressão dentro dos módulos, definido pelos *hub genes*, foi diferente, indicando diferenças importantes, em termos de interações gênicas, entre as duas espécies. Conclui-se que esta análise inicial dos transcriptomas de oócitos e blastocistos bubalinos indicou a presença de genes relevantes para a qualidade oocitária e embrionária. E a comparação dos transcriptomas de búfalos e bovinos indicoudiferenças de expressão dentro de módulos conservados que justificam a elaboração de protocolos de PIVE específicos para búfalos.

Palavras-chave: transcriptoma, búfalo, *Bubalus bubalis*, PIVE, NGS, RNA-seq, oócitos, blastocistos, expressão diferencial, redes de co-expressão.

ABSTRACT

The buffalos have economic and industrial potential, which is, however, underutilized due, among other factors, to the low efficiency of reproductive biotechnologies such as In Vitro Embryo Production (PIVE). The efficiency of PIVE, in turn, is related to knowledge about the functioning of oocytes and embryos. In this sense, the use of New Generation Sequencing (NGS) of RNA (RNA-seq) has provided advances for the understanding of the mechanisms of oocyte maturation and embryo development in several domestic species, and were reviewed in **chapter 1**. This work is dedicated To study the molecular aspects of oocyte maturation and embryonic development in buffalos. In **chapter 2**, the description and analysis of the transcriptomes of mature oocytes and blastocysts produced in vitro of buffalos was made. For this, mRNA was extracted for the construction of barcoded libraries, which were sequenced on the Proton ™ Ion platform. The reads were aligned to the bovine reference genome (Bos taurus UMD3.1), using the Cufflinks program to calculate the relative abundance of the transcripts and the DESeq2 package to analyze expression. It was observed the expression of 13,976 genes, of which there were shared (62%) and exclusive genes with specialized functions in oocytes (1.6%) and blastocysts (15.7%). There were 4,153 genes differentially expressed in blastocysts in relation to oocytes, of which some genes with greater variation in expression were related to lipid metabolism, implantation and oocyte maturation. A transcriptomic panel of oocytes and in vitro blastocysts of buffalos was elaborated, and the importance of promising target genes was discussed in future strategies for enhancement of PIVE in the buffalo species. In **chapter 3**, transcripts of oocytes and buffalo blastocysts were compared to those of cattle reported in the literature. Since the PIVE protocols in buffalos are largely bovine inspired, this comparison was used to indicate the mimicry of the transcriptomes of these species. The analysis of coexpression networks and preservation of modules of the WGCNA package of the R program was used. As a result, a great similarity of buffalo and bovine transcriptomic profiles was observed. Of the 7 co-expression modules identified in buffalo, 4 were strongly preserved (Z-summary> 10) in cattle, with their genetic ontologies related to the embryonic development program. However, the expression profile within the modules, defined by the genes hub, was different, indicating important differences in terms

of gene interactions between the two species. It is concluded that this initial analysis of oocyte and buffalo blastocyst transcripts indicated the presence of genes relevant for oocyte and embryonic quality. And the comparison of buffalo and bovine transcriptomes indicated differences of expression within conserved modules that justify the elaboration of specific PIVE protocols for buffalos.

Key words: transcriptome, buffalo, *Bubalus bubalis*, IVEP, NGS, RNA-seq, oocytes, blastocysts, differential expression, co-expression network.

LISTA DE ABREVIATURAS

%	Porcentagem
°C	Grau Celsius
µg	Micrograma
Na+.	Cátion Sódio
BSA	Albumina Sérica Bovina
CCO	Complexo-*cumulus*-oócito
CIV	Cultivo *In Vitro*
CO_2	Gás Carbônico
et al.	*et alli* (e colaboradores)
FIV	Fecundação *In Vitro*
FSH	Hormônio Folículo Estimulante
HEPES	*N-2-hydroxyethylpiperazine-N'-ethanesulfonic acid*
RNAi	Ácido ribonucléico de interferência
LH	Hormônio Luteinizante
MIV	Maturação *In Vitro*
MPF	fator promotor da fase M
NGS	*Next Generation Sequence* (Sequenciamento de nova geração)
O_2	Oxigênio
p	Probabilidade
PIVE	Produção *in vitro* de embrião
PBS	Solução Salina Tamponada
RNAm	Ácido ribonucléico mensageiro
RNA-seq	*RNA-sequencing* (Sequenciamento de RNA)
miRNA	*micro* Ácido ribonucléico
PCR	Reação em cadeia da polimerase
SFB	Soro Fetal Bovino
SOF	*Sintetic Ouviduct Fluid* (Fluido do Oviduto Sintético*)*
TCM	*Tissue Culture Medium* (Meio de Cultura de Tecidos)
TALP	TALP (Tyrodes com Albumina, Lactato e Piruvato)

SUMÁRIO

1. INTRODUÇÃO

A espécie bubalina é de grande potencial econômico e industrial, no entanto esse potencial está subaproveitado. Isso se deve, entre outros fatores, a baixa eficiência das biotecnologias reprodutivas usadas em búfalos, com destaque para a produção *in vitro* de embrião (PIVE) que, apesar das limitações, têm se mostrado a mais promissora para a multiplicação de rebanhos bubalinos (GALLI et al. 2014).

Para mudar esse panorama, é essencial conhecer o funcionamento dos gametas e embriões bubalinos. Através das análises de perfil de expressão usando PCR tempo real e microarranjos foram feitos muitos avanços no entendimento do metabolismo e mecanismos de desenvolvimento de embriões de espécies domésticas relacionadas, como os bovinos (NIEMANN; WRENZYCKI 2000; CORCORAN et al. 2006; RIZOS et al. 2003). E, atualmente, ganharam grandes proporções com o uso da tecnologia de sequenciamento de nova geração de transcritos gênicos (RNA-seq) (GRAF et al. 2014; GILCHRIST et al. 2016; KROPP; KHATIB 2015; JIANG et al. 2014; REYES, et al., 2015).

Consequentemente, o entendimento de aspectos moleculares da maturação oocitária e do desenvolvimento embrionário *in vivo* e *in vitro* têm contribuído para a elaboração de meios de cultivo mais adequados as necessidades dos embriões, para a compreensão dos mecanismos genéticos de regulação da maturação oocitária e de desenvolvimento embrionário e dos seus aspectos moleculares (RUAN et al. 2012; SPATE et al. 2015; REDEL et al. 2015; BUNEL et al. 2015; LUO et al. 2016).

Dessa forma, este trabalho propõe o uso de RNA-seq como uma ferramenta para estudar o transcriptoma de oócitos maturados e blastocistos bubalinos produzidos *in vitro.* O conhecimento do perfil de expressão de oócitos e embriões de búfalo poderá auxiliar na elaboração de futuras estratégias de aprimoramento da PIVE nessa espécie.

2. OBJETIVOS

2.1 OBJETIVO GERAL

- Estudar os aspectos genéticos da maturação oocitária e de blastocistos de búfalos (*Bubalus bubalis*) produzidos *in vitro.*

2.2 OBJETIVOS ESPECÍFICOS

- Descrever o perfil transcriptômico de oócitos maturados e de blastocistos produzidos *in vitro* de búfalos;

- Analisar os genes diferencialmente expressos nos blastocistos de búfalos em relação aos oócitos maturados *in vitro*;

- Comparar o perfil transcriptômico de oócitos maturados e blastocistos de búfalos produzidos *in vitro* ao de bovinos (Bos *taurus)* descrito na literatura.

3. REVISÃO DE LITERATURA

3.1 IMPORTÂNCIA DA BUBALINOCULTURA

No Brasil, a população de bubalinos é de aproximadamente 1,19 milhões de animais (ANUALPEC, 2013) e cerca de 38,6% está no estado do Pará. De forma geral, os búfalos domésticos têm sido muito apreciados na agropecuária em diferentes países do mundo, devido a sua versatilidade para a produção de leite, carne e tração. Possuem características como grande rusticidade, tolerância, fácil adaptação ao terreno e resistência a doenças contagiosas e parasitárias, que são consideradas vantajosas, pois tornam a manutenção do rebanho muito econômica. Essas características particulares permitiram que os bubalinos povoassem extensas áreas na Amazônia (EMBRAPA, 2007).

Quanto à importância econômica da espécie bubalina, apesar de possuir dupla aptidão, seu potencial é especialmente notável para a indústria de laticinios, pois o leite de búfala comparado ao de vaca apresenta maiores níveis de gordura, proteína, vitamina A entre outros (VERRUMA; SALGADO, 1994). Por isso pode ser considerado mais nutritivo além de proporcionar maior rendimento na fabricação de derivados lácteos apresentando uma diferença de rendimento, comparado ao leite de vaca, que pode variar de 20 a 40% dependendo do produto produzido (TEIXEIRA et al., 2005).

A indústria pecuária é uma atividade de grande importância econômica para o Brasil. Especialmente com relação à exportação de carne bovina e o abastecimento do mercado interno com leite e derivados. Como exemplo, a produção de leite com uso de alta tecnologia foi à atividade agropecuária com a maior rentabilidade no ano de 2012 (7,7%) no ranking de investimentos financeiros e apresentou aumento em 2013 (10,1%) se mantendo como a atividade mais lucrativa (Scot consultoria).

3.2 USO DA PRODUÇÃO *IN VITRO* DE EMBRIÕES (PIVE) EM BUBALINOS

A Produção *In Vitro* de Embriões (PIVE) é uma biotecnologia da reprodução, cujo principal objetivo é promover a multiplicação mais rápida de genótipos zootecnicamente superiores (incluindo com maior produtividade), pois a cada vez que é realizada pode resultar no nascimento de vários bezerros (50 a 100 embriões/fêmea/ano), sendo assim encurta o tempo de geração nos rebanhos (Figura 1).

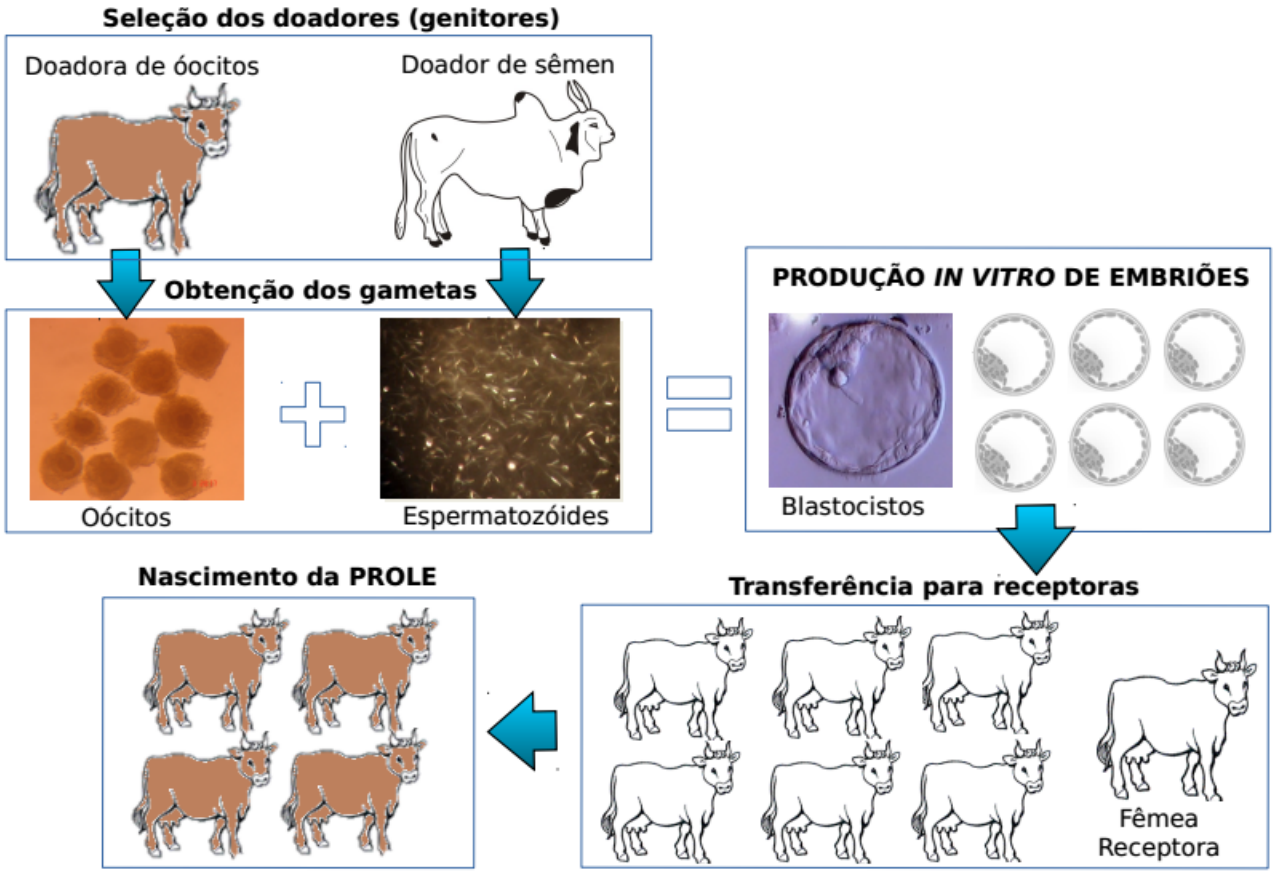


Figura 1. Uso da PIVE para a multiplicação mais rápida de genótipos zootecnicamente superiores. A seleção dos genitores é feita com base em suas características genéticas, como por exemplo, maior produtividade de carne ou leite. Os gametas são obtidos e utilizados na PIVE, gerando blastocistos que são transferidos para fêmeas receptoras. O resultado é o nascimento de prole com características genéticas selecionadas.

No Brasil, a PIVE é a biotecnologia reprodutiva mais empregada para acelerar a reprodução de espécies domésticas com interesse zootécnico (VIANA et al. 2010). Em bovinos, a PIVE é considerada uma biotécnica consolidada, enquanto que em búfalos representa a alternativa mais promissora para melhorar a produção dos rebanhos (VARAGO et al., 2008; GALLI et al. 2014). Em bovinos, as condições da PIVE são constantemente aperfeiçoadas e proporcionam taxa de blastocisto de 42 a 50% (COSTA et al., 2013; SANTANA et al., 2014a; SANTANA et al., 2014b). Em bubalinos, porém a eficiência da PIVE é menor e varia de 16 a 26% (GASPARRINI et al. 2014; DI FRANCESCO et al. 2011; NEGLIA et al. 2003). A figura 2 compara a eficiência da PIVE para as duas espécies.

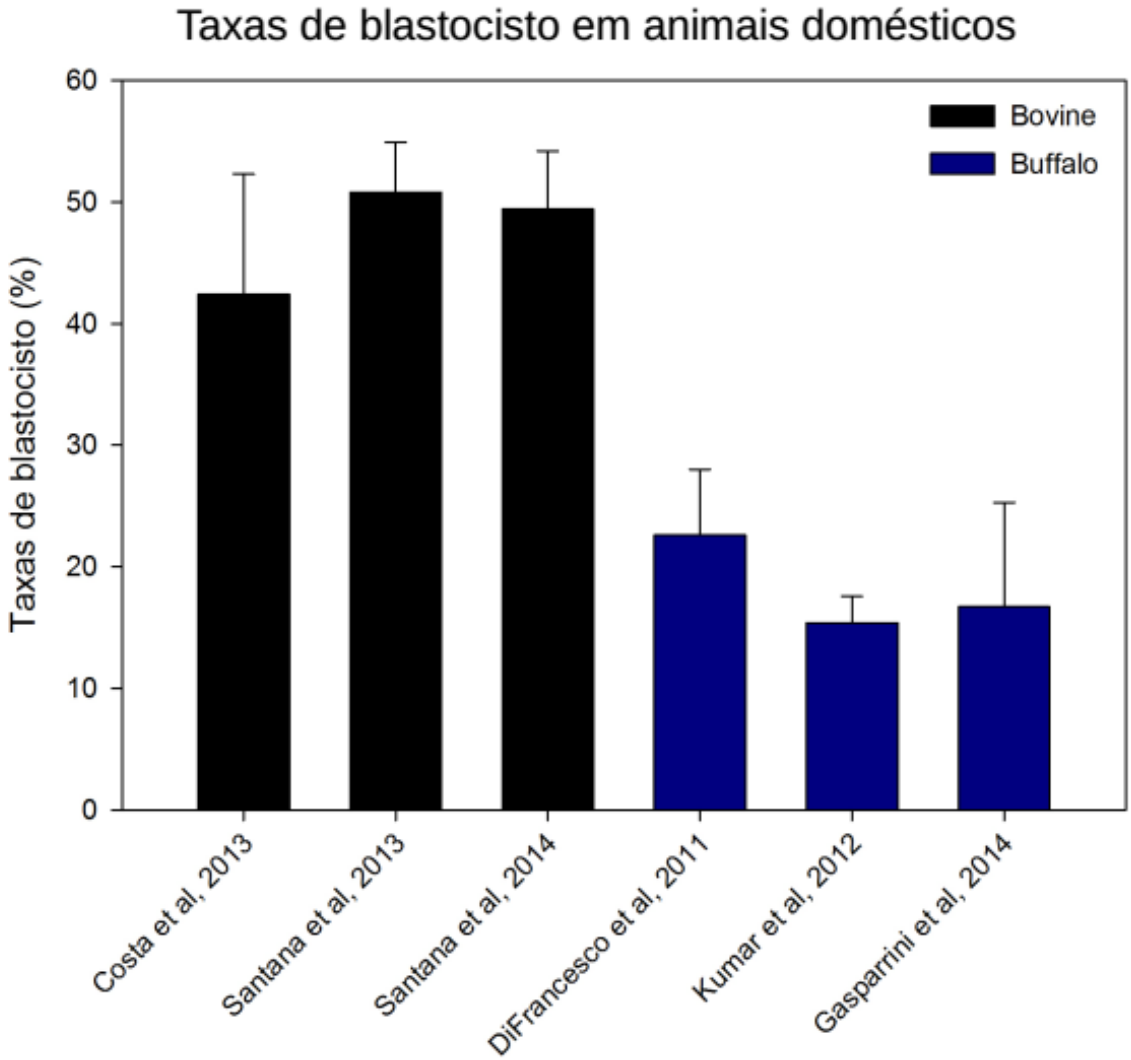


Figura 2. Eficiência da PIVE, medida em termos de taxas de blastocisto (número de blastocistos dividido pelo total de oócitos maturados), para búfalos e bovinos.

A baixa eficiência na PIVE em bubalinos tem sido relacionada à baixa qualidade oocitária e dos embriões produzidos *in vitro* indicando a necessidade de aprimorar as etapas da PIVE, que até o momento são baseadas em protocolos pré-existentes de espécie relacionada, o bovino (NEGLIA et al. 2003; DI FRANCESCO et al. 2011, 2012).

Para isso, umas das medidas seria a elaboração de meios de cultivo *in vitro* específicos para o processo de maturação oocitária, capacitação espermática, e cultivo *in vitro* dos embriões bubalinos (OHASHI et al., 2006). Para melhorar a PIVE, portanto, o primeiro passo é compreender os aspectos biológicos dos gametas e embriões bubalinos para, em seguida, buscar medidas para aprimorar as condições de cultivo *in vitro.*

3.3 ASPECTOS DA PRODUÇÃO *IN VITRO* DE EMBRIÕES (PIVE)

A PIVE pode ser dividida em 4 etapas. A primeira etapa é a colheita de oócitos, em seguida a maturação *in vitro* (MIV), a fecundação *in vitro* (FIV) e a última etapa é o cultivo *in vitro* (CIV) de embriões do estágio de zigoto até blastocisto. Então, os embriões podem ser criopreservados ou transferidos para fêmeas (PARAMIO; IZQUIERDO, 2014).

Cada etapa da PIVE é realizada com o uso de meios de cultivo específicos. Em geral, são compostos por íons, substratos energéticos, tampões e aminoácidos (BIGGERS; SUMMERS, 2008). Além desses componentes básicos, são adicionados hormônios como o FSH (hormônio estimulante folicular), LH (hormônio luteinizante) entre outros, no caso dos meios para indução da maturação oocitária (Blaschka et al., 2016). Ou são adicionados agentes indutores da capacitação do espermatozóide como heparina, penicilamina, hipotaurina, entre outros, como é o caso dos meios de FIV (PARRISH et al. 1988; MILLER et al. 2015). Ou podem ter combinações de BSA (Albumina Sérica Bovina), SFB (Soro Fetal Bovino) e agentes antioxidantes, como no caso dos meios de cultivo para suportar o desenvolvimento embrionário (KRISHER et al., 1999; FEUGANG et al. 2004).

Cada etapa da PIVE tem o objetivo de simular os ambientes *in vivo* (Figura 3), contudo este é um grande desafio tendo em vista a complexidade da tuba uterina e endométrio. Por isso, as condições da PIVE vêm sendo constantemente aperfeiçoadas através da investigação de novos componentes para os meios de cultivo *in vitro* (COSTA et al., 2016, 2013; SANTANA et al., 2014a; SANTANA et al., 2014b; SANTANA et al., 2017). E através de adaptações técnicas como o uso de atmosfera gasosa de alta ou baixa tensão de oxigênio (ELAMARAN et al., 2012) e como o preparo do sêmen e a

indução da capacitação espermática antes da FIV (LIMA et al. 2013; VELASCO et al. 2014).

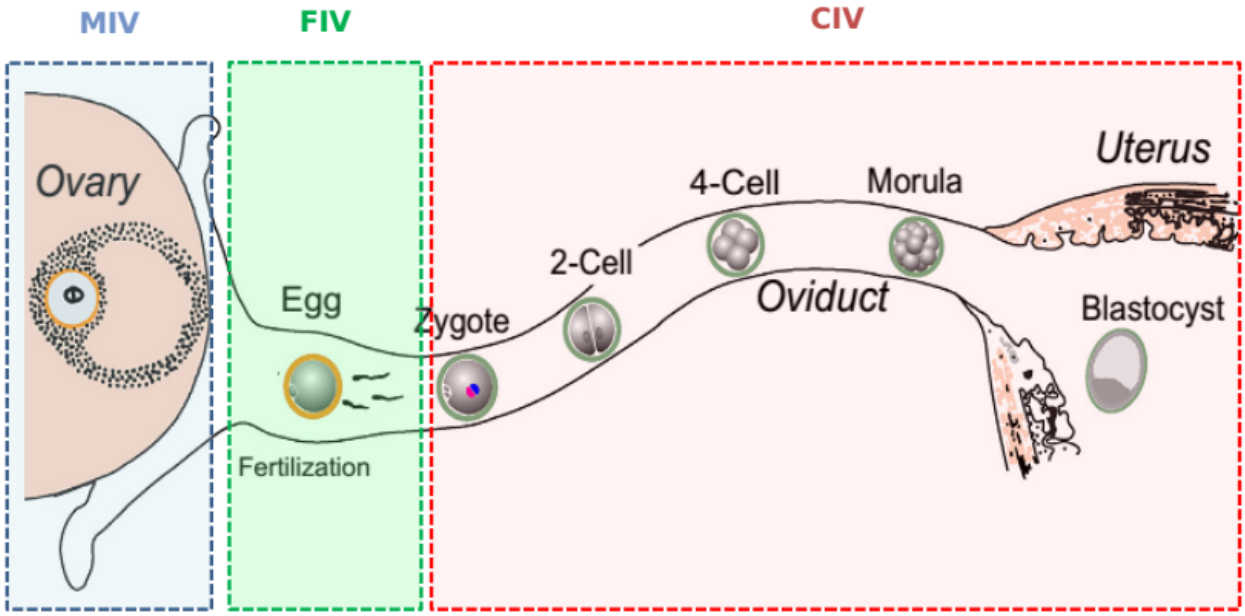


Figura 3. Etapas da Produção *In Vitro* De Embriões (PIVE). Cada etapa da PIVE mimetiza as condições do ambiente *in vivo* correspondente. A maturação *in vitro* (MIV) tenta reproduzir as condições do folículo ovariano. A fecundação *in vitro* (FIV) as condições da tuba uterina. E o cultivo *in vitro* (CIV) de embriões tenta reproduzir as condições da tuba uterina e endométrio.

3.4 USO DE RNA-seq PARA O ESTUDO DOS TRANSCRIPTOMAS

O conjunto de transcritos, ou transcriptoma, pode gerar informações sobre o metabolismo e funcionamento celular, além de fornecer evidências para interpretar os elementos funcionais do genoma (WANG et al., 2009). Atualmente, os transcriptomas são gerados pelo uso de tecnologias de hibridização (microarranjos) e de Seqüenciamento de Próxima Geração de RNA (RNA-seq). E as principais aplicações dos estudos de transcriptoma são para catalogar mRNAs e RNAs não codificantes, para inferir eventos de splicing alternativo, para identificar novos genes, para a análise de expressão diferencial e para os estudos de redes de co-expressão (HAN et al., 2015).

Os RNA-seq, geram milhões de sequências, ou *reads,* por isso são analisados empregando conjuntos de ferramentas de bioinformática. Atualmente, existe uma variedade de ferramentas disponíveis para cada etapa de processamento de RNA-seq

(STEIJGER et al. 2013). As principais etapas de processamento de RNA-seq para análise da expressão são: filtrar as reads com boa qualidade, alinhar com o genoma de referência (abordagem de montagem com referência), a etapa de contagem de reads alinhadas com transcritos, de conversão para valores relativos de expressão, de quantificação da expressão diferencial e de ontologia gênica (Figura 4) (CONESA et al. 2016).

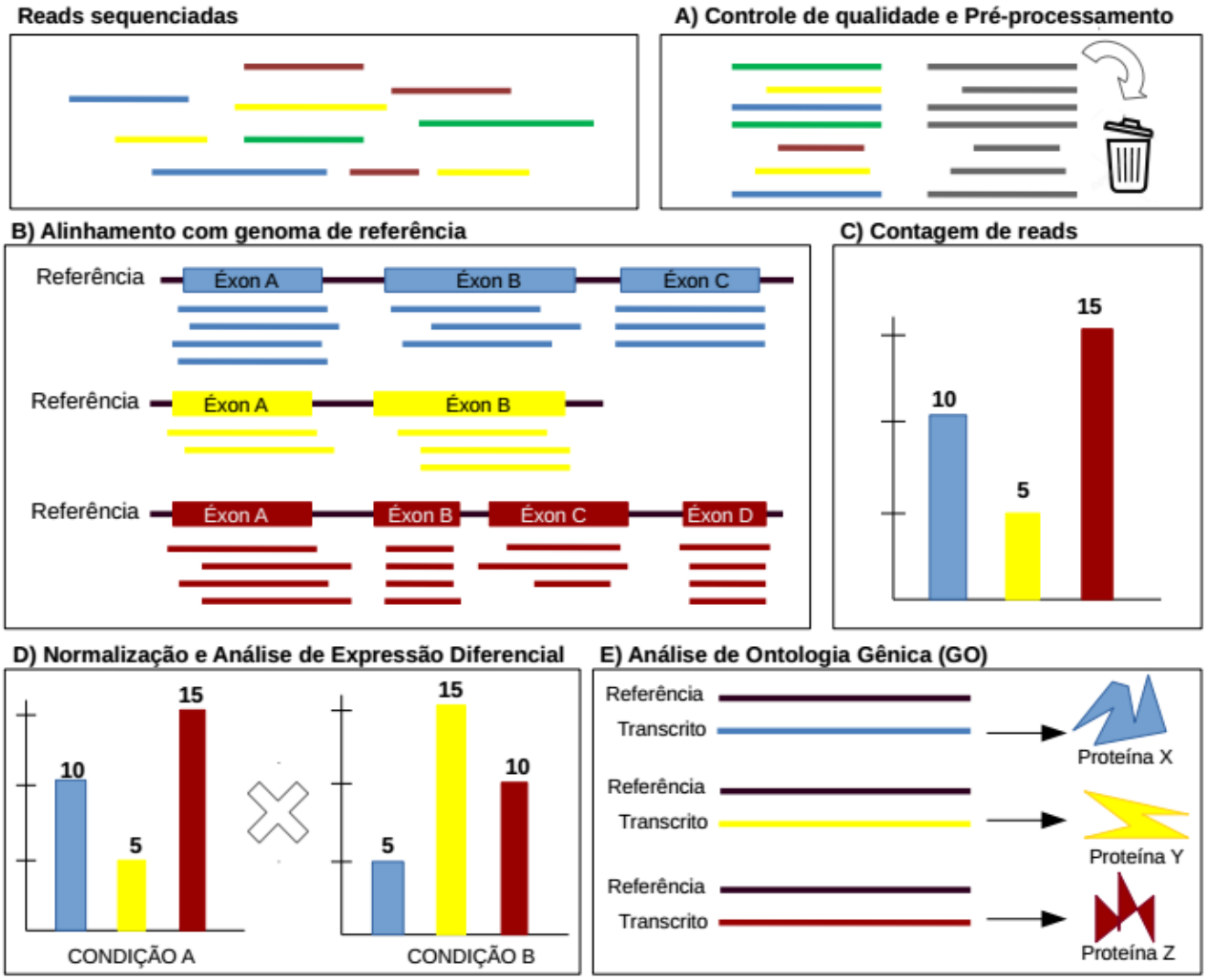


Figura 4. As principais etapas de processamento de RNA-seq para análise da expressão diferencial. A primeira etapa é filtrar as reads com boa qualidade (A), em seguida, alinhar com o genoma de referência (B), a etapa de contagem de reads alinhadas com transcritos na referência (C), etapa e normalização dos dados de expressão gênica e quantificação da expressão diferencial (D) e, por fim, análise de ontologia gênica (E).

Em RNA-seq a expressão gênica pode ser mensurada através da quantidade de reads que mapeam em uma região gênica, então, a contagem de reads alinhadas com um transcrito pode ser feita. Em seguida, existe uma variedade de métodos estatísticos para normalizar os níveis de expressão relativa, considerando o tamanho do gene (medido em kilobases) e o tamanho da biblioteca sequenciada (medido em milhões) (MORTAZAVI et al. 2008). E também os métodos de normalização que consideram além do tamanho a composição da biblioteca (ANDERS; HUBER, 2010; LOVE et al., 2014). Por fim, a etapa de ontologia gênica atribuí funções biológicas aos transcritos.

Comparada aos microarranjos a tecnologia de RNA-seq apresenta vantagens como a detecção de genes com expressão pouco abundante, e até novas variantes genéticas (HUANG; KHATIB, 2010; DYCK et al. 2014). Usando RNA-seq têm sido sequenciados os transcriptomas de várias espécies de procariotos e eucariotos (SCHOLZ et al., 2012). Além disso, os custos de RNA-seq têm diminuido tornando esta tecnologia mais acessível, por isso a quantidade de trabalhos empregando RNA-seq é cada vez maior (DRIVER et al., 2012; CHITWOOD, et al. 2013; GRAF, et al. 2014; GILCHRIST, et al. 2016; KROPP; KHATIB, 2015).

Tendo em vista estas vantagens, a tecnologia de RNA-seq é considerada a mais robusta para o estudo de transcriptomas, mesmo quando o genoma da espécie ainda não está totalmente sequenciado ou anotado, como é o caso do búfalo.

3.5. USO DE REDES DE CO-EXPRESSÃO NO ESTUDO DOS TRANSCRIPTOMAS

A análise de correlação de rede ponderada (*Weighted correlation network analysis,* WGCNA) é um método de biologia de sistemas usado para descrever os padrões de correlação entre um grande número de genes (LANGFELDER; HORVATH, 2008), tornando-a de grande utilidade para a comparação de transcriptomas e de relações evolutivas entre espécies (XUE et al., 2013; OLDHAM et al., 2008; JIANG et al., 2015).

Uma aplicação da análise de redes de co-expressão é exemplificada pelo estudo de Oldham et al. (2006), cujo objetivo foi comparar as bases moleculares da organização do cérebro de humanos e chimpanzés, fornecendo evidências do impacto das mudanças

evolutivas nos níveis de expressão gênica, além de genes candidatos de contribuírem para as especializações cognitivas em humanos (OLDHAM et al., 2006).

A análise de redes de co-expressão também fá foi aplicada para correlacionar os transcriptomas de embriões pré-implantacionais em humanos e camundongos (XUE et al., 2013) e em camundongos, humanos e bovinos (JIANG et al., 2014), com o objetivo de comparar os programas de desenvolvimento embrionário nessas espécies. Os resultados levaram a conclusão que muitos aspectos genéticos do programa de desenvolvimento são conservados nessas espécies de mamíferos. Para chegar nesses resultados, porém, as redes de co-expressão empregam conceitos que serão brevemente descritos a seguir.

Uma rede de co-expressão consisti em um grafo no qual os vértices representam os genes, e cada par de vértices é conectado por arestas que representam a relação de co-expressão entre os genes (Figura 5). Quando os genes tem relação de co-expressão, significa que seus níveis de expressão gênica tenderão a modificar juntos em função de uma determinada condição experimental.

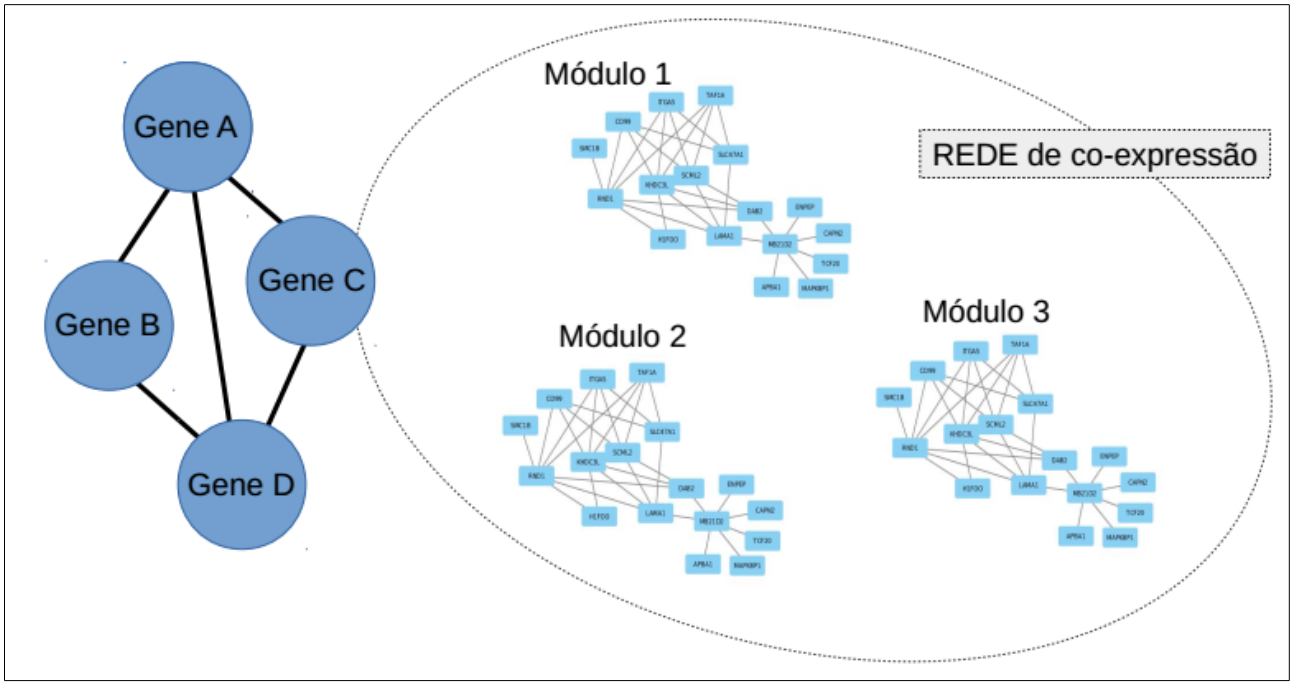


Figura 5. Conceitos de rede de co-expressão. A rede de co-expressão é um grafo no qual os vértices representam os genes (A, B, C e D) e cada par de vértices é conectado por arestas que representam a relação de co-expressão entre eles. No exemplo acima, os genes A, B, C e D estão conectados, logo tem relação de co-expressão. Com base na similaridade das correlações de co-expressão os genes são agrupados em módulos. Uma rede de co-expressão é formada pelo conjunto de vários módulos de genes co-expressos.

Ao determinar a correlação de co-expressão entre os genes, são estabelecidas conexões entre eles, e a partir disso, eles poderão ser agrupados em módulos, de acordo com a similaridade dessas correlações. Portanto, entende-se como “módulo” qualquer subgrupo de vértices que formam uma subrede dentro de uma rede maior (ZHANG; HORVATH, 2005; LANGFELDER; HORVATH, 2008). Sendo assim, uma rede de co-expressão pode ser formada, e geralmente é, pelo conjunto de vários módulos de genes co-expressos.

A representação de transcriptomas em forma de módulos de genes co-expressos é uma forma de interpretar a grande quantidade de informação, através da construção de uma “arquitetura” do transcriptoma, na qual são visualizadas as múltiplas relações de expressão de cada gene, na forma de conexões (LANGFELDER; HORVATH, 2008).

Uma análise posterior, é a comparação de módulos preservados em duas redes de co-expressão, por exemplo, de duas espécies distintas (Figura 6). Ao fim, esta análise indica a similaridade das duas arquiteturas transcriptômicas, podendo ser revelados módulos de co-expressão e também funções biológicas presentes em ambos os transcriptomas e importantes para os seus funcionamentos (LANGFELDER et al., 2011).

A ferramenta mais empregada para realizar análises de redes de co-expressão é o pacote estatístico WGCNA (*Weighted correlation network analysis*) do programa R. Esta ferramenta realiza a construção das redes de co-expressão, a identificação dos módulos de co-expressão, e a análise de módulos preservados entre duas redes de co-expressão (LANGFELDER et al., 2008; LANGFELDER; HORVATH, 2007; LANGFELDER et al., 2011; ZHANG; HORVATH, 2005).

REFERÊNCIAS BIBLIOGRÁFICAS

ANDERS, S.; HUBER, W. Differential expression analysis for sequence count data. **Genome Biol.** 11, R106, 2010.

BIGGERS, J.D.; SUMMERS, M.C. Choosing a culture medium: making informed choices. **Fertil Steril** 90, 473–483, 2008.

BUNEL, A.; JORSSEN, E.P.; MERCKX, E.; LEROY, J.L.; BOLS, P.E.; SIRARD, M. A. Individual bovine in vitro embryo production and cumulus cell transcriptomic analysis to distinguish cumulus-oocyte complexes with high or low developmental potential. **Theriogenology** 83, 228–237, 2015.

CHITWOOD, J.L.; RINCON, G.; KAISER, G.G.; MEDRANO, J.F.; ROSS, P.J. RNA-seq analysis of single bovine blastocysts. **BMC Genomics** 14, 350, 2013.

CONESA, A.; MADRIGAL, P.; TARAZONA, S.; GOMEZ-CABRERO, D.; CERVERA, A.; MCPHERSON, A.; SZCZEŚNIAK, M.W.; GAFFNEY, D.J.; ELO, L.L.; ZHANG, X.; MORTAZAVI, A. A survey of best practices for RNA-seq data analysis. **Genome Biol.** 17, 13, 2016.

CORCORAN, D.; FAIR, T.; PARK, S.; RIZOS, D.; PATEL, O.V.; SMITH, G.W.; COUSSENS, P.M.; IRELAND, J.J.; BOLAND, M.P.; EVANS, A.C.O.; LONERGAN, P. Suppressed expression of genes involved in transcription and translation in in vitro compared with in vivo cultured bovine embryos. **Reproduction** 131, 651–60, 2006.

COSTA, N.; BRITO, K.; SANTANA, P.; CORDEIRO, M.; SILVA, T.; SANTOS, A.; RAMOS, P.; SANTOS, S.; KING, W.; MIRANDA, M.; OHASHI, O. Effect of cortisol on bovine oocyte maturation and embryo development in vitro. **Theriogenology** 85, 323–9, 2016.

COSTA, N.N.; CORDEIRO, M.S.; SILVA, T.V.G.; SASTRE, D.; SANTANA, P.P.B.; SÁ, A.L. A; SAMPAIO, R.V.; SANTOS, S.S.D.; ADONA, P.R.; MIRANDA, M.S.; OHASHI, O.M. Effect of triiodothyronine on developmental competence of bovine oocytes. **Theriogenology** 80, 295–301, 2013.

DI FRANCESCO, S.; BOCCIA, L.; CAMPANILE, G.; DI PALO, R.; VECCHIO, D.; NEGLIA, G.; ZICARELLI, L.; GASPARRINI, B. The effect of season on oocyte quality and

developmental competence in Italian Mediterranean buffaloes (Bubalus bubalis). **Anim. Reprod. Sci.** 123, 48–53, 2011.

DI FRANCESCO, S.; NOVOA, M.V.S.; VECCHIO, D.; NEGLIA, G.; BOCCIA, L.; CAMPANILE, G.; ZICARELLI, L.; GASPARRINI, B. Ovum pick-up and in vitro embryo production (OPU-IVEP) in Mediterranean Italian buffalo performed in different seasons. **Theriogenology** 77, 148–154, 2012.

DRIVER, A.M.; PENAGARICANO, F.; HUANG, W.; AHMAD, K.R.; HACKBART, K.S.; WITLBANK, M.C.; KHATIB, H. RNA-Seq analysis uncovers transcriptomic variations between morphologically similar in vivo- and in vitro-derived bovine blastocysts. **BMC Genomics** 13, 118, 2012.

DYCK, M.K.; ZHOU, C.; TSOI, S.; GRANT, J.; DIXON, W.T.; FOXCROFT, G.R. Reproductive technologies and the porcine embryonic transcriptome. **Anim. Reprod. Sci.** 149, 11–18, 2014.

ELAMARAN, G.; SINGH, K.P.; SINGH, M.K.; SINGLA, S.K.; CHAUHAN, M.S.; MANIK, R.S.; PALTA, P. Oxygen Concentration and Cysteamine Supplementation During In vitro Production of Buffalo (Bubalus bubalis) Embryos Affect mRNA Expression of BCL-2, BCL-XL, MCL-1, BAX and BID. **Reprod. Domest. Anim.** 47, 1027–1036, 2012.

FEUGANG, J.M.; DE ROOVER, R.; MOENS, A.; LEONARD, S.; DESSY, F.; DONNAY, I. Addition of beta-mercaptoethanol or Trolox at the morula/blastocyst stage improves the quality of bovine blastocysts and prevents induction of apoptosis and degeneration by prooxidant agents. **Theriogenology** 61, 71–90, 2004.

GALLI, C.; DUCHI, R.; COLLEONI, S.; LAGUTINA, I.; LAZZARI, G. Ovum pick up, intracytoplasmic sperm injection and somatic cell nuclear transfer in cattle, buffalo and horses: from the research laboratory to clinical practice. **Theriogenology** 81, 138–151, 2014.

GASPARRINI, B.; NEGLIA, G.; DI PALO, R.; VECCHIO, D.; ALBERO, G.; ESPOSITO, L.; CAMPANILE, G.; ZICARELLI, L. Influence of oocyte donor on in vitro embryo production in buffalo. **Anim. Reprod. Sci.** 144, 95–101, 2014.

GILCHRIST, G.C.; TSCHERNER, A.; NALPATHAMKALAM, T.; MERICO, D.; LAMARRE, J. MicroRNA Expression during Bovine Oocyte Maturation and Fertilization. **Int. J. Mol. Sci.,** 2016.

GRAF, A.; KREBS, S.; ZAKHARTCHENKO, V.; SCHWALB, B.; BLUM, H.; WOLF, E. Fine mapping of genome activation in bovine embryos by RNA sequencing. **Proc. Natl. Acad. Sci. U. S. A.** 111, 1–6, 2014.

HUANG, W.; KHATIB, H. Comparison of transcriptomic landscapes of bovine embryos using RNA-Seq. **BMC Genomics** 11, 711, 2010.

JIANG, Z.; DONG, H.; ZHENG, X.; MARJANI, S.L.; DONOVAN, D.M.; CHEN, J.; TIAN, X.C. mRNA Levels of Imprinted Genes in Bovine In Vivo Oocytes, Embryos and Cross Species Comparisons with Humans, Mice and Pigs. **Sci. Rep.** 5, 17898, 2015.

JIANG, Z.; SUN, J.; DONG, H.; LUO, O.; ZHENG, X.; OBERGFELL, C.; TANG, Y.; BI, J.; O'NEILL, R.; RUAN, Y.; CHEN, J.; TIAN, X. Transcriptional profiles of bovine in vivo pre-implantation development. **BMC Genomics** 15, 756, 2014.

KRISHER, R.L.; LANE, M.; BAVISTER, B.D. Developmental Competence and Metabolism of Bovine Embryos Cultured in Semi-Defined and Defined Culture Media. **Biol. Reprod.** 60, 1345–1352, 1999.

KROPP, J.; KHATIB, H. mRNA fragments in in vitro culture media are associated with bovine preimplantation embryonic development. **Front. Genet.** 6, 273, 2015.

LANGFELDER, P.; HORVATH, S. WGCNA: an R package for weighted correlation network analysis. **BMC Bioinformatics** 9, 559, 2008.

LANGFELDER, P.; HORVATH, S. Eigengene networks for studying the relationships between co-expression modules. **BMC Syst. Biol.** 1, 2007.

LANGFELDER, P.; LUO, R.; OLDHAM, M.C.; HORVATH, S. Is My Network Module Preserved and Reproducible? **PLOS Comput. Biol.** 7, 1–29, 2011.

LANGFELDER, P.; ZHANG, B.; HORVATH, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. **Bioinformatics** 24, 2008.

LIMA, J.S.; LEÃO, D.L.; SAMPAIO, R.V; BRITO, A.B.; SANTOS, R.R.; MIRANDA, M.S.; OHASHI, O.M.; DOMINGUES, S.F.S. Embryo production by parthenogenetic activation and fertilization of in vitro matured oocytes from Cebus apella. **Zygote** 21, 162–6, 2013.

LOVE, M.I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biol.** 15, 550, 2014.

LUO, F.; JIA, R.; YING, S.; WANG, Z.; WANG, F. Analysis of genes that influence sheep follicular development by different nutrition levels during the luteal phase using expression profiling. **Anim. Genet.** 47, 354–64, 2016.

MATTIOLI, M.; BACCI, M.L.; GALEATI, G.; SEREN, E. Effects of LH and FSH on the maturation of pig oocytes in vitro. **Theriogenology** 36, 95–105, 2017.

MILLER, G.F.; GLIEDT, D.W.; RAKES, J.M.; RORIE, R.W. Addition of penicillamine, hypotaurine and epinephrine (PHE) or bovine oviductal epithelial cells (BOEC) alone or in combination to bovine in vitro fertilization medium increases the subsequent embryo cleavage rate. **Theriogenology** 41, 689–696, 2015.

MORTAZAVI, A.; WILLIAMS, B.A.; MCCUE, K.; SCHAEFFER, L.; WOLD, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. **Nat Meth** 5, 621–628, 2008.

NEGLIA, G.; GASPARRINI, B.; CARACCIOLO DI BRIENZA, V.; DI PALO, R.; CAMPANILE, G.; PRESICCE, G.A.; ZICARELLI, L. Bovine and buffalo in vitro embryo production using oocytes derived from abattoir ovaries or collected by transvaginal follicle aspiration. **Theriogenology** 59, 1123–1130, 2003.

NIEMANN, H.; WRENZYCKI, C. Alterations of expression of developmentally important genes in preimplantation bovine embryos by in vitro culture conditions: implications for subsequent development. **Theriogenology** 53, 21–34, 2000.

OLDHAM, M.; HORVATH, S.; GESCHWIND, D. Conservation and Evolution of Gene Co-expression Networks in Human and Chimpanzee Brains. **Proc Natl Acad Sci USA** 103, 2006.

OLDHAM, M.C.; KONOPKA, G.; IWAMOTO, K.; LANGFELDER, P.; KATO, T.; HORVATH, S.; GESCHWIND, D.H. Functional organization of the transcriptome in human brain. **Nat. Neurosci**. 11, 2008.

PARAMIO, M.T.; IZQUIERDO, D. Current Status of *In Vitro* Embryo Production in Sheep and Goats. **Reprod. Domest. Anim.** 49, 37–48, 2014.

PARRISH, J.J.; SUSKO-PARRISH, J.; WINER, M.A.; FIRST, N.L. Capacitation of bovine sperm by heparin. **Biol. Reprod.** 38, 1171–1180, 1988.

REDEL, B.K.; TESSANNE, K.J.; SPATE, L.D.; MURPHY, C.N.; PRATHER, R.S. Arginine increases development of in vitro-produced porcine embryos and affects the protein

arginine methyltransferase-dimethylarginine dimethylaminohydrolase-nitric oxide axis. **Reprod. Fertil. Dev.** 27, 655–66, 2015.

REYES, J.M.; CHITWOOD, J.L.; ROSS, P.J. RNA-Seq profiling of single bovine oocyte transcript abundance and its modulation by cytoplasmic polyadenylation. **Mol. Reprod. Dev.** 82, 103–14, 2015.

RIZOS, D.; GUTIERREZ-ADAN, A.; PEREZ-GARNELO, S.; DE LA FUENTE, J.; BOLAND, M.P.; LONERGAN, P. Bovine embryo culture in the presence or absence of serum: implications for blastocyst development, cryotolerance, and messenger RNA expression. **Biol Reprod** 68, 236–243, 2003.

RUAN, Y.C.; GUO, J.H.; LIU, X.; ZHANG, R.; TSANG, L.L.; DONG, J.D.A.; CHEN, H.; YU, M.K.; JIANG, X.; ZHANG, X.H.; FOK, K.L.; CHUNG, Y.W.; HUANG, H.; ZHOU, W.L.; CHAN, H.C. Activation of the epithelial Na+ channel triggers prostaglandin E_2 release and production required for embryo implantation. **Nat. Med.** 18, 1112–7, 2012.

SANTANA, P.D.P.B.; SILVA, T.V.G.; DA COSTA, N.N.; DA SILVA, B.B.; CARTER, T.F.; CORDEIRO, M.D.S.; DA SILVA, B.J.M.; SANTOS, S.D.S.D.; HERCULANO, A.M.; ADONA, P.R.; OHASHI, O.M.; MIRANDA, M.D.S. Supplementation of bovine embryo culture medium with L - arginine improves embryo quality via nitric oxide production. **Mol. Reprod. Dev.** 81, 918–927, 2014.

SANTANA, P.P.B.; CARVALHO, C.M.F.; DA COSTA, N.N.; SILVA, T.V.G.; RAMOS, P.C. A; CORDEIRO, M.S.; SANTOS, S.S.D.; KHAYAT, A.S.; OHASHI, O.M.; MIRANDA, M.S. Effect of dexamethasone on development of in vitro-produced bovine embryos. **Theriogenology** 1–7, 2014.

SANTANA, P.P.B.; DA SILVA, B.B.; SILVA, T.V.G.; COSTA, N.N.; CORDEIRO, M.S.; SANTOS, S.S.D.; OHASHI, O.M.; MIRANDA, M.S. Addition of L-arginine to the fertilization medium enhances subsequent bovine embryo development rates. **Theriogenology** 85, 1132–1138, 2017.

SANTOS, S.S.D.; DANTAS, J.K.; MIRANDA, M.S.; BIONDI, F.C.; OHASHI, O.M. Cinética da maturação nuclear in vitro de oócitos bubalinos. **Brazilian J. Vet. Res. Anim. Sci.** 39, 266–270, 2002.

SCHOLZ, M.B.; LO, C.C.; CHAIN, P.S.G. Next generation sequencing and bioinformatic bottlenecks: The current state of metagenomic data analysis. **Curr. Opin. Biotechnol.** 23, 9–15, 2012.

SPATE, L.D.; BROWN, A.; REDEL, B.K.; WHITWORTH, K.M.; PRATHER, R.S. PS48 can replace bovine serum albumin in pig embryo culture medium, and improve in vitro embryo development by phosphorylating AKT. **Mol. Reprod. Dev.** 82, 315–20, 2015.

STEIJGER, T.; ABRIL, J.F.; ENGSTRÖM, P.G.; KOKOCINSKI, F; The RGASP Consortium; Hubbard, T.J.; Guigó, R.; Harrow, J.; Bertone, P. Assessment of transcript reconstruction methods for RNA-seq. **Nat. Methods** 10, 1177–1184, 2013.

VARAGO, F.C.; MENDONÇA, L.F.; LAGARES, M.A. Produção in vitro de embriões bovinos: estado da arte e perspectiva de uma técnica em constante evolução. Rev. **Bras. Reprod. Anim.** 32, 100–109, 2008.

VELASCO, T.S.G.; BARAÚNA, B.; LUIZ, A.S.A.; NOGUEIRA, N. Influence of L-arginine during bovine in vitro fertilization. **Indian Journal of Experimental Biology** 52, 1159–1164, 2014.

VIANA, J.H.M.; SIQUEIRA, L.G.B.; PALHÃO, M.P.; CAMARGO, L.S.A. Use of in vitro Fertilization Technique in the Last Decade and its Effect on Brazilian Embryo Industry and Animal Production. **Acta Sci. Vet.** 38, 661–674, 2010.

XUE, Z.; HUANG, K.; CAI, C.; CAI, L.; JIANG, C.; FENG, Y.; LIU, Z.; ZENG, Q.; CHENG, L.; SUN, Y.E.; LIU, J.; HORVATH, S.; FAN, G. Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. **Nature** 500, 593–597, 2013.

ZHANG, B.; HORVATH, S. A General Framework for Weighted Gene Co-Expression Network Analysis. **Stat. Appl. Genet. Mol. Biol.,** 2005.

4. CAPÍTULO 1 - CONTRIBUIÇÕES DOS ESTUDOS DE TRANSCRIPTOMA PARA O APRIMORAMENTO DA PRODUÇÃO *IN VITRO* DE EMBRIÕES.

Artigo 1: O manuscrito foi submetido para o periódico Biology of Reproduction e segue as normas de formatação da revista

Fator de impacto (2016/2017): 3.432

Qualis Capes A2 (área de Biotecnologia)

REVIEW

Title: CONTRIBUTIONS OF TRANSCRIPTOMIC STUDIES FOR IMPROVEMENTS IN *IN VITRO* EMBRYO PRODUCTION.

Running Title: Transcriptomic analysis in *in vitro* embryo production (IVEP)

Summary Sentence: The transcriptome data has generated large amount of biological knowledge about oocytes and embryos, which is being employed to improve the efficiency of IVEP in domestic animals.

Key words: Transcriptome – RNA-seq – microarray - *In Vitro* Embryo Production – oocyte maturation – embryo development – blastocyst rate – quality biomarkers.

Authors and Affiliations: Priscila P.B. Santana[1], Artur L.C. da Silva[2], Rommel T.J. Ramos[2], Marcielly F.C. Lobato[3], Nathalia N. Costa[3], Priscilla C.A. Ramos[3], Thiago V.G. Silva[3], Marcela S. Cordeiro[4], Simone S.D. Santos[3], Otávio Mitio Ohashi[3], Moysés S. Miranda[3]

[1] Federal Rural University of Amazon, Capitão-Poço, Pará, Brazil;

[2] Laboratory of Genomics and Bioinformatics, Federal University of Pará;

[3] Laboratory of *In Vitro* Fertilization, Institute of Biological Science, Federal University of Pará;

[4] Federal Institute of Education, Science and Technology of Pará;

Grant Support: CAPES, Bubrás.

Correspondence: Santana PPB.

Abstract

In Vitro Embryo Production (IVEP) is widely used to increase reproductive efficiency in livestock animals, however, is still a challenge to increase the efficiency of IVEP in terms of embryo development rates, embryo quality, pregnancy rates and birth of healthy offspring. For that, the main limitation is the lack of biological knowledge about the gametes and embryos, which makes difficult to develop specie specific IVEP protocols. Herein, it was shown the contributions of high throughput techonologies to generate transcriptomic profiles of oocytes and embryos, and next, how this knowlegde is being used to improve the efficiency of IVEP in many livestock animals.

Introduction

In vitro embryo production (IVEP) is a reproductive biotechnology, that may be used to increase genetic gain in short generation intervals (1), to preserve genetic diversity and to treat infertility (2,3). In many species, however, improving the efficiency of IVEP is a challenge due to biological and technical aspects (4). A technical aspect is the composition of the *in vitro* culture medium, and some biological aspects are the oocyte and embryo quality (5). The efficiency of IVEP relies on the adaptation of the technical aspects to meet the biological requirements of each specie, and may be compromised by the lack of biological knowledge about the gametes and embryos (6,7).

That statement can be exemplified by comparing the rate of blastocysts obtained in IVEP systems of different domestic animals (Figure 1). In cattle, the domestic specie that there is a huge biological knowledge, the efficiency of IVEP in terms of blastocyst rate varies from 42 to 50% (8–10). However, the efficiency is lower for sheep (26-35%) (11–13), buffalos (15-22%) (14–16), pigs (13-22%) (17–19), and goats (10-19%) (20–22). Even considering that IVEP protocols may vary among laboratories, becomes clear the need to know better the requirements of gametes and embryos in order to establish more efficient protocols for each specie.

In this context, the transcriptome of oocytes and embryos has helped to overcome the lack of biological knowledge by the quantification of the expression of millions of genes. Therefore, it is used to investigate changes in the expression profile of messenger RNAs (mRNA) and even microRNAs (miRNAs). Currently, the biological knowledge generated by transcriptomes has been used to develop *in vitro* culture media more suitable for oocyte maturation and embryo culture (23) . Also to define markers of oocyte and embryo quality (24–26).

In this study, is shown the use of transcriptomic analyzes in the field of IVEP. The current biological knowledge of gametes and embryos derived from transcriptomes, and finally, how this knowledge has contributed to the improve of IVEP in many species.

High throughput technologies to generate transcriptomes

The set of transcripts, or transcriptome, can generate information about cell metabolism and functioning, as well as provide evidence to the interpretation of the functional genomic elements. Currently, transcripts are generated by the use of hybridization (microarray) and Next Generation Sequencing (NGS) of RNA (RNA-seq). The main applications of transcriptome studies are to catalog non-coding mRNAs and miRNAs, to infer events of alternative splicing, to identify new genes, for differential expression analysis and for studies of coexpression networks (27,28).

Basically, the microarray consists of single-stranded DNA oligonucleotides (probes) bound individually on the surface of histological-like slides. Each probe is builded based on the nucleotide sequence of the target gene, that must be previously known. Each microarray slide can contain a limited number of custom probes. In practice, the whole mRNA is extracted and hybridized in a complementary and specific manner to the cDNA probes. Thus, the gene expression profile is revealed by the fluorescence reading, each time the hybridization occurs. Then, one of the limitations of microarray is the need to previously know the nucleotide sequence of the target genes for a given organism, in order to build the probes (29).

When a specie-specific microarrays are not available, a microarray of closely related specie is often used to investigate the expression profile, a technique called heterologous hybridization. For example, in the buffalo specie for which the genome is not annotated (30) and there is also no specific microarray, yet microarray builded for cattle are used to analyze the gene expression profile of oocytes and buffalo embryos (31,32). However, this heterologous hybridization has important limitations because it fails to detect rare transcripts and new gene variants (33).

NGS is a high-throughput technology because each replicate can generate more information than an entire human genome (2.9 Gb). Using this approach, the genome of many domestic animals were sequenced such as goat (34), sheep (35), cow (36) and pig (37), available in online databases such as NCBI (*National Center for Biotechnology Information*, www.ncbi.nlm.nih.gov) and Ensemble (www.ensembl.org/). The use of NGS is continuously growing, as exemplified by the number of sequenced genomes (see www.ncbi.nlm.nih.gov/genbank/statistics/), as well as the number of transcriptomes available in public repositories such as GEO (*Gene Expression Omnibus*,

www.ncbi.nlm.nih.gov/geo/), SRA (*Sequence Read Archive,* www.ncbi.nlm.nih.gov/sra) and *Array Express* (www.ebi.ac.uk/arrayexpress/).

RNA-seq has been made, predominantly, on the SOLiD™, Ion™ and Illumina™ platforms, that differ in many technical respects such as the method of libraries preparation, chemistry and sequencing efficiencies (38). However, a common feature between them is the generation of millions of reads. In addition, for each read is calculated a quality value which is based on the sum of the quality of each nucleotide base, the called Phred metric. Then the accuracy of the sequencing can be measured. For example, a Phred value of 20 represents the accuracy of 99% and means that every 100bp sequenced there may be 1 error (Ewing et al., 1998). Generally, NGS platforms produce reads with Phred value equal to or above 20, that is, high quality.

Due to the large number of sequenced reads, RNA-seq data is analyzed using bioinformatics tools. Currently, a variety of tools are available for each step of RNA-seq processing (39). The main processing steps for differential expression are summarized in figure 2, such as to filter the reads with good quality, to align with a reference genome (assembly based on reference), the step of counting reads aligned with transcripts, and the conversion for relative expression values, the differential expression quantification, and, finally, the gene ontology analyse, in which the biological and molecular functions of the transcripts are attributed (40).

Using RNA-seq approach, the gene expression can be measured by the number of reads that align in a genomic region, further identified as a transcript. Next, statistical methods are used to count reads and normalize relative gene expression. To exemplify, it is known that some transcripts may be longer than others due to the gene size. Therefore, if a gene has a large number of reads it may means that it has more transcripts or that it is a longer gene than the others. One way to normalize this expression data is to use the Reads per Kilobase Million (RPKM) method that represents the expression value normalized by the gene size (measured in kilobases) and by the number of reads sequenced (measured in millions). RPKM and other similar methods such as FPKM (Fragments Per Kilobase Million) and TPM (Transcripts Per Kilobase Million) can adjust expression levels between different sized libraries (41).

There are other statistic normalization methods for adjusting expression levels between libraries with size and also composition of transcripts. For example, a X gene is

expressed in sample A (reads other than 0) and not expressed in sample B (reads equal to 0), which means that the gene is differentially expressed between the samples. However, as a consequence, the reads of X absent in B will be distributed among the other genes in sample B, leading B to have increased reads in comparison to sample A, but that not not means difference in expression. Statistical packages such as DESeq2 and edgeR apply a series of calculations to normalize libraries of different size and composition and to determine differentially expressed genes (42,43).

Finally, the biological information of the transcripts may be described in the gene ontology step. Firstly, the nucleotide sequence of the transcripts are submitted to online databases that contain the gene annotations. Thus, the biological and molecular functions of the transcripts can be defined by their homology with the sequences in the databases. The main databases for gene ontology are the UniProt (*Universal Protein Resource*, www.uniprot.org/), KEGG (*Kyoto Encyclopedia of Genes and Genomes*, www.genome.jp/kegg/), InterPro (*InterPro Consortium*, www.ebi.ac.uk/interpro/) and GO Database (*Gene Ontology Consortium*, www.geneontology.org/).

In summary, the RNA-seq approach allows the detection of low abundance transcripts (small number of reads) and new genetic variants, that are advantages over microarray (29,44,45). Another advantage is that RNA-seq can be used even when the genome is not available (46), in this case, the *de novo* assembly approach can be applied, or assembly using the genome of a closely related specie as a reference (47). In addition, NGS technologies and RNA-seq analysis are continuously upgraded through new sequencing chemistries and new bioinformatic tools. Due to that, RNA-seq is been gradually more accessible (48) and thus the number of studies in the field of reproductive biotechnology of domestic species has increased over the years (Table 1).

Transcriptomic profile of oocytes and embryos

The use of RNA-seq and microarray to generate the transcriptomic profiles of oocytes and embryos, are helping to understand important issues in reproductive biology. Next, it will be shown the novel remarks on genetic mechanisms of oocyte maturation and preimplantation embryo development and metabolism. It will also be shown how the transcriptomic data is helping to identify molecular markers of oocyte and embryo quality, to understand the effects of *in vitro* culture conditions on embryo development and quality. Then, how this knowledge has been used to improve IVEP.

Understanding of the molecular aspects of maturation and oocyte quality

It is known that during oogenesis occurs the dynamic modulation of transcripts in the cytoplasm. At the germinal vesicle stage, mammalian oocytes store a variety of transcripts. Next, at the metaphase II stage, during the oocyte maturation process, the amount of transcripts decreases (49). This modulation of RNA storage was investigated by a RNA-seq study in bovines. Was found 10,494 expressed genes, of which 2,455 were differentially expressed between immature and mature *in vitro* oocytes, about 78% (1,952 genes) was repressed. It was also reported that the transcripts stored in the oocyte's cytoplasm may undergo post-transcriptional regulation since they exhibited motifs for regulatory elements like cytoplasmic polyadenylation, in the 3 'region (50).

Oocyte transcriptomic studies also contributed to identify the miRNAs regulating oocyte maturation. The miRNAs are small and non-coding RNAs that regulate gene expression in a post-transcriptional manner, mostly through translation repression and mRNA degradation, thereby decreasing the performance of the target transcripts (51). In cattle, a RNA-seq identified 400 distinct miRNAs in immature and mature oocytes and early zygotes. Using bioinformatic tools was performed the prediction of target genes regulated by the miRNAs, which were related to transcription, protein kinase signaling and tissue formation (26). These findings suggested a regulatory role of miRNA during oocyte maturation and early embryo development. Also reinforce the evidence that the transcripts storages in oocytes are regulated by many mechanisms that remain activated in the zygotes at least during the initial period of development.

Cumulus cell transcriptomic studies also contributed to elucidate the molecular aspects of oocyte maturation. It is known that follicular cells, especially the sorrounding

cumulus cells, play an important role in the oocyte maturation by the bidirectional communication stablished with oocytes for the exchange of amino acids, metabolites, growth factors and signaling molecules through gap junctions (52,53). In cattle, it has been reported that *cumulus* cells also produce transcripts that are stored in oocytes. The transfer of mRNA from *cumulus* to oocytes occurs through transzonal projections (cytoplasmic projections that penetrate the zona pellucida and make contact with the oocyte membrane) was visualized by transmission electron microscopy (54). Notably, removal of *cumulus* cells significantly decreased the maturation rate in CCOs (35.4%) compared to intact CCOs (79.5%), showing its importance for oocyte maturation (55).

The transcriptomic studies of *cumulus* cells have also been carried out in bovines (56), sheep (57,58) and humans (59) as a noninvasive strategy for the evaluation of oocyte quality. This is particularly important in the context of IVEP and assisted human reproduction, as it allows the selection of oocytes without making it unfeasible for *in vitro* fertilization. In bovines, biopsy of *cumulus* cells were carried out in CCOs before in vitro maturation, after that the oocytes were fertilized and cultured *in vitro* to evaluate the blastocyst rate. Microarray analyzes revealed the differential expression of 68 genes between biopsies taken from oocytes that developed in blastocysts *versus* biopsies taken from the ones that failed to develop. As a result, were reported genes of oocyte quality, such as AGPAT9 (1-acylglycerol-3-phosphate O-acyltransferase) related to lipid metabolism, KRT8 (Keratin 8), and GATM (L-arginine: glycine amidinotransferase) related to amino acid metabolism and free radical sequestration (56).

Furthermore, microarray evidence in buffalos showed that the morphological and molecular quality of oocytes can influence embryo development and quality. It was reported that buffalo oocytes matured *in vivo* and exposed to high temperatures during spring and summer seasons (38 to 40°C) showed lower morphological and molecular quality and resulted in a decreased blastocyst rate compared to oocytes exposed to the autumn and winter temperatures (8 to 15°C) (60). These results, sugested that the oocyte quality is closely related to the embryo development potential (61). Therefore, the selection of good quality oocytes may result in increased developmental rates and / or embryonic quality, which are important aspects of IVEP efficiency.

Understanding of the genetic mechanisms of embryo development

After fertilization the zygotes are grown to the blastocyst stage, named *in vitro* culture (IVC) step of IVEP. During IVC important genetic events occurs such as the maternal-zygotic transition (MZT) and the small and large waves of embryo genome activation (EGA). In fact, after fertilization, the maternal genome controls all aspects of embryo development through mRNA and proteins stored in the oocyte's cytoplasm. As the embryo develops, the MZT occurs, then, fisrtly, the maternal mRNA is gradually eliminated and, secondly, the zygotic transcription is initiated (reviewed by (62).

Understand the processes of MZT and EGA is particularly useful for IVEP, since the embryo's transcriptomic profile changes completely after the start of EGA. It is assumed that know the expression profile may help to adress the embryo metabolic requirements at this period. However, EGA occurs at different times for each specie, thus, the first important information is the definition of the time of EGA.

For that, experiments with α-amanitin, an inhibitor of RNA polymerase II, induced blockage of transcription as well as the arrest of embryo development at the stage of 8 to 16 cells in cattle and sheep, and at the stage of 2 to 4 cells in buffalos (63,64), leading to the conclusion that the onset of EGA occurred at that time in these species. While in pigs, the onset of EGA showed to occur at the 4 to 8 cells stage, when the appearance of new transcripts was detected by the incorporation of [3H] uridine (65).

The RNA-seq approach introduced new strategies to identify the onset of EGA. The first approach was the detection of genes expressed in embryos and not in oocytes. The second approach was the detection of transcripts from the paternal allele using the identification of specific SNPs. And the third was the detection of incompletely processed transcripts, that is, the appearance of newly synthesized transcripts detected by the presence of intronic sequences due to incomplete co-transcriptional splicing (66). In order to generate these studies, it was sequenced 6 stages of pre-implantation development, that means a great sequence effort and analysis difficulties due to the large amount of bioinformatic data. This may explain the few species studied for the onset of EGA so far (cattle, mice, swine and humans) using RNA-seq.

The novel strategies of RNA-seq, helped to specifically determine that the onset of EGA occurred more prematurely than previously thought in bovines. Because the first wave of EGA (first wave of gene transcription) occurs at the 4 to 8 cells, as well as the

second wave of EGA (second wave of gene activation) at the 8 to 16 cells (67). In this study, was compared the transcriptomic profile of oocytes in metaphase II, embryos with 2-, 4-, 8-, 16-cells, morula and blastocyst produced *in vivo*. Were described about 2,845 genes differentially expressed between that stages, with the first major alteration of expression levels at the 4 and 8 cells, confirming the onset of minor EGA at this period.

In pigs, a RNA-seq confirmed the onset of EGA at the 2 and 4-cells by the identification of 2,101 genes differentially expressed between the stages. Gene ontology found that most of the genes were related to the "RNA splicing and mRNA processing" and "DNA metabolic processes" in agreement with the start of transcription activity (68).

Currently, the availability of transcriptomic data in online databases is allowing the comparison of the transcriptomic profile, then the embryo development programm between distant species. Surprisingly, the transcriptomic profiles of bovine and human embryos were more similar than those of mouse and human embryos, suggesting that bovine is a most suitable model to study the embryogenesis of humans (69).

Understanding of the metabolism of pre-implantational embryo

IVC is a critical step of IVEP, due to specific chemical and physical conditions that allow the embryo development from the zygote to the blastocyst stage. For that, the intrauterine and oviduct microenvironment is considered as a "gold standard" model for IVC (70), because its composition quickly changes in response to physiological regulation in order to provide the substances required by each stage of embryo development (71), such as morula compaction, blastocyst formation and hatching (72). Given the multiplicity of embryonic development and the complexity of the fluid, it is reasonable to think that the IVC environmental conditions are not entirely appropriate.

One strategy to adapt the *in vitro* culture medium to the ideal conditions (*in vivo*) has been the use of supplementations to the IVC medium followed by the gene expression analysis of the *in vitro* produced embryo compared to the *in vivo* counterpart (5,73). Therefore, tools like real-time PCR remain extremely useful for study the gene expression profiles, and helped to understand the role of leptin (74), the insulin-1-IGF-1-linked growth factor (75), relaxin (76), vascular endothelial growth factor – VEGF (77), cysteamine (78), triiodothyronine (8), L-arginine (10), among other substances supplemented in *in vitro*

culture medium. All the cited substances have been shown to increase the embryo development rates and quality in cattle, pigs and buffalos.

In contrast, real-time PCR analyzes only a few number of genes (79), given that it is need to carefully select the target genes according to their close relation with the biological condition investigated. Alternatively, the use of high throughput tecnhologies allow the quantification of a large number of genes simultaneously, thus is possible to adress the effect of a substance upon the overall expression profile of the cell.

To exemplify this point of view, using RNA-seq, Jiang et al. (2014) reported a specific metabolic behavior of bovine *in vivo* embryos, different from differentiated tissues and more similar to those of cancer cells. Based on the expression of 11,488 to 12,729 genes, the authors confirmed the speculations in the literature (80,81), because they were able to describe the expression of many metabolic pathways related to the Warburg effect such as the pentoses-phosphate, glycolysis, oxidative phosphorylation, and the tricarboxylic acid cycle (67).

The behavior known as the Warburg effect was originally proposed in 1953. According to the model, cancer cells depend on aerobic glycolysis mainly instead of mitochondrial oxidative phosphorylation. In consequence, that facilitates the uptake and incorporation of nutrients into the biomass (nucleotides, amino acids and lipids) and reduces the generation of ATP. As a result, a smaller amount of ATP is generated in comparison to the oxidative phosphorylation. On the other hand, great amount of biomass is generated and becomes the fuel needed to produce new cells (81) and to trigger rapid cell proliferation. That consists in an advantageous metabolic adaptation to the cancer cells and also to the developing embryo. Simultaneously, while glucose is shifted to the Warburg effect, the hypothesis is that the embryo depends on the β-oxidation of the fatty acids to provide the required ATP for the cells (82).

That transcriptomic studies revolutionized the knowledge about the metabolism of carbohydrates and fatty acids of pre-implantational embryos. An example that the transcriptomic findings has introduced improvements in IVEP is that novel supplements have been tested in order to induce the Warburg effect in embryos produced *in vitro*.

In pigs, supplementation of the IVC medium with 5- (4-chloro-phenyl) -3-phenyl-pent-2-enoic acid (PS48) was reported to increase blastocyst formation, as well as the total number of cells. It was observed that PS48, a PDK1 (phosphoinositide-dependent

protein kinase 1) activator, increased phosphorylation of protein kinase B (PKB / Akt). Considering the role of PKB / Akt in mediating the activation of the Warburg effect, it was speculated that the PS48 mechanism would be via its activation (19).

Also in pigs, supplementation of the IVC medium with 1.69 mM arginine resulted in increased blastocyst rate and the total number of cells (83). Considering that arginine may induce phosphorylation of mTOR (mammalian target of the rapamycin complex) and its cell signaling pathway, the authors speculated that arginine would play the role of activate the Warburg effect. Similarly, supplementation with 1 mM arginine in bovine IVC also increased blastocyst quality and hatching rate (10). Taken together, these results suggested the role of Warburg effect on the embryo development of different species of domestic animals. Also demonstrated that the knowledge derived from transcriptomes has encouraged improvements in IVEP.

Understanding of the *in vitro* culture influence on embryo quality

Embryo quality is considered as a set of characteristics related to morphology, gene expression, metabolism and cryotolerance (5). Morphological characterization has been the method of choice to predict quality, since it is non-invasive and preserves the integrity of the embryo. Is based on the time of the cellular divisions (84), number and assimetry of blastomers, cytoplasmic aspect, kinectics of development, diameter of the embryo (85,86), the cytoplasmic fragmentation (87,88) among other criteria. Based on all these aspects, embryos are classified into quality categories. Generally, only high quality embryos are selected for transfer to recipient females (86).

However, the selection of embryos based only on morphology is limited. Therefore, there is a great interest in to correlate the morphological and molecular aspects of embryo quality with the goal of select embryos theoretically more competent for implantation. As the *in vivo* embryo is considered the best quality reference, many transcriptomic studies have focused on the comparison of embryos *in vivo* and *in vitro*.

In cattle, a RNA-seq study found 793 differentially expressed genes between *in vivo* and *in vitro* embryos, mainly related to the transcription and translation (89). These results were considerably larger than a previous microarray, that identified 384 expressed genes, of which 85% (326) were differentially expressed in blastocysts *in vitro* versus *in vivo* (90). In pigs, a RNAseq showed the differential expression of 1,143 genes in *in vitro versus in*

vivo embryos at the 2-cells stage, of which 993 (86%) were induced in *in vitro* embryos (68).

It is known that the morphological and transcriptomic distinction between *in vitro* and *in vivo* embryos is associated with sub-optimal *in vitro* culture conditions (5,91–93). The components added to the culture medium can modify gene expression. Although the use of Fetal Bovine Serum (FBS) and Bovine Serum Albumin (BSA) is considered essential for the production of good quality embryos (94,95) concentrations of both components only decreases survival after vitrification (cryotolerance). For that, one of the effects of SFB is the accumulation of lipid droplets in the cytoplasm which is the most probable mechanism responsible for the decrease in cryotolerance (96,97). However, its molecular mechanisms on embryo development are still unclear.

In cattle, a RNA-seq study shed a light on the functions of SFB by reporting the induced expression of the cholesterol biosynthesis pathway in embryos *in vitro* cultured in a SFB-free medium compared to the *in vivo* counterparts. The authors speculated that the absence of SFB may lead to the activation of cholesterol biosynthesis, as a mechanism to compensate the *in vitro* system, distinct from the uterine environment, where cholesterol would be supplied unlimited (89). This led to the conclusion that it is necessary to provide a certain amount of lipids in the medium, what makes sense once lipids are required to produce phospholipids and energy via the β oxidation pathway.

Morphological studies about the differences in size, number, spatial distribution and chemical composition of lipid droplets support the notion that oocytes and embryos differentially regulate the fatty acid storage and oxidation (98,99). Thus, possibly different amounts of lipids are required for oocyte *in vitro* maturation and *in vitro* embryo culture, however, future studies on the lipid profile of oocytes and embryos may adress this question. Taken togheter, these evidences demonstrate the need to establish the correlation between morphological and molecular aspects of embryo quality.

Understanding of the embryo quality and its relation to implantation

The “gold standard” and undoubted test of embryo quality is the successful implantation and pregnancy (86). But, how is it possible to know which embryo is capable of implanting and generating pregnancy before the transfer? Regarding to the implantation, during the whole process occurs a complex interaction between embryo and

endometrium through hormonal regulation and cross-talk of molecular signals. Depending on the specie, the implantation may be invasive or non-invasive, which means that different levels of interaction can be established between embryo and endometrial cells (100). Surprisingly, evidences has shown that good-quality embryos can signal their developmental competence, so the endometrium selects them positively and pregnancy is established, the notion named selective endometrium (101).

The first evidence supporting the notion of selective endometrium emerged from the microarray of carunculate and endometrial intercaruncular areas obtained from pregnant bovine females (102). The goal was to address the endometrial response and transcriptomic profile after the transfer of embryos produced by Nuclear Somatic Cell Transfer (TNCS) compared to *In Vitro* Fertilization (IVF) and Artificial Insemination (AI), considering that TNCS embryos have low potential to establish pregnancy (103,104). The results of the microarray showed many differentially expressed genes in the endometrium of pregnant animals of TNCS embryos in comparison to the control pregnancies (IVF and AI). These results indicated that the type or qulity of the embryos influence the response of the endometrium and the establishment of pregnancy.

The microarray of the human endometrium also support the selective notion (105). Surprisingly, endometrial cells exposed to culture medium conditioned by good quality embryos showed only 15 differentially expressed genes in comparison to 449 differentially expressed genes in endometrial cells exposed to the medium conditioned by non-viable embryos. It suggested a correlation between embryo quality and the hability to induce response in the endometrium. Furthermore, this evidence also raises a significant question: how do embryos signal their development potential?

Notably, was showed that a serine protease released by the embryos (trypsin), elicits cascade effects ending up with the release of prostaglandin E2 in mouse endometrial epithelial cells, which in turn led to decidualization and implantation (106). The mechanism described above may be only part of the explanation, it is likely that embryos are able to release many signal molecules to induce endometrial response. Certainly, the identification of embryo signals and the understanding of how to induce its production and release in an *in vitro* culture system may open new perspectives in reproductive biology, particularly for the IVEP technology.

Other molecular biomarkers of embryo quality have been reported in *in vitro* culture medium, that is particularly important as a non-invasive strategy to select good quality embryos. For that, RNA-seq of mRNA and miRNA in conditioned medium were performed. In bovines, it was observed the differential expression of 11 miRNAs (107) and 17 mRNAs (24) in the medium conditioned by arrested embryos *versus* medium conditioned by fully developed blastocysts. Additionally, in the last one, POSTN and VSNL-1 mRNAs were found strongly induced, and were speculated as biomarkers of good quality embryos. There is also evidence that the mRNAs and miRNAs secreted by embryos may play an important role in embryo development and implantation. To exemplify, the inhibition of POSTN mRNAs translation, using the RNA interfering technique, resulted in a significant decrease of the blatocyst rate (24). Future studies may clarify the regulatory role of RNAs secreted by embryos and its effects on embryo quality, possibly related to the communication with the endometrium.

The correlations between morphological and molecular aspects of embryo quality are not easy to establish, but the transcriptomic studies are helping to establish them. Undoubtedly, the understanding of the molecular aspects of embryo quality and the development of non-invasive strategies to select good quality embryos for transfer has already contributed considerably to select embryos, theorically, most able to implant. Looking foward, transcriptomic studies may contribute to applied the strategy of the non-invasive selection to other species too. That, sooner, may increase pregnancy rates and the efficiency of IVEP in many livestock animals.

Future Perspectives

The knowledge about the molecular aspects of oocyte maturation and embryo development has led to the improvement of *in vitro* culture medium and the description of molecular biomarkers for selection of competent oocytes and embryos. In parallel, other reproductive biotechnologies like TNCS and transgenesis can take advantage of the improvements in IVEP (108). Similarly, studies about the genetic mechanisms of embryo development may provide insights into related research fields like stem cell reprogramming and genetics of cancer cells.

Future challenges to the transcriptomic analysis will be the study of the gene biological functions and its correlations with the embryo development rates. With this aim, the inhibition of mRNA translation has been used to verify the impact of the loss of the gene function on embryo development (109). As well as the use of supplementation on the *in vitro* culture medium with substances to inhibit and induce the expression of specific target genes. Next, proteomic and metabolomic studies may help to clarify the mechanisms of post-transcriptional and translational regulation of the gene expression.

Acknowledgment

We thanks all members of the *In Vitro* Fertilization Laboratory in Federal University of Pará for their valuable opinions on this topic.

References

1. Kasinathan P, Wei H, Xiang T, Molina JA, Metzger J, Broek D, et al. Acceleration of genetic gain in cattle by reduction of generation interval. *Sci Rep.* 2015; 5:8674.
2. Comizzoli P, Songsasen N, Wildt DE. Protecting and extending fertility for females of wild and endangered mammals. *Cancer Treat Res.* 2010; 156:87–100.
3. Lima JS, Leão DL, Sampaio R V, Brito AB, Santos RR, Miranda MS, et al. Embryo production by parthenogenetic activation and fertilization of in vitro matured oocytes from Cebus apella. *Zygote* 2013; 21(2):162–6.
4. Paramio M-T, Izquierdo D. Current Status of *In Vitro* Embryo Production in Sheep and Goats. *Reprod Domest Anim* 2014; 49(s4):37–48.
5. Lonergan P, Fair T, Corcoran D, Evans AC. Effect of culture environment on gene expression and developmental characteristics in IVF-derived embryos. *Theriogenology* 2006; 65(1):137–52.
6. de Souza-Fabjan JMG, Panneau B, Duffard N, Locatelli Y, de Figueiredo JR, Freitas VJ de F, et al. In vitro production of small ruminant embryos: late improvements and further research. *Theriogenology* 2014; 81(9):1149–62.
7. Grupen CG. The evolution of porcine embryo in vitro production. *Theriogenology* 2014; 81(1):24–37.
8. Costa NN, Cordeiro MS, Silva TVG, Sastre D, Santana PPB, Sá a. L a, et al. Effect of triiodothyronine on developmental competence of bovine oocytes. *Theriogenology*. 2013; 80(4):295–301.
9. Santana PPB, Carvalho CMF, da Costa NN, Silva TVG, Ramos PC a, Cordeiro MS, et al. Effect of dexamethasone on development of in vitro-produced bovine embryos. *Theriogenology* 2014; 1–7.
10. Santana PDPB, Silva TVG, da Costa NN, da Silva BB, Carter TF, Cordeiro MDS, et al. Supplementation of bovine embryo culture medium with L-arginine improves embryo quality via nitric oxide production. *Mol Reprod Dev* 2014; 81(10):918–27.
11. Mara L, Sanna D, Casu S, Dattena M, Mayorga Muñoz IM. Blastocyst rate of in vitro embryo production in sheep is affected by season. *Zygote* 2013; 22:366–71.

12. Shirazi A, Ardali MA, Ahmadi E, Nazari H, Mamuee M, Heidari B. The Effect of Macromolecule Source and Type of Media During in vitro Maturation of Sheep Oocytes on Subsequent Embryo Development. *J Reprod Infertil* 2012; 13(1):13–9.

13. Naderi MM, Borjian Boroujeni S, Sarvari A, Heidari B, Akhondi MM, Zarnani A-H, et al. The Effect of Media Supplementation with Angiotensin on Developmental Competence of Ovine Embryos Derived from Vitrified-warmed Oocytes. *Avicenna J Med Biotechnol* 2016; 8(3):139–44.

14. Gasparrini B, Neglia G, Di Palo R, Vecchio D, Albero G, Esposito L, et al. Influence of oocyte donor on in vitro embryo production in buffalo. *Anim Reprod Sci* 2014; 144(3–4):95–101.

15. Di Francesco S, Boccia L, Campanile G, Di Palo R, Vecchio D, Neglia G, et al. The effect of season on oocyte quality and developmental competence in Italian Mediterranean buffaloes (Bubalus bubalis). *Anim Reprod Sci* 2011; 123(1–2):48–53.

16. Kumar P, Verma A, Roy B, Rajput S, Ojha S, Anand S, et al. Effect of Varying Glucose Concentrations during In Vitro Maturation and Embryo Culture on Efficiency of In Vitro Embryo Production in Buffalo. *Reprod Domest Anim* 2012; 47(2):269–73.

17. Yoshioka K, Noguchi M, Suzuki C. Production of piglets from in vitro-produced embryos following non-surgical transfer. *Anim Reprod Sci.* 2012; 131(1):23–9.

18. Nguyen BX, Kikuchi K, Uoc NT, Dang-Nguyen TQ, Linh NV, Men NT, et al. Production of Ban miniature pig embryos by in vitro fertilization: A comparative study with Landrace. *Anim Sci J* 2015; 86(5):487–93.

19. Spate LD, Brown A, Redel BK, Whitworth KM, Prather RS. PS48 can replace bovine serum albumin in pig embryo culture medium, and improve in vitro embryo development by phosphorylating AKT. *Mol Reprod Dev* 2015; 82(4):315–20.

20. Rodríguez-Dorta N, Cognié Y, González F, Poulin N, Guignot F, Touzé J-L, et al. Effect of coculture with oviduct epithelial cells on viability after transfer of vitrified in vitro produced goat embryos. *Theriogenology* 2007; 68(6):908–13.

21. Bormann CL, Ongeri EM, Krisher RL. The effect of vitamins during maturation of caprine oocytes on subsequent developmental potential in vitro. *Theriogenology* 2017; 59(5):1373–80.

22. Kątska-Książkiewicz L, Opiela J, Ryńska B. Effects of oocyte quality, semen donor and embryo co-culture system on the efficiency of blastocyst production in goats. *Theriogenology* 2007; 68(5):736–44.

23. Prather R, Redel B, Whitworth K, Zhao M. Genomic profiling to improve embryogenesis in the pig. *Anim Reprod Sci* 2014; 149(1–2):39–45.

24. Kropp J, Khatib H. mRNA fragments in in vitro culture media are associated with bovine preimplantation embryonic development. *Front Genet* 2015; 6:273.

25. Reyes JM, Chitwood JL, Ross PJ. RNA-Seq profiling of single bovine oocyte transcript abundance and its modulation by cytoplasmic polyadenylation. *Mol Reprod Dev* 2015; 82(2):103–14.

26. Gilchrist GC, Tscherner A, Nalpathamkalam T, Merico D, LaMarre J. MicroRNA Expression during Bovine Oocyte Maturation and Fertilization. *Int J Mol Sci* 2016; 17(3).

27. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009; 10(1):57–63.

28. Han Y, Gao S, Muegge K, Zhang W, Zhou B. Advanced Applications of RNA Sequencing and Challenges. *Bioinform Biol Insights* 2015; 9(Suppl 1):29–46.

29. Zhao S, Fung-Leung W-P, Bittner A, Ngo K, Liu X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One* 2014; 9(1):e78644.

30. Tantia MS, Vijh RK, Bhasin V, Sikka P, Vij PK, Kataria RS, et al. Whole-genome sequence assembly of the water buffalo (Bubalus bubalis). *Indian J Anim Sci.* 2011; 81(5).

31. Kandil OM, Ghanem N, Abdoon a SS, Hölker M, Phatsara C, Schellander K, et al. Transcriptional analysis of buffalo (Bubalus bubalis) oocytes during in vitro maturation using bovine cDNA microarray. *Reprod Domest Anim.* 2010; 45(1):63–74.

32. Strazzullo M, Gasparrini B, Neglia G, Balestrieri ML, Francioso R, Rossetti C, et al. Global transcriptome profiles of italian mediterranean buffalo embryos with normal and retarded growth. *PLoS One* 2014; 9(2):1–9.

33. Pariset L, Chillemi G, Bongiorni S, Romano Spica V, Valentini A. Microarrays and high-throughput transcriptomic analysis in species with incomplete availability of genomic sequences. *N Biotechnol* 2009; 25(5):272–9.

34. Dong Y, Xie M, Jiang Y, Xiao N, Du X, Zhang W, et al. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). *Nat Biotech* 2013; 31(2):135–41.

35. Bahcall O. Sheep genome. *Nat Genet* 2014; 46(7):665.

36. Elsik CG, Tellam RL, Worley KC, Gibbs R a., Muzny DM, Weinstock GM, et al. The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. *Science* 2009; 324(5926):522–8.

37. Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, et al. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 2012; 491(7424):393–8.

38. Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. Comparison of next-generation sequencing systems. *J Biomed Biotechnol* 2012; 2012:251364.

39. Steijger T, Abril JF, Engström PG, Kokocinski F, The RGASP Consortium, Hubbard TJ, et al. Assessment of transcript reconstruction methods for RNA-seq. *Nat Methods* 2013; 10(12):1177–84.

40. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. *Genome Biol* 2016; 26;17:13.

41. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Meth* 2008; 5(7):621–8.

42. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol* 2010; 11(10):R106.

43. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; 15(12):550.

44. Dyck MK, Zhou C, Tsoi S, Grant J, Dixon WT, Foxcroft GR. Reproductive technologies and the porcine embryonic transcriptome. *Anim Reprod Sci.* 2014; 149(1):11–8.

45. Huang W, Khatib H. Comparison of transcriptomic landscapes of bovine embryos using RNA-Seq. *BMC Genomics* 2010; 11(1):711.

46. Salvemini M, Arunkumar KP, Nagaraju J, Sanges R, Petrella V, Tomar A, et al. De Novo Assembly and Transcriptome Analysis of the Mediterranean Fruit Fly Ceratitis capitata Early Embryos. *PLoS One* 2014; 9(12):e114191.

47. Martin JA, Wang Z. Next-generation transcriptome assembly. *Nat Rev Genet* 2011; 12(10):671–82.

48. Glenn TC. Field guide to next-generation DNA sequencers. *Mol Ecol Resour.* 2011; 11(5):759–69.

49. Labrecque R, Sirard M-A. The study of mammalian oocyte competence by transcriptome analysis: progress and challenges. *Mol Hum Reprod* 2014; 20(2):103–16.

50. Reyes JM, Chitwood JL, Ross PJ. RNA-Seq profiling of single bovine oocyte transcript abundance and its modulation by cytoplasmic polyadenylation. *Mol Reprod Dev* 2015; 82(2):103–14.

51. Ambros V. microRNAs: tiny regulators with great potential. *Cell* 2001; 28(107):823–6.

52. Gilchrist RB, Lane M, Thompson JG. Oocyte-secreted factors: regulators of cumulus cell function and oocyte quality. *Hum Reprod Update* 2008; 14(2):159–77.

53. Wigglesworth K, Lee K-B, O'Brien MJ, Peng J, Matzuk MM, Eppig JJ. Bidirectional communication between oocytes and ovarian follicular somatic cells is required for meiotic arrest of mammalian oocytes. *Proc Natl Acad Sci U S A* 2013; 110(39):E3723-9.

54. Macaulay AD, Gilbert I, Caballero J, Barreto R, Fournier E, Tossou P, et al. The gametic synapse: RNA transfer to the bovine oocyte. *Biol Reprod* 2014; 91(4):90.

55. Macaulay AD, Gilbert I, Scantland S, Fournier E, Ashkar F, Bastien A, et al. Cumulus Cell Transcripts Transit to the Bovine Oocyte in Preparation for Maturation. *Biol Reprod* 2016; 94(1):16.

56. Bunel A, Jorssen EP, Merckx E, Leroy JL, Bols PE, Sirard MA. Individual bovine in vitro embryo production and cumulus cell transcriptomic analysis to distinguish

cumulus-oocyte complexes with high or low developmental potential. *Theriogenology* 2015; 83(2):228–37.

57. Luo F, Jia R, Ying S, Wang Z, Wang F. Analysis of genes that influence sheep follicular development by different nutrition levels during the luteal phase using expression profiling. *Anim Genet* 2016; 47(3):354–64.

58. Wu Y, Lin J, Li X, Han B, Wang L, Liu M, et al. Transcriptome profile of one-month-old lambs' granulosa cells after superstimulation. *Asian-Australasian J Anim Sci*. 2016;

59. Feuerstein P, Puard V, Chevalier C, Teusan R, Cadoret V, Guerif F, et al. Genomic assessment of human cumulus cell marker genes as predictors of oocyte developmental competence: impact of various experimental factors. *PLoS One* 2012; 7(7):e40449.

60. Abdoon AS, Gabler C, Holder C, Kandil OM, Einspanier R. Seasonal variations in developmental competence and relative abundance of gene transcripts in buffalo (Bubalus bubalis) oocytes. *Theriogenology* 2014; 82(8):1055–67.

61. Keefe D, Kumar M, Kalmbach K. Oocyte competency is the key to embryo potential. *Fertil Steril* 2015; 103(2):317–22.

62. Tadros W, Lipshitz HD. The maternal-to-zygotic transition: a play in two acts. *Development* 2009; 136(18):3033–42.

63. Telford NA, Watson AJ, Schultz GA. Transition from maternal to embryonic control in early mammalian development: a comparison of several species. *Mol Reprod Dev* 1990; 26(1):90–100.

64. Kumar P, Verma a., Roy B, Rajput S, Ojha S, Anand S, et al. Effect of Varying Glucose Concentrations during In Vitro Maturation and Embryo Culture on Efficiency of In Vitro Embryo Production in Buffalo. *Reprod Domest Anim.* 2012; 47(2):269–73.

65. Memili E, First NL. Developmental changes in RNA polymerase II in bovine oocytes, early embryos, and effect of alpha-amanitin on embryo development. *Mol Reprod Dev* 1998; 51(4):381–9.

66. Graf A, Krebs S, Zakhartchenko V, Schwalb B, Blum H, Wolf E. Fine mapping of genome activation in bovine embryos by RNA sequencing. *Proc Natl Acad Sci U S A* 2014; 111(11):1–6.

67. Jiang Z, Sun J, Dong H, Luo O, Zheng X, Obergfell C, et al. Transcriptional profiles of bovine in vivo pre-implantation development. *BMC Genomics* 2014; 15(1):756.

68. Østrup O, Olbricht G, Østrup E, Hyttel P, Collas P, Cabot R. RNA profiles of porcine embryos during genome activation reveal complex metabolic switch sensitive to in vitro conditions. *PLoS One* 2013; 8(4):e61547.

69. Jiang Z, Dong H, Zheng X, Marjani SL, Donovan DM, Chen J, et al. mRNA Levels of Imprinted Genes in Bovine In Vivo Oocytes, Embryos and Cross Species Comparisons with Humans, Mice and Pigs. *Sci Rep* 2015; 5:17898.

70. Tervit HR, Whittingham DG, Rowson LE. Successful culture in vitro of sheep and cattle ova. *J Reprod Fertil* 1972; 30(3):493–7.

71. Knobil E, Neill JD, Hess AP, Nayak NR, Giudice LC. Knobil And Neill's Physiology Of Reproduction. In: Neill JD, editor. 3ª ed. St. Louis, USA: Elsevier Academic Press; 2006. p. 337.

72. Badr H, Bongioni G, Abdoon ASS, Kandil O, Puglisi R. Gene expression in the in vitro-produced preimplantation bovine embryos. *Zygote* 2007; 15(4):355–67.

73. VanGuilder HD, Vrana KE, Freeman WM. Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques* 2008; 44(5):619–26.

74. Boelhauve M, Sinowatz F, Wolf E, Paula-Lopes FF. Maturation of bovine oocytes in the presence of leptin improves development and reduces apoptosis of in vitro-produced blastocysts. *Biol Reprod.* 2005; 73(4):737–44.

75. Block J. Use of insulin-like growth factor-1 to improve post-transfer survival of bovine embryos produced in vitro. *Theriogenology* 2007; 68 Suppl 1:S49-55.

76. Feugang JM, Greene JM, Willard ST, Ryan PL. In vitro effects of relaxin on gene expression in porcine cumulus-oocyte complexes and developing embryos. *Reprod Biol Endocrinol* 2011; 9(1):15.

77. Biswas D, Jung EM, Jeung EB, Hyun SH. Effects of vascular endothelial growth factor on porcine preimplantation embryos produced by in vitro fertilization and somatic cell nuclear transfer. *Theriogenology* 2011; 75(2):256–67.

78. Elamaran G, Singh KP, Singh MK, Singla SK, Chauhan MS, Manik RS, et al. Oxygen Concentration and Cysteamine Supplementation During In vitro Production of Buffalo (Bubalus bubalis) Embryos Affect mRNA Expression of BCL-2, BCL-XL, MCL-1, BAX and BID. *Reprod Domest Anim.* 2012; 47(6):1027–36.

79. Godfrey TE, Kelly LA. Development of quantitative reverse transcriptase PCR assays for measuring gene expression. *Methods Mol Biol* 2005; 291:423–45.

80. Redel BK, Brown AN, Spate LD, Whitworth KM, Green JA, Prather RS. Glycolysis in preimplantation development is partially controlled by the Warburg Effect. *Mol Reprod Dev* 2012; 79(4):262–71.

81. Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* 2009; 324(5930):1029–33.

82. Krisher RL, Prather RS. A role for the Warburg Effect in preimplantation embryo development: Metabolic modification to support rapid cell proliferation. *Mol Reprod Dev* 2012;

83. Redel BK, Tessanne KJ, Spate LD, Murphy CN, Prather RS. Arginine increases development of in vitro-produced porcine embryos and affects the protein arginine methyltransferase–dimethylarginine dimethylaminohydrolase–nitric oxide axis. 2015;

84. Milazzotto MP, Goissis MD, Chitwood JL, Annes K, Soares CA, Ispada J, et al. Early cleavages influence the molecular and the metabolic pattern of individually cultured bovine blastocysts. *Mol Reprod Dev* 2016; 83(4):324–36.

85. Stringfellow DA, Seidel SM. Manual of the International Embryo Transfer Society. 3rd ed. International Embryo Transfer Society, IETS; 1998.

86. Van Soom A, Mateusen B, Leroy J, De Kruif A. Assessment of mammalian embryo quality: what can we learn from embryo morphology? *Reprod Biomed Online* 2003; 7(6):664–70.

87. Alikani M, Schimmel T, Willadsen SM. Cytoplasmic fragmentation in activated eggs occurs in the cytokinetic phase of the cell cycle, in lieu of normal cytokinesis, and in response to cytoskeletal disorder. *Mol Hum Reprod* 2005; 11(5):335–44.

88. Alikani M, Cohen J, Tomkin G, Garrisi GJ, Mack C, Scott RT. Human embryo fragmentation in vitro and its implications for pregnancy and implantation. *Fertil Steril* 1999; 71(5):836–42.

89. Driver AM, Penagaricano F, Huang W, Ahmad KR, Hackbart KS, Witlbank MC, et al. RNA-Seq analysis uncovers transcriptomic variations between morphologically similar in vivo- and in vitro-derived bovine blastocysts. *BMC Genomics* 2012; 13(1):118.

90. Corcoran D, Fair T, Park S, Rizos D, Patel O V, Smith GW, et al. Suppressed expression of genes involved in transcription and translation in in vitro compared with in vivo cultured bovine embryos. *Reproduction* 2006; 131(4):651–60.

91. Lonergan P, Rizos D, Gutierrez-Adan A, Fair T, Boland MP. Effect of culture environment on embryo quality and gene expression - experience from animal studies. *Reprod Biomed Online* 2003; 7(6):657–63.

92. Rizos D, Clemente M, Bermejo-Alvarez P, de La Fuente J, Lonergan P, Gutiérrez-Adán A. Consequences of in vitro culture conditions on embryo development and quality. *Reprod Domest Anim* 2008; 43 Suppl 4:44–50.

93. Niemann H, Wrenzycki C. Alterations of expression of developmentally important genes in preimplantation bovine embryos by in vitro culture conditions: implications for subsequent development. *Theriogenology* 2000; 53(1):21–34.

94. Krisher RL, Lane M, Bavister BD. Developmental Competence and Metabolism of Bovine Embryos Cultured in Semi-Defined and Defined Culture Media. *Biol Reprod* 1999; 60(6):1345–52.

95. McElroy S, Kim J, Kim S, Jeong Y, Lee E, Park S, et al. Effects of culture conditions and nuclear transfer protocols on blastocyst formation and mRNA expression in pre-implantation porcine embryos. *Theriogenology* 2008; 69(4):416–25.

96. Abe H, Yamashita S, Satoh T, Hoshi H. Accumulation of cytoplasmic lipid droplets in bovine embryos and cryotolerance of embryos developed in different culture

systems using serum-free or serum-containing media. *Mol Reprod Dev* 2002; 61(1):57–66.

97. Sastre D, Da Costa N, De Sá A, Conceição S, Chiaratti M, Adona P, et al. Expression of PLIN2 and PLIN3 during oocyte maturation and early embryo development in cattle. *Theriogenology* 2014; 81(2):326–31.

98. Bradley J, Pope I, Masia F, Sanusi R, Langbein W, Swann K, et al. Quantitative imaging of lipids in live mouse oocytes and early embryos using CARS microscopy. *Development* 2016; 143(12):2238–47.

99. Kikuchi K, Ekwall H, Tienthai P, Kawai Y, Noguchi J, Kaneko H, et al. Morphological features of lipid droplet transition during porcine oocyte fertilisation and early embryonic development to blastocyst in vivo and in vitro. *Zygote* 2002; 10(4):355–66.

100. Bazer FW, Wu G, Spencer TE, Johnson GA, Burghardt RC, Bayless K. Novel pathways for implantation and establishment and maintenance of pregnancy in mammals. *Mol Hum Reprod* 2010; 16(3):135–52.

101. Macklon NS, Brosens JJ. The human endometrium as a sensor of embryo quality. *Biol Reprod* 2014; 91(4):98.

102. Mansouri-Attia N, Sandra O, Aubert J, Degrelle S, Everts RE, Giraud-Delville C, et al. Endometrium as an early sensor of in vitro embryo manipulation technologies. *Proc Natl Acad Sci U S A* 2009; 106(14):5687–92.

103. Huang Y, Ouyang H, Yu H, Lai L, Pang D, Li Z. Efficiency of porcine somatic cell nuclear transfer - a retrospective study of factors related to embryo recipient and embryos transferred. *Biol Open* 2013; 2(11):1223–8.

104. Smith SL, Everts RE, Tian XC, Du F, Sung L-Y, Rodriguez-Zas SL, et al. Global gene expression profiles reveal significant nuclear reprogramming by the blastocyst stage after cloning. *Proc Natl Acad Sci U S A* 2005; 102(49):17582–7.

105. Brosens JJ, Salker MS, Teklenburg G, Nautiyal J, Salter S, Lucas ES, et al. Uterine selection of human embryos at implantation. *Sci Rep* 2014; 4:3894.

106. Ruan YC, Guo JH, Liu X, Zhang R, Tsang LL, Dong J Da, et al. Activation of the epithelial Na+ channel triggers prostaglandin E_2 release and production required for embryo implantation. *Nat Med* 2012; 18(7):1112–7.

107. Kropp J, Salih SM, Khatib H. Expression of microRNAs in bovine and human pre-implantation embryo culture media. *Front Genet* 2014; 5:91.

108. Tan WS, Carlson DF, Walton MW, Fahrenkrug SC, Hackett PB. Precision editing of large animal genomes. *Adv Genet* 2012; 80:37–97.

109. Betts DH, Bain NT, Madan P. The p66(Shc) adaptor protein controls oxidative stress response in early bovine embryos. *PLoS One* 2014; 9(1):e86978.

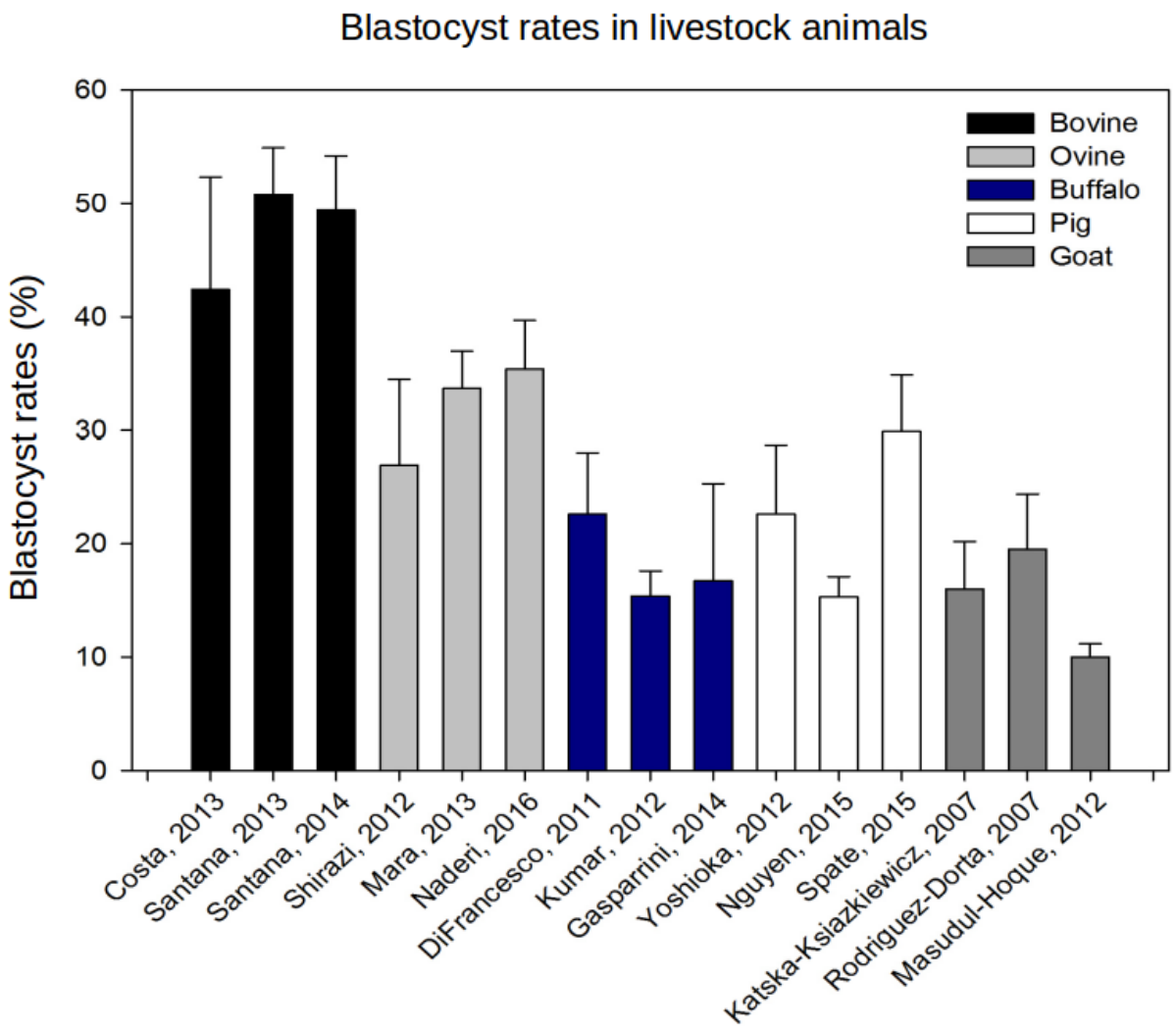


Figure 1. Efficiency of IVEP, measured in terms of blastocyst rates (total number of blastocysts divided by the total number of CCOs) in different livestock animals.

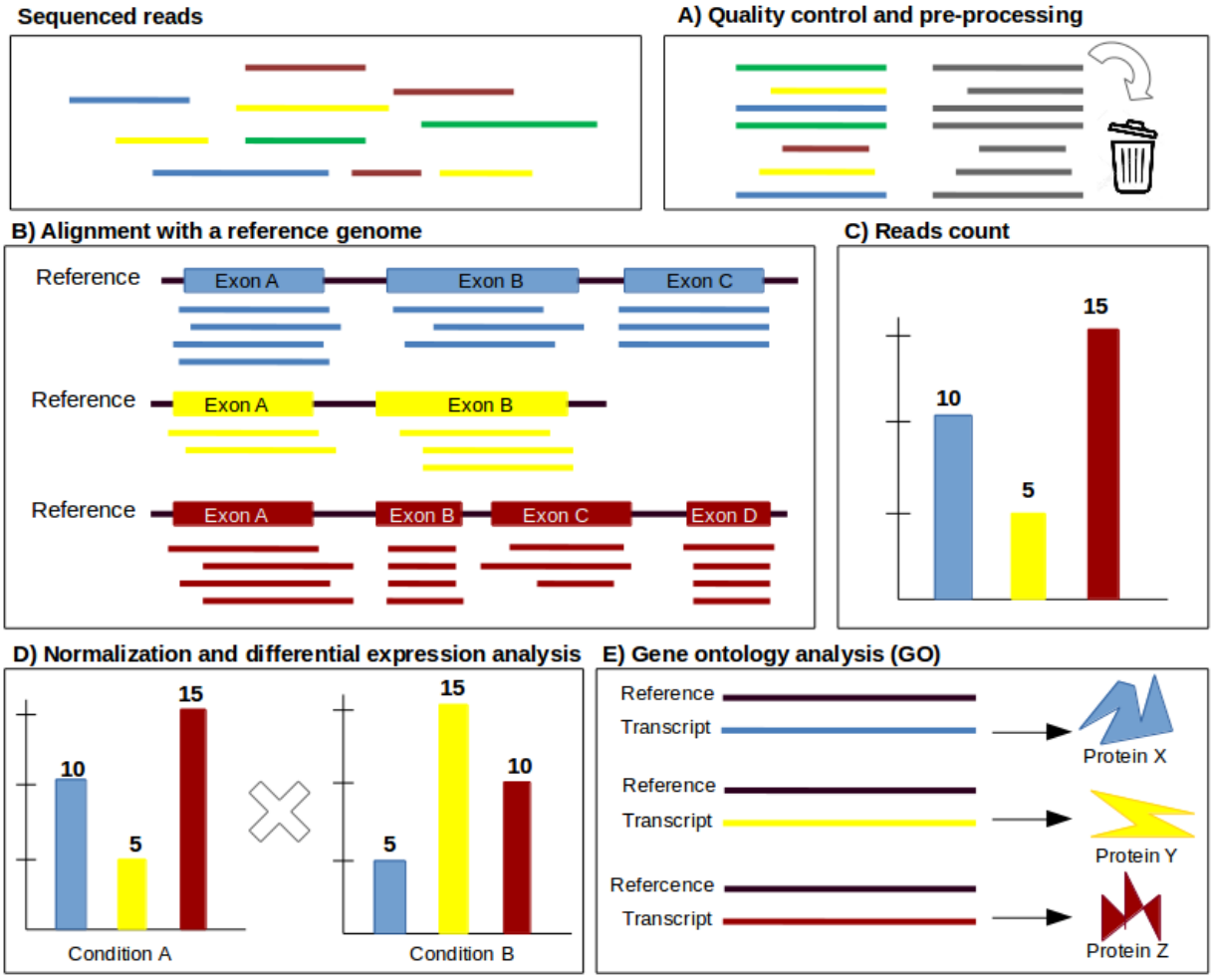


Figure 2. The main steps of RNA-seq processing for differential expression analysis. The first step is to filter the good quality reads (A), then align with the reference genome (B), the step of counting reads aligned with the reference (C), normalization of gene expression data and quantification of differential expression (D) finally, the gene ontology analysis (E).

Table 1. Summary of studies using RNA-seq to adress the transcriptomic profile of oocytes and embryos of different livestock animals.

Reference	Specie	RNAseq plataform	Stage of development	Applications of RNAseq
Bauer et al. (2010)	Porcine	Illumina, GAIIx	embryos collected from the uterus at 2-4-cells and cultured in vitro until blastocyst stage versus embryos enterely cultured *in vivo*	Study the IVF effects on embryo development
Huang and Khatib (2010)	Bovine	Illumina, GAIIx	*in vitro* blastocyst and degenerative embryos (failed to develop from morula to blastocyst stage)	Study the embryo development
Robert et al. (2011)	Bovine	Roche 454, Titanium	germinal vesicle oocyte, 2-, 4-, 8-, 16-cells, morula, blastocyst and hatched blastocyst	Produce an EST dataset for selection of probes in a microarranjo platform
Driver et al. (2012)	Bovine	Illumina, Hiseq 2000	*in vitro* and *in vivo* blastocyst	Study the IVF effects on embryo development
Tsoi et al (2012)	Porcine	Genome Sequencer FLX, 454 Life Sciences	germinal vesicle, metaphase II oocyte, 2-, 4-, 8-, morula, expanded and hatched blastocyst	Produce an EST dataset for selection of probes in a microarranjo platform
Chitwood et al. (2013)	Bovine	Illumina, GAIIx	single *in vitro* blastocyst	Study SNP detection and characterization of allele specific expression
Ostrup et al. (2013)	Porcine	Applied Biosystems, SOLiD	2- and 4-cells embryos produced *in vitro* and *in vivo*	Study the EGA onset and IVF effects on embryo development
Cao et al. (2014)	Porcine	Applied Biosystems, SOLiD 4	1-,2-, 4-, 8-cells, morula and blastocyst produced *in vivo* and produced by TNCS	Study the IVF effects on embryo development and EGA onset
Graf et al. (2014)	Bovine	Illumina, GAIIx	germinal vesicle oocyte, metaphase II oocyte, 4-, 8-, 16-cells and blastocyst produced *in vitro*	Study the EGA onset

Macaulay et al. (2014)	Bovine	Illumina, HiSeq2000	germinal vesicle oocyte and metaphase II oocyte	Study the oocyte competence
Jiang et al. (2014)	Bovine	Applied Biosystems, SOLiD	metaphase II oocyte, 2-, 4-, 8-, 16-cells, early morula, late morula and blastocyst produced *in vivo*	Study the embryo genome activation onset
Jiang et al. (2015)	Bovine, human, mice and pig	RNAseq data downloaded from online database	metaphase II oocyte, 2-, 4-, 8-cells, early morula, compact morula and blastocyst produced *in vivo*	Study the genetic mechanisms of embryo development
Reyes et al. (2015)	Bovine	Illumina, HiSeq 2000	germinal vesicle oocyte and metaphase II oocyte	Study the oocyte competence
Kropp and Khatib (2015)[a]	Bovine	Illumina, GAIIx	media conditioned by blastocyst and embryos that failed to develop from the morula to blastocyst stage	Study non-invasive mRNAs biomarkers for embryo competence
Kropp and Khatib (2015)[b]	Bovine	Illumina, GAIIx	media conditioned by blastocyst and embryos that failed to develop from the morula to blastocyst stage	Study non-invasive miRNAs biomarkers for embryo competence
Luo et al (2016)	Sheep	Illumina, HiSeq 2000	granulosa cells from animals with different levels of nutrition	Study non-invasive biomarkers for oocyte competence
Macaulay et al. (2016)	Bovine	Illumina, HiSeq2000	germinal vesicle oocyte and metaphase II oocyte	Study the oocyte competence
Gilchrist et al. (2016)	Bovine	Illumina, HiSeq 2500	germinal vesicle oocytes, metaphase II oocytes and presumptive zygotes	Study miRNA profile for oocyte competence and embryo development

* EST: *Expressed Sequence Tags.* EGA: *Embryo Genome Activation*.

5. CAPÍTULO 2 - PERFIL TRANSCRIPTÔMICO DE OÓCITOS MATURADOS E BLASTOCISTOS BUBALINOS (*Bubalus bubalis*) PRODUZIDOS *IN VITRO*.

Artigo 2: O manuscrito será submetido para o periódico Reproduction e segue as normas de formatação da revista

Fator de impacto: 3,49

Qualis Capes A2 (área de Biotecnologia)

Title: TRANSCRIPTOMIC PROFILE OF BUFFALO (*Bubalus bubalis*) MATURED OOCYTES AND BLASTOCYSTS PRODUCED *IN VITRO*

Running title: Transcriptomic analyzes of oocytes and buffalo blastocysts

Key-words: buffalo – *Bubalus bubalis* - RNA-seq – In Vitro Embryo Production – IVEP – matured oocytes – blastocyst – embryo – embryo development – oocyte maturation – differential expression – transcriptome – metabolism – implantation.

Authors and affiliations: Priscila P.B. Santana[1], Kenny C. Pinheiro[2], Lino C.S. Pereira[2], Soraya S. Andrade[2], Flávia F. Aburjaile[2], Vanessa C. Brito[3], Nathalia N. Costa[3], Alessandra X. Santos[3], Priscilla C.A. Ramos[3], Marcela S. Cordeiro[3], Simone S.D. Santos[3], Otávio M. Ohashi[3], Moysés S. Miranda[3], Rommel T.J. Ramos[2], Artur L.C. Silva[2].

[1] Federal Rural University of Amazon, Capitão-Poço, Pará, Brazil;

[2] Laboratory of Genomics and Bioinformatics, Federal University of Pará;

[3] Laboratory of In Vitro Fertilization, Institute of Biological Science, Federal University of Pará;

Finantial support: CAPES, Bubrás.

Correspondence: Santana PPB.

ABSTRACT

To know how oocytes and embryos worked is essential to promote Embryo In Vitro Production (IVEP). In this sense, RNA-seq is a powerful tool for accessing large amounts of biological information. In this study, the transcriptomes of mature oocytes and blastocysts produced in vitro from buffalos were sequenced. It was observed the expression of 13,976 genes, and we identified shared genes (62%), exclusive with biological functions specialized in oocytes (1.6%) and embryos (15.7%), besides 4,153 genes differentially expressed in blastocysts in relation to oocytes, of which some genes with greater variation in expression were related to metabolism, implantation and oocyte maturation. This study discussed the importance of promising target genes in future strategies for enhancement of IVEP in the buffalo species and contributed to the elaboration of a transcriptomic panel of oocytes and blastocysts in vitro of buffalos.

INTRODUCTION

Transcriptomic studies in cattle, swine, humans and mice (Graf et al. 2014; Jiang et al. 2015; Xue et al. 2013) contributed to expand the knowledge on the molecular aspects of gametes and embryos. In particular, using the RNA-seq approach to estimate the levels of gene expression, through the number of reads derived from each gene (Mortazavi et al. 2008). The RNA-seq of oocytes and embryos was applied to elucidate the genetic aspects of oocyte maturation (Gilchrist et al. 2016; Reyes, Chitwood, and Ross 2015) and embryo development (Graf et al. 2014; Østrup et al. 2013). To report quality marker genes, and to design noninvasive strategies for evaluating oocyte (Bunel et al. 2015; Feuerstein et al. 2012; Luo et al. 2016) and embryo quality (Kropp and Khatib 2015a, 2015b).

In the buffalo specie, low oocyte quality and embryos produced in vitro are reported as a limitation for the performance of reproductive biotechnologies such as Embryo In Vitro Production, IVEP (Di Francesco et al. 2012; Gasparrini et al. 2014; Yindee et al. 2011). However, there are few studies on oocyte transcriptomic profiles (Kandil et al. 2010) and buffalo embryos (Abdoon et al. 2012; Strazzullo et al. 2014) to support strategies aimed at improving the quality of mature oocytes and embryos produced in vitro in this specie.

This work proposes the study of transcripts of oocytes and buffalo embryos using RNA-seq to better understand their molecular aspects and thus contribute to the elaboration of future strategies to improve IVEP in this species.

MATERIALS AND METHODS

In Vitro Embryo Production (IVEP)

Buffalo oocytes obtained from abattoir-derived ovaries were *in vitro* matured in tissue culture medium 199 (TCM199) supplemented with 50µg/ml gentamycin, 10mg/mL pyruvate, 0.5µg/ml follicule-stimulating hormone (Folltropin, Bioniche Animal Health, Belleville, Ont., Canada), 50µg/ml luteinizing hormone (Lutropin, Bioniche Animal Health), and 10% fetal bovine serum (FBS, Gibco BRL, Grand Island, NY), for 19h at 38.5°C in a 5% CO_2 incubator with humidified air. Were selected for in vitro maturation only *complexos cumulus oophorus* (CCOs) grade 1 and 2 according (Leibfried and First 1979).

For *in vitro* fertilization, frozen sperm from one bull was thawed, processed by Percoll gradient and 2x106 sperm/mL of spermatozoa were incubated in 80µL droplets containing 20 oocytes for 24 hours, in modified TALP-FERT medium (Parrish et al. 1988). After fertilization, presumptive zygotes were cultivated in 100µL droplets of modified Synthetic Oviduct Fluid (SOF) medium supplemented with 0.2 mM pyruvate, 50 mg/mL gentamycin, 5 mg/mL bovine serum albumine (BSA), and 10% FBS for 6 days at 38.5°C and 5% CO_2 incubator with humidified air. Only hatched blastocysts grade 1 (good quality) were chosen based on morphological criteria established by International Society of Embryo Transfer, *IETS* (Stringfellow and Seidel 1998).

In average 72% (±3.34 sd) of oocytes presented the first polar body after 19h of incubation, which is the indicative of metaphase II resumption, or oocyte maturation. After confirm the nuclear maturation, the oocytes were enzimatically treated with pronase to remove the pellucida zone. The rate of embryo development was in average 21.3% (±4.18 sd), calculated from the total number of CCOs selected. Each biological replicate was obtained from three IVEP. Matured oocytes and hatched blastocysts were stored in RNAlater® solution and kept in -80°C until the mRNA extraction.

Library Preparation, Sequencing and Data analysis

Each biological replicate contained pools of 35 buffalo embryos and 205 *in vitro* matured oocytes. The mRNA isolation was performed using Dynabeads© mRNA Direct Micro Kit (Life Technologies). The isolated mRNA from the oocyte and blastocyst were used to prepare the single-end barcoded libraries, with the Ion Total RNA-seq Kit v2 (Life Technologies). All kits were performed according to the manufacturer's instructions.

Each library was amplified, quantified on Qubit® 2.0 then sequenced on an Ion Proton™ platform (Life Technologies). Sequencing reads of low quality (Phred < 20) were trimmed and filtered using FASTX-Toolkit (http://hannonlab.cshl.edu), then visuzalized using the FastQC tool (http://www.bioinformatics.babraham.ac.uk). Just filtered reads were mapped to the *Bos taurus* genome (Bos_taurus.UMD3.1, Ensemble, release 87).

The Bowtie2 and TMAP Alignment (Life Technologies) were used for assembly the reads to the reference genome. Were adopted the standard parameters, and mismatches allowed were 1 (Bowtie2) and 2 (TMAP). Bowtie2 (http://bowtie-bio.sourceforge.net) apply the *Burrows-Wheeler transform, BWT* algorithm (Langmead and Salzberg 2012). TMAP (https://github.com/iontorrent/TMAP) with the function “mapall”, apply the combination of *Burrows-Wheeler Aligner (BWA), Sequence Search and Alignment by Hashing Algorithm (SSAHA)* and *Super-maximal exact matches (SMEM)*. The *CLC Genomics Workbench 4.7.2 software* (QIAGEN Bioinformatics) was used to visualize the mapping and coverage.

Gene expression and GO enrichment analysis

To determine the total number of genes expressed in oocytes and embryos, each replicate was individually submitted to the Cufflinks for an estimation of the transcripts relative abundances (Trapnell et al. 2010). The default parameters and the use of the reference Bos taurus UMD3.1 were adopted. The assemblies of each replicate were pooled into the merged.gtf file using the Cuffmerge tool. Next, merged.gtf was used to run the Cuffdiff tool. Reads counts were normalized by the Reads Per Kilobase Million (RPKM) method and were obtained from the gene_exp_diff file, the genes with RPKM > 0.4 were considered as expressed. This value was adopted based on (Ramsköld et al. 2009).

To determine the differentially expressed genes in blastocysts related to matured oocytes was used the HTSeq Count (http://www-huber.embl.de) for the reads count with the *–union* mode (Anders, Pyl, and Huber 2015). The reads count data were normalized and tested for differential expression using the DESeq2 package in Bioconductor (https://bioconductor.org/packages) which is based on negative binominal distribution (Love, Huber, and Anders 2014). In DESeq2, the false discovery rate was adjusted to 0.05. Genes that showed adjusted p-value according (Benjamini and Hochberg 1995) ≤ 0.05 were considered as differentially expressed. The similarity analysis among the samples was based on the Euclidean distance calculation. The Euclidean distance and hierarchical gene cluster were generated using the transformed values in *regularized logarithm, rlog*.

To find enriched gene ontology categories, firstly the CDS (coding DNA sequences) were obtained using the BioMart tool in Ensembl (http://www.ensembl.org/biomart/). Then, the CDS were submitted to the Gofeat (http://www.computationalbiology.ufpa.br/gofeat/) platform, that seeks gene annotation by searching, based on sequence homology, in the NCBI, Kegg, InterPro e Uniprot databases (unpublished data).

RESULTS

Sequenced and mapped data

The RNA-seq of buffalo oocytes and embryos (*Bubalus bubalis*) in vitro produced were performed. For this, two biological replicates consisted of pools of hatched blastocysts and oocytes in vitro matured were sequenced. From the total number of reads sequenced for embryos (27,902,704) and oocytes (8,014,809) about 90% were mapped using the bovine reference genome (*Bos taurus*) in Bowtie2 and TMAP programs (Tab. 1).

Bowtie2 apply the BWT algorithm and TMAP was adjusted to use the combination of the BWA, SSHA, and SMEM algorithms. Both were executed with default parameters (except for the number of mismatches), in order to obtain the largest number of positive alignments. The reads aligned by both TMAP and Bowtie2 were considered positive. Therefore, to investigate the presence of false positive alignments the reads mapped exclusively by each program were obtained, and then realigned to the reference genome.

The visualization of the realignments evidenced regions frequently located in areas without gene annotation, according to the reference Bos taurus UMD3.1 (Supl. Fig. 1). Increased number of non-gene regions (nongene_areas) were observed in TMAP realignments. Later, the Blast of these areas showed that they represent predicted buffalo genes (*Bubalus bubalis*), and in this way, they were discarded for false positive alignments (Supl. Fig. 2). Based on this result and the higher number of reads aligned by TMAP, these asssemblies were selected for the further analyzes.

Analysis of genes shared by oocytes and embryos

Oocytes and blastocysts in vitro produced of buffalos expressed 13,976 genes, which corresponds to 63% of the bovine genome (22,000 genes) and the estimated buffalo genome (Tantia et al., 2011). In total, 12,576 and 10,049 genes were expressed in buffalo blastocysts and oocytes, respectively. There was a predominance of protein coding genes

in embryos (91.48%) and oocytes (94.41%). Proportionally, the blastocysts presented a higher number of non-coding RNAs (4.1%) compared to oocytes (1.72%), as small nuclear and nucleolar RNAs, that may be related to their higher transcriptional activity (Fig. 1a).

It was observed that 62% (8,649) of the expressed genes (13,976) were shared by the stages. However, the fraction of shared genes corresponded to 69% of the expression in embryos (12,576) and 83% in oocytes (10,049). Thus, proportionally, they represent a more significant fraction of oocyte transcriptome compared to embryos. The shared genes were mostly protein coders (94.17%), but there were also non-coding RNAs (1.75%), pseudogenes (3.76%) and likely new transcripts (0.32%). Thus, only 14% (1,400) and 32% (3,927) of the genes were exclusively expressed in oocytes and buffalo embryos, respectively, so called exclusive genes (Fig 1b).

Biological functions of the shared genes

According to the ontology analysis, most protein coding genes shared by oocytes and buffalo embryos consisted of plasma membrane (42%), cytoplasmic (27%), and cellular components of Golgi apparatus (8%) and endoplasmic reticulum (7%). Other fraction comprised 16% and were components of nuclear membrane, nuclear chromatin, transcription complexes, ribonucleoproteins, microtubules and kinetochores (Fig. 2a).

The biological functions were associated to maintenance of cellular machinery such as translation, transcription, intracellular transport of proteins, signal transduction pathways mediated by GTPase, regulation of apoptosis, cytoskeletal organization, DNA repair and replication, chromatin remodeling, among others (Supl. Tab. 1).

Biological functions of the unique genes

The specialized biological functions of oocytes and embryos were investigated. For that, just the genes found uniquely expressed in each stage were considered. However, it was observed that a fraction of the unique genes were related to biological functions common to the stages (n = 47), such as the regulation of transcription and apoptosis, gametogenesis, cell adhesion, mediation of signal transduction pathways, intracellular transport of proteins, and transcription factors responsible for development (genes HOX). That is, 44% of the unique genes analyzed in embryos and 84% of the exclusive genes in oocytes were compromised with common functions, but were exerted by different gene groups at each stage (Supl. Tab. 2). Then, the unique genes linked to specialized functions were effectively 224 (16%) in mature oocytes and 2200 (56%) in blastocysts.

To describe the specialized functions, we sought to categorize them even if they had a reduced number of genes (functions with less than 3 genes were not shown). It was considered that the regulation of expression can often depend on the nature of the induced response than on the amount of genes expressed in each pathway, such as in the case of transcription factors, which when activated in small numbers, can trigger drastic changes in the expression profiles of cells (Takahashi and Yamanaka 2006).

It was observed that most of the specialized functions analyzed in embryos were related to the RNA and protein processing, differentiation, cellular proliferation, embryonic development, signaling pathways such as TGFβ and BMP, energetic metabolic pathways mainly of fatty acids and lipids. And functions related to embryo implantation, such as the regulation of cytokines (Table 2). In addition, at least 35% of the embryo-unique genes (versus 26% in the oocyte) consisted of proteins contained in exosomes, denoting the increased transit of molecules between the embryoblasts and/or between the embryo and the medium external activity, in other words, activity of cell signaling (Fig. 2b).

The unique genes analyzed in mature oocytes (Table 3) corresponded to specific functions related mainly to the embryonic development (SLC18A2, SOX*, CDKN1C), cellular differentiation (CCDC88A, SFRP1, MEF2C), the regulation of signaling cascades such as JAK-STAT (FLRT*) and MAPK (PELI2) and the regulation of transmembrane transport (CFTR, AKAP6, GABR*, GRIN2A, SORT1). In addition, at least 33% of the oocyte-unique genes (versus 22% in the embryo) consisted of proteins located on the plasma membrane in accordance with the aforementioned functions of signaling and transmembrane transport (Fig. 2b).

Analysis of differentially expressed genes

The transcriptomic profile of the buffalo blastocyst was analyzed in comparison to the buffalo matured oocyte. It was observed that the embryos exhibited 4,153 genes differentially expressed in relation to oocytes, which means that half of the genes shared by the stages were induced or repressed (Fig. 3a). Their transcriptomic profiles showed dissimilarity, as observed in the Euclidean distance map that separated oocytes and embryos into two distinct clusters (Fig. 3b). Thus, despite the proportion of shared genes (62%), oocytes and embryos profiles are sufficiently different to cluster separately.

Buffalo embryos showed more induced (3309) than repressed (844) genes in relation to the oocyte. According to the ontology analyzes, 200 biological functions were associated, such as: regulation of transcription, intracellular transport of proteins, signal

transduction pathways and cytoskeletal organization. In all cited functions there was a greater number of induced genes, suggesting that they are more positively regulated in embryos relative to oocytes. However, other biological functions such as gametogenesis, cell adhesion and regulation of apoptosis exhibited a close number of induced and repressed genes suggesting that their regulation are relevant for both stages (Tab. 4).

Analysis of gene clusters

Gene Hierarchical Cluster was performed to identify differentially expressed genes that most strongly characterize embryos and oocytes as two distinct stages. For that, were identified the genes whose expression varies most intensely between the stages, same gene may be induced in one stage and repressed in the other. It has been shown that the transcriptional profile of oocytes and buffalo embryos may be distinguished by two groups of genes: group 1 was induced in embryos and repressed in oocytes; and group 2 that behaves in opposite way, that is, repressed in embryos and induced in oocytes (Fig. 4).

Table 5 shows some genes of group 1 (highest fold change), induced in embryos, which the biological functions were related to metabolism (APO1), regulation of apoptosis (KRT18, AHNAK e ANXA6), to embryo development (ENSBTAT*) and implantation (S100A14, SLC34A2, PRSS8, ANXA2). The APO1 is involved in the metabolism of lipids and glucocortidóides, that are important energetic sources for the embryo. The KRT18, AHNAK and ANXA6 genes are involved in the down regulation of apoptosis through calcium sequestration and regulation of RNA splicing. The transcripts with phosphatase activity (ENSBTAT*) are possibly related to the formation of the trophoblast (McDougall et al. 2002). The genes related to implantation showed variable molecular functions, such as the activation of metalloproteinases (S100A14) and thus stimulating cellular invasiveness (Chen et al., 2012). Also the cell adhesion activity (ANXA2), serine protease (PRSS8) and transmembrane transport (SLC34A2) that indicate to be important for implantation in mice (Ruan et al. 2012; Shibasaki et al. 2009; B. Wang et al. 2015).

Genes in group 2, induced in oocytes, were related to cell signaling (BMP15), cell cycle (UCHL1, WEE1, NLRPs) and fertilization (ZP2, ZP4). The ZP1 and ZP2 genes encode membrane receptors for sperm binding to the zona pellucida, essential to the fertilization (Yanagimachi 1981). The BMP15 belongs to the family of transforming growth factors, is secreted by oocytes and acts on follicular development and ovulation (Sudiman et al. 2014). Other genes regulate the cell cycle through the MAP kinase (UCHL1) and

cyclins (WEE2, NLRPs), promoting maintenance of oocyte arrest in metaphase II in which the oocyte remains until the fertilization event (Tripathi, Kumar, and Chaube 2010).

In oocytes, was also induced genes to regulate transcription, translation and RNA stability (KPNA7, ENSBTAT*), as well as receptors for translocation through nuclear pores, karyopherin (KPNA7), which have been related to oocyte competence in pigs (Xin Wang et al. 2012).

DISCUSSION

This is the first report, so far, of the expression profiles of mature oocytes and in vitro buffalo blastocysts using the RNA-seq approach. Bubaline oocytes and blastocysts expressed number of genes very close to the number reported for other related species. In cattle, the expression in oocytes matured in vitro varies from 10,494 (Reyes, Chitwood, and Ross 2015) to 13,327 genes (Graf et al. 2014), and in blastocyst varies from 11,501 (Chitwood et al. 2013) to 13,724 genes (Graf et al. 2014).

In this regard the reports in the literature may vary for biological reasons, such as the use of replicates, and/or for technical reasons related to the data analysis pipeline, such as the use of different programs for alignment and reads counting. In addition, there are variations regarding the criteria for determination of the expressed genes, generally adopting RPKM of 0.1 to 0.4 (Jiang et al. 2014; Reyes, Chitwood, and Ross 2015). Therefore, oocytes and embryos are thought to express about half of the genome of the specie, which has been showed in cattle (Jiang et al. 2014) and bufalloes too.

The part of the genome that the cell type express reflect its functions within the tissue or organism. It is known that oocytes after fertilization originate the zygote and then the blastocyst, so oocytes and blastocysts represents the stages in which the fertilization and implantation processes occurs, respectively. In this work, their transcriptomic profiles in the buffalo specie were investigated in order to identify the genes related to these processes in vitro. Although, it is not possible to extrapolate the results from the in vitro to the in vivo context (Lonergan et al. 2006). Then, the transcriptomes of buffalo oocytes and blastocysts produced in vitro reflect their metabolic requirements in the context of IVEP.

Fraction of transcripts devoted to specific functions

It was observed that although oocytes and embryos are different cell types, namely gamete and embryonic stem cells, respectively, they share about 60% of the expressed genes considering as the minimum limit for RPKM> 0.4 detection. Only 14 to 32% of the expressed genes are actually unique to each stage. However, considering the biological functions, it was observed that part of the unique genes in oocytes (84%) and embryos (44%) were related to common functions of maintenance of the cellular machinery. Therefore, only 1.6% of the genes were effectively related to specialized oocyte functions, as well as 15.7% to specialized functions of the buffalo embryos.

Therefore, most part of the machinery for cellular maintenance is common to gametes and embryonic stem cells, furthermore the fraction of the shared transcriptome is much larger than the fraction related to specialized functions in oocytes and embryos. Ramskold et al. (2009) reported a similar result when compared the transcriptome of several tissues in humans and mice, such as brain, liver, heart, skeletal muscle, testes, among others. It was found that those tissues with very well-specialized functions shared 75% of the mRNAs encoding proteins.

However, the differences in expression found between the shared and, in particular, differentially expressed genes were enough to distinguish embryos and oocytes in two distinct groups. About half of the shared genes were repressed or induced in embryos relative to oocytes, showing that there may be differences in expression even among endogenous genes. These differentially expressed genes added to the unique genes thus represented the differences found in the transcriptomes of oocytes and buffalo embryos.

Energy metabolism of buffalo blastocysts produced in vitro

The specialized functions of buffalo blastocysts were related to the metabolism, proliferation, differentiation and embryo implantation, which are essential at this stage of preimplantation embryonic development. In addition, they are correlated with each other, since, in general, genes linked to metabolism are also linked to the cell growth and proliferation, as they aid in the production of macromolecules and metabolic energy for the formation of new cells (Vander Heiden, Cantley, and Thompson 2009). An example is the mTOR complex activator (LAMTOR1) which regulates lipid metabolism, also the activation of cell proliferation pathways (Rebsamen et al. 2015).

Proliferation is an essential feature of embryonic stem cells. Accordingly, it has been reported predominant expression of genes related to proliferation and maintenance of

pluripotency in bovine blastocysts (Jiang et al. 2014), mice and humans (Xue et al. 2013). Similarly, buffalo blastocysts expressed genes of cell proliferation and differentiation, especially progression to the gastrula stage (genes LRP5, NODAL, MEGF8).

Among the embryo unique genes, the lipid metabolism was especially represented. Leptin (LEP), transmembrane receptors for the uptake of low density lipoproteins (LRP5), enzymes for the modification (FA2H) and oxidation (ACOT8) of fatty acids are expressed. In addition, the gene hierarchical cluster showed that the polyprotein (APO1) was one of the genes with the greatest variation of expression between oocytes and blastocysts, and is related to the metabolism of lipids and glucocorticoids. It is known that in tissue cells APO1 induces cholesterol efflux, in cancer cells it is part of the FAS/APO1 complex that induces apoptosis (Park et al. 2017), but in embryonic cells its function is unknown. These results suggest that lipid metabolism may play an important role in the maturation and in vitro culture of buffalos. Thus, it would be important to investigate the relationship of lipid metabolism and the developmental potential of buffalo embryos produced in vitro.

Transcriptomic profile related to embryo implantation

Regarding the embryo implantation, was identified unique genes for the regulation of cytokines such as interferon-gamma and interleukin (WNT5A, DDX58). The role of cytokines in the implantation of mammals is not yet fully known, but it seems that they influence the maternal immune system avoiding recognition of the embryo as a foreign element and thus favor the implantation (Geisert et al. 2012).

However, even when the embryos develop to blastocyst, they may not be able to implant because the implantation involves complex interactions between the embryo and the endometrium (Bazer et al. 2011). In mice, a serine protease secreted by the embryos (trypsin) triggered cell signaling that culminated in the decidualization of endometrial cells, which is necessary for implantation in this specie (Ruan et al. 2012). In cattle, was showed that blastocysts can secrete mRNAs in culture medium such as VSNL1 and PUM2 (Kropp and Khatib 2015b), also their resulting proteins stimulated endometrial cells and promoted implantation in sheep (Ahn et al, 2009). Further studies may elucidate these mechanisms of interaction, these results indicates that embryos uses several molecules to establish communication with the endometrial cells. In this study, buffalo blastocysts expressed VSNL-1 and PUM2, and also a serine protease protein (PRSS8) that may correspond to a specie specific gene to the interaction mechanisms for implantation in buffalos.

The KRT18, ANXA2 and S100A14 genes were also induced in buffalo blastocysts and play roles in embryonic development according studies using the RNA interference to inhibit gene expression and translation. In cattle, the inhibition of keratin 18 (KRT18) decreased the blastocyst rate (Goossens et al. 2010). In mice, annexin A2 inhibition (ANXA2) reduced the adhesion of blastocysts to the endometrial tissue in vitro, as well as the number of sites of implantation in vivo (B. Wang et al. 2015). The proposed mechanism of action of ANXA2 involves its interaction with S100A14, since they form a protein complex on the cell surface (Myrvang et al. 2013) which supposedly facilitates the cell adhesion interactions related to implantation. Taken togheter, these evidences indicate that the mentioned genes play important roles for embryo implantation. Therefore, it is possible that the regulation of their expression levels positively influences the implantation capacity of buffalo embryos produced in vitro.

Metabolism of buffalo oocytes during the in vitro maturation

Many of the specialized functions in oocytes were related to plasma membrane cellular components. In the context of IVEP, in vitro oocyte maturation is induced by the addition of gonadotropic hormones to the maturation media (Wang et al. 2014), which is also composed of energy sources, mainly glucose and pyruvate, mixture of amino acids, and often Bovine Fetal Serum (Tervit, Whittingham, and Rowson 1972). Therefore, the ability of oocytes to respond to hormonal stimulation is the first step in initiating nuclear and cytoplasmic maturation events that make it suitable for fertilization. Buffalo oocytes in vitro matured expressed ligand-dependent receptors, such as estrogen (SFRP1) and gamma-aminobutyric acid (GABR), protein transport channels (SORT1), amino acids (GLRA3), cholesterol (CFTR) and calcium (AKAP6, GRIN2A), indicating that these substances are transported intensely through the membrane.

This results corroborates the observation of Ramskold et al (2009) that suggested that a large part of the differences between differentiated tissues lies in the expression of membrane receptors mediating cellular communication. This is the case of the oocyte maturation process, which requires an intense cellular communication between oocytes and granulosa cells. That is one of the roles of BMP15, a growth factor secreted by oocytes that acts on granulosa cells promoting metabolic changes and cell proliferation. In bovine IVEP, the addition of BMP15 to the in vitro maturation medium increased the blastocyst rate, possibly through the modulation of glycolysis, tricarboxylic acid cycle, and

glutathione pathway (Sudiman et al. 2014). Since BMP15 has been strongly induced in buffalo matured oocytes, it is essential to study its role in oocyte maturation of buffalos.

The accumulation of transcripts in the oocyte cytoplasm during the maturation process is a molecular aspect associated with oocyte quality, since the transcripts are used to promote early embryonic development mechanisms (Labrecque and Sirard 2014; Tadros and Lipshitz 2009). In buffalo in vitro matured oocytes, were identified several strongly induced transcripts of transcriptional regulation, translation and stability of RNA, that we hypothesized to be related to the regulation of mRNA deposits in buffalo oocytes. Thus, it becomes important to investigate their roles during the oocyte maturation, and as a promissing biomarker genes of oocyte quality in buffalo.

The karyopherin KPNA7, was other strongly induced gene in buffalo oocytes. This result corroborates the study of Wang et al. (2012) which reported the high expression of a variety of karyopherins in in vitro matured oocytes of pigs, particularly KPNA7. In addition, levels of karyotyferin expression decreased gradually to the blastocyst stage, which is in accordance with the result in blastocysts of buffalos, in which KPNA7 was strongly repressed. Furthermore, inhibition of KPNA7 translation by RNA interfering, decreased the blastocyst rate, indicating its role in embryonic development (Xin Wang et al. 2012).

As a result, we speculate that KPNA7 transcripts join the set of those accumulated in oocyte cytoplasm and inherited by the embryo, which play an important role in early embryonic development in buffalos. Thus, we suggest that KPNA7 may be considered as a promising target for future studies of IVEP enhancement in buffalo specie.

CONCLUSION

This study contributed to the elaboration of a general panel of transcriptomic profile of oocytes and blastocysts produced in vitro of buffalos (Fig. 6). They showed a greater proportion of genes related to the maintenance of cellular machinery, and to a lesser extent to specialized functions during oocyte maturation and embryonic development.

The most variable genes in blastocyst were related to lipid metabolism and embryo implantation, and in oocytes to the fertilization and oocyte quality. The results indicate that these genes are important for the molecular aspect of oocyte and embryonic quality in buffalos. Thus, we suggest that they can be considered as promising target genes for future strategies to improve IVEP in buffalos.

REFERENCES

Abdoon, A S et al. 2012. “cDNA Microarray Analysis of Gene Expression in Parthenotes and in Vitro Produced Buffalo Embryos.” *Theriogenology* 77(6): 1240–51. http://www.sciencedirect.com/science/article/pii/S0093691X1100584X.

Anders, Simon, Paul Theodor Pyl, and Wolfgang Huber. 2015. “HTSeq—a Python Framework to Work with High-Throughput Sequencing Data.” *Bioinformatics* 31(2): 166–69. http://dx.doi.org/10.1093/bioinformatics/btu638.

Ahn HW, Farmer JL, Bazer FW, Spencer TE. Progesterone and interferon tau-regulated genes in the ovine uterine endometrium: identification of periostin as a potential mediator of conceptus elongation. Reproduction. 2009 Nov;138(5):813-25. doi: 10.1530/REP-09-0208. Epub 2009 Jul 27.

Bazer, F W, T E Spencer, G A Johnson, and R C Burghardt. 2011. “Uterine Receptivity to Implantation of Blastocysts in Mammals.” *Front Biosci (Schol Ed)* 3: 745–67. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=21196409.

Benjamini, Yoav, and Yosef Hochberg. 1995. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1): 289–300. http://www.jstor.org/stable/2346101.

Bunel, a. et al. 2015. “Individual Bovine in Vitro Embryo Production and Cumulus Cell Transcriptomic Analysis to Distinguish Cumulus-Oocyte Complexes with High or Low Developmental Potential.” *Theriogenology* 83(2): 228–37. http://linkinghub.elsevier.com/retrieve/pii/S0093691X14005093.

Chen H, Yuan Y, Zhang C, Luo A, Ding F, Ma J, Yang S, Tian Y, Tong T, Zhan Q, Liu Z. Involvement of S100A14 protein in cell invasion by affecting expression and function of matrix metalloproteinase (MMP)-2 via p53-dependent transcriptional regulation. J Biol Chem. 2012 May 18;287(21):17109-19. doi: 10.1074/jbc.M111.326975. Epub 2012 Mar 26.

Chitwood, James L et al. 2013. “RNA-Seq Analysis of Single Bovine Blastocysts.” *BMC genomics* 14(1): 350. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3668197&tool=pmcentrez&rendertype=abstract.

Feuerstein, Prisca et al. 2012. “Genomic Assessment of Human Cumulus Cell Marker Genes as Predictors of Oocyte Developmental Competence: Impact of Various Experimental Factors.” *PloS one* 7(7): e40449. http://www.ncbi.nlm.nih.gov/pubmed/22848380 (July 26, 2016).

Di Francesco, Serena et al. 2012. “Ovum Pick-up and in Vitro Embryo Production (OPU-IVEP) in Mediterranean Italian Buffalo Performed in Different Seasons.” *Theriogenology* 77(1): 148–54. http://dx.doi.org/10.1016/j.theriogenology.2011.07.028.

Gasparrini, B. et al. 2014. “Influence of Oocyte Donor on in Vitro Embryo Production in Buffalo.” *Animal Reproduction Science* 144(3–4): 95–101. http://linkinghub.elsevier.com/retrieve/pii/S0378432013003485.

Geisert, Rodney, Asgerally Fazleabas, Mathew Lucy, and Daniel Mathew. 2012. “Interaction of the Conceptus and Endometrium to Establish Pregnancy in Mammals: Role of Interleukin 1β.” *Cell and tissue research* 349(3): 825–38. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5069968/.

Gilchrist, Graham C et al. 2016. “MicroRNA Expression during Bovine Oocyte Maturation and Fertilization.” *International journal of molecular sciences* 17(3). http://www.ncbi.nlm.nih.gov/pubmed/26999121 (July 20, 2016).

Goossens, Karen et al. 2010. “Suppression of Keratin 18 Gene Expression in Bovine Blastocysts by RNA Interference.” *Reproduction, Fertility and Development* 22(2): 395–404. https://doi.org/10.1071/RD09080.

Graf, Alexander et al. 2014. “Fine Mapping of Genome Activation in Bovine Embryos by RNA Sequencing.” *Proceedings of the National Academy of Sciences of the United States of America* 111(11): 1–6. http://www.ncbi.nlm.nih.gov/pubmed/24591639.

Vander Heiden, M G, L C Cantley, and C B Thompson. 2009. “Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation.” *Science* 324(5930): 1029–33. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19460998.

Jiang, Zongliang et al. 2014. “Transcriptional Profiles of Bovine in Vivo Pre-Implantation Development.” *BMC Genomics* 15(1): 756. http://www.biomedcentral.com/1471-2164/15/756.

———. 2015. “mRNA Levels of Imprinted Genes in Bovine In Vivo Oocytes, Embryos and Cross Species Comparisons with Humans, Mice and Pigs.” *Scientific reports* 5: 17898. http://www.ncbi.nlm.nih.gov/pubmed/26638780 (July 20, 2016).

Kandil, O M et al. 2010. “Transcriptional Analysis of Buffalo (Bubalus Bubalis) Oocytes during in Vitro Maturation Using Bovine cDNA Microarray.” *Reproduction in domestic animals = Zuchthygiene* 45(1): 63–74.

Kropp, Jenna, and Hasan Khatib. 2015a. “Characterization of microRNA in Bovine in Vitro Culture Media Associated with Embryo Quality and Development.” *Journal of Dairy Science* 98(9): 6552–63.

———. 2015b. “mRNA Fragments in in Vitro Culture Media Are Associated with Bovine Preimplantation Embryonic Development.” *Frontiers in Genetics* 6: 273. http://journal.frontiersin.org/Article/10.3389/fgene.2015.00273/abstract (August 8, 2016).

Labrecque, Rémi, and Marc-André Sirard. 2014. “The Study of Mammalian Oocyte Competence by Transcriptome Analysis: Progress and Challenges.” *Molecular human reproduction* 20(2): 103–16. http://www.ncbi.nlm.nih.gov/pubmed/24233546.

Langmead, Ben, and Steven L Salzberg. 2012. “Fast Gapped-Read Alignment with Bowtie 2.” *Nat Meth* 9(4): 357–59. http://dx.doi.org/10.1038/nmeth.1923.

Leibfried, L, and N L First. 1979. “Characterization of Bovine Follicular Oocytes and Their Ability to Mature in Vitro.” *Journal of animal science* 48(1): 76–86.

Lonergan, P, T Fair, D Corcoran, and A C Evans. 2006. “Effect of Culture Environment on Gene Expression and Developmental Characteristics in IVF-Derived Embryos.” *Theriogenology* 65(1): 137–52. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16289260.

Love, Michael I, Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15(12): 550. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4302049/.

Luo, F et al. 2016. “Analysis of Genes That Influence Sheep Follicular Development by Different Nutrition Levels during the Luteal Phase Using Expression Profiling.” *Animal genetics* 47(3): 354–64. http://www.ncbi.nlm.nih.gov/pubmed/26970339 (July 25, 2016).

McDougall, K, C Plumb, W A King, and A Hahnel. 2002. “Inhibitor Profiles of Alkaline Phosphatases in Bovine Preattachment Embryos and Adult Tissues.” *Journal of Histochemistry & Cytochemistry* 50(3): 415–22. http://dx.doi.org/10.1177/002215540205000312.

Mortazavi, Ali et al. 2008. “Mapping and Quantifying Mammalian Transcriptomes by RNA-Seq.” *Nat Meth* 5(7): 621–28. http://dx.doi.org/10.1038/nmeth.1226.

Myrvang, Helene K., Xiaoxia Guo, Chan Li, and Lodewijk V. Dekker. 2013. “Protein Interactions between Surface Annexin A2 and S100A10 Mediate Adhesion of Breast Cancer Cells to Microvascular Endothelial Cells.” *FEBS Letters* 587(19): 3210–15. http://doi.wiley.com/10.1016/j.febslet.2013.08.012 (June 23, 2017).

Østrup, Olga et al. 2013. “RNA Profiles of Porcine Embryos during Genome Activation Reveal Complex Metabolic Switch Sensitive to in Vitro Conditions.” *PloS one* 8(4): e61547. http://www.ncbi.nlm.nih.gov/pubmed/23637850 (July 19, 2016).

Park, In-Ja et al. 2017. “Cryptotanshinone Sensitizes DU145 Prostate Cancer Cells to Fas(APO1/CD95)-Mediated Apoptosis through Bcl-2 and MAPK Regulation.” *Cancer Letters* 298(1): 88–98. http://dx.doi.org/10.1016/j.canlet.2010.06.006.

Parrish, J J, J Susko-Parrish, M a Winer, and N L First. 1988. “Capacitation of Bovine Sperm by Heparin.” *Biology of reproduction* 38(5): 1171–80.

Ramsköld, Daniel, Eric T Wang, Christopher B Burge, and Rickard Sandberg. 2009. “An Abundance of Ubiquitously Expressed Genes Revealed by Tissue Transcriptome Sequence Data.” *PLOS Computational Biology* 5(12): e1000598. https://doi.org/10.1371/journal.pcbi.1000598.

Rebsamen, Manuele et al. 2015. “SLC38A9 Is a Component of the Lysosomal Amino Acid Sensing Machinery That Controls mTORC1.” *Nature* 519(7544): 477–81. http://dx.doi.org/10.1038/nature14107.

Reyes, Juan M, James L Chitwood, and Pablo J Ross. 2015. “RNA-Seq Profiling of Single Bovine Oocyte Transcript Abundance and Its Modulation by Cytoplasmic Polyadenylation.” *Molecular reproduction and development* 82(2): 103–14. http://www.ncbi.nlm.nih.gov/pubmed/25560149 (July 20, 2016).

Ruan, Ye Chun et al. 2012. “Activation of the Epithelial Na+ Channel Triggers Prostaglandin E_2 Release and Production Required for Embryo Implantation.” *Nature*

medicine 18(7): 1112–17. http://www.ncbi.nlm.nih.gov/pubmed/22729284 (July 13, 2016).

Shibasaki, Yuri et al. 2009. 381 Biochemical and Biophysical Research Communications *Targeted Deletion of the Tybe IIb Na+-Dependent Pi-Co-Transporter, NaPi-IIb, Results in Early Embryonic Lethality*. http://www.sciencedirect.com/science/article/pii/S0006291X09003362 (June 17, 2017).

Strazzullo, Maria et al. 2014. “Global Transcriptome Profiles of Italian Mediterranean Buffalo Embryos with Normal and Retarded Growth.” *PLoS ONE* 9(2): 1–9.

Stringfellow, D A, and S M Seidel. 1998. *Manual of the International Embryo Transfer Society*. 3rd ed. International Embryo Transfer Society, IETS.

Sudiman, Jaqueline et al. 2014. “Bone Morphogenetic Protein 15 in the Pro-Mature Complex Form Enhances Bovine Oocyte Developmental Competence.” *PLOS ONE* 9(7): e103563. https://doi.org/10.1371/journal.pone.0103563.

Tadros, Wael, and Howard D Lipshitz. 2009. “The Maternal-to-Zygotic Transition: A Play in Two Acts.” *Development (Cambridge, England)* 136(18): 3033–42. http://www.ncbi.nlm.nih.gov/pubmed/19700615 (July 8, 2016).

Takahashi, Kazutoshi, and Shinya Yamanaka. 2006. “Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors.” *Cell* 126(4): 663–76. http://dx.doi.org/10.1016/j.cell.2006.07.024.

Tantia MS, Vijh RK, Bhasin V, Poonam Sikka, Vij PK, Kataria RS, Mishra BP, Yadav SP, Pandey AK, Sethi RK, Joshi BK, Gupta SC and Pathak KML. Whole-genome sequence assembly of the water buffalo (Bubalus bubalis). Indian Journal of Animal Sciences 81 (5): 38–00, May 2011.

Tervit, H R, D G Whittingham, and L E Rowson. 1972. “Successful Culture in Vitro of Sheep and Cattle Ova.” *J Reprod Fertil* 30(3): 493–97. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=4672493.

Trapnell, Cole et al. 2010. “Transcript Assembly and Abundance Estimation from RNA-Seq Reveals Thousands of New Transcripts and Switching among Isoforms.” *Nature biotechnology* 28(5): 511–15. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3146043/.

Tripathi, Anima, K.V. Prem Kumar, and Shail K. Chaube. 2010. “Meiotic Cell Cycle Arrest in Mammalian Oocytes.” *Journal of Cellular Physiology* 223(3): n/a-n/a. http://doi.wiley.com/10.1002/jcp.22108 (June 22, 2017).

Wang, Bing et al. 2015. “Annexin A2 Acts as an Adhesion Molecule on the Endometrial Epithelium during Implantation in Mice.” *PLOS ONE* 10(10): e0139506. https://doi.org/10.1371/journal.pone.0139506.

Wang, Xin et al. 2012. “*KPNA7*, an Oocyte- and Embryo-Specific Karyopherin α Subtype, Is Required for Porcine Embryo Development.” *Reproduction, Fertility and Development* 24(2): 382–91. https://doi.org/10.1071/RD11119.

Wang, Xuemei et al. 2014. “Impact of Gonadotropins on Oocyte Maturation, Fertilisation and Developmental Competence in Vitro.” *Reproduction, Fertility and Development* 26(5): 752–57. https://doi.org/10.1071/RD13024.

Xue, Zhigang et al. 2013. “Genetic Programs in Human and Mouse Early Embryos Revealed by Single-Cell RNA[thinsp]sequencing.” *Nature* 500(7464): 593–97. http://dx.doi.org/10.1038/nature12364.

Yanagimachi, R. 1981. “Mechanisms of Fertilization in Mammals.” In *Fertilization and Embryonic Development In Vitro SE - 6*, eds. Luigi Mastroianni Jr. and JohnD. Biggers. Springer US, 81–182. http://dx.doi.org/10.1007/978-1-4684-4016-4_6.

Yindee, M. et al. 2011. “Maturation Competence of Swamp Buffalo Oocytes Obtained by Ovum Pick-Up and from Slaughterhouse Ovaries.” *Reproduction in Domestic Animals* 46(5): 824–31.

FIGURES

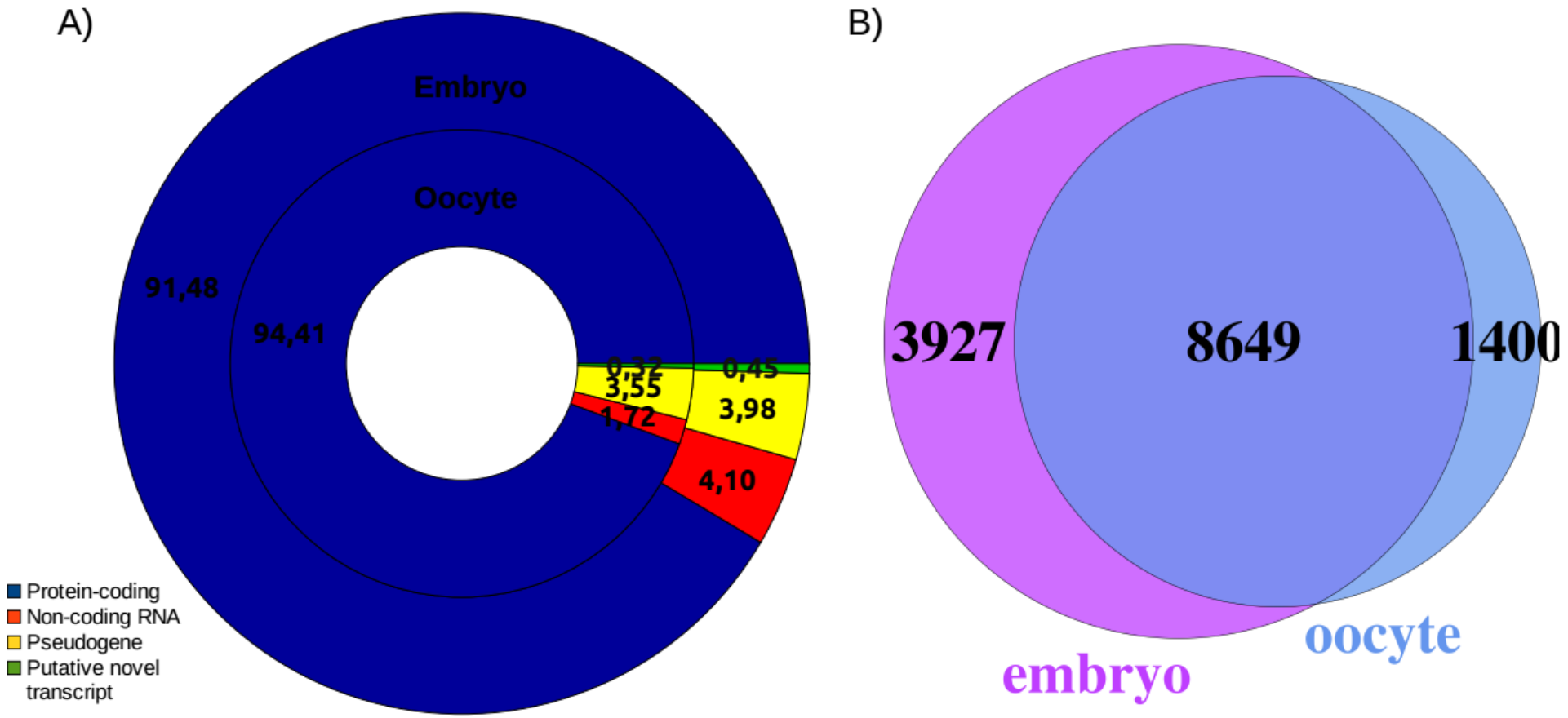


Figure 1: Classification and distribution of genes expressed in matured oocytes and buffalo embryos produced in vitro (RPKM > 0.4). (A) Classification of the genes according to the biotype and (B) Venn diagram representing the shared and unique genes.

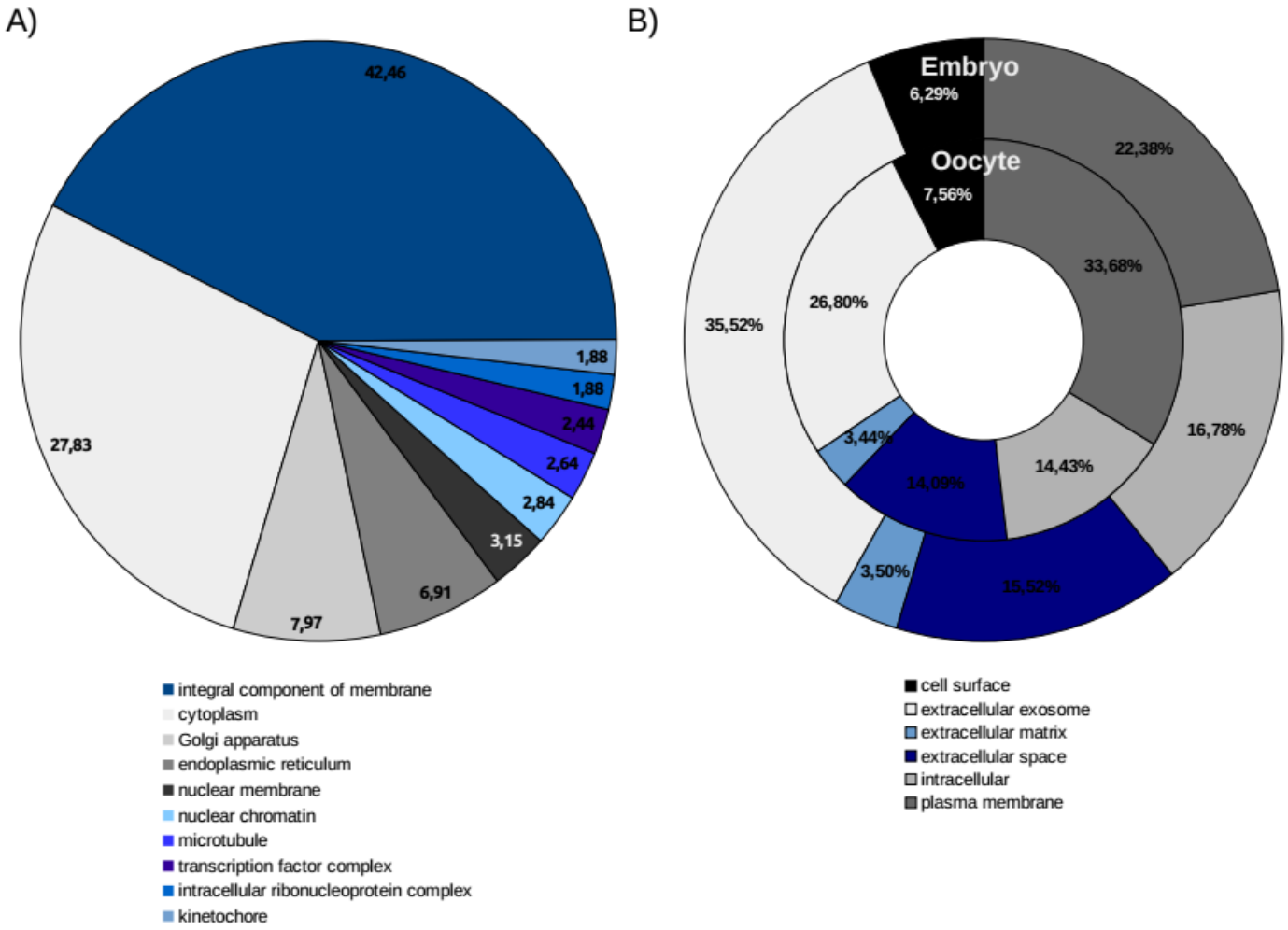


Figure 2: Classification of protein coding genes, according to the ontology of cellular components. (A) The distribution of the ubiquitously expressed genes in embryos and oocytes, gene ontology level 4. (B) Unique genes of embryos and oocytes were shown with the gene ontology level 3.

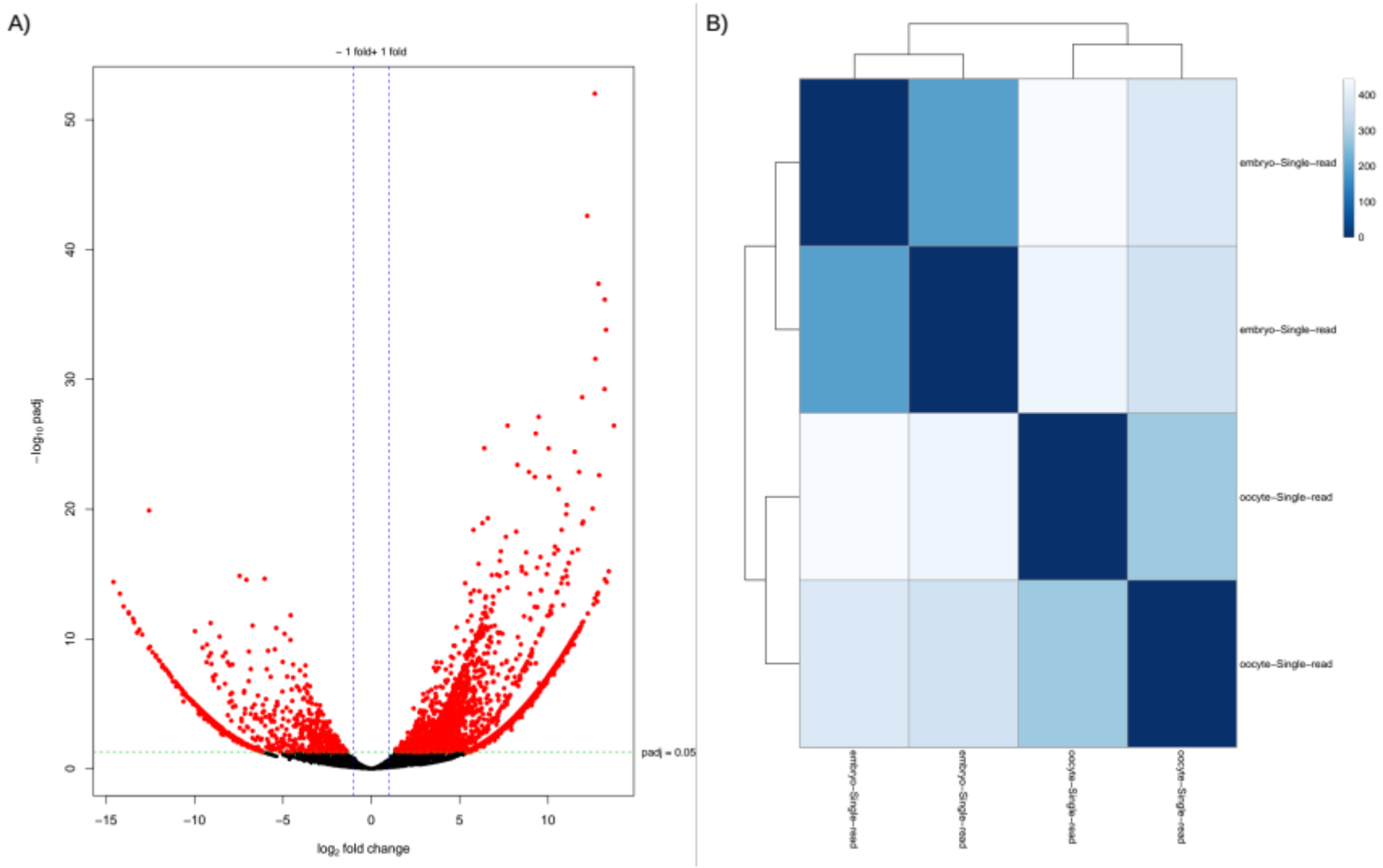


Figure 3: Analysis of differentially expressed genes in buffalo blastocysts in comparison to in vitro matured oocytes. (A) Volcano plot showing differentially expressed genes (red dots) and non-differentially expressed (black dots), besides the induced genes (positive fold change) and repressed genes (negative fold chenage). (B) Heatmap of the Euclidean distance and hierarchical cluster of the biological replicates of oocytes and embryos, showing the separation of two distinct clusters.

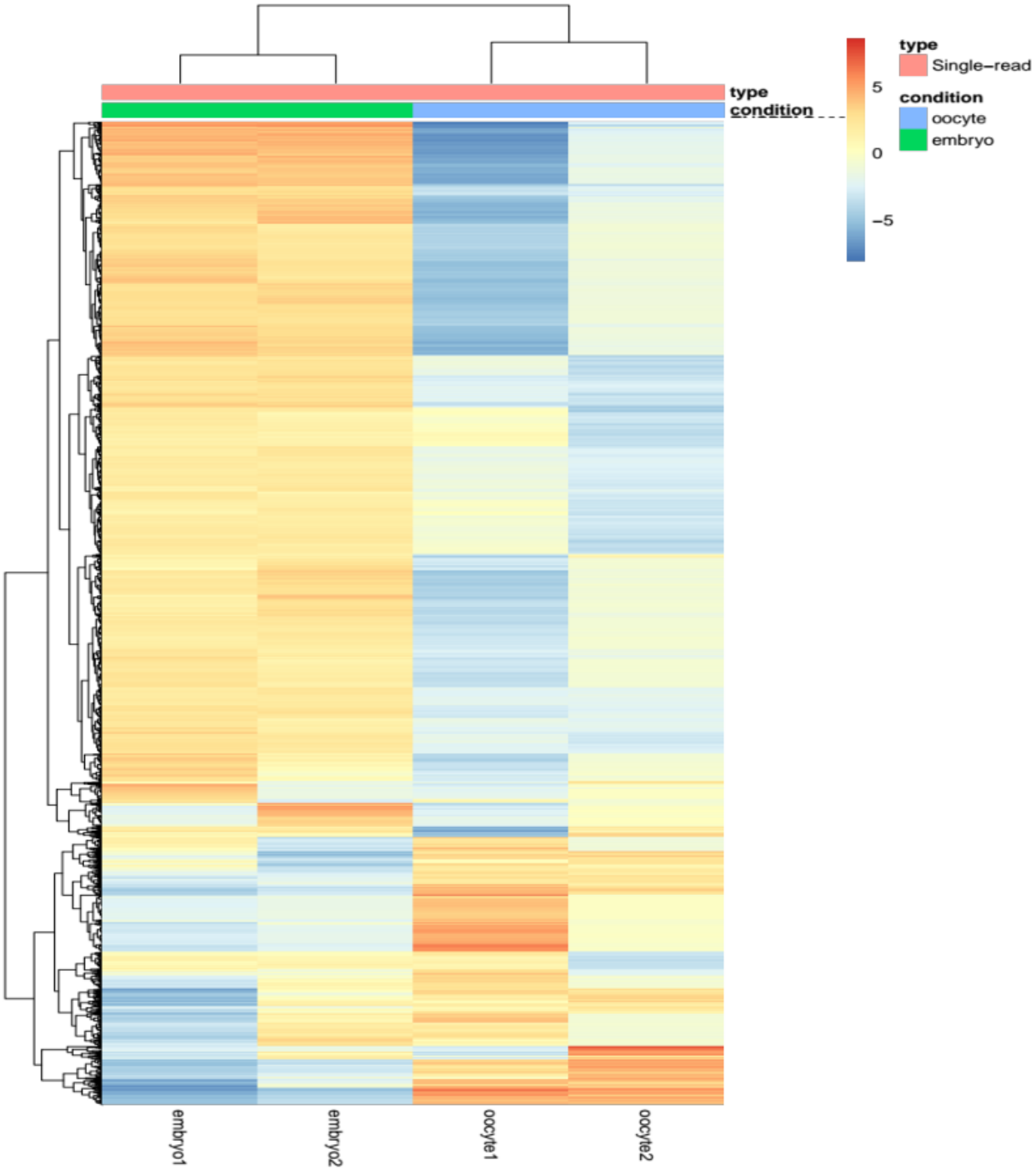


Figure 4: Heatmap and hierarchical cluster of 4,154 genes with the most variable expressed genes showing two clusters in oocytes and buffalo embryos produced in vitro.

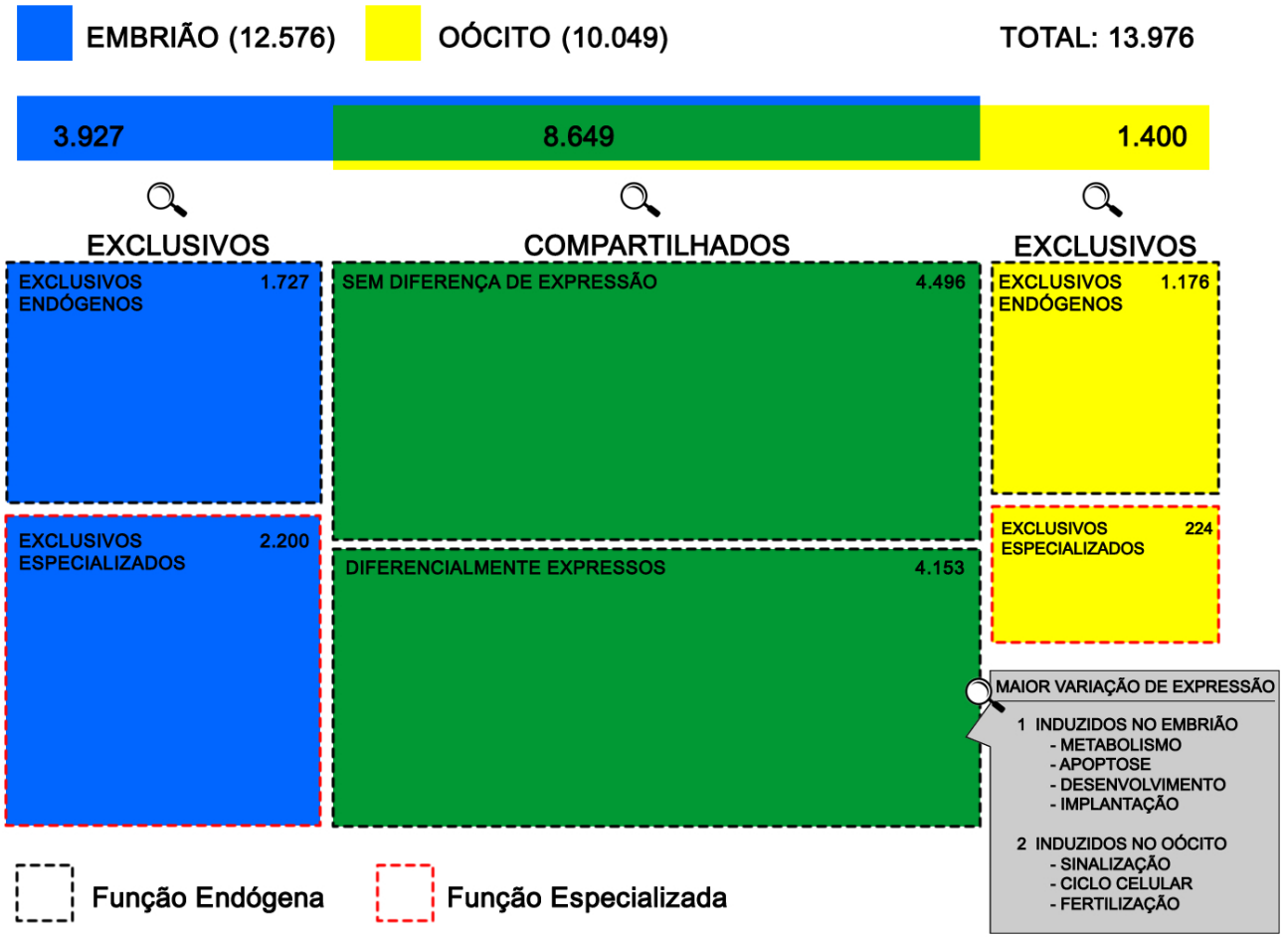


Figure 5: Global panel of the transcriptomes of the matured oocyte and blastocyst produced in vitro of buffalos. The gene ontology of the shared genes (green blocks) and unique genes of embryos (blue blocks) and oocytes (yellow blocks) were analyzed. The gene ontology of shared genes and a fraction of unique (exclusive endogenous) presented endogenous functions (black dash). Another fraction of the unique genes presented specific biological functions, called as exclusive specialized (red dashed). Among the shared genes were identified the differentially expressed and the ones with most variable expression in embryos and oocytes (gray frame).

TABLES

Table 1: Number of reads sequenced and mapped.

	In vitro matured oocytes	In vitro produced embryos
Total number of sequenced reads	8,014,809	27,902,704
Total number of mapped reads (Bowtie2)	6,216,429	21,426,031
Total number of mapped reads (TMAP)	7,252,174	24,321,010

Table 2: Unique and specialized functions of blastocysts produced in vitro of buffalos.

*Gene Ontology – Biological Fuctions	**Gene Names
Cytokine and inflamatory response	
positive regulation of interferon-gamma production	PDE4B, CD226, WNT5A, ENSBTAG00000020535
positive regulation of interleukin-6 production	ARHGEF2, WNT5A, ENSBTAG00000020535, DDX58
cellular response to cytokine stimulus	MME, FOXH1, DPYSL3
negative regulation of cytokine secretion	FFAR4, LRRC32, RGCC
positive regulation of inflammatory response	TRPV4, WNT5A, GPRC5B
positive regulation of interleukin-8 production	MAVS, PRKD2, DDX58
positive regulation of macrophage chemotaxis	TRPV4, RARRES2, MAPK3
Differentiation, Development and Proliferation	
negative regulation of fat cell differentiation	SIRT2, ZADH2, ZFP36L2, JDP2, WNT5A, ZFPM1, TGFB1, GPER1
endodermal cell differentiation	COL4A2, MMP14, ITGA5, COL6A1, NODAL, MMP2, COL7A1
fat cell differentiation	FFAR4, INHBB, PRLH, TTC8, SDF4, ENSBTAG00000020147, STEAP4
hematopoietic progenitor cell differentiation	DHTKD1, ENSBTAG00000005475, AGPAT5, PDGFRA, ZNF784, TGFB1
keratinocyte differentiation	ST14, WNT5A, DSG4, SCEL, DSP
melanocyte differentiation	HPS6, GLI3, EDN3, HPS4
negative regulation of osteoblast differentiation	SEMA4D, LRP5, CHRD, HAND2
chondrocyte differentiation	NOV, GLI2, TGFB1
negative regulation of chondrocyte differentiation	NKX3-2, GLI2, ADAMTS7
embryonic skeletal system development	NKX3-2, WNT5A, SULF1, PCSK5
negative regulation of neuron projection development	APOE, DPYSL3, LPAR1, ENSBTAG00000039129
keratinocyte development	EXPH5, BCL11B, KRT2
proximal/distal pattern formation	HOXC10, GLI1, HOXC11, GLI2, GLI3
negative regulation of endothelial cell proliferation	APOE, RGCC, VASH1, SULF1
positive regulation of smooth muscle cell proliferation	SULF1, NOTCH3, PDGFRB

ventricular cardiac muscle tissue morphogenesis	ENSBTAG00000040053, PROX1, SMAD7, MYBPC3, TGFB1, MYH7
branching morphogenesis of an epithelial tube	MMP14, GLI2, GDF7
cardiac muscle tissue morphogenesis	ACTC1, TBX20, ZFPM1
embryonic digestive tract morphogenesis	PDGFRA, OVOL2, GLI3
epithelial to mesenchymal transition	WNT5A, LOXL3, LOXL3, TGFB1
cell migration involved in gastrulation	LRP5, NODAL, MEGF8
gastrulation with mouth forming second	LRP5, CHRD, ZBTB17
G1/S transition of mitotic cell cycle	CCNE2, CDKN1A, CDK3, CDKN3
G2/M transition of mitotic cell cycle	CHEK2, CDKN1A, BRSK2
mitotic spindle assembly	TUBGCP6, CHEK2, MZT1
Metabolic and biosynthetic Process	
cholesterol metabolic process	LRP5, CLN6, LEP, CH25H, ANGPTL3, LCAT
fatty acid metabolic process	FA2H, ENSBTAG00000010270, ENSBTAG00000009788, NDUFS6, ANGPTL3
phospholipid metabolic process	PLPPR5, ENSBTAG00000008584, PLPP3, PLPP1, PLA2G5
acyl-CoA metabolic process	ENSBTAG00000010270, ENSBTAG00000009788, ACOT8
bile acid metabolic process	LEP, NR5A2, AMACR
oligosaccharide metabolic process	MOGS, ENSBTAG00000017656, ST6GALNAC2
cellular response to amino acid stimulus	LAMTOR1, PDGFRA, COL6A1, RRAGB, MMP2
response to insulin	TRPV4, LEP, PRLH, CRY2
glycogen biosynthetic process	PHKG2, AGL, PHKG1
fatty acid biosynthetic process	ENSBTAG00000004248, FA2H, ENSBTAG00000045728, ENSBTAG00000047957
positive regulation of collagen biosynthetic process	TGFB3, RGCC, F2R, TGFB1
dolichol-linked oligosaccharide biosynthetic process	ENSBTAG00000004586, PQLC3, ALG12
heparan sulfate proteoglycan biosynthetic process	EXTL1, EXTL2, B3GAT3
hydroxylysine biosynthetic process	PLOD3, PLOD1, PLOD2
mRNA and Protein Processing	

translation	ENSBTAG00000024125, PDF, ENSBTAG00000001794, ENSBTAG00000002565, ENSBTAG00000027930, ENSBTAG00000002330, RPL37A, RPL10L, MRPL35, METTL17, ENSBTAG00000030490, ENSBTAG00000006963, ENSBTAG00000011704, ENSBTAG00000010680, ENSBTAG00000013866, ENSBTAG00000012898, ENSBTAG00000014518, ENSBTAG00000007394, ENSBTAG00000014449, RPL17, MRPS18C, ENSBTAG00000019701, ENSBTAG00000030199, , ENSBTAG00000034503, ENSBTAG00000022534, ENSBTAG00000027015, ENSBTAG00000030185, ENSBTAG00000019007, ENSBTAG00000046949, ENSBTAG00000040435, , MRPL36, , MRPS18A, ENSBTAG00000046820, ENSBTAG00000047858, ENSBTAG00000037991, ENSBTAG00000038027, ENSBTAG00000047226
protein processing	PCSK4, CASP7, ENSBTAG00000010057, ENSBTAG00000010828, ENSBTAG00000013055, CPM, PSEN2, FKRP, IFT172, GLI3, ENSBTAG00000018000, RHBDL2
proteolysis involved in cellular protein catabolic process	CTSL, SCPEP1, CTSO, CTSF, CASP8, ADAMTS7, CTSK
positive regulation of protein catabolic process	LPCAT1, TNFSF12, RHBDD3, WNT5A, RILP
establishment of protein localization to plasma membrane	ROCK2, EFR3B, TTC8, CDH2
positive regulation of protein binding	PLXND1, CTHRC1, DACT1, WNT5A
regulation of proteolysis	LTBP4, SPOPL, ABTB1, BTBD1
rRNA processing	WDR36, ENSBTAG00000014449, ENSBTAG00000016510, DIS3L2
positive regulation of proteolysis	AURKAIP1, CLN6, FGFR4
protein N-linked glycosylation via asparagine	ENSBTAG00000017656, ST6GALNAC2, DERL3
protein O-linked mannosylation	TMEM5, LARGE2, FKRP
RNA processing	LARP6, TRUB2, DHX34
RNA secondary structure unwinding	AGO4, DDX51, ENSBTAG00000048317
Signaling Pathway	
BMP signaling pathway	FAM83G, TGFB3, RGMB, NODAL, RGMA, DSG4, MAPK3, MEGF8, ENSBTAG00000046807, GDF7
transforming growth factor beta receptor signaling pathway	LTBP4, FOXH1, TGFB3, SRC, DUSP15, SMAD7, PML, ZYX, ENSBTAG00000039513, TGFB1
extrinsic apoptotic signaling pathway	SGPP1, TNFSF12, SIVA1, PML, TNFRSF1B, TGFB1
phospholipase C-activating G-protein coupled receptor signaling pathway	GRPR, P2RY2, F2R, CYSLTR2, ADRA1B

protein kinase B signaling	ILK, SIRT2, RPS6KB2, GAS6, TGFB1
extrinsic apoptotic signaling pathway in absence of ligand	BAK1, BAD, ERBB3, BCL2L11
lipopolysaccharide-mediated signaling pathway	LYN, NOS3, MAPK3, TGFB1
positive regulation of apoptotic signaling pathway	ENSBTAG00000002534, PRKCD, MAGED1, INHBB
positive regulation of Notch signaling pathway	NOV, PDCD10, SLC35C2, ZMIZ1
Wnt signaling pathway	WNT6, WNT11, WNT5A, ENSBTAG00000038204
negative regulation of fibroblast growth factor receptor signaling pathway	PRDM14, WNT5A, SULF1
negative regulation of Wnt signaling pathway	BARX1, GSC, NKD2
positive regulation of Wnt signaling pathway	DACT1, ATP6V1C2, SULF1
regulation of MAPK cascade	GDF11, TGFB3, NODAL, ENSBTAG00000046807, GDF7
regulation of GTPase activity	PLXND1, SYDE2, ENSBTAG00000008858, PLXNB1, EPHA5, ENSBTAG00000020535
cellular response to growth factor stimulus	INSR, NOS1, CPNE3, GAS6
positive regulation of phosphatidylinositol 3-kinase activity	LYN, TGFB1, PDGFRB

* Gene ontologies correspond to GO level 6. Terms with less than 3 hits were not shown.

** Gene without gene_symbol available were identified by ensembl_id.

Table 3: Unique and specialized functions of in vitro matured oocytes of buffalos.

*Gene Ontology – Biological Fuctions	**Gene Names
Signaling Pathway	
gamma-aminobutyric acid signaling pathway	GABRA3, GABRB3, GABRG1, GABRA4
fibroblast growth factor receptor signaling pathway	FGF16, FLRT2, FLRT1
neuropeptide signaling pathway	SORT1, GPR19, GLRA3
Toll signaling pathway	PELI2, PELI3, PALM3
cellular response to cAMP	RAPGEF3, CFTR, AKAP6
response to amphetamine	GRIN2A, SLC18A2, ADORA2A
Development and Differentiation	
neuronal action potential	SCN9A, SCN4A, SCN1A, SCN2A, SCN10A
excitatory postsynaptic potential	SEZ6, MEF2C, DGKI
neuron maturation	IRX5, FEV, CDKN1C
positive regulation of chondrocyte differentiation	ZBTB16, SOX5, SOX6
regulation of neuron projection development	FRMD7, CCDC88A, SFRP1

* Gene ontologies correspond to GO level 6. Terms with less than 3 hits were not shown.

** Gene without gene_symbol available were identified by ensembl_id.

Table 4: Biological functions of differentially expressed genes in matured oocytes and blastocysts produced in vitro from buffalo.

*Gene Ontology – Biological Functions	INDUCED GENES		REPRESSED GENES	
	Number of Genes	**Gene names	Number of genes	**Gene names
intracellular protein transport	48	SELENOS, AP1B1, ANKRD50, CLTA, VTI1B, GGA2, SNX1, STX12, TOM1L2, STAM, DSCR3, AP1G1, STX16, SEC24C, TBC1D16, TOM1L1, TMED10, COPZ1, KPNB1, TBC1D12, CD74, COPG1, TIMM17B, CHML, NAPA, CLTC, SNX6, CLTB, SNX33, VPS18, AP2M1, AP2S1, TBC1D13, TOM1, GRTP1, TBC1D22A, AP1S1, SNX2, AP2A2, BCAP31, TBC1D9B, TBC1D9, ENSBTAG00000034871, ENSBTAG00000048306, AP2A1, AP1G2, PDCD6, EVI5	9	RIC1, SEC23A, APBA1, ENSBTAG00000008495, RABGAP1, TBC1D15, RPH3AL, SYTL2, ARFIP1
transcription, DNA-templated	33	HES1, CREBZF, COL4A2, SUPT4H1, TFDP2, GTF2IRD1, ZSCAN16, ASXL1, GCM1, DIDO1, CCNT2, ATXN7L3, ZNF394, ENSBTAG00000015007, TEAD2, NR2F6, RXRA, HDAC1, MAPK3, ESRRB, MGA, TEAD3, MAF1, POLR2I, TEAD4, TBX3, ZNF496, RUVBL1, ZSCAN10, SMAD7, ENSBTAG00000021111, ZNF131, ENSBTAG00000011789,	16	MAFG, HEY2, PLAG1, NR3C2, STAT1, VDR, RORA, FOXR1, SMAD5, PLAGL1, HDAC7, ASXL3, ENSBTAG00000047164, ZSCAN31, IRF8, CUX2
small GTPase mediated signal transduction	27	RHOG, CDC42, ARL3, ENSBTAG00000006785, ARL8A, DOCK6, RAC1, ARFRP1, ARHGAP18, CHML, RHOV, RAC2, RGL2, SH2D3A, ARL1, RHOQ, RHOC, RALGDS, ITSN1, RND3, DOCK1, ARL10, RHOB, LRRK2, RHOD, ARF6, RAPGEF5	6	RASGEF1A, DNMBP, DOCK2, ENSBTAG00000047340, RND1, DAB1
spermatogenesis	15	PGM3, TYRO3, KDM2B, ZNF296, SIRT1, LIMK2, MYCBPAP, PUM1, BCAP31, DNMT3A, ENSBTAG00000046900, ACOX1, BAG6, SGPL1, SBF1	11	SCMH1, NLRP14, TDRD1, TEX15, APOB, TDRP, ENSBTAG00000018955, ROS1, SOX30, CLOCK, MEIOC
actin cytoskeleton organization	18	PALLD, SDAD1, CORO1C, INF2, CAP1, ENSBTAG00000013834, FLII, ENSBTAG00000018312, MAP3K1, PACSIN2, NF2, KLHL17, ENSBTAG00000018000, FLNB, LIMK1, ARHGEF17, ENSBTAG00000047424, BCR	7	ENSBTAG00000004836, DOCK2, NISCH, WASF1, FGD1, WASF3, MTSS1
positive regulation of gene expression	18	TLE1, DNMT3B, ROCK2, FN1, PIK3CB, PPM1F, FUBP1, ERBB3, CALR, SNF8, FGFR4, RPS7, EPHX2, PDCD10, ARRB2, KDM5B, CNN2, ZBTB7B	7	SEC16B, CASR, TLR2, APOB, CDH3, SLC6A4, CUX2
homophilic cell adhesion via plasma membrane adhesion molecules	13	IGSF9, DSG3, DSC2, PIK3CB, NECTIN2, PCDH1, PLXNB2, CDH1, CDH24, DSC3, CDH2, CADM1, DSG2	9	FAT3, PCDHGC5, CELSR2, CDH23, PCDH9, CDH3, ROBO1, CD200, ROBO2
regulation of apoptotic process	13	SLK, CHEK2, LRP5, TNFRSF21, RBM25, ENSBTAG00000013956, HS1BP3, DAPK3, BCL2L14, ENSBTAG00000039129, TRAF4, ENSBTAG00000031396	7	ENSBTAG00000005429, ENSBTAG00000015646, GLS2, BCL10, FBXO10, BMP6, BCL2L2

apoptotic process	12	PRKCD, CUL1, DLC1, MAP3K1, BRAT1, PRUNE2, VDAC1, PIDD1, BCLAF1, ENSBTAG00000037571, BCL2L14, PDCD6	4	BCL10, ENSBTAG00000047294, OPA1, BCL2L2
axon guidance	9	RELN, SEMA4F, FEZ2, CSF1R, DPYSL2, NFASC, LAMB2, ENSBTAG00000039129, KIF5B	7	MAPK8IP3, BOC, LHX2, ROBO1, CHL1, UNC5D, ROBO2
neural tube closure	10	GRHL2, SETD2, SDC4, ST14, PLXNB2, TEAD2, RPS7, TGFB1, SPINT1, MTHFD1L	6	PRICKLE1, ALX1, CECR2, LHX2, APAF1, ABL1
negative regulation of gene expression	9	PICALM, MIF, CDKN1A, NOS2, RACK1, INPPL1, SLC35C2, PDCD10, TGFB1	4	RBL2, FYN, ROS1, TP53INP1
neuron apoptotic process	8	DIABLO, CASP7, BOK, AIFM1, ERBB3, TNFRSF21, ATN1, APP	5	RB1, SCN2A, USP53, APAF1, NLRP1
negative regulation of inflammatory response	8	TYRO3, APOA1, METRNL, OTULIN, MVK, NLRX1, NFKB1, BCR	3	NR1H4, SOCS5, RORA,
positive regulation of protein catabolic process	7	LPCAT1, RHBDD3, NKD1, SNF8, OAZ1, ATG7, VPS28	3	MYLIP, SORL1, GPC3,
Notch signaling pathway	4	HES1, HEYL, TGFB1, EPN1	5	NOTCH2, KCNA5, NR1H4, HHEX, DTX3
anterior/posterior pattern specification	5	LRP5, HIPK2, RING1, PCSK5, HOXB4	3	SCMH1, ALX1, MSX1,
BMP signaling pathway	5	TWSG1, FAM83G, DDX5, MAPK3, MEGF8	3	BMPR1B, TGFBR3, BMP6,
positive regulation of MAPK cascade	5	LAMTOR1, KSR1, WWC1, CDH2, PSAP	3	PELI2, FAM58A, MOS,
positive regulation of Notch signaling pathway	4	HES1, ZMIZ1, SLC35C2, PDCD10	4	TSPAN5, NOTCH1, EYA1, STAT3
cellular response to retinoic acid	4	LYN, TEAD2, PTK7, ABCA1	3	BRINP3, BRINP2, TESC,
defense response to bacterium	4	PRKCD, NOS2, MAVS, RAB14	3	ANKRD17, NR1H4, IRF8,
positive regulation of interleukin-6 production	3	IL6R, RIPK2, DDX58	3	TLR2, NOD1, TICAM1,

* Gene ontologies correspond to GO level 6. Terms with less than 3 hits were not shown.

** Gene without gene_symbol available were identified by ensembl_id.

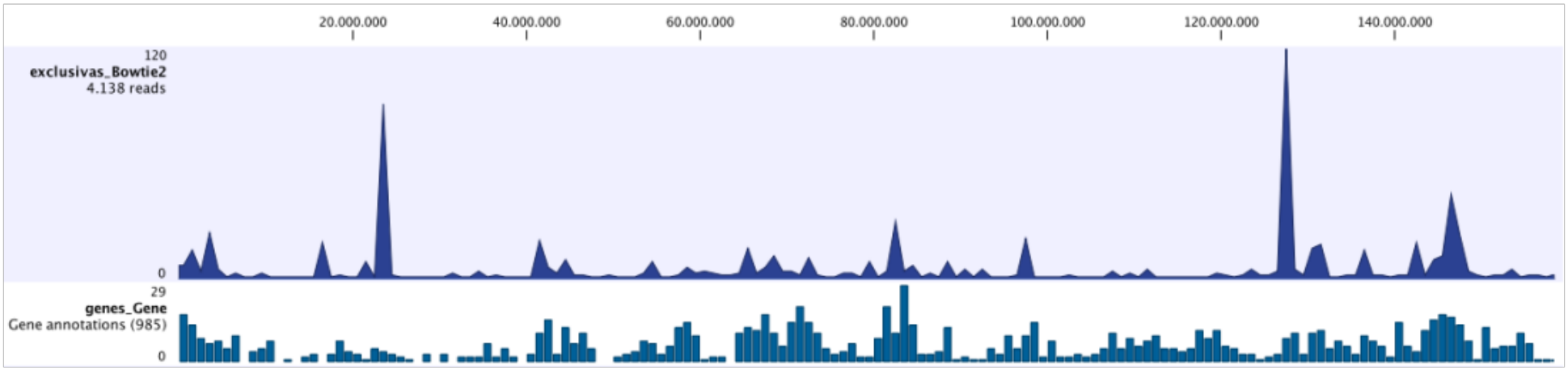
Table 5: Ontology of the most variable genes of groups 1 and 2 visualized in the gene hierarchical cluster. Molecular and biological functions are shown with Gene Ontology levels 4 and 6, respectively.

	GROUP1 – Up-regulated in Blastocysts		
	Molecular function	**Biological function**	**Gene_symbol or Gene_id**
1	calcium ion binding	homophilic cell adhesion via plasma membrane adhesion molecules	ANXA2, ANXA6, DSC2, S100A14
2	RNA binding	negative regulation of apoptotic process, regulation of RNA splicing	KRT18, AHNAK
3	phosphatase activity	regulation of phosphatase activity	ENSBTAT00000022731.4, ENSBTAT00000022269.3
4	cholesterol transporter activity	glucocorticoid metabolic process, integrin-mediated signaling pathway	APOA1
5	high-density lipoprotein particle binding	lipoprotein biosynthetic process, high-density lipoprotein particle assembly	APOA1
6	ligand-gated ion channel activity	apoptotic signaling pathway, negative regulation of sequestering of calcium ion	ANXA6
7	phospholipase inhibitor activity	phospholipase inhibitor activity	ANXA2
8	serine-type endopeptidase activity	positive regulation of sodium ion transport	PRSS8
9	sodium-dependent phosphate transmembrane transporter activity	In utero embryonic development	SLC34A2
	GROUP2 – Up-regulated in Matured oocytes		
	Molecular function	**Biological function**	**Gene_symbol or Gene_id**
1	magnesium ion binding	mitotic cell cycle, negative regulation of cyclin-dependent	WEE2, ATP10D
2	nuclear localization sequence binding	NLS-bearing protein import into nucleus	KPNA7
3	ribonuclease activity	regulation of RNA stability	ENSBTAT00000034504.3
4	telomeric RNA binding	telomere maintenance via telomerase	ENSBTAT00000034504.3
5	thiol-dependent ubiquitin-specific protease activity	ubiquitin-dependent protein catabolic, negative regulation of MAP kinase activity process	UCHL1
6	transforming growth factor beta receptor binding	BMP signaling pathway, granulosa cell development	BMP15
7	translation factor activity, RNA binding	negative regulation of cytoplasmic translation	ENSBTAT00000000819.5
8	transcription factor activity, sequence-specific DNA binding	regulation of transcription, DNA-templated	ENSBTAT00000065334.1
9	ATP binding	Spermatogenesis, negative regulation of cyclin-dependent protein serine/threonine kinase activity	NLRP14, NLRP8, WEE2
10	-	binding of sperm to zona pellucida	ZP2, ZP4

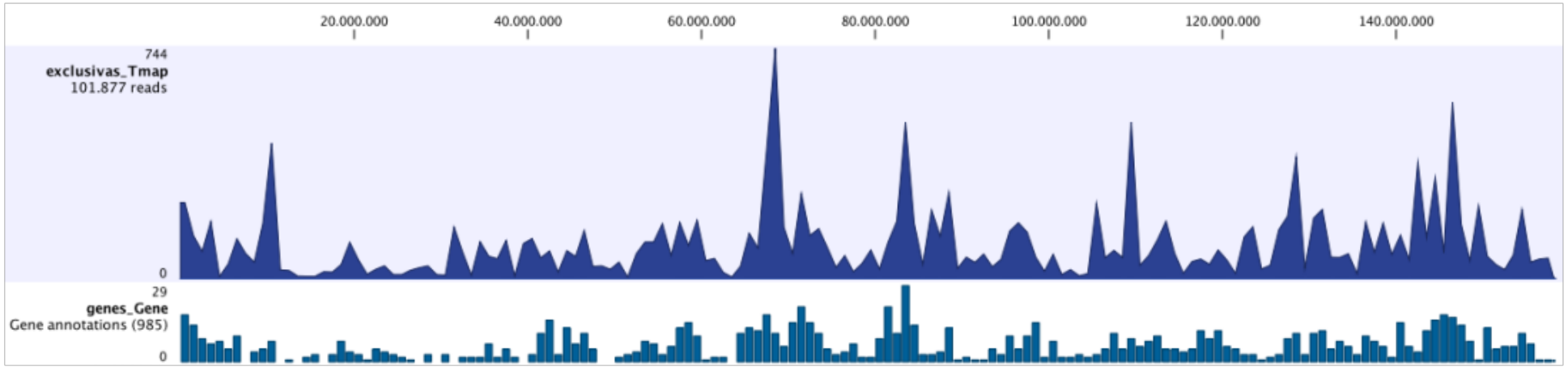
SUPPLEMENTARY MATERIAL

Supplementary Figure 1: Visualization of the realignments in TMAP and Bowtie2. The reads exclusively aligned (A) by TMAP and (B) by Bowtie2 were visualized together with the Bos taurus UMD3.1 reference annotation.

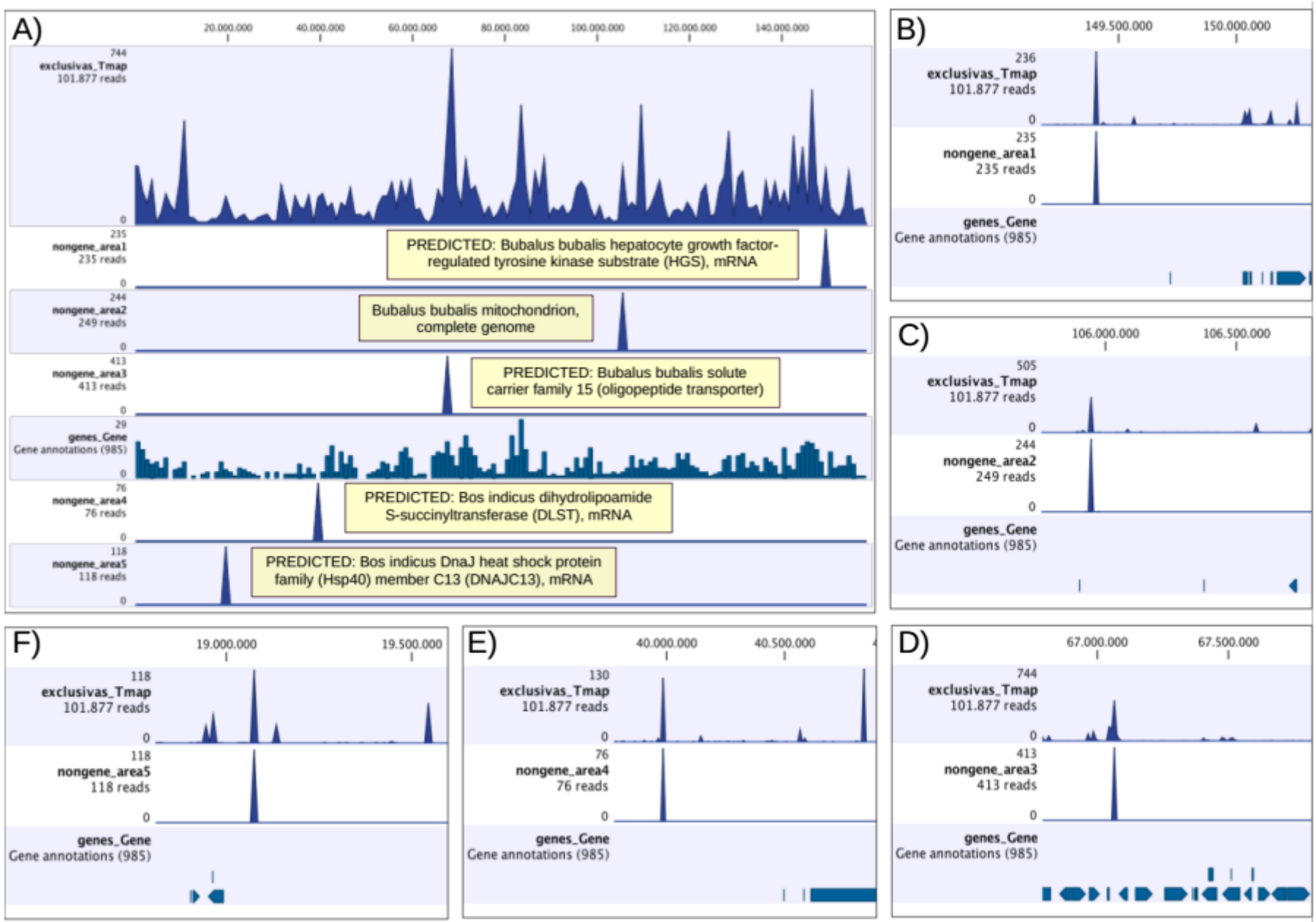
A)



B)



Supplementary Figure 2: Investigation of false positive alignments in TMAP. The reads are mapped exclusively by the TMAP followed by the representation of the Bos taurus UMD3.1 annotation and five groups of reads that mapped in non-genic regions (nongene_areas) with their respective locations in the genome. (A) The Blast result of the five nongene_areas is shown in the yellow boxes. In the lateral frames (B, C, D, E, F) the nongene_areas were enlarged for a better visualization of their location, coinciding with the absence of genes in the annotation.



Supplementary Table 1: Biological functions of the shared genes in matured oocytes and blastocysts produced *in vitro* from buffalos.

Gene Ontology – Biological Functions	Numberof genes	Gene names
translation	126	ENSBTAG00000000168, ENSBTAG00000000560, ENSBTAG00000000417, RPS17, ENSBTAG00000024604, RPL14, MRPL15, RPL19, RPS12, ENSBTAG00000001648, RSL24D1, ENSBTAG00000001538, MRPL19, MRPL21, RPS13, ENSBTAG00000038949, ENSBTAG00000003205, MRPL47, RPS28, RPS3, ENSBTAG00000000425, RPL23, ENSBTAG00000005349, RPL12, ENSBTAG00000027610, RPS9, ENSBTAG00000007454, ENSBTAG00000006383, MRPL49, MRPS5, MRPS18B, ENSBTAG00000008517, UBA52, ENSBTAG00000017441, ENSBTAG00000010970, ENSBTAG00000014662, COA1, RPL26L1, ENSBTAG00000012344, ENSBTAG00000016391, ENSBTAG00000022275, RPL18, ENSBTAG00000034433, RPL18A, ENSBTAG00000012276, RPS5, ENSBTAG00000005495, ENSBTAG00000024839, RPS11, RPL35A, ENSBTAG00000009908, ENSBTAG00000025564, RPS7, ENSBTAG00000027075, ENSBTAG00000013358, MRPL34, ENSBTAG00000011963, RPL32, ENSBTAG00000011963, ENSBTAG00000033080, ENSBTAG00000034973, MRPS23, RPS6, ENSBTAG00000008570, MRPL41, ENSBTAG00000045619, ENSBTAG00000031723, ENSBTAG00000025595, ENSBTAG00000020532, ENSBTAG00000046620, ENSBTAG00000032775, ENSBTAG00000046952, ENSBTAG00000027213, ENSBTAG00000031082, ENSBTAG00000047547, ENSBTAG00000030834, ENSBTAG00000047548, RPS15, ENSBTAG00000022498, ENSBTAG00000033887, RPL36A-HNRNPH2, ENSBTAG00000047029, , ENSBTAG00000020733, , ENSBTAG00000047136, , ENSBTAG00000038379, , ENSBTAG00000027878, , ENSBTAG00000019253, , ENSBTAG00000024378, MRPS21, ENSBTAG00000040167, , ENSBTAG00000047132, ENSBTAG00000023792, , ENSBTAG00000048169, , ENSBTAG00000030164, , ENSBTAG00000035660, , ENSBTAG00000021649, ENSBTAG00000040051, ENSBTAG00000019147, RPS20, ENSBTAG00000015438, ENSBTAG00000031205, ENSBTAG00000047787, ENSBTAG00000022278, ENSBTAG00000046394, ENSBTAG00000023274, ENSBTAG00000048179, , RPS4X, ENSBTAG00000045504, ENSBTAG00000034656, ENSBTAG00000027172, , RPS21, ENSBTAG00000034844, ENSBTAG00000040367, ENSBTAG00000026327, ENSBTAG00000030941, ENSBTAG00000046623, , RPS15A, ENSBTAG00000022902, ENSBTAG00000020807, ENSBTAG00000015473, RPS27A, ENSBTAG00000039456, ENSBTAG00000032022, ENSBTAG00000018426, ENSBTAG00000032432, ENSBTAG00000047647, ENSBTAG00000021093, ENSBTAG00000046531, ENSBTAG00000023343, RPL28, , RPL11
transcription, DNA-templated	114	MAFG, HES1, SSRP1, ASH2L, EPC1, MORF4L2, POLR2C, ENSBTAG00000010255, SETD7, JUN, NR3C2, SMAD1, PLAG1, IRF6, ZNF215, ZNF165, SMAD4, IKBKAP, ENSBTAG00000007746, ABLIM3, ENSBTAG00000008642, ASXL1, GTF2H1, HOXC9, FOXO4, CCNT1, DIDO1, SWT1, NCOA3, POLR3H, MYBBP1A, BCL3, RORA, NR2C1, ENSBTAG00000012816, ZNF449, E2F5, HDAC4, RARB, ZKSCAN2, INO80, ZNF202, MGA, IRF8, FOXR1, ZSCAN29, ENSBTAG00000015007, ENSBTAG00000009406, ZNF394, NR2C2, ENSBTAG00000010988, SMAD3, PPARD, CALCOCO1, POLR2D, THRA, CNOT8, RARA, TCEANC2, ZSCAN12, ENSBTAG00000008810, MAF1, ENSBTAG00000010252, CCNT2, SIX1, POLR2I, BPTF, CUX1, HOXD9, HDAC1, ENSBTAG00000017945, FOXP4, POLR3K, ENSBTAG00000021111, ZSCAN25, RUVBL2, IRF7, ENSBTAG00000009566, EPC2, ENSBTAG00000019359, , , MSL3, , HDAC7, , PHF3, , HSF1, , NCOA2, , MAFK, , KMT2A, , PPP1R13L, ENSBTAG00000011789, , SATB1, , IRF9, , ZNF483, , POLR2B, , ZNF518A, , NFIX, ENSBTAG00000023338, , NR6A1, , RUVBL1, , HDAC8, , ZNF770, , IRF1, , LIN9, , DDIT3, , SKI, , SMAD5, , TCF3, , ZNF664, , TEAD3, ENSBTAG00000039871, , POLR3C, ENSBTAG00000047954, , HIRA, , ZSCAN26, , ZKSCAN1,

intracellular protein transport	79	RIC1, CLTA, VTI1B, VPS16, STX12, AP1B1, SNX1, AP1G1, SEC24D, TBCK, AP4B1, TOM1L2, STAM, SEC24C, GGA2, KPNB1, STX4, VPS11, STX18, STX17, SYTL4, IPO9, CSE1L, COPZ1, CTTN, TBC1D17, VPS39, STX3, XPO6, STX7, WLS, TBC1D30, AP1S1, SYTL3, TBC1D10B, AP2M1, VPS18, SAR1A, RUNDC1, CHML, SNAPIN, AP1M1, CLTB, COPB2, TBC1D2B, AP2S1, XPO1, COPG1, SAR1B, TIMM17B, KIF13A, RAB3GAP2, PDCD6, ENSBTAG00000020316, TGFBRAP1, NAPB, RABGAP1, SNX2, BCAP31, TBC1D15, STX2, USP6NL, AP1S2, ENSBTAG00000034871, AP4E1, SEC23B, TBC1D9B, STX6, BCAP29, STX5, AP3M1, ENSBTAG00000018086, ACD, SNX11, AP2A1, ARFIP1, EVI5L, RAN, SEC24B
spermatogenesis	48	MNS1, BRCA2, PGM3, TYRO3, BCL6, SPATA6, AURKC, WDR48, KDM2B, ASPM, FANCF, PHC2, PLEKHA1, JAG2, PAIP2, TDRD1, UBR2, RAD51C, YBX3, RACGAP1, MLH1, ARNTL, SIRT1, PIWIL2, XRN2, LIMK2, HERC4, GOLGA3, KIAA1524, BAG6, H2AFX, DNMT3A, PUM1, BCAP31, TDRKH, ZGLP1, YY2, CCNYL1, ARID4A, SGPL1, ATRX, SBF1, ENSBTAG00000010711, MAEL, TDRP, TDRD9, INPP5B, NDC1
small GTPase mediated signal transduction	44	RHOG, ENSBTAG00000000682, CDC42, DOCK3, ARL5A, ARL4A, VAV1, BCAR3, SOS2, VAV2, ARL13B, RND2, SOS1, ARHGEF37, DOCK6, RAC1, ARF4, GDI1, DOCK5, ARL1, IFT22, CHML, VAV3, MFHAS1, RHOC, ARL8B, RABIF, ENSBTAG00000008967, DNMBP, ARL6, RGL2, RHOB, DOCK1, ENSBTAG00000035958, RGL1, ARL5B, RHOT1, RASGEF1B, ARHGAP1, ARL16, ARF6, RAPGEF5, RALGPS2, ARL15
actin cytoskeleton organization	34	PALLD, BCL6, CORO6, ARHGEF5, TBCK, MAP3K2, ABLIM3, MAP3K14, DAAM1, MAP3K3, GAS2L3, WASF3, MOB2, FMNL3, NISCH, WASF1, NF1, PFN2, FGD1, FLII, NF2, SSH1, MAP3K4, FLNB, BCR, WASF2, ENSBTAG00000019601, ENSBTAG00000021656, AMOT, ENSBTAG00000047868, SDCBP, DAAM2, IQSEC2, ENSBTAG00000018312
DNA repair	34	RDM1, POLB, SSRP1, APEX1, BACH1, ENSBTAG00000008133, DMAP1, INTS3, TRIM28, FANCL, POLM, FANCI, INO80, CHEK1, PDS5B, PDS5A, RAD9A, FAAP24, DOT1L, ENSBTAG00000040065, UPF1, NSMCE1, INIP, RUVBL2, ZRANB3, RAD1, NABP1, RBM17, RUVBL1, ENSBTAG00000047347, MMS19, MBD4, UBE2A, NSMCE3
regulation of apoptotic process	31	SLK, SRA1, TRIM24, BCL2L1, MITF, PEA15, BCL3, ENSBTAG00000015154, RBM25, ENSBTAG00000015646, ENSBTAG00000013284, CARD10, JAK2, EGR1, RELT, TRAF3, BCL10, ENSBTAG00000013956, DUSP1, MALT1, DEDD, PHLPP1, NGFR, STK26, FRS2, CASP2, PINK1, DAPK3, ENSBTAG00000023192, TRAF4, RASSF5
apoptotic process	29	TAF10, SH3RF1, SAV1, CLPTM1L, BCL2L1, CUL1, RIPK1, PRUNE2, MAP1S, BCL10, VDAC1, SIX1, ING4, MCM2, ENSBTAG00000013535, ENSBTAG00000047294, KREMEN1, OGT, PDCD6, TRAF7, STK26, HIP1R, ENSBTAG00000037571, PPP1R13L, ENSBTAG00000043949, IRF1, DDIT3, OPA1, BCLAF1
positive regulation of gene expression	29	TLE1, UBAP2, PRDM1, UBAP2L, WNT16, KDM6A, ACTA1, MAPK8, MAPK9, FN1, RBM4B, DROSHA, CALR, EP300, RAB3GAP1, MAPK1, RPS7, HMGB2, UBR5, STOX1, ARRB2, YY2, RPS6KA2, BRAF, ACVR1B, RNF207, MED1, SNF8, QKI
chromatin remodeling	25	KAT2B, ARID1B, DMAP1, TOP1, ENSBTAG00000012816, HDAC4, INO80, DEK, DAXX, ENSBTAG00000012426, ACTR8, TP63, PIH1D1, BPTF, SMARCA2, RSF1, TTF1, MSL3, RERE, CHD1, CHD7, SATB1, ATRX, RIOX1, NPM2,
DNA replication	22	SSRP1, ORC6, TICRR, RPA2, SLX4, TOP1, NUP98, PCNA, RECQL5, POLD3, SSBP1, RPA1, RPA3, RFC3, RRM1, ING4, WRNIP1, LIG3, RECQL4, KAT7, NFIX, TBRG1
RNA processing	21	DHX35, HNRNPUL1, CCNL1, EXOSC9, GRSF1, DHX36, CHERP, SSB, TRUB1, SUGP2, DHX33, ADARB1, LARP7, SUGP1, HELB, DHX37, ZFC3H1, DDX54, TDRD9, TRMT2A, MRM3

Supplementary Table 2: Biological functions in common between the unique genes of oocytes and blastocysts produced in vitro from buffalos.

*Ontologia Gênica – Funções Biológicas	EMBRIÕES IN VITRO		OÓCITOS IN VITRO	
	N° de genes	**Nome dos genes	N° de genes	**Nome dos genes
transcription, DNA-templated	41	CREBZF, PPARG, NEUROD1, SUPT4H1, ZSCAN16, POU4F2, POLR2F, COL4A2, NFE2L3, HOXD3, GTF2IRD1, ENSBTAG00000007074, ENSBTAG00000039628, ENSBTAG00000026286, HES6, NR2F6, GCM1, ENSBTAG00000024648, BHLHE40, ESRRB, PAX8, ENSBTAG00000021684, , SMAD7, , ATXN7L3, , LBH, , ZNF213, , ZFPM1, , ZSCAN10, , ZNF219, , E2F3, , TEAD2, , ZNF131, NR0B1, ABLIM2, MAPK3, TEAD4, HES2, TBX3, ZNF496, ZNF500, ENSBTAG00000040021	26	RFX3, HEY2, ZSCAN2, HDAC9, PAX6, HDAC10, NPAS1, NR0B2, PAX3, PPARA, NFIA, TBX19, E2F2, PLAGL1, ITGB3BP, ZNF641, ENSBTAG00000025847, TBX18, CUX2, ENSBTAG00000011476, VDR, FOXO6, EBF1, ONECUT1, ENSBTAG00000047164, ASXL3
positive regulation of gene expression	17	TRPV4, LRRC32, PPM1F, HOXD3, ROCK2, ACTC1, AMH, FGFR4, WHRN, ERBB3, KDM5B, MST1, ZBTB7B, GAS6, PDCD10, SERPINB9, EPHX2	12	STAP1, NKX3-1, CASR, VSTM2A, TLR2, NR0B2, SH3PXD2B, AGR2, FEV, CUX2, APOB, EPB41L4B
regulation of apoptotic process	20	TNFRSF9, LRP5, GDF11, CHEK2, TGFB3, RASSF6, NODAL, CASP8, ENSBTAG00000031396, LCK, TNFRSF1B, LTBR, HS1BP3, TNFRSF6B, BCL2L14, ENSBTAG00000039129, ENSBTAG00000046807, DIP2A, GDF7, ENSBTAG00000048226	6	ENSBTAG00000005429, GLS2, TRAF5, BMP6, ENSBTAG00000033967, BCL2L2
homophilic cell adhesion via plasma membrane adhesion molecules	16	IGSF9, DSG3, IGSF9B, CD226, DSC2, CDH24, CDH26, NECTIN2, DSC3, DSG4, CDH2, PCDHA13, CADM4, PCDHB16, SDK2, FAT2	9	CDHR1, FAT3, CDH4, CDH11, PCDH9, ROBO2, DSCAML1, CDH7, CADM3
intracellular protein transport	19	ANKRD50, TBC1D16, STX16, TOM1L1, SNX31, TBC1D22A, TOM1, STX11, TBC1D13, SNX33, TBC1D12, NAPA, TSNARE1, SYTL1, CD74, RAMP3, TBC1D9, ENSBTAG00000048306, AP1G2,	5	AP3M2, APBA1, SYTL2, SGSM2, ENSBTAG00000008495
small GTPase mediated signal transduction	17	DOCK8, ENSBTAG00000004110, ARL2, ARFRP1, ENSBTAG00000006785, ARL8A, RHOH, RAC1, ARL10, ARL13A, RHOD, RHOV, RHOF, RASGRF1, RHOBTB1, RND3, ARL4C	7	RASGEF1A, DOCK2, ENSBTAG00000033083, DOCK4, DAB1, RND1, ENSBTAG00000047340
axon guidance	11	FEZ1, GLI2, WNT5A, EPHA5, TTC8, DPYSL2, GLI3, LAMB2, ENSBTAG00000039129, NFASC, GDF7,	8	MAPK8IP3, CDH4, LHX2, BOC, UNC5D, ROBO2, CNTN4, FLRT2
spermatogenesis	7	GLI1, MYCBPAP, SUN5, BCL2L11, CRTAP, ENSBTAG00000046520	12	CNBD2, NLRP14, SOHLH1, C14orf39, TEX15, MOV10L1, MORC1, SOX30, MEIOC, APOB, ENSBTAG00000018955, ENSBTAG00000048283
anterior/posterior pattern specification	12	HOXC10, HOXC4, LRP5, HOXD3, BARX1, HOXC11, HOXB2, GLI2, PCGF2, GLI3, PCSK5, RING1	5	TSHZ1, ZBTB16, MLLT3, ALX1, HOXA6

* As ontologias gênicas correspondem ao nível GO 6. Os termos com menos de 3 hits não foram mostrados.

** Genes sem gene_symbol disponível foram identificados pelo ensembl_id.

5.1. ANEXO DO CAPÍTULO 2

METODOLOGIA DAS ANÁLISES DE BIOINFORMÁTICA

O capítulo 2 consiste na análise dos transcriptomas de oócitos maturados e embriões produzidos *in vitro* de búfalo. A seguir é descrita detalhamente a metodologia usada para as análises de bioinformática. Os tópicos foram organizados de acordo com a sequência dos experimentos e contém uma breve introdução, as linhas de comandos para execução e os locais de armazenamento dos dados.

SUMÁRIO

1. DADOS BRUTOS DE SEQUENCIAMENTO

Os dados de sequenciamento foram obtidos através da construção de single-end barcoded libraries que foram sequenciadas na plataforma Ion Próton. Foram realizadas dois sequenciamentos cada um contendo 1 chip preeenchido com bibliotecas barcoded de oócito maturado e blastocisto produzidos *in vitro* (RUN1 e RUN2, respectivamente). As replicatas biológicas de embriões foram identificadas com os barcode 008 (RUN 1)e 005 (RUN 2) e oócitos maturados *in vitro* com os barcode 013 (RUN1) e 010 (RUN 2).

Os arquivos estão armazenadas no servidor Paula nos seguintes diretórios:

RUN 1: /home/priscila/storage/0.data_bruto_RUN1/

RUN 2: /home/priscila/storage/0.data_bruto_RUN2/

2. AVALIAÇÃO DA QUALIDADE E PRÉ-PROCESSAMENTO DOS DADOS

A qualidade dos sequenciamentos foi avaliada com a ferramenta FastQC versão 0.11.3 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). No gráfico do FastQC são mostrados os valores de PHRED de cada base nucleotídica na forma de barras de quartil em que a linha vermelha representa a mediana. O processamento dos dados consistiu em trimar e filtrar as sequências usando o FASTX-Toolkit versão 0.0.12 (http://hannonlab.cshl.edu/fastx_toolkit/).

Primeiramente as sequências foram trimadas para manter somente os nucleotideos com valor PHRED igual ou maior a 20 (-t 20). As sequências trimadas foram em seguida filtradas para que mostrassem no minimo 50% de sua extensao (-p 50) com PHRED igual ou maior a 20 (-q 20).

Trimagem: *fastq_quality_trimmer -Q 33 -t 20 -v -i [input.fastq] -o [output.fastq]*
Filtro: *fastq_quality_filter -Q 33 -q 20 -p 50 -v -i [trimmerOutput.fastq] -o [output.fastq]*

A qualidade dos sequenciamentos melhorou após a trimagem e filtro em comparação com a qualidade inicial (Figura 1 e 2). Assim, os arquivos .fatqc gerados foram usados para as análises posteriores.

ANTES DO PRÉ-PROCESSAMENTO APÓS O PRÉ-PROCESSAMENTO

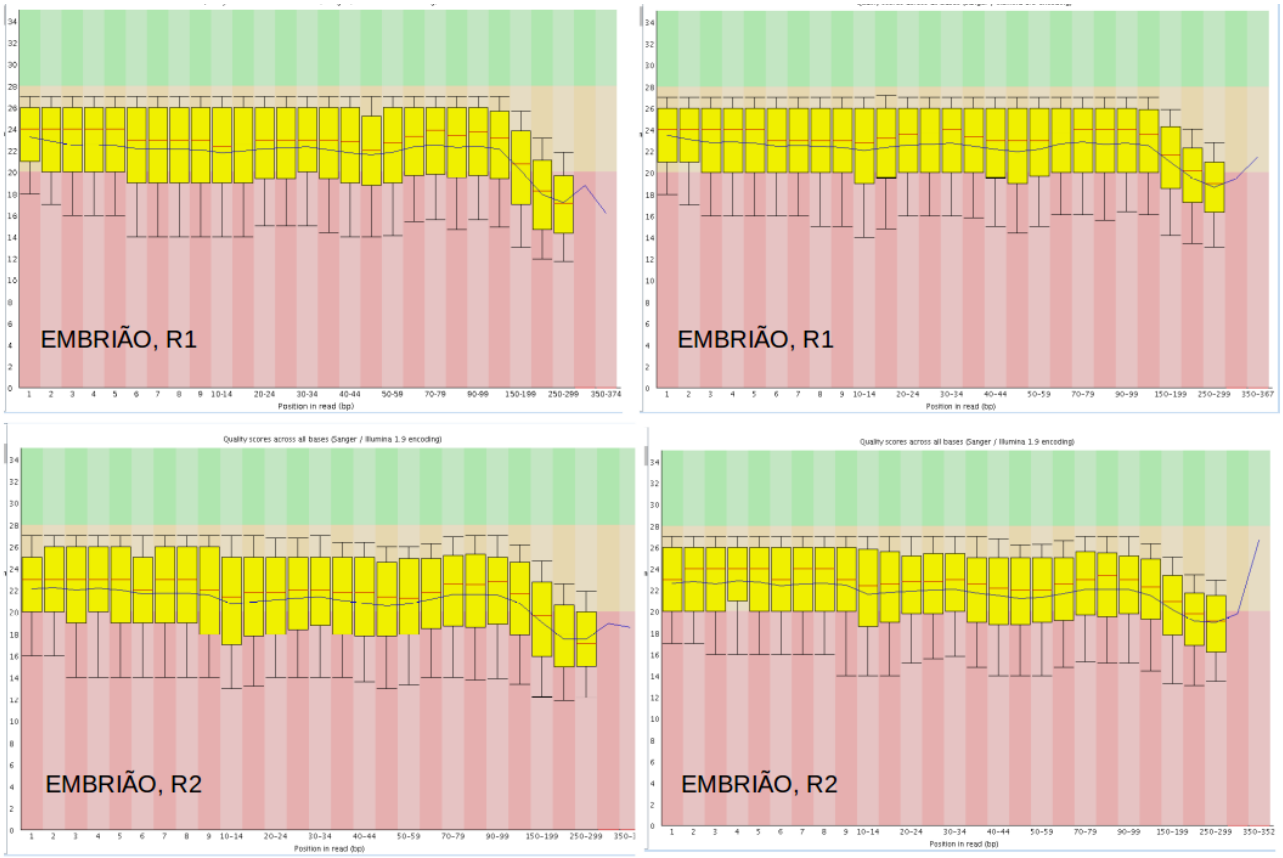


Figura 1. Qualidade dos sequenciamentos de blastocistos *in vitro* de búfalo antes (à esquerda) e após (à direita) o pré-processamento com trimagem e filtro.

ANTES DO PRÉ-PROCESSAMENTO APÓS O PRÉ-PROCESSAMENTO

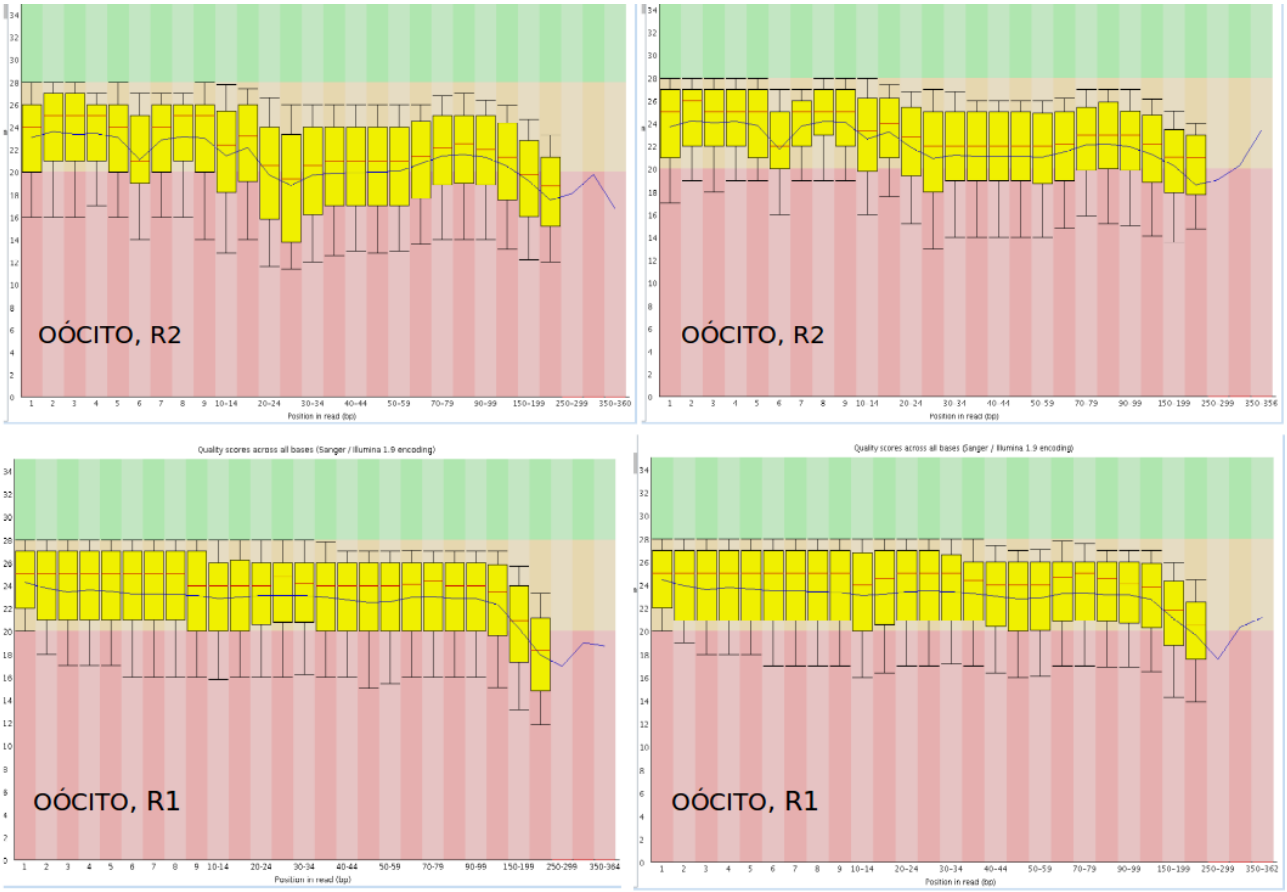


Figura 2. Qualidade dos sequenciamentos de oócitos maturados *in vitro* de búfalos antes (à esquerda) e após (à direita) o pré-processamento com trimagem e filtro.

Os arquivos estão disponíveis nos seguintes diretórios:
Dados da RUN 1: /home/priscila/storage/1.filter_RUN1/
Dados da RUN 2: /home/priscila/storage/1.filter_RUN2/

3. ALINHAMENTO POR ABORDAGEM COM REFERÊNCIA

Foi escolhida a abordagem por alinhamento com referência por causa da disponibilidade do genoma completo e anotado do bovino (*Bos taurus*) que é uma espécie relacionada ao búfalo. Foram usados o genoma e a anotação do Ensembl,

Bos_taurus.UMD3.1 release 87, release 08-12-2016 (ftp://ftp.ensembl.org/pub/release-87/fasta/bos_taurus/).

Após descompactados os cromossomos .fa foram reunidos em 1 único arquivo usando a função cat no prompt de comando. O arquivo foi denominado Bos_taurus.UMD3.1.dna.ALLchromosomes.fa. O mesmo procedimento foi feito para a anotação, nomeada Bos_taurus.UMD3.1.dna.ALLchromosomes.gff3.

Diretórios de armazenamento das referências .fa e .gff3 respectivamente:
/home/priscila/storage/0.2016_Ensembl_Bos_taurus_UMD3.1_DNA/
/home/priscila/storage/0.2016_Ensembl_Bos_taurus_UMD3.1_annotation/

Foram usados dois programas alinhadores com o objetivo de aumentar o número de sequências alinhadas, o TMAP e o Bowtie2. Em ambos os programas, o procedimento inicial é indexar o genoma de referência. No index são gerados arquivos que serão utilizados na etapa seguinte de alinhamento.

Index no TMAP: *tmap index -f [referencia.fa]*
Index no Bowtie2: *bowtie2-build [ref.fa] [output.bowtie2Index].*

O arquivo .fastq de cada amostra foi alinhado separadamente ao genoma indexado. No TMAP foi usado o modulo 'mapall' com os parâmetros default exceto pela permissão de no máximo 2 mismatches (--max mismatches 2) em cada sead de alinhamento (default 32). Enquanto no Bowtie2 também foram usados os parâmetros default exceto pela permissão do máximo 1 mismatche (-N 1) nas seads (default 32).

Alinhamento no TMAP: *tmap mapall -f Bos_taurus.UMD3.1.dna.ALLchromosomes.fa -r [input.fastq] -n 5 -v stage1 map1 –max- mismatches 2 map2 map3 > [output.sam]*

Alinhamento no Bowtie2: *bowtie2 -q -N 1 -x Bos_taurus.UMD3.1.dna.ALLchromosomes -U [input.fastq] -S [output.sam]*

Os arquivos de alinhamento estão nos diretórios a seguir:
Alinhamentos do TMAP: /home/priscila/storage/2.mapping_Tmap/
Alinhamentos do Bowtie2: /home/priscila/storage/2.mapping_Bowtie2/

4. INVESTIGAÇÃO DE ALINHAMENTOS FALSO POSITIVOS

O objetivo deste experimento foi comparar os alinhamentos do Bowtie2 e TMAP para investigar a presença de alinhamentos falso positivos. Dessa forma, selecionar o resultado que mostrasse o maior número de alinhamentos não inespecíficos.

Para isso, foram obtidas dos arquivos de alinhamento .bam somente as sequencias mapeadas ao genoma de referência (mapped.bam), eliminando, portanto as sequências não mapeadas. Em seguida, o arquivo mapped.bam foi convertido para .fasta e dele extraídos somente os cabecalhos das sequências.

4.1. Extração das sequências mapeadas por TMAP e Bowtie2

samtools view -S -b [input.sam] > [output.bam]

samtools view -b -F 4 [output.bam] > [mapped.bam]

samtools fasta [mapped.bam] > [mapped.fasta]

4.2. Extração dos cabeçalhos dos arquivos *[mapped.fasta]*

grep ">" [mapped.fasta] > [cabecalho.txt]

4.3 Comparação dos cabeçalhos dos arquivos *[mapped.fasta]*

As listas com os cabecalhos de sequências mapeadas por cada alinhador foram usadas para encontrar as sequências alinhadas por ambos TMAP e Bowtie2 e exclusivamente por cada programa. Com este objetivo, foram usados os scripts head_compara_comum.pl e head_compara_diferente.pl.

Os scripts tem a função de comparar duas listas de objetos e imprimir todos os que estão presentes em ambas as listas (head_compara_comum.pl) ou ausentes na lista 1 e presentes na lista 2 (head_compara_diferente.pl). Como resultado, foram obtidas listas com os cabecalhos das sequencias alinhadas por ambos TMAP e Bowtie2 e outras listas com cabecalhos de sequencias exclusivas de cada programa.

perl head_compara_comum.pl cabecalho_Bowtie2.txt cabecalho_Tmap.txt > ambos.txt

Resultado = alinhamentos feitos por ambos TMAP e Bowtie2

perl head_compara_diferente.pl cabecalho_Bowtie2.txt cabecalho_Tmap.txt > cabecalho_BxT.txt

Resultado = Alinhamentos ausentes no Bowtie2 e presentes no TMAP (exclusivas do TMAP)

perl head_compara_diferente.pl cabecalho_Tmap.txt cabecalho_Bowtie2.txt > cabecalho_TxB.txt

Resultado = Alinhamentos ausentes no TMAP e presentes no Bowtie2 (exclusivas do Bowtie2)

head [cabecalho.txt]
>8W32V:01332:11640
>8W32V:01335:11621
>8W32V:01335:11624

4.4. Obtenção dos exclusivos.fasta a partir das listas de cabeçalhos de sequências

As listas de cabecalhos das sequencias exclusivas do TMAP e Bowtie2 foram usadas para gerar arquivos .fasta. Para isso, as listas foram submetidas ao script meus_fastas.pl, que tem a função de encontrar os cabecalhos contidos na lista e imprimir as sequencias nucleotídicas correspondentes.

sed 's/^.//' [cabecalho.txt] > [cabecalho.txt]
head [cabecalho.txt]
8W32V:01332:11640
8W32V:01335:11621
8W32V:01335:11624
perl meusfastas.pl [mapped.fasta] [cabecalho_.txt]
Arquivo criado: novo.fasta!!

4.5 Realinhamento das exclusivas.fasta com a referência

As sequências exclusivas foram realinhadas ao genoma de referência com o objetivo de visualizar se correpondiam à regiões gênicas conforme a anotação de referência usada, ou seja, alinhamentos positivos. Em cada programa foram mantidos os parâmetros de alinhamento aplicados anteriormente.

Realinhamento das reads exclusivas no TMAP: *tmap mapall -f Bos_taurus.UMD3.1.dna.ALLchromosomes.fa -r exclusivas_Tmap.fasta -n 5 -v stage1 map1 --max-mismatches 2 map2 map3 > exclusivas_Tmap.sam*

Realinhamento das reads exclusivas no Bowtie2: *bowtie2 -f -N 1 -x bowtie2Index_UMD3.1 -U exclusivas_Bowtie2.fasta -S exclusivas_Bowtie2.sam*

4.6. Visualização em um *genome browser*

Os alinhamentos foram visualizados no CLCGenomics Workbenck e verificado se correspondiam a regiões de genes de acordo com a anotação de referência.

CLCGenomics workbench: Importar o alinhamento .BAM
"Import" --> "SAM/BAM mapping files" --> selecionar .BAM --> selecionar genoma
CLCGenomics: Visualizar o alinhamento com a anotacao do genoma
abrir alinhamento --> criar track list --> selecionar tracks de genes e transcritos

Os arquivos citados nas etapas acima são encontrados no seguinte diretório: /home/priscila/storage/4.mapped_comparison/. E nas seguintes subpastas:
Mapped reads: /4.mapped_comparison/1.mapped_reads
Cabecalhos de mapped reads: /4.mapped_comparison/2.cabecalhos_mapped
Exclusivas do TMAP: /4.mapped_comparison/3.exclusivasTmap.fa
Exclusivas do Bowtie2: /4.mapped_comparison/3.exclusivasBowtie2.fa
Re-alinhamento das exclusivas: /4.mapped_comparison/4.Re_mapping_exclusivas

5. CONTAGEM DE READS E NORMALIZAÇÃO COM O CUFFLINKS

Para a identificação dos genes expressos em oócitos e embriões foi utilizado o programa Cufflinks (Trapnell et al. 2011), que normaliza os dados de contagem de reads usando o cálculo de RPKM (*Reads Per Kilobase Million*). No cálculo de RPKM a contagem das reads de cada gene é normalizada pelo número total de reads sequenciadas (medido em milhões) e pelo tamanho do gene (medido em kilobases).

Primeiramente os arquivos .sam de alinhamento foram convertidos para .sam.sorted (arquivo de entrada no cufflinks), conforme indicado no manual do programa quando se trata de arquivo proveniente de programas diferentes do Bowtie2-TopHat (http://cole-trapnell-lab.github.io/cufflinks/cufflinks).

sort -k 3,3 -k 4,4n 008_t20p50m2.sam > embryo1_t20p50m2.sam.sorted
sort -k 3,3 -k 4,4n 013_t20p50m2.sam > oocyte1_t20p50m2.sam.sorted
sort -k 3,3 -k 4,4n 005_t20p50m2.sam > embryo2_t20p50m2.sam.sorted
sort -k 3,3 -k 4,4n 010_t20p50m2.sam > oocyte2_t20p50m2.sam.sorted

O protocolo usado neste trabalho consistiu no uso de três ferramentas: cufflinks, cuffmerge e cuffdiff (Trapnell et al. 2012). O cufflinks realiza a montagem das sequências (transcritos) e normalização através do cálculo do RPKM em cada amostra. Foram usados os parametros default, selecionado o tipo de bilbioteca (fr-secondstrand) e a opção -G, que utilizou a referência *Bos taurus* UMD3.1 para a montagem.

cufflinks -p 6 --library-type fr-secondstrand -o [output.dir] -G [referencia.gff3] [input.sam.sorted]

Em seguida a ferramenta cuffmerge faz a união do arquivo de saída do Cufflinks, o *transcripts.gtf* de cada replicata para gerar um único arquivo chamado *merged.gtf*, e que corresponde a um índice geral dos transcritos identificados em todas as amostras e replicatas. Para isso, e inserida na linha de comando do cuffmerge uma lista dos diretórios de cada *transcripts.gtf,* chamada aqui de *assemblies.txt.*

cuffmerge -o [output.dir] -g [referencia.gff3] -s [referencia.fa] -p 6 assemblies.txt

Por fim a ferramenta cuffdiff executa o cálculo da expressão diferencial, usando o *merged.gtf.* Foram usados os parâmetros *default* e adotada a distribuição de Poisson. Porém, o Cuffdiff foi usado apenas para a geração do arquivo *gene_exp_diff* no qual são mostrados os RPKM das amostras e não somente de cada replicata, como no arquivo de saída *genes.fpkm_tracking* do Cufflinks.

A ordem de entrada dos arquivos é importante na execução do Cuffdiff, pois como neste trabalho pretende-se calcular os genes diferencialmente expressos em embriões em relação aos oócitos, neste caso as amostras *oocyte* devem ser inseridas em primeiro lugar seguidas das amostras *embryo.*

cuffdiff -o [output.dir] -b [referencia.fa] --library-type fr-secondstrand --dispersion-method poisson -p 5 -L oocyte,embryo -u merged.gtf \
> [oocyte1.sam.sorted],[oocyte2.sam.sorted] \
> [embryo1.sam.sorted],[embryo2.sam.sorted]

Os arquivos estão armazenados nos seguintes diretórios:

Input.sam.sorted: /home/priscila/storage/2.mapping_Tmap/

Output cufflinks: /home/priscila/storage/3.CufflinksTmap_oocXemb/

Output cuffmerge: /home/priscila/storage/3.CufflinksTmap_oocXemb/

Output cuffdiff: /home/priscila/storage/3.CufflinksTmap_oocXemb/Poisson/

6. OBTENÇÃO DOS RESULTADOS DE EXPRESSÃO

A partir dos valores de RPKM foram obtidos os genes expressos em oócitos e embriões, exclusivos de cada um e compartilhados por ambos. Foi considerado expresso todo gene que apresentou RPKM>0.4, que corresponde a um valor pouco mais estringente do que 0.3 referido por Ramskold et al. como sendo o valor mínimo para considerar um gene expresso (Ramsköld et al. 2009).

Os valores de RPKM foram extraídos do arquivo de saída *gene_exp_diff* do Cuffdiff, pois ele apresenta o cálculo de RPKM para cada gene em cada amostra. A seguir é mostrado passo-a-passo como foram obtidas as listas de genes exclusivos e compartihados por oócitos e embriões.

6.1. Obtenção das listas de genes expressos em cada amostra (RPKM>0.4)

Primeiramente foi obtida uma lista de todos os genes com RPKM > 0.4. Foi usado o arquivo *gene_exp_diff* que contém o cálculo de RPKM para ambas as replicatas biológicas de oócitos (coluna 8 chamada value1) e embriões (coluna 9 chamada value2). As colunas 1 (código criado pelo cufflinks), coluna 3 (gene_symbol) e coluna 8 ou 9 (RPKM) foram imprimidas somente se a condição acima fosse satisfeita.

```
cat gene_exp.diff | awk -F "\t" '{if($8>0.4) print $1 "\t" $3 "\t" $8}' > oocyte_RPKM04.txt
cat gene_exp.diff | awk -F "\t" '{if($9>0.4) print $1 "\t" $3 "\t" $9}' > embryo_RPKM04.txt
```

Assim foi criada uma lista para cada amostra contendo os códigos do cufflinks, nome dos genes (gene_symbol) e valores de RPKM. Em seguida, somente as colunas 1 das listas foram isoladas e comparadas, pois cada gene tem apenas um código, e pode apresentar mais de um gene_symbol.

```
cat embryo_RPKM04.txt | awk -F "\t" '{print $1}' > embryo_RPKM04_code.txt
cat oocyte_RPKM04.txt | awk -F "\t" '{print $1}' > oocyte_RPKM04_code.txt
```

6.2. Obtenção das listas de genes expressos em ambas as amostras

As listas de genes compartilhados (RPKM>04) por oócitos e embriões foram obtidas usando o script abaixo, que compara a coluna 1 de dois arquivos e imprime somente os objetos presentes em ambos.

perl head_compara_comum.pl oocyte_RPKM04_code.txt embryo_RPKM04_code.txt > ambos_RPKM04.txt

6.3. Obtenção das listas de genes exclusivos de cada amostra

As listas de genes exclusivos (RPKM>04) de oócitos e embriões foram obtidas usando outro script, que por sua vez compara a coluna 1 de dois arquivos e imprime somente os objetos ausentes no arquivo 1 e presentes no arquivo 2, ou seja, exclusivos do arquivo 2.

perl head_compara_diferente.pl oocyte_RPKM04_code.txt embryo_RPKM04_code.txt > exclusivos_embryo_RPKM04.txt
perl head_compara_diferente.pl embryo_RPKM04_code.txt oocyte_RPKM04_code.txt > exclusivos_oocyte_RPKM04.txt

6.4. Obtenção das listas de genes com informação de “gene_symbol”

Os mesmos resultados podem ser obtidos usando os scripts head_compara_comum_3cols.pl e head_compara_diferente_3cols.pl, que tem as mesmas funções dos anteriores (head_compara_comum.pl e head_compara_diferente.pl, respectivamente), ou seja, comparam a primeira coluna de 2 arquivos. Porém, os arquivos de entrada poderão ter até 3 colunas. As comparações acima foram repetidas com os scripts citados com o objetivo de obter as informações de gene_symbol e RPKM.

6.5. Obtenção do “gene_symbol” dos genes compartilhados pelas amostras

A comparacao também foi feita usando a coluna 1 (código do cufflinks), e os scripts “*comum.pl” e “*comum_3cols.pl” encontraram o mesmo número de genes (8663). Assim foi obtida a lista de genes compartilhados por oócitos e embriões (ambos.txt) com gene_symbol e RPKM. Desta lista foram retirados somente os RNAs ribossomais, de forma que o número total de genes compartilhados diminuiu para 8649.

cat gene_exp.diff | awk -F "\t" '{if($8>0.4) print $1 "\t" $3 "\t" $8}' > oocyte_RPKM04.txt
cat gene_exp.diff | awk -F "\t" '{if($9>0.4) print $1 "\t" $3 "\t" $9}' > embryo_RPKM04.txt

perl head_compara_comum_3cols.pl oocyte_RPKM04.txt embryo_RPKM04.txt > ambos_oe_RPKM04.txt
perl head_compara_comum_3cols.pl embryo_RPKM04.txt oocyte_RPKM04.txt > ambos_eo_RPKM04.txt

6.6. Obtenção do “gene_symbol” dos genes exclusivos de cada amostra

Da mesma forma os scripts “*diferente.pl” e “*diferente_3cols.pl” encontraram o mesmo número de genes. Assim foi obtida a lista de genes exclusivos de oócitos e embriões (*exclusivos.txt) com os repectivos gene_symbol e valores de RPKM. Das listas de exclusivos foram retirados somente os RNAs ribossomais, de forma que o número total de genes diminuiu em embriões (3927) e oócitos (1400).

perl head_compara_diferente_3cols.pl oocyte_RPKM04.txt embryo_RPKM04.txt > embryo_exclusivos.txt
perl head_compara_diferente_3cols.pl embryo_RPKM04.txt oocyte_RPKM04.txt > oocyte_exclusivos.txt

OBS: Existe a alternativa de selecionar os genes exclusivos através dos valores "inf" e "-inf" na coluna log2Fold-change do arquivo *gene_exp_diff*. Os valores correspondem aos genes com 0 RPKM. Especificamente "-inf” significa 0 RPKM em oócito, portanto gene exclusivo de embrião. Assim como "inf” siginifica 0 RPKM em embrião, portanto gene exclusivo de oócito.

Porém, observar que o critério usado nesse trabalho considerou os genes expressos quando o RPKM>0.4, ou seja, genes com RPKM diferente de 0 e abaixo de 0.4 foram considerados não expressos. Dessa forma, comparado a alternativa citada acima o critério adotado no trabalho é mais estringente.

7. CLASSIFICAÇÃO DOS GENES EXPRESSOS EM BIOTIPOS

Após obter os genes expressos em oócitos e embriões foi feita a classificação enquanto o biotipo dos genes (“gene_biotype”), como sendo codificantes de proteínas, não-codificantes, pseudogenes entre outros. O “gene_biotype” consta na referência Bos_taurus.UMD3.1.gff3 e foi adicionada às listas de genes usando o biomaRt do Bioconductor (https://bioconductor.org/packages/release/bioc/html/biomaRt.html).

O biomaRt permite acessar os dados de anotação e extrair informações específicas (Durinck et al., 2009). Primeiramente o biomaRt baixou a anotação de referência e construiu uma query com os atributos solicitados. Então, foram incluídas as listas de genes exclusivos e compartilhados por oócitos e embriões contendo “ensembl_ID”, “gene_symbol” ou “transcripts_ID” para correlacionar ao “gene_biotype”.

```
library("biomaRt")
ensembl = useEnsembl(biomart="ensembl", dataset="btaurus_gene_ensembl", version =
87)
annotation <- getBM(attributes=c('hgnc_symbol', 'gene_biotype'), mart = ensembl)
head(annotation)
annot.table <- data.frame(annotation)
ensembl_gene_id <- character(length = 0L)
genes.table <- read.csv(file="List_Of_Genes_Expressed_Symbol.csv")
annot.table <- merge(x = annot.table, y = genes.table,
                     by.x = "hgnc_symbol", by.y =
"Gene", all.x = T, all.y = F )
head(annot.table)
symbol <- as.data.frame(annot.table)
write.csv(symbol, file= "List_Of_Gene_Biotype_symbol.csv")
ensembl = useEnsembl(biomart="ensembl", dataset="btaurus_gene_ensembl", version =
87)
annotation <- getBM(attributes=c('ensembl_gene_id', 'gene_biotype',), mart = ensembl)
head(annotation)
annot.table <- data.frame(annotation)
ensembl_gene_id <- character(length = 0L)
genes.table <- read.csv(file="List_Of_Genes_Expressed_EnsemblID.csv")
annot.table <- merge(x = annot.table, y = genes.table,
                   by.x = "ensembl_gene_id", by.y =
                     "Gene", all.x = T, all.y = F )
head(annot.table)
gene <- as.data.frame(annot.table)
write.csv(gene, file= "List_Of_Gene_Biotype_ensemblID.csv")
ensembl = useEnsembl(biomart="ensembl", dataset="btaurus_gene_ensembl", version =
87)
annotation <- getBM(attributes=c('ensembl_transcript_id', 'gene_biotype'), mart =
ensembl)
head(annotation)
annot.table <- data.frame(annotation)
ensembl_gene_id <- character(length = 0L)
genes.table <- read.csv(file="List_Of_Genes_Expressed_transcriptID.csv")
annot.table <- merge(x = annot.table, y = genes.table,
                   by.x = "ensembl_transcript_id", by.y =
                     "Gene", all.x = T, all.y = F )
head(annot.table)
transcript <- as.data.frame(annot.table)
write.csv(transcript, file= "List_Of_Gene_Biotype_transcriptID.csv")
```

Foram gerados gráficos de “pizza” com a classificação dos genes em 4 biotipos: codificantes de proteínas, não-codificantes, pseudogenes e prováveis novos transcritos.

8. ANÁLISE DE EXPRESSÃO DIFERENCIAL COM O HTSEQ-DESEQ2

A contagem de reads no HTSeq foi realizada para cada amostra separadamente. Os arquivos de alinhamento .sam foram usados como entrada adotando os parametros default para contagem das reads: *htseq-count -f sam -s yes -a 10 -t exon -i gene_id -m union [input.sam] [referencia.gtf] > [output.txt]*

htseq-count -f sam -s yes -a 10 -t exon -i gene_id -m union [embryo1.sam] [referencia.gtf] > embryo1.txt
htseq-count -f sam -s yes -a 10 -t exon -i gene_id -m union [embryo2.sam] [referencia.gtf] > embryo2.txt
htseq-count -f sam -s yes -a 10 -t exon -i gene_id -m union [oocyte1.sam] [referencia.gtf] > oocyte1.txt
htseq-count -f sam -s yes -a 10 -t exon -i gene_id -m union [oocyte2.sam] [referencia.gtf] > oocyte2.txt

As contagens de reads de cada amostra foram reunidas em uma única planilha e carregada no DESeq2 como uma matriz (arquivo countData.csv) e convertida na função DESeqDataSetFromMatrix. Também foram atribuídas as características das amostras (biblioteca single-end e tipo celular) por meio do arquivo colData.csv. A ordem das colunas em countData foi a mesma ordem das amostras em colData (oócito 1, oócito 2, embrião 1 e embrião 2). Sendo que as replicatas da amostra referência para comparação (oócito) consistiram nas primeiras colunas do arquivo countData.csv.

Em seguida foram filtrados os genes com contagem de reads igual a 0. Esta etapa é indicado para aumentar a velocidade da análise e filtrar melhor os dados. Para aumentar a estringência das análises o FDR (False Discovery Rate) foi ajustado para 0.05 (alpha 0.05) e o valor de significância adotado foi de p-value ajustado ≤ 0.05 (padj ≤ 0.05). Esses resultados foram armazenados na variável res.05, e em seguida foi feita uma lista dos genes com padj ≤ 0.05, ou seja, diferencialmente expressos (Res05padj005.csv).

> library("DESeq2")
> countData <- as.matrix(read.csv(file="countData.csv",sep=",",row.names="gene_id"))
> colData <- read.csv(file="colData.csv",sep=",",row.names=1)

```
> colData <- colData[,c("condition","type")]
> head(countData)
                 oocyte1 oocyte2 embryo1 embryo2
ENSBTAG00000000005     193       0       0       0
ENSBTAG00000000008       0       0       0       0
ENSBTAG00000000009       0       0       0       0
ENSBTAG00000000010      25       0     103       2
ENSBTAG00000000011      38       0       0       0
ENSBTAG00000000012       0       0      16       1
> head(colData)
              condition      type
oocyte1      oocyte   Single-read
oocyte2      oocyte   Single-read
embryo1     embryo  Single-read
embryo2     embryo  Single-read
> countData <- countData[, rownames(colData)]
> all(rownames(colData) == colnames(countData))
> dds <- DESeqDataSetFromMatrix(countData = countData,
+                               colData = colData,
+                               design = ~ condition)
> dds
class: DESeqDataSet
dim: 24601 4
metadata(0):
assays(1): counts
rownames(24601): ENSBTAG00000000005 ENSBTAG00000000008 ... __not_aligned
__alignment_not_unique
rowRanges metadata column names(0):
colnames(4): oocyte1 oocyte2 embryo1 embryo2
colData names(2): condition type
> dds$condition <- factor(dds$condition, levels=c("oocyte","embryo"))
> dds <- DESeq(dds)
estimating size factors
estimating dispersions
gene-wise dispersion estimates
mean-dispersion relationship
final dispersion estimates
fitting model and testing
> res <- results(dds)
> nrow(dds)
[1] 24601
> dds <- dds[ rowSums(counts(dds)) > 1, ]
> nrow(dds)
[1] 13478
> res <- results(dds)
> res.05 <- results(dds, alpha=.05)
> head(res.05)
log2 fold change (MAP): condition embryo vs oocyte
Wald test p-value: condition embryo vs oocyte
DataFrame with 6 rows and 6 columns
```

```
                   baseMean log2FoldChange     lfcSE       stat           pvalue       padj
                  <numeric>      <numeric> <numeric>  <numeric>    <numeric>  <numeric>
ENSBTAG00000000005   9.539065     -7.0117919 2.4983930 -2.8065208 0.005007967 0.01669066
ENSBTAG00000000010   7.554728      1.5001880 2.0970997  0.7153632 0.474384628 0.61754450
ENSBTAG00000000011   1.878158     -3.7144995 2.9096255 -1.2766246 0.201734814         NA
> table(res.05$padj <= 0.05)
FALSE  TRUE
 6709  4156
> write.csv(as.data.frame(res.05[which( res.05$padj <= 0.05),]),
file="/home/priscila/Res05padj005.csv")
```

9. ANÁLISES DE PCA (*PRINCIPAL COMPONENT ANALYSIS*) E CLUSTERIZAÇÃO

Primeiramente os dados de *reads count* foram transformados em *rlog* (*regularized-logarithm transformation*), que é uma das normalizações oferecidas pelo DESeq2 para estabilizar a variância através da média (Love, Huber, and Anders 2014).

A transformação de rlog foi escolhida pois o conjunto de dados neste trabalho é pequeno (n <30) e foi observada variedade de profundidade de seqüenciamento nas amostras (http://www.bioconductor.org/help/workflows/rnaseqGene/).

```
> rld <- rlog(dds, blind=FALSE)
> head(assay(rld), 3)
                     oocyte1    oocyte2      embryo1   embryo2
ENSBTAG00000000005 4.226959 -0.2878531 -2.852744 -2.006005
ENSBTAG00000000010 2.278463  1.2779260  3.964727  1.333813
ENSBTAG00000000011 1.387824 -0.3324173 -1.203263 -1.110715
```

Em seguida, para observar a distância entre as amostras, ou seja o grau de similaridade entre oócitos e embriões foi gerado o gráfico de PCA, que consiste em uma análise de covariância que mostra a correlação dos componentes, no caso, oócitos e embriões e as respectivas replicatas biológicas.

Outra forma de visualizar a distância entre as amostras é através do cálculo da distância euclidiana, que foi representada como um gráfico de heatmap, em que as amostras mais similares foram agrupadas juntas.

```
> plotPCA(rld, intgroup=c("condition"))
> sampleDists <- dist(t(assay(rld)))
> sampleDists
            oocyte1      oocyte2      embryo1
oocyte2   277.4723
embryo1 445.4862   382.2122
embryo2 425.2703   361.0830    197.0448
> library("RColorBrewer")
> sampleDistMatrix <- as.matrix(sampleDists)
> rownames(sampleDistMatrix) <- paste(rld$condition, rld$type, sep="-")
> colnames(sampleDistMatrix) <- paste(rld$condition, rld$type, sep="-")
> colors <- colorRampPalette( rev(brewer.pal(9, "Blues")) )(255)
> library("pheatmap")
> pheatmap(sampleDistMatrix,
+          clustering_distance_rows=sampleDists,
+          clustering_distance_cols=sampleDists,
+          col=colors)
```

Para identificar os principais grupos de genes responsáveis pela similaridade ou dissimilaridade entre as amostras foi feita a anáĺise de agrupamento ou clusterização dos genes mais variáveis. No DESeq2, o gene variável é identificado calculando-se o quanto o valor rld do gene em uma amostra específica desvia da média do gene para todas as amostras. Dessa forma, os genes mais variáveis são diferencialmente expressos, porém não são necessariamente os mais induzidos ou reprimidos (já que esta definição é baseada no cálculo de log2FoldChange).

```
> library("genefilter")
> topVarGenes <- head(order(rowVars(assay(rld)),decreasing=TRUE),4156)
> mat <- assay(rld)[ topVarGenes, ]
> mat <- mat - rowMeans(mat)
> head(mat)
                          oocyte1  oocyte2   embryo1   embryo2
ENSBTAG00000045782  6.231605 6.803476 -7.303414 -5.731667
ENSBTAG00000006965  6.278934 6.395636 -7.118895 -5.555675
ENSBTAG00000004658  6.953498 5.475607 -6.993365 -5.435741
> df <- as.data.frame(colData(rld)[,c("condition", "type")])
> head(df)
            condition      type
oocyte1      oocyte   Single-read
oocyte2      oocyte   Single-read
embryo1    embryo  Single-read
embryo2    embryo  Single-read
> library("pheatmap")
> pheatmap(mat, annotation_col=df)
```

Para a visualização dos genes induzidos e reprimidos foi gerado o gráfico volcano, que expressa os valores de log2FoldChange e de padj transformados para escala logarítimica. Todos os genes considerados diferencialmente expressos no embrião em relação ao oócito (padj ≤ 0.05) são visualizados como pontos vermelhos, acima da linha tracejada verde.

Os genes induzidos (FoldChange negativo) são localizadas à esquerda da primeira linha tracejada azul (-1 fold). Enquanto os genes reprimidos (FoldChange positivo) são localizados à direita da segunda linha tracejada azul (+1 fold).

```
> tab = data.frame(logFC = res.05$log2FoldChange, negLogPadj = -log10(res.05$padj))
> head(tab)
      logFC negLogPadj
1 -7.0117919  1.7775264
2  1.5001880  0.2093317
3 -3.7144995         NA
> par(mar = c(5, 4, 4, 4))
> plot(tab, pch = 16, cex = 0.6, xlab = expression(log[2]~fold~change), ylab = expression(-
log[10]~padj))
> lfc = 1
> padj = 0.05
> signGenes = (abs(tab$logFC) > lfc & tab$negLogPadj > -log10(padj))
> #signGenes
> points(tab[signGenes, ], pch = 16, cex = 0.8, col = "red")
> abline(h = -log10(padj), col = "green3", lty = 2)
> abline(v = c(-lfc, lfc), col = "blue", lty = 2)
> mtext(paste("padj =", padj), side = 4, at = -log10(padj), cex = 0.8, line = 0.5, las= 1)
> mtext(c(paste("-", lfc, "fold"), paste("+", lfc, "fold")), side = 3, at = c(-lfc, lfc), cex = 0.8,
line = 0.5)
```

10. CLASSIFICAÇÃO DOS GENES DIFERENCIALMENTE EXPRESSOS EM BIOTIPOS

As listas de genes expressos diferencialmente expressos usadas na análise de ontologia gênica foram obtidas a partir do arquivo de saída do DESeq2 (Res05padj005.csv). O HTSeq e DESeq2 usam o ensembl_ID para as análises, pois são identificadores únicos dos genes. Então foram adicionados às tabelas, os respectivos “gene_symbol” e “gene_biotype” usando o pacote BiomaRt do R.

```
> library("biomaRt")
> ensembl = useEnsembl(biomart="ensembl", dataset="btaurus_gene_ensembl", version
= 87)
> annotation <- getBM(attributes=c('ensembl_gene_id', 'hgnc_symbol', 'gene_biotype'),
mart = ensembl)
> head(annotation)
    ensembl_gene_id        hgnc_symbol   gene_biotype
1 ENSBTAG00000047958       DPP7    protein_coding
2 ENSBTAG00000020495      SH3YL1   protein_coding
3 ENSBTAG00000003540                protein_coding
> annot.table <- as.data.frame(annotation)
> ensembl_gene_id <- character(length = 0L)
> genes.table <- read.csv(file="/home/priscila/Documentos/UPpadj005.csv")
> head(genes.table)
                X  baseMean log2FoldChange    lfcSE      stat           pvalue       padj
1 ENSBTAG00000045782 1619.3352      -14.59094 1.722930 -8.468681 2.481883e-17
4.085705e-15
2 ENSBTAG00000006965 1197.5547      -14.22606 1.734860 -8.200119 2.401488e-16
3.261521e-14
3 ENSBTAG00000004658 1124.8269      -14.01596 1.773994 -7.900792 2.771365e-15
3.136550e-13
> annotation <- annotation[match(genes.table$X, annot.table$ensembl_gene_id), ]
> all(genes.table$X == annotation$ensembl_gene_id)
[1] TRUE
> head(annotation)
       ensembl_gene_id              hgnc_symbol   gene_biotype
8951  ENSBTAG00000045782       BMP15       protein_coding
13906 ENSBTAG00000006965       NLRP8       protein_coding
8132  ENSBTAG00000004658       WEE2        protein_coding
```

11. PREPARO DOS ARQUIVOS PARA ONTOLOGIA GÊNICA

Para a busca das funções biológicas e moleculares dos genes expressos e diferencialmente expressos foi usada a plataforma Gofeat (http://www.computationalbiology.ufpa.br/gofeat/) para ontologia gênica, que realiza o blast das CDS (Sequências de DNA Codificantes) nos bancos de dados do NCBI, Uniprot, Interpro e Kegg (dados não publicados).

Os arquivos de entrada no Gofeat foram multifastas obtidos usando a ferramenta online BioMart do Ensembl (http://www.ensembl.org/biomart/martview/), que é uma interface web para facilitar a seleção de filtros e atributos (Smedley et al. 2015), e tem as mesmas funções do pacote biomaRt do biocondutor, citado anteriormente (Item 7).

Como cada gene possui somente um único “ensembl_id”, foi definido usar este atributo para a seleção das CDS no BioMart. Como nos arquivos de saída do HTSeq-DESeq2 os genes são identificados pelos seus “ensembl_id”, então não foi necessário preparar as listas de genes para submissão no BioMart. Porém, os arquivos de saída do Cufflinks necessitaram do procedimento de preparo descrito a seguir.

11.1. Conversão dos “gene_symbol” para “ensembl_id” nos arquivos do Cufflinks

Os genes exclusivos e compartilhados por oócitos e embriões foram provenientes do Cufflinks (ver Item 6) em que os genes são identificados por meio de códigos elaborados pelo Cufflinks (coluna 1) e por meio do atributo “gene_symbol” (coluna 3). Nesta etapa de preparo os “gene_symbol” foram convertidos para “ensembl_id” usando o pacote biomaRt no ambiente R.

```
> library("biomaRt")
> ensembl = useEnsembl(biomart="ensembl", dataset="btaurus_gene_ensembl", version
= 87)
> annotation <- getBM(attributes=c('hgnc_symbol', 'ensembl_gene_id', 'gene_biotype'),
mart = ensembl)
> head(annotation)
  hgnc_symbol    ensembl_gene_id        gene_biotype
1        DPP7  ENSBTAG00000047958 protein_coding
2      SH3YL1         ENSBTAG00000020495 protein_coding
3                     ENSBTAG00000003540 protein_coding
> annot.table <- data.frame(annotation)
> hgnc_symbol <- character(length = 0L)
> genes.table <-
read.csv(file="/home/priscila/Documentos/exclusivo_RPKM04_oocyte_symbol.csv")
> annot.table <- merge(x = annot.table, y = genes.table,
+                      by.x = "hgnc_symbol", by.y =
+ "Gene", all.x = T, all.y = F )
> head(annot.table)
  hgnc_symbol    ensembl_gene_id        gene_biotype     RPKM
1                      ENSBTAG00000003540 protein_coding <NA>
2                     ENSBTAG00000035144 protein_coding <NA>
3                     ENSBTAG00000038537 protein_coding <NA>
> symbol <- as.data.frame(annot.table)
> write.csv(symbol, file= "/home/priscila/Documentos/List_Of_symbol_for_BioMart.csv")
```

11.2. Obtenção dos arquivos multifastas no BioMart

No BioMart, foram obtidos os arquivos multifasta contendo as CDS dos genes exclusivos, e compartilhados por oócitos e embriões, e também dos diferencialmente expressos entre as amostras.

Para isso, as listas de genes contendo os “ensembl_id” dos genes foram submetidas à ferramenta BioMart que selecionou as CDS correspondentes, usando a referência Bos_tauros.UMD3.1.

Na interface BioMart foram feitos os procedimentos a seguir:

- Seleção do banco de dados: “Choose Database” --> Ensembl Genes 89
- Seleção da referência: “Choose Dataset” --> Cow genes (UMD3.1)
- Seleção dos filtros: “Filters” / “Gene” --> inserir lista de ensembl_id dos genes
- Seleção dos atributos: “Sequences” / “Sequences” --> selecionar “Coding sequence” / “Header information” --> selecionar “Gene stable ID” e “Gene name”
- Selecionar “Results” em seguida “Unique results only”

OBS: Alguns “ensembl_id” das listas de genes não foram encontrados pois consistem em RNAs não codificantes ou pseudogenes, sendo assim não possuem CDS disponível.

12. SCRIPTS - FERRAMENTAS CRIADAS EM LINGUAGEM DE PROGRAMAÇÃO PERL

Os scripts usados nesta metodologia estão disponíveis na plataforma online GitHub por meio do link https://github.com/PriscilaSantana/RNAseq. A seguir foram sumarizados o tipo de arquivo de entrada, breve descrição da função e linha de comando para execução de cada script.

12.1. Script head_compara_comum.pl

Arquivo de entrada: 2 listas .txt com 1 coluna

Função: Comparar 2 listas de objetos e imprimir objetos presentes em ambas.

Linha de comando: *perl head_compara_comum.pl [lista1.txt] [lista2.txt] > output.txt*

12.2. Script head_compara_diferente.pl

Arquivo de entrada: 2 listas .txt com 1 coluna cada.

Função: Comparar 2 listas de objetos e imprimir os objetos ausentes na primeira lista e presentes na segunda.

Linha de comando: *perl head_compara_diferente.pl [lista1.txt] [lista2.txt] > output.txt*

12.3. Script head_compara_comum_3cols.pl

Arquivo de entrada: 2 lista .txt com até 3 colunas cada.

Função: Comparar a primeira coluna de 2 arquivos e imprimir objetos presentes em ambas as listas. Em seguida, imprimir as colunas 2 e 3 correspondentes aos objetos selecionados. A ordem dos arquivos de entrada é importante, pois as colunas 2 e 3 são impressas a partir do arquivo 1.

Linha de comando: *perl head_compara_comum_3cols.pl [lista1.txt] [lista2.txt] > output.txt*

12.4. Script head_compara_diferente_3cols.pl

Arquivo de entrada: 2 listas .txt com até 3 colunas cada.

Função: Comparar a primeira coluna de 2 arquivos e imprimir objetos ausentes na 1ª lista e presentes na 2ª. Em seguida, imprimir as colunas 2 e 3 correspondentes aos objetos selecionados.

Linha de comando: *perl head_compara_diferente_3cols.pl [lista1.txt] [lista2.txt] > output.txt*

12.5. Script meusfastas.pl

Arquivo de entrada: 1 dados.fasta e 1 lista .txt ou .csv

Função: Localizar os objetos da lista .txt no arquivo de entrada fasta. Em seguida, selecionar as sequências fasta correspondentes aos objetos e direciona-las para um arquivo de saída multifasta.

Linha de comando: *perl meusfastas.pl [dados.fasta] [lista.txt]*

REFERÊNCIAS BIBLIOGRÁFICAS

Durinck S, Spellman P, Birney E and Huber W (2009). “Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt.” Nature Protocols, **4**, pp. 1184–1191.

Ramsköld, D., Wang, E. T., Burge, C. B., & Sandberg, R. (2009). An Abundance of Ubiquitously Expressed Genes Revealed by Tissue Transcriptome Sequence Data. *PLOS Computational Biology*, *5*(12), e1000598.

Smedley, D., Haider, S., Durinck, S., Pandini, L., Provero, P., Allen, J., … Kasprzyk, A. (2015). The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Research*, *43*(W1), W589–W598.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., … Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protocols*, *7*(3), 562–578.

Trapnell, C., Williams, B. a, Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., … Pachter, L. (2011). Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. *Nature Biotechnology*, *28*(5), 511–515.

6. CAPÍTULO 3 – COMPARAÇÃO DAS REDES DE CO-EXPRESSÃO GÊNICAS EM OÓCITOS E EMBRIÕES PRODUZIDOS *IN VITRO* DE BUBALINOS E BOVINOS.

Artigo 3: O manuscrito será submetido para o periódico Reproduction e segue as normas de formatação da revista

Fator de impacto: 3,49

Qualis Capes A2 (área de Biotecnologia)

Title: COMPARISON OF GENE CO-EXPRESSION NETWORKS IN OOCYTES AND BLASTOCYSTS PRODUCED *IN VITRO* OF BUFFALO AND BOVINE.

Running title: Co-expression networks in buffalo and bovine.

Key-words: buffalo – *Bubalus bubalis* - RNAseq – networks – co-expression – system biology – preservation – WGCNA – In Vitro Embryo Production – IVEP – matured oocytes – blastocyst – embryo – embryo development – transcriptome

Authors and Affiliations: Priscila P.B. Santana[1], Kenny C. Pinheiro[2], Flávia F. Aburjaile[2], Eduardo B. de Souza[3], Thiago V.G. Silva[3], Alessandra X. Santos[3], Marcela S. Cordeiro[3], Simone S.D. Santos[3], Otávio M. Ohashi[3], Moysés S. Miranda[3], Artur L.C. Silva[2], Rommel T.J. Ramos[2].

[1] Federal Rural University of Amazon, Capitão-Poço, Pará, Brazil;

[2] Laboratory of Genomics and Bioinformatics, Federal University of Pará;

[3] Laboratory of In Vitro Fertilization, Institute of Biological Science, Federal University of Pará;

Finantial Support: CAPES, Bubrás.

Author correspondence: Santana PPB.

ABSTRACT

In vitro embryo production (IVEP) protocols in buffalo are largely inspired by bovine, however, there are morphological and molecular differences between their oocytes and embryos. In this study, the transcriptomes of mature oocytes and blastocysts produced in vitro from buffalo and bovine were compared with the aim of indicating the similarities of the co-expression relations of their transcriptomes. For this, were applied the analysis of coexpression network and module preservation in RNAseq data obtained in the literature. Of the 7 co-expression modules identified in buffalo, 4 were strongly preserved (Z-summary > 10) in bovine. The gene ontology was related to the embryonic development program. However, the expression profile within the modules, defined by the hub genes, was largely different, indicating important differences in terms of gene interactions between the two species, that justify the elaboration of specific IVEP protocols for buffalos.

INTRODUCTION

In cattle, transcriptome studies using the RNAseq approach have been important to elucidate the molecular aspects of oocyte maturation (Kropp et al., 2014; Kropp and Khatib, 2015; Reyes et al., 2015) and embryo development (Chitwood et al., 2013; Driver et al., 2012; Graf et al., 2014; Xue et al., 2013). In buffalos, however, few transcriptome studies in oocytes (Kandil et al., 2010) and embryos (Abdoon et al., 2012; Strazzullo et al., 2014) were carry out to investigate the maturation and embryo development in the specie.

In contrast to cattle, however, in vitro matured oocytes of buffalo show different morphology (Neglia et al., 2003) and kinetics of nuclear maturation (Santos et al., 2002). Likewise, buffalo embryos show different developmental kinetics and morphologies in relation to cattle (Gasparrini et al., 2014; Neglia et al., 2003).In addition, studies of expression profile evidenced that oocytes and buffalo embryos have peculiar metabolic aspects (Abdoon et al., 2014; Kandil et al., 2010; Kumar et al., 2012). Considering that, although buffalos and cattle are related species, there are morphological and molecular differences between their gametes and embryos.

Using the RNAseq approach it is possible to investigate the molecular differences between species trhough the comparison of the transcriptomes then find similarities between them (Jiang et al., 2014; Oldham et al., 2008; Xue et al., 2013). This is one of the applications of the *Weighted correlation network analysis (*WGCNA) that consists of a system biology method used to describe the correlation patterns between a large number of genes (Langfelder and Horvath, 2008).

Briefly, this analysis considers that the genes establish multiple co-expression relationships, then represent them in a graph in which the genes (vertices) are connected by their relations (edges). Groups of related genes tend to cluster in modules, in turn a set of modules makes up a network of co-expression. In a way, this approach summarizes the vast information of the transcriptome, through the construction of an "architecture", in which the relations between the genes are visualized (Langfelder and Horvath, 2008). Thus, the comparative preservation analysis of the modules between two coexpression networks, in this case of two different species, indicates the similarity between their transcritpomic "architectures" (Langfelder et al., 2011).

In this study, the WGCNA analysis was used to describe clusters of co-expressed genes in oocytes and blastocysts produced in vitro from buffalos and cattle, in order to indicate similarities in their transcriptomic profiles.

MATERIALS AND METHODS

Selection of RNAseq data for buffalos and bovine

The data of buffalo RNAseq of item 2 (item 5), and of bovine RNAseq of Graf et al. (2014). For the selection of bovine data, it was checked whether the IVEP conditions (see Hiendleder et al., 2004) and sequencing were similar as the buffalos (pools of oocytes and blastocysts), and if were aligned using the same reference genome (*Bos taurus* UMD3.1). The expression data normalized using the VST method to bovine oocytes and blastocysts produced *in vitro* were obtained from GEO platform (acession number GSE52415, file *GSE52415_VST_normalized.txt.gz*). The data of buffalo were obtained using the same normalization method in DESeq2 package (see Annex).

Building of co-expression networks

The WGCNA package in R programm (Langfelder and Horvath, 2008) was used to build independent co-expression networks for buffalo and bovine. Firstly, the adjacency matrices were builded using the soft-threshold of 20 and the adjacency function (Zhang and Horvath 2005). Then the adjacency matrices were used to calculate the similarity between the forces of coexpression, resulting in a topological overlap matrix (TOM).

The Dynamic Hybrid Tree Cut algorithm was used to define the branches of the clustering tree, that is the co-expression modules. The main components of each module (eigengene modules) were used to quantify the similarity between the expression profiles of the modules. Modules with very similar expression profiles (correlation of 0.75, default value) were joined and represented in a dendrogram.

The coexpression networks were analyzed for the correlation of the eingengenes modules and the stages of development. The modules that presented correlation greater than 0.9 and p-value < 0.05 were considered specific stages.

Preservation module statistics

To access the preservation of buffalo co-expression modules in bovine, was applied the modulePreservation function of the WGCNA package (Langfelder et al., 2011). As a result, the Z-summary value that indicates the preservation of the modules was calculated. In general, Z-summary value > 10 means a strong preservation, that is the buffalo and bovine modules are densely connected. The Z-summary value between 2 and 10 means moderate preservation and Z-summary < 2 means poor preservation.

Gene Ontology Analysis

The gene ontology analyzes of the coexpression modules were performed using the Bioconductor packages *(*https://www.bioconductor.org/packages/*), GO.db* (Carlson, 2017) and *AnnotationDBI* (Pages et al., 2017) *in* R programm.

Hub genes Identification and visualization

To identify genes with central role and highly connected within the modules, defined as hub genes, we calculated the correlation between each gene and the eingengene modules. Were considered hub genes only those genes that exhibited significant ($p < 0.05$) correlation (ModuleMembership > 0.9). The 150 strongest connections (intramodular connectivity > 0.7) were selected in the specific stage modules for visualization in the Cytoscape program (Shannon et al., 2003).

RESULTS

Modules of co-expressed genes of buffalo and bovine

To understand the co-expression relationships between buffalo and bovine genes, the analysis of weighted gene co-expression networks (WGCNA) was performed. Seven co-expression modules were identified in buffalo oocytes and blastocysts produced in vitro (Fig. 1A), of which 1 (brown module, see Fig.Supl.1) was significantly ($r = 0.97$, $p < 0.05$) correlated to embryos, therefore considered module stage specific. In the bovine counterpart, 27 modules were observed (Fig. 1B), of which only 1 (orangered3 module, see Fig.Supl.1) was significantly ($r = 1$, $p < 0.05$) related to embryos.

In this study, no specific buffalo oocyte modules were identified, but the number of samples used may have been a limiting factor for identification, since two modules (darkseagreen4 and sienna3) were significantly observed in the bovine counterpart ($r > 0.9$, $p < 0.05$) related to oocytes matured in vitro.

Preservation of co-expression modules

The preservation module uses a statistical test summarized in the Z-summary value, that indicates the preservation of the module in independent datasets. It was observed that of the 7 modules identified in buffalos, 4 indicated strong preservation (Z-summary > 10) and 3 indicated weak preservation (Z-summary < 2) in bovine. Considering the total number of genes analyzed in the mentioned modules (8,729), thus there is strong evidence that 75% of the buffalo transcriptome (6,560) was preserved in cattle (Fig. 2).

According to the gene ontology analyzes, the specific module of buffalo embryos that is preserved in bovine (brown), shares functional similarity with the bovine specific module (orangered3), since they were both related to the function of exosome components (Tab .Supl.1), indicating its importance in embryonic development. The other modules also strongly preserved were related to steroid metabolism, cell proliferation and morphogenesis, indicating that they are similarly regulated in buffalos and bovine. The low preserved modules were related to the cell cycle and amino acid transport, indicating that these functions are differentially regulated in the species studied (Fig. 2).

Hub genes highly connected in preserved modules

The *hub genes* characterize the expression profile of the module. The analyse of the hub genes composition in the specific modules of embryos (brown and orangered3) it was observed that 61% (1,451 hub genes) were different between buffalo and bovine, indicating that the expression profiles into the modules are different.

The strongest co-expression relationships within the specific module of blastocysts in buffalo (Fig. 3) and bovine (Fig. 4) shown the genes with central role into the module. The comparison of both graphs evidenced the lasgest difference in the co-expression relationships between the genes within the modules. Additionally, it was found that at least 11% of the hub genes of buffalo (268) and bovine (322) reported in this study were previously reported in bovine (Jiang et al., 2014), validating its importance for embryonic development even in different experimental conditions.

DISCUSSION

This study is pioneer in reporting the presence of co-expressed modules genes in matured oocytes and blastocyst produced in vitro from buffalos. In addition, we investigated the preservation of these modules in bovine coexpression network, with the goal of indicate important biological functions for the functioning of oocytes and embryos in these species. Comparison of buffalo and bovine transcriptomes is particularly important for In Vitro Embryo Production (IVEP), since these species are considered very related, however, it is unknown how much they are similar in the molecular aspect (Neglia et al., 2003). It is assumed that there are marked transcriptomic differences, as well as obvious morphological differences between oocytes and embryos between the species. In order to understand how much their transcriptomes are similar, the coexpression networks were builded, followed by the detection of the preserved modules.

It was possible to build the co-expression network for buffalos, although the amount of transcriptomic data in the specie is restricted (Abdoon et al., 2014; Kandil et al., 2010), thus the sample number of the WGCNA analysis was limited. This aspect consisted of a limitation for the detection of the specific stage modules in buffalo oocytes. However, for blastocysts, the sample number was not limiting, since only 1 specific stage was detected in both buffalo and bovine. The difference in the number of modules detected for the two species (7 versus 27) was probably related to the number of biological replicates in each dataset, which were higher for bovines compared to buffalos.

We have conserved functions such as cell proliferation and morphogenesis that are related to the processes of tissue differentiation and formation in the embryo (Basson, 2012), and directly correlated to steroid metabolism, also found strongly preserved. For, steroid biosynthesis, like cholesterol, is necessary for cell growth and division (Singh et al., 2013) and embryo development. In oocytes, steroid biosynthesis is related to production and response to hormones such as progesterone and estradiol, which are important inducers of the oocyte maturation process (Norman et al., 2004; Wang et al., 2014).

On the other hand, the cell cycle modules in buffalos were poorly preserved in bovine, explaining the differences in kinetics of oocyte nuclear maturation (Santos et al., 2002) and embryo development (Gasparrini et al., 2014) between the species. Likewise, the transmembrane amino acid transport module was poorly preserved and is related to the maintenance of cellular homeostasis (Zhang et al., 2017). Particularly, the transport of arginine is associated with the activation of cell signaling pathways (Rebsamen et al.,

2015) to promote cell growth and proliferation in pigs (Redel et al., 2015) and sheep (Kim et al., 2011). The poor preservation of these modules suggests that the related functions are regulated differently in buffalo and bovine. Therefore it may be important to investigate their roles in embryo development of buffalo.

The coexpression networks and preservation modules have already been used to correlate the transcriptomes of pre-implantation embryos in human and mice (Xue et al., 2013) and mice, human and bovine (Jiang et al., 2014), with the aim of comparing embryonic development programs. Strong conservation of the modules and their biological functions has been reported, showing that embryonic development is a well conserved program among these mammalian species. In this study, a great similarity was found between buffalo and bovine transcriptomes indicating that the embryonic development programs are strongly preserved in these two species as well.

However, the characterization of expression profile within the modules, throught the identification of the hub genes, indicated that the molecular differences between buffalo and bovine blastocysts produced in vitro may reside in the connections, or co-expression relationships between the genes, because about 60% of the hub genes in buffalo was different from bovine. In addition, the comparison of the hub genes identified in this study, with other previously reported (Jiang et al., 2014) corroborated with the identification of central genes of the specific module of embryos, leading to the belief that this study was able to identify them, despite of sample number limitations. On the other hand, it is necessary to increase the dataset to validate the hub genes in buffalos.

CONCLUSION

We conclude that the transcriptomes of matured oocytes and blastocysts produced in vitro from buffalo and bovine showed strong evidence of preservation indicating that their functions are similar, except in terms of development kinetics and amino acid transport. However, the expression profiles within the blastocyst modules showed greater dissimilarity, showing that even within conserved modules there are important expression differences between buffalo and bovine, which justify the elaboration of specific IVEP protocols for buffalos.

REFERENCES

Abdoon, A.S., Gabler, C., Holder, C., Kandil, O.M., Einspanier, R., 2014. Seasonal variations in developmental competence and relative abundance of gene transcripts in buffalo (Bubalus bubalis) oocytes. Theriogenology 82, 1055–1067. doi:10.1016/j.theriogenology.2014.07.008

Abdoon, A.S., Ghanem, N., Kandil, O.M., Gad, A., Schellander, K., Tesfaye, D., 2012. cDNA microarray analysis of gene expression in parthenotes and in vitro produced buffalo embryos. Theriogenology 77, 1240–1251. doi:http://dx.doi.org/10.1016/j.theriogenology.2011.11.004

Basson, M.A., 2012. Signaling in Cell Differentiation and Morphogenesis. Cold Spring Harb. Perspect. Biol. 4, a008151. doi:10.1101/cshperspect.a008151

Carlson M (2017). *GO.db: A set of annotation maps describing the entire Gene Ontology*. R package version 3.4.1.

Chitwood, J.L., Rincon, G., Kaiser, G.G., Medrano, J.F., Ross, P.J., 2013. RNA-seq analysis of single bovine blastocysts. BMC Genomics 14, 350. doi:10.1186/1471-2164-14-350

Di Francesco, S., Novoa, M.V.S., Vecchio, D., Neglia, G., Boccia, L., Campanile, G., Zicarelli, L., Gasparrini, B., 2012. Ovum pick-up and in vitro embryo production (OPU-IVEP) in Mediterranean Italian buffalo performed in different seasons. Theriogenology 77, 148–154. doi:10.1016/j.theriogenology.2011.07.028

Driver, A.M., Penagaricano, F., Huang, W., Ahmad, K.R., Hackbart, K.S., Witlbank, M.C., Khatib, H., 2012. RNA-Seq analysis uncovers transcriptomic variations between morphologically similar in vivo- and in vitro-derived bovine blastocysts. BMC Genomics 13, 118. doi:10.1186/1471-2164-13-118

Gasparrini, B., Neglia, G., Di Palo, R., Vecchio, D., Albero, G., Esposito, L., Campanile, G., Zicarelli, L., 2014. Influence of oocyte donor on in vitro embryo production in buffalo. Anim. Reprod. Sci. 144, 95–101. doi:10.1016/j.anireprosci.2013.11.010

Graf, A., Krebs, S., Zakhartchenko, V., Schwalb, B., Blum, H., Wolf, E., 2014. Fine mapping of genome activation in bovine embryos by RNA sequencing. Proc. Natl. Acad. Sci. U. S. A. 111, 1–6. doi:10.1073/pnas.1321569111

Hiendleder, S., Mund, C., Reichenbach, H.-D., Wenigerkind, H., Brem, G., Zakhartchenko, V., Lyko, F., Wolf, E., 2004. Tissue-Specific Elevated Genomic Cytosine Methylation Levels Are Associated with an Overgrowth Phenotype of Bovine Fetuses Derived by In Vitro Techniques1. Biol. Reprod. 71, 217–223.

Jiang, Z., Sun, J., Dong, H., Luo, O., Zheng, X., Obergfell, C., Tang, Y., Bi, J., O'Neill, R., Ruan, Y., Chen, J., Tian, X., 2014. Transcriptional profiles of bovine in vivo pre-implantation development. BMC Genomics 15, 756. doi:10.1186/1471-2164-15-756

Kandil, O.M., Ghanem, N., Abdoon, a S.S., Hölker, M., Phatsara, C., Schellander, K., Tesfaye, D., 2010. Transcriptional analysis of buffalo (Bubalus bubalis) oocytes during in vitro maturation using bovine cDNA microarray. Reprod. Domest. Anim. 45, 63–74. doi:10.1111/j.1439-0531.2008.01238.x

Kim, J.Y., Burghardt, R.C., Wu, G., Johnson, G.A., Spencer, T.E., Bazer, F.W., 2011. Select nutrients in the ovine uterine lumen. VII. Effects of arginine, leucine, glutamine, and glucose on trophectoderm cell signaling, proliferation, and migration. Biol Reprod 84, 62–69. doi:biolreprod.110.085738 [pii]10.1095/biolreprod.110.085738

Kropp, J., Khatib, H., 2015. mRNA fragments in in vitro culture media are associated with bovine preimplantation embryonic development. Front. Genet. 6, 273. doi:10.3389/fgene.2015.00273

Kropp, J., Salih, S.M., Khatib, H., 2014. Expression of microRNAs in bovine and human pre-implantation embryo culture media. Front. Genet. 5, 91. doi:10.3389/fgene.2014.00091

Kumar, P., Verma, a., Roy, B., Rajput, S., Ojha, S., Anand, S., Yadav, P., Arora, J., De, S., Goswami, S.L., Datta, T.K., 2012. Effect of Varying Glucose Concentrations during In Vitro Maturation and Embryo Culture on Efficiency of In Vitro Embryo Production in Buffalo. Reprod. Domest. Anim. 47, 269–273. doi:10.1111/j.1439-0531.2011.01849.x

Langfelder, P., Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559. doi:10.1186/1471-2105-9-559

Langfelder, P., Luo, R., Oldham, M.C., Horvath, S., 2011. Is My Network Module Preserved and Reproducible? PLOS Comput. Biol. 7, 1–29. doi:10.1371/journal.pcbi.1001057

Neglia, G., Gasparrini, B., Caracciolo Di Brienza, V., Di Palo, R., Campanile, G., Presicce, G.A., Zicarelli, L., 2003. Bovine and buffalo in vitro embryo production using oocytes derived from abattoir ovaries or collected by transvaginal follicle aspiration. Theriogenology 59, 1123–1130. doi:10.1016/S0093-691X(02)01170-6

Norman, A.W., Mizwicki, M.T., Norman, D.P.G., 2004. Steroid-hormone rapid actions, membrane receptors and a conformational ensemble model. Nat Rev Drug Discov 3, 27–41.

OHASHI OM, CORDEIRO MS E MIRANDA MS (2006) biotecnologia da reprodução aplicada a bubalino. Rev. Ciên. Agrárias, 45.

Oldham, M.C., Konopka, G., Iwamoto, K., Langfelder, P., Kato, T., Horvath, S., Geschwind, D.H., 2008. Functional organization of the transcriptome in human brain. Nat. Neurosci. 11. doi:10.1038/nn.2207

Pagès H, Carlson M, Falcon S and Li N (2017). *AnnotationDbi: Annotation Database Interface*. R package version 1.38.1.

Rebsamen, M., Pochini, L., Stasyk, T., de Araujo, M.E.G., Galluccio, M., Kandasamy, R.K., Snijder, B., Fauster, A., Rudashevskaya, E.L., Bruckner, M., Scorzoni, S., Filipek, P.A., Huber, K.V.M., Bigenzahn, J.W., Heinz, L.X., Kraft, C., Bennett, K.L., Indiveri, C., Huber, L.A., Superti-Furga, G., 2015. SLC38A9 is a component of the lysosomal amino acid sensing machinery that controls mTORC1. Nature 519, 477–481.

Redel, B.K., Tessanne, K.J., Spate, L.D., Murphy, C.N., Prather, R.S., 2015. Arginine increases development of in vitro-produced porcine embryos and affects the protein arginine methyltransferase–dimethylarginine dimethylaminohydrolase–nitric oxide axis. doi:10.1071/RD14293

Reyes, J.M., Chitwood, J.L., Ross, P.J., 2015. RNA-Seq profiling of single bovine oocyte transcript abundance and its modulation by cytoplasmic polyadenylation. Mol. Reprod. Dev. 82, 103–14. doi:10.1002/mrd.22445

Santos, S. do S.D., Dantas, J.K., Miranda, M. dos S., Biondi, F.C., Ohashi, O.M., 2002. Cinética da maturação nuclear in vitro de oócitos bubalinos. Brazilian J. Vet. Res. Anim. Sci. 39, 266–270. doi:10.1590/S1413-95962002000500009

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. 13. doi:10.1101/gr.1239303

Singh, P., Saxena, R., Srinivas, G., Pande, G., Chattopadhyay, A., 2013. Cholesterol Biosynthesis and Homeostasis in Regulation of the Cell Cycle. PLoS One 8, 1–12. doi:10.1371/journal.pone.0058833

Strazzullo, M., Gasparrini, B., Neglia, G., Balestrieri, M.L., Francioso, R., Rossetti, C., Nassa, G., De Filippo, M.R., Weisz, A., Di Francesco, S., Vecchio, D., D'Esposito, M., D'Occhio, M.J., Zicarelli, L., Campanile, G., 2014. Global transcriptome profiles of italian mediterranean buffalo embryos with normal and retarded growth. PLoS One 9, 1–9. doi:10.1371/journal.pone.0090027

Wang, X., Tsai, T., Qiao, J., Zhang, Z., Feng, H.L., 2014. Impact of gonadotropins on oocyte maturation, fertilisation and developmental competence in vitro. Reprod. Fertil. Dev. 26, 752–757.

Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C., Feng, Y., Liu, Z., Zeng, Q., Cheng, L., Sun, Y.E., Liu, J., Horvath, S., Fan, G., 2013. Genetic programs in human and mouse early embryos revealed by single-cell RNA[thinsp]sequencing. Nature 500, 593–597.

Zhang, B., Horvath, S., 2005. A General Framework for Weighted Gene Co-Expression Network Analysis. Stat. Appl. Genet. Mol. Biol. doi:10.2202/1544-6115.1128

Zhang, S., Zeng, X., Ren, M., Mao, X., Qiao, S., 2017. Novel metabolic and physiological functions of branched chain amino acids: a review. J. Anim. Sci. Biotechnol. 8, 10. doi:1s0.1186/s40104-016-0139-z

FIGURES

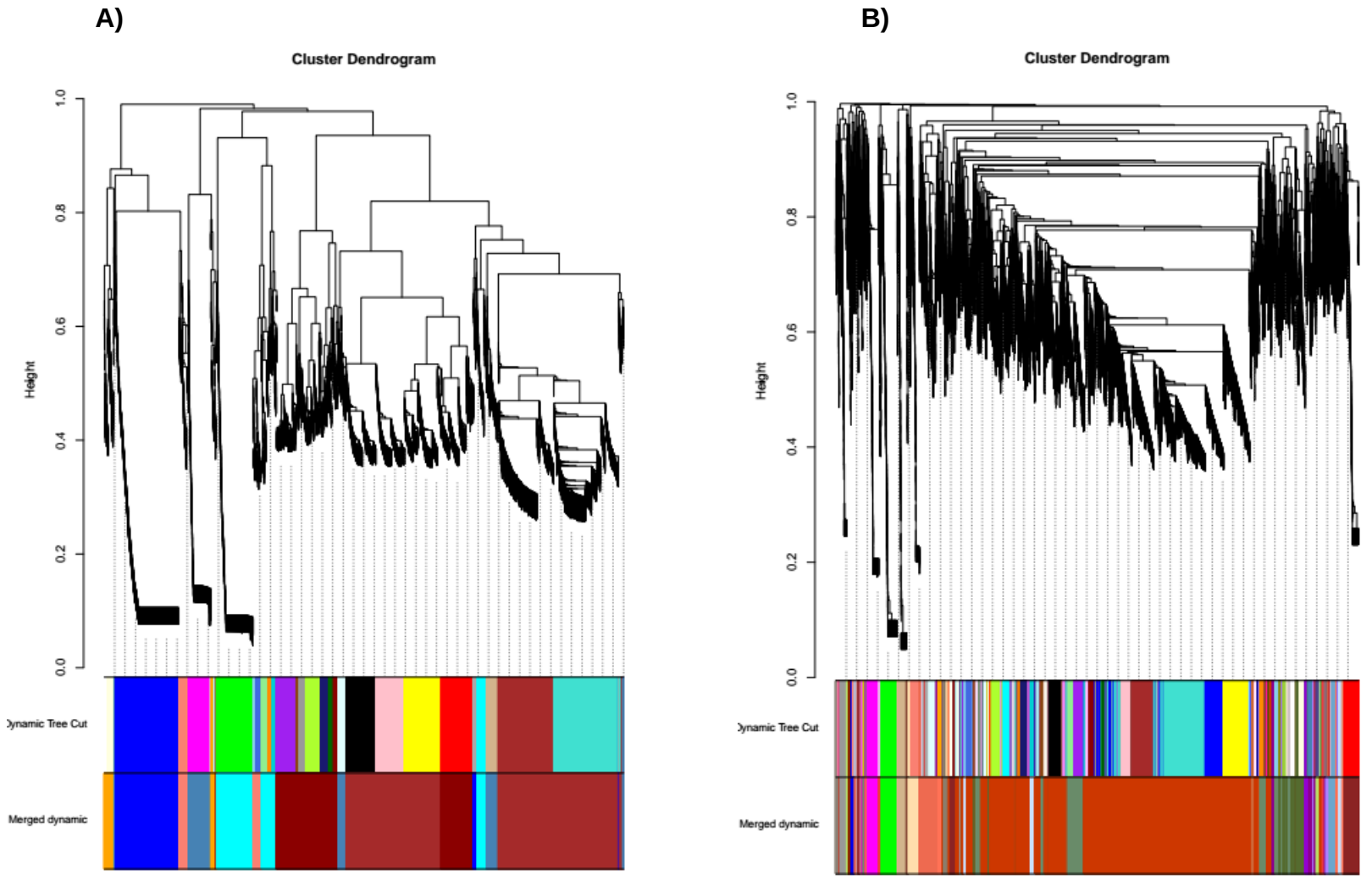
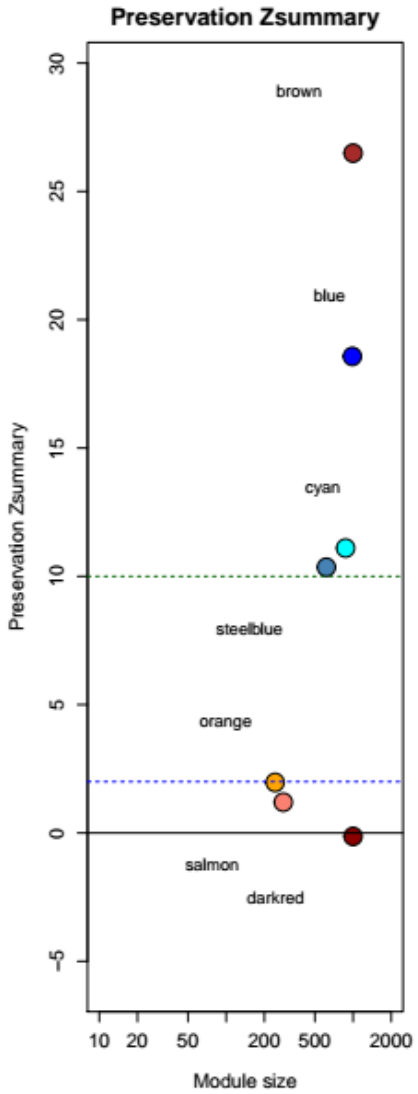


Figure 1. Dendogram of gene clustering in oocytes and embryos of buffalo (A) and bovine (B) with dissimilarity based on the topological overlap method (TOM), followed by the color code of the originally identified modules and the color codes of the joined modules based on similarity of eingengenes expression profile.



brown	**Componentes de exossomo**	GO:0003723 GO:0031988 GO:0065010 GO:0070062 GO:0044822 GO:0031982 GO:1903561
blue	**Metabolismo de esteróides**	GO:0008202 GO:0019218 GO:0006694 GO:0007186 GO:0003707
cyan	**Proliferação celular**	GO:0034341 GO:0071346 GO:0050877 GO:0010633 GO:0042102
steelblue	**Morfogênese**	GO:0055008 GO:0061384 GO:0060415 GO:0048644 GO:0004888 GO:0045746
orange	**Ciclo celular**	GO:0090068 GO:0072686 GO:0035064 GO:0045787 GO:0071158 GO:0007049
salmon	**Transporte de AA**	GO:0015825 GO:0015194 GO:0015293 GO:0042541 GO:0008028 GO:0015804
darkred	**Ciclo celular**	GO:0007067 GO:0000278 GO:0007049 GO:0051276 GO:0005819

Figure 2. Preservation of buffalo coexpression modules in bovine and their respective gene ontology results. If Z-summary > 10 there is strong evidence of preservation. If Z-summary < 2 indicates weak evidence of preservation.

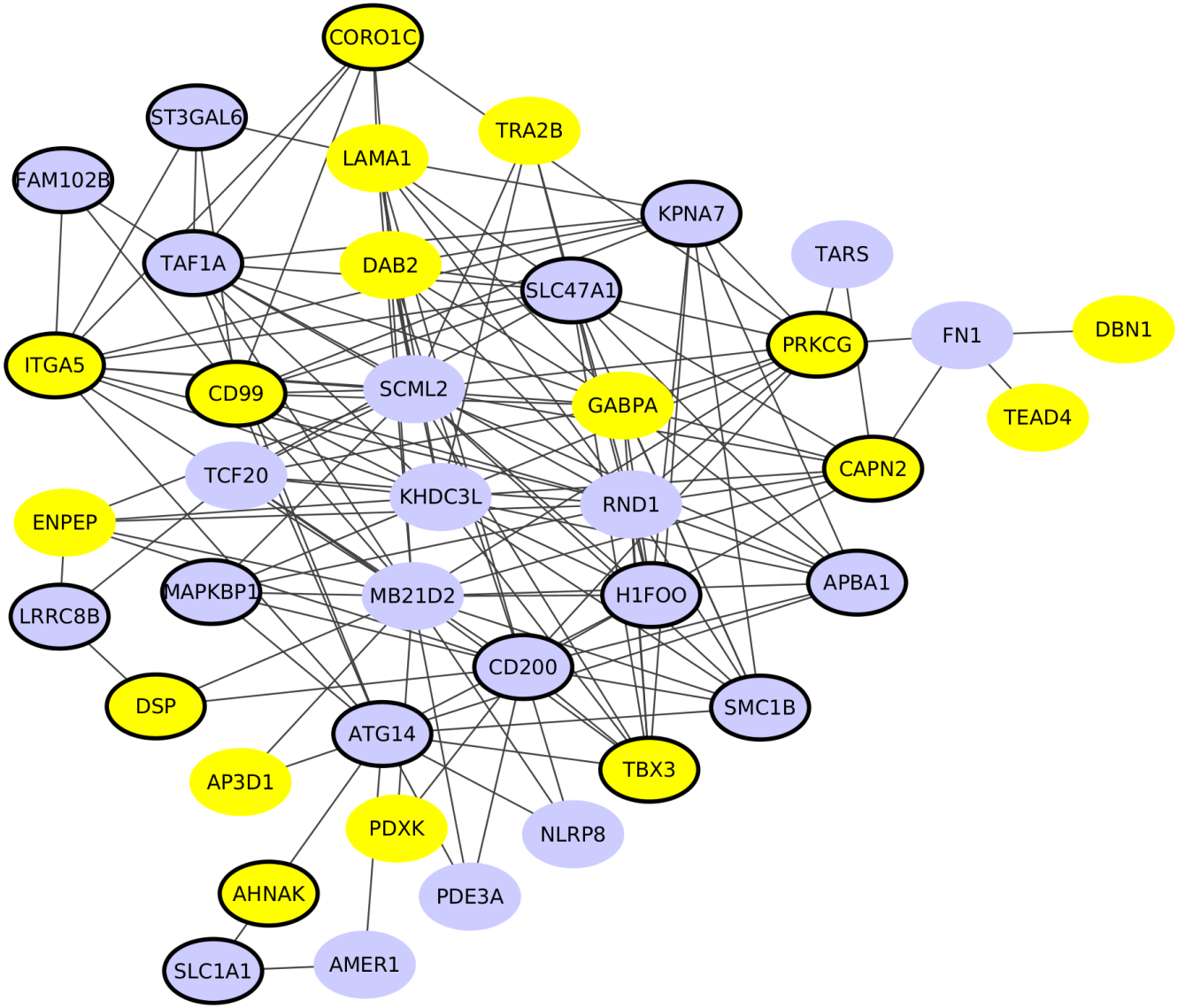


Figure 3. Visualization of the 150 strongest connections in the buffalo blastocyst specific module (brown module). The connections between the hub genes (yellow boxes) and other connected genes (blue boxes) were shown. Genes circled in bold were validated as hub genes in previou study in bovine (Jiang et al., 2014).

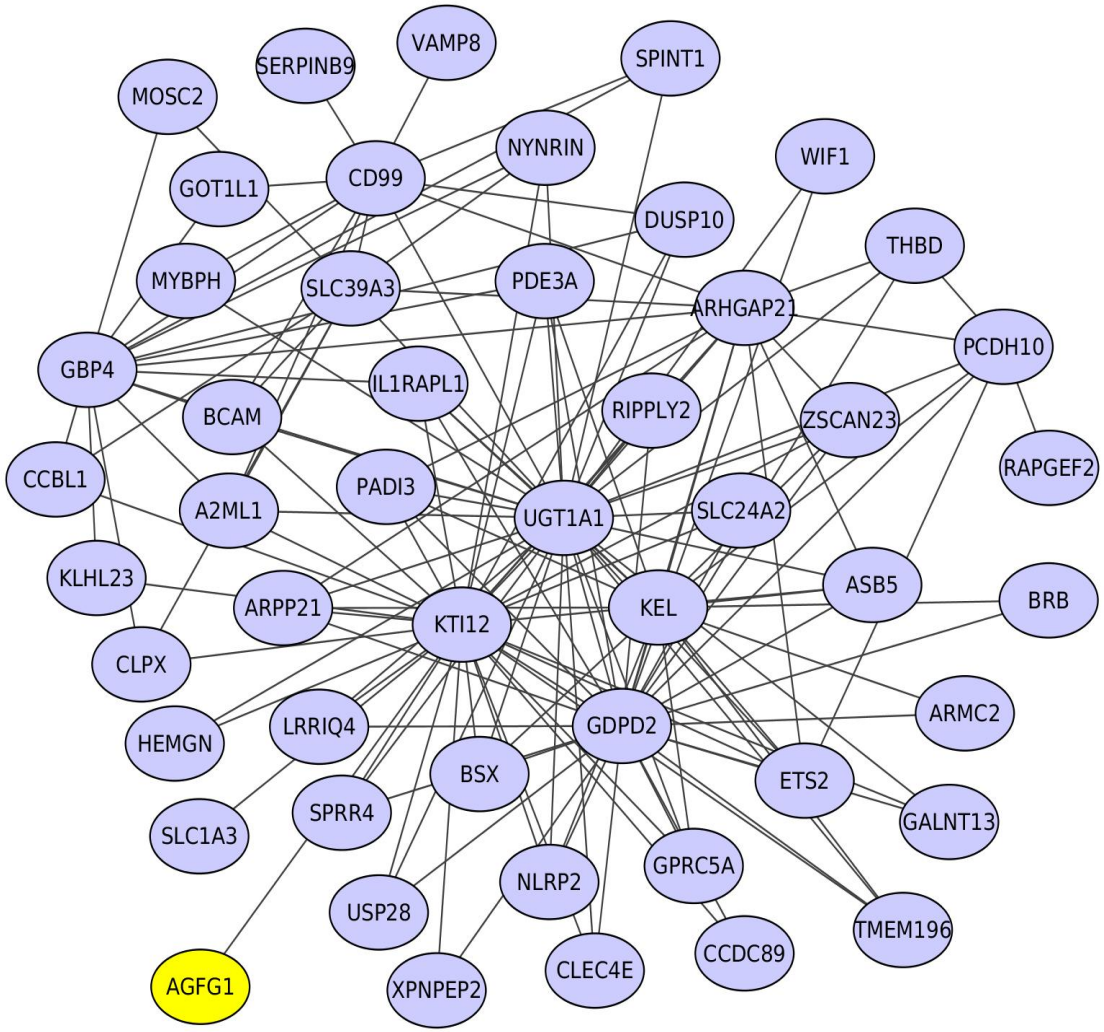
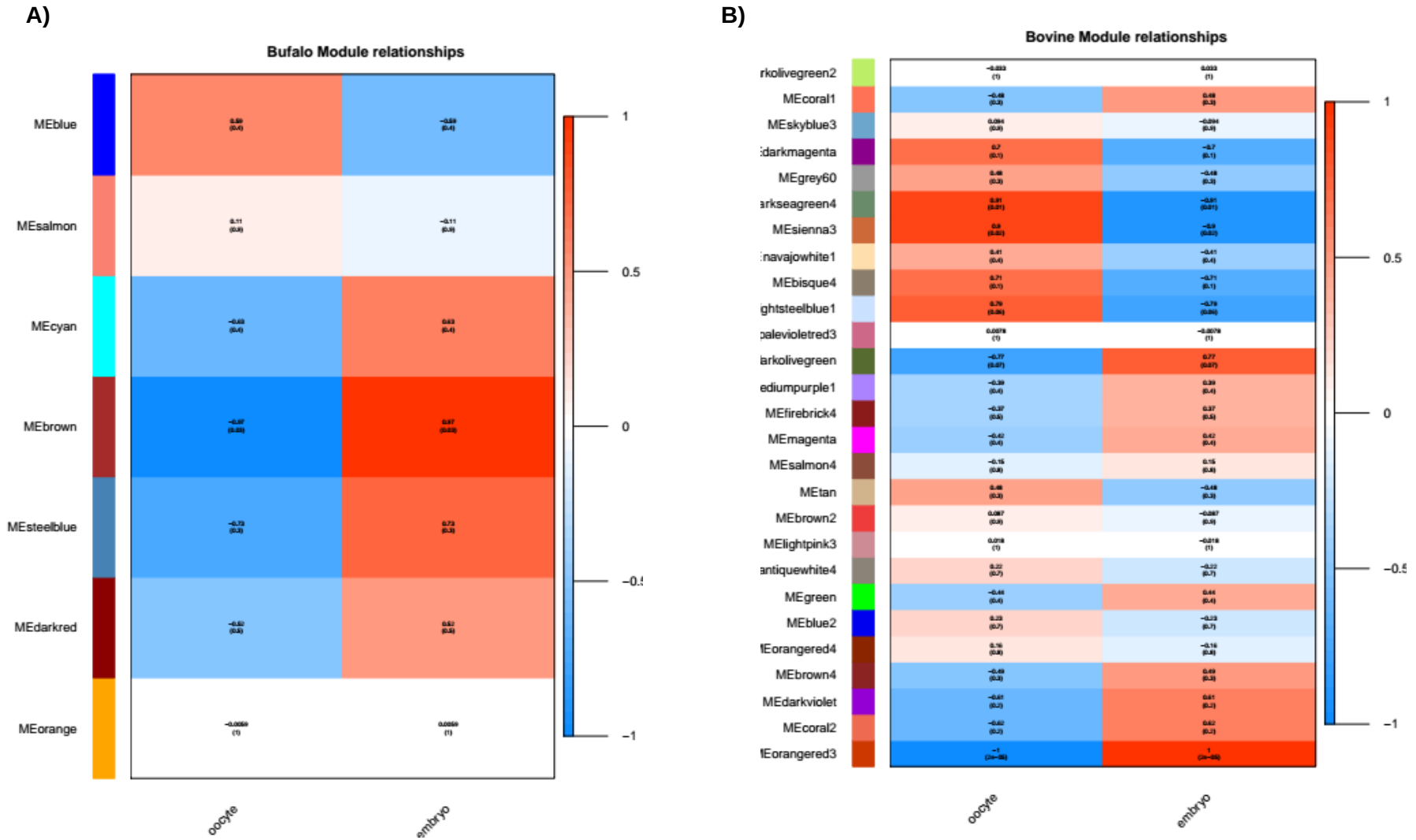


Figure 4. Visualization of the 150 strongest connections in the bovine blastocyst specific module (orangered3 module). The connections between the hub genes (yellow box) and other connected genes (blue boxes) were shown.

SUPPLEMENTARY FIGURE



Supplementary Figure 1. Heatmaps illustrating the correlations (and corresponding p-values) between the modules and stages of development (oocyte and embryo) in buffalo (A) and bovine (B). Each module is represented by eingengene (main component). Modules (lines) can be highly correlated (induced expression) at different stages (columns). The color legend indicates the level of correlation (induced or repressed) between the co-expression modules and the stages.

SUPPLEMENTARY TABLE

Table 1: Gene ontology analysis of the stage-specific modules of buffalo and bovine blastocysts.

Specie/Stage	Module	* enrichmentP	GO term ID	Term name
BUFFALO – EMBRYO	brown	0,0000001	GO:0003723	RNA binding
	brown	0,0000003	GO:0005737	cytoplasm
	brown	0,0000004	GO:0031988	membrane-bounded vesicle
	brown	0,0000006	GO:0065010	extracellular membrane-bounded organelle
	brown	0,0000006	GO:0070062	extracellular exosome
	brown	0,0000008	GO:0044822	poly(A) RNA binding
	brown	0,0000012	GO:0031982	vesicle
	brown	0,0000012	GO:1903561	extracellular vesicle
	brown	0,0000058	GO:0005622	intracellular
	brown	0,0000104	GO:0043229	intracellular organelle
BOVINE – EMBRYO	orangered3	0,0005850	GO:1903561	extracellular vesicle
	orangered3	0,0005903	GO:0065010	extracellular membrane-bounded organelle
	orangered3	0,0005903	GO:0070062	extracellular exosome
	orangered3	0,0009551	GO:0022625	cytosolic large ribosomal subunit
	orangered3	0,0010154	GO:0031982	vesicle
	orangered3	0,0015337	GO:0031988	membrane-bounded vesicle
	orangered3	0,0017703	GO:0030117	membrane coat
	orangered3	0,0023516	GO:0043229	intracellular organelle
	orangered3	0,0030978	GO:0016810	hydrolase activity
	orangered3	0,0032777	GO:0043231	intracellular membrane-bounded organelle

* *enrichmentP* means the p-value calculated by the *GoenrichmentAnalysis* function in WGCNA package

6.1. ANEXO DO CAPÍTULO 3

METODOLOGIA DAS ANÁLISES DE BIOINFORMÁTICA

O capítulo 3 consiste em comparar o perfil transcriptômico de oócitos e embriões de búfalos *versus* bovinos, usando as análises de redes de co-expressão e preservação de módulos, que foram descritas detalhamente neste anexo. Os tópicos seguem a sequência dos experimentos e contém uma breve introdução, as linhas de comando e alguns resultados parciais não mostrados no capítulo 3.

SUMÁRIO

INTRODUÇÃO

A análise de rede de co-expressão gênica é um método de biologia de sistemas usado para descrever os padrões de correlação entre um número muito grande de genes (Peter Langfelder and Horvath 2008). Neste trabalho foi utilizada para encontrar módulos de co-expressão, para correlaciona-los com as amostras (oócitos e embriões de búfalo), e por fim para definir a preservação dos módulos de búfalos em bovinos. A seguir, é mostrada uma lista dos termos usados nesta metodologia e suas respectivas definições.

Termo	**Definição**
WGCNA	*Weighted correlation network analysis* (WGCNA) é uma análise de biologia de sistemas e também um pacote do programa R com a função de construir redes de co-expressão entre outras
Rede de co-expressão	São grafos não direcionados nos quais os vértices correspondem a genes, cada par de vértices é conectado por arestas que indicam a relação de co-expressão entre os genes.
Módulo de co-expressão	São clusters de genes altamente conectados entre si, ou seja, grupos de genes com alta relação de co-expressão
Matriz de adjacência	Definida como uma matriz simétrica cujas entradas representam as forças de conexão (co-expressão) entre os nodos (genes)
Conectividade	Para cada gene a conectividade é definida como a soma das forças das conexões com os demais genes da rede.
Conectividade intramodular	A conectividade intramodular é medida pela conectividade dos genes dentro de um módulo específico.
Módulos Eigengene	Definido como o primeiro componente principal de um determinado módulo. Pode ser considerado a representação do perfil de expressão gênica daquele módulo.
Hub gene	Definido como o gene que possui um alto número de conexões.

1. OBTENÇÃO DOS ARQUIVOS DE ENTRADA DO WGCNA

O pacote WGCNA (*Weighted correlation network analysis*) do programa R foi usado para encontrar módulos, ou grupos, de genes preservados entre oócitos e embriões de búfalos e bovinos (Peter Langfelder et al. 2011). Foram usados os dados de RNAseq de búfalo gerados no Capítulo 2, e os de RNAseq bovino (Graf et al. 2014) obtidos da plataforma GEO (*Gene expression Omnibus,* número de acesso GSE52415).

Foi certificado que os dois conjuntos de dados, búfalos e bovinos, foram alinhados usando o genoma de referência *Bos taurus UMD.3.1.*, com realização de contagem de *reads* no Htseq e a normalização dos dados de expressão no DESeq2. Foi adotado o método de transformação VST (*Variance Stabilizing Transformation*), que é indicado para análises de clusterização (http://www.bioconductor.org/help/workflows/rnaseqGene/).

A tabela com os dados de expressão normalizados por método VST de bovinos foi obtida da plataforma GEO (GSE52415_VST_normalized.txt.gz), e a de búfalos foi gerada como descrito no Capítulo 2. A seguir é mostrado o arquivo de saída do DESeq2.

```
library("DESeq2")
countData <- as.matrix(read.csv(file="countData.csv",sep=",",row.names="gene_id"))
colData <- read.csv(file="colData.csv",sep=",",row.names=1)
colData <- colData[,c("condition","type")]
head(countData)
head(colData)
countData <- countData[, rownames(colData)]
all(rownames(colData) == colnames(countData))
dds <- DESeqDataSetFromMatrix(countData = countData,
                              colData = colData,
                              design = ~ condition)
dds
dds$condition <- factor(dds$condition, levels=c("oocyte","embryo"))
dds <- DESeq(dds)
nrow(dds)
dds <- dds[ rowSums(counts(dds)) > 1, ]
nrow(dds)
vsd <- varianceStabilizingTransformation(dds, blind=FALSE)
head(assay(vsd), 3)
data <- as.data.frame(assay(vsd))
write.csv(data, file= "teste.csv")
```

O Htseq e DESeq2 utilizam o "*ensembl_id*" como identificador dos genes, por isso o arquivo de saída acima foi manipulado para converter em "*gene_symbol*", usando a ferramenta biomaRt, conforme descrito anteriormente no "Anexo do capítulo2" (item 11.1).

Esta conversão foi necessária, pois as tabelas foram comparadas em análise posterior. Como búfalos e bovinos foram alinhados com o mesmo genoma de referência, então os *gene_symbol* foram correspondentes. Os cabeçalhos das tabelas com os dados de expressão de búfalo (A) e bovino (B) são mostrados a seguir.

A)

Gene	*Buf_oocyte1*	*Buf_oocyte2*	*Buf_embryo1*	*Buf_embryo2*
A1BG	*4.37835926957161*	*4.37835926957161*	*4.37835926957161*	*5.10836311890393*
A1CF	*4.37835926957161*	*4.37835926957161*	*4.37835926957161*	*4.37835926957161*
A2M	*4.37835926957161*	*4.37835926957161*	*7.81884349589515*	*6.90173345091496*
A3GALT2	*4.37835926957161*	*4.37835926957161*	*4.37835926957161*	*4.37835926957161*
A4GALT	*4.37835926957161*	*4.37835926957161*	*6.80729787364605*	*7.82882560532151*
A4GNT	*4.37835926957161*	*4.37835926957161*	*4.37835926957161*	*4.37835926957161*
AAAS	*5.69584681756271*	*4.37835926957161*	*6.25252416898603*	*6.06300012493195*

B)

Gene	*Bov_oocyte1*	*Bov_oocyte2*	*Bov_oocyte3*	*Bov_blastocyst1*	*Bov_blastocyst2*	*Bov_blastocyst3*
20ALPHA-HSD	*1.09708741270023*	*1.09708741270023*	*1.09708741270023*	*1.09708741270023*	*2.10863279144669*	*1.09708741270023*
5htr2c	*1.09708741270023*	*1.09708741270023*	*1.09708741270023*	*1.09708741270023*	*3.0170265509284*	*2.42244302990797*
A1BG	*1.09708741270023*	*1.09708741270023*	*1.09708741270023*	*1.09708741270023*	*1.09708741270023*	*1.09708741270023*
A2LD1	*1.09708741270023*	*1.09708741270023*	*3.64329626263893*	*7.93197370197891*	*7.30025349371749*	*7.19517646880733*
A2M	*1.99415343818173*	*2.43508091720697*	*1.09708741270023*	*10.0589619739727*	*10.8388159428059*	*10.9854062305407*
A2ML1	*1.09708741270023*	*1.09708741270023*	*1.09708741270023*	*6.88527834828485*	*6.74428552685041*	*6.6165843306564*
A4GNT	*6.84887968897*	*6.42438398379138*	*5.39005824169158*	*1.09708741270023*	*1.09708741270023*	*1.09708741270023*

2. PRÉ-TRATAMENTO DE CADA CONJUNTO DE DADOS (BÚFALOS E BOVINOS)

As tabelas de búfalos e bovinos foram trabalhadas separadamente. Brevemente: 1) foram retirados os genes “missing values”; 2) as amostras mais similares entre si foram agrupadas ou clusterizadas, resultando em um dendograma para cada espécie; 3) os dados pré-tratados foram armazenados na variável *datExpr;* 4) foi formado um *dataframe* (*datTraits*) com o *datExpr* e as informações biológicas das amostras. Por fim, os dados pré-tratados foram salvos (*_dataInput.Rdata).

```
library("WGCNA")
options(stringsAsFactors = FALSE);
BufData = read.csv("VSD_name.csv", header=TRUE)
dim(BufData)
names(BufData)
datExpr0 = as.data.frame(t(BufData[, -1]))
names(datExpr0) = BufData$Gene
rownames(datExpr0) = names(BufData)[-1]
```

```
names(datExpr0)
gsg = goodSamplesGenes(datExpr0, verbose = 3);
gsg$allOK
if (!gsg$allOK)
{
  if (sum(!gsg$goodGenes)>0)
    printFlush(paste("Removing genes:", paste(names(datExpr0)[!gsg$goodGenes], collapse = ",
")))
  if (sum(!gsg$goodSamples)>0)
    printFlush(paste("Removing samples:", paste(rownames(datExpr0)[!gsg$goodSamples],
collapse = ", ")));
  datExpr0 = datExpr0[gsg$goodSamples, gsg$goodGenes]
}
dim(datExpr0)
sampleTree = hclust(dist(datExpr0), method = "average")
sizeGrWindow(12,9)
par(cex = 0.6)
par(mar = c(0,4,2,0))
plot(sampleTree, main = "Sample clustering to detect outliers", sub="", xlab="", cex.lab = 1.5,
     cex.axis = 1.5, cex.main = 2)
abline(h = 130, col = "red");
clust = cutreeStatic(sampleTree, cutHeight = 130, minSize = 2)
table(clust)
keepSamples = (clust==c(1,1,2,2))
datExpr = datExpr0[keepSamples, ]
dim(datExpr)
nGenes = ncol(datExpr)
nSamples = nrow(datExpr)
traitData = read.csv("colData.csv", header=TRUE)
dim(traitData)
head(traitData)
names(traitData)
bufalosamples = rownames(datExpr)
traitRows = match(bufalosamples, traitData$cell)
datTraits = traitData[traitRows, -1]
rownames(datTraits) = colData[traitRows, 1]
collectGarbage()
sampleTree2 = hclust(dist(datExpr), method = "average")
traitColors = numbers2colors(datTraits, signed = FALSE)
plotDendroAndColors(sampleTree2, traitColors,
                    groupLabels = names(datTraits),
                    main = "Sample dendrogram and trait heatmap")
save(datExpr, datTraits, file = "Bufalo_dataInput.RData")
```

Esta etapa foi realizada conforme as instruções do manual do WGCNA no link https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/Tutorials O aspecto dos dendogramas de búfalos (*Bubalus bubalis*) e bovinos (*Bos taurus*) são mostrados a seguir. Observou-se que oócítos e embriões formam “clusters” diferentes nas duas espécies, indicando que os seus perfis de expressão são diferentes.

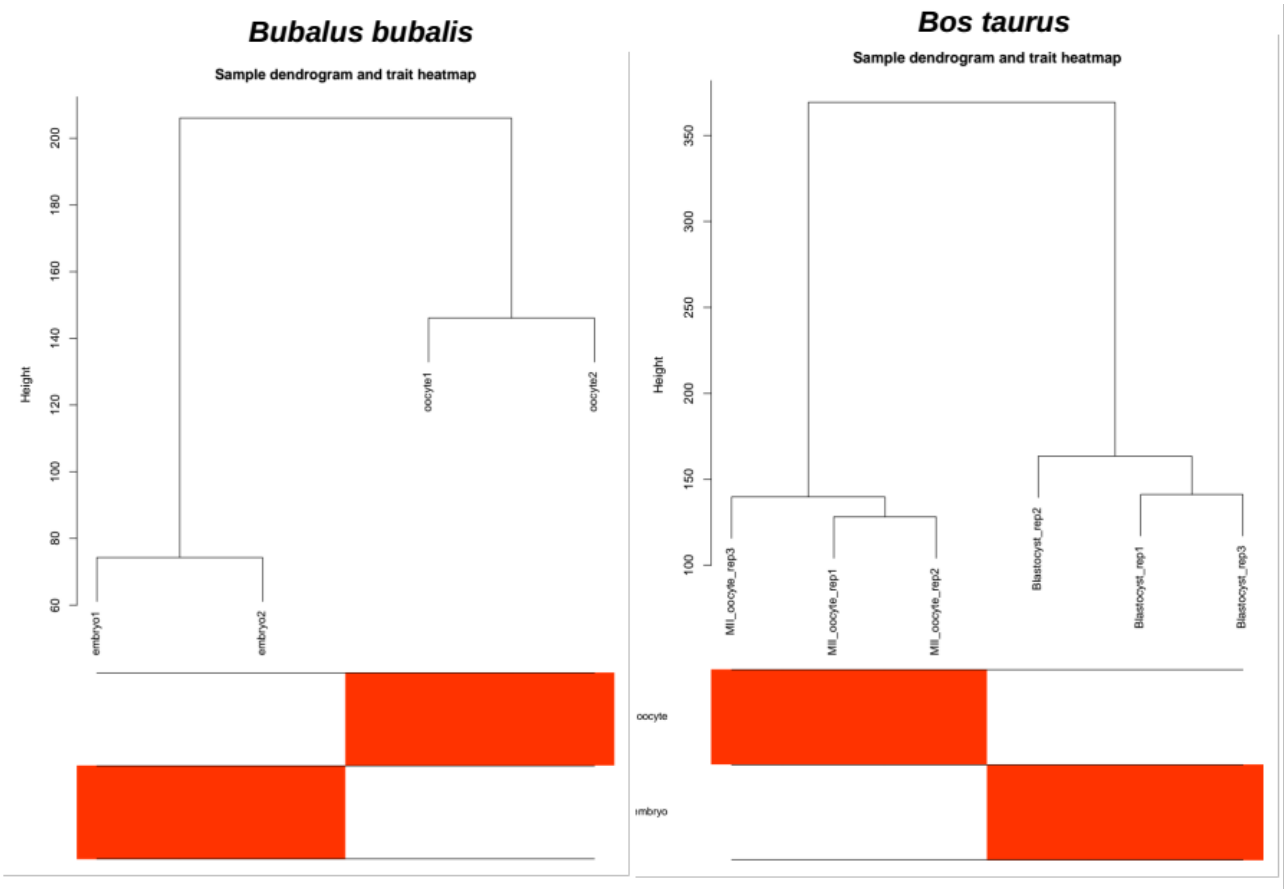


Figura 1. Dendogramas mostrando a análise de clusterização de oócitos e embriões de BÚFALOS (*Bubalus bubalis*) e BOVINOS (*Bos taurus*).

3. CONSTRUÇÃO DAS REDES DE CO-EXPRESSÃO E DETECÇÃO DOS MÓDULOS

Os dados pré-tratados obtidos na etapa anterior (*_dataInput.Rdata) dos clusters de oócitos e embriões, foram usados para construir redes de co-expressão independentes de búfalos e bovinos. Optou-se pelo tutorial de construção passo-a-passo da rede em https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/Tutorials/FemaleLiver-02-networkConstr-man.pdf.

3.1. Escolha do *soft-threshold* e construção da matriz de adjacência

Primeiramente, foi escolhido o *soft-threshold,* ou índice de topologia livre, que será usado para construir a matriz de adjacência. Este índice auxilia no próximo passo da construção da rede de co-expressão, que é a conversão dos dados de expressão em medidas de correlação (correlação de Pearson), em seguida, em forças de conexão entre os genes através de uma função de adjacência. A função *pickSoftThreshold* calcula o

índice de topologia livre e resulta em um gráfico, sendo indicada a escolha do menor índice que apresenta saturação da curva de topologia livre (B Zhang and Horvath 2005).

Como foi observada uma tendência de saturação da curva no índice de 20, para búfalos e bovinos, então foi escolhido para a construção das matrizes. Além disso, o índice escolhido coincidiu com o recomendado para grupos de dados pequenos https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/faq.html. As linhas de comando e os gráficos de índice de topologia livre são mostrados a seguir.

```
library("WGCNA")
options(stringsAsFactors = FALSE)
enableWGCNAThreads()
lnames = load(file ="Bufalo_dataInput.RData")
powers = c(c(1:10), seq(from = 12, to=20, by=2))
sft = pickSoftThreshold(datExpr, powerVector = powers, verbose = 5)
sizeGrWindow(1, 0.20)
par(mfrow = c(1,2))
cex1 = 0.9
plot(sft$fitIndices[,1], -sign(sft$fitIndices[,3])*sft$fitIndices[,2],
    xlab="Soft Threshold (power)",ylab="Scale Free Topology Model Fit,signed R^2",type="n",
    main = paste("Scale independence"))
text(sft$fitIndices[,1], -sign(sft$fitIndices[,3])*sft$fitIndices[,2],
    labels=powers,cex=cex1,col="red")
abline(h=0.,col="red")
plot(sft$fitIndices[,1], sft$fitIndices[,5],
    xlab="Soft Threshold (power)",ylab="Mean Connectivity", type="n",
    main = paste("Mean connectivity"))
text(sft$fitIndices[,1], sft$fitIndices[,5], labels=powers, cex=cex1,col="red")
```

A) *B)*

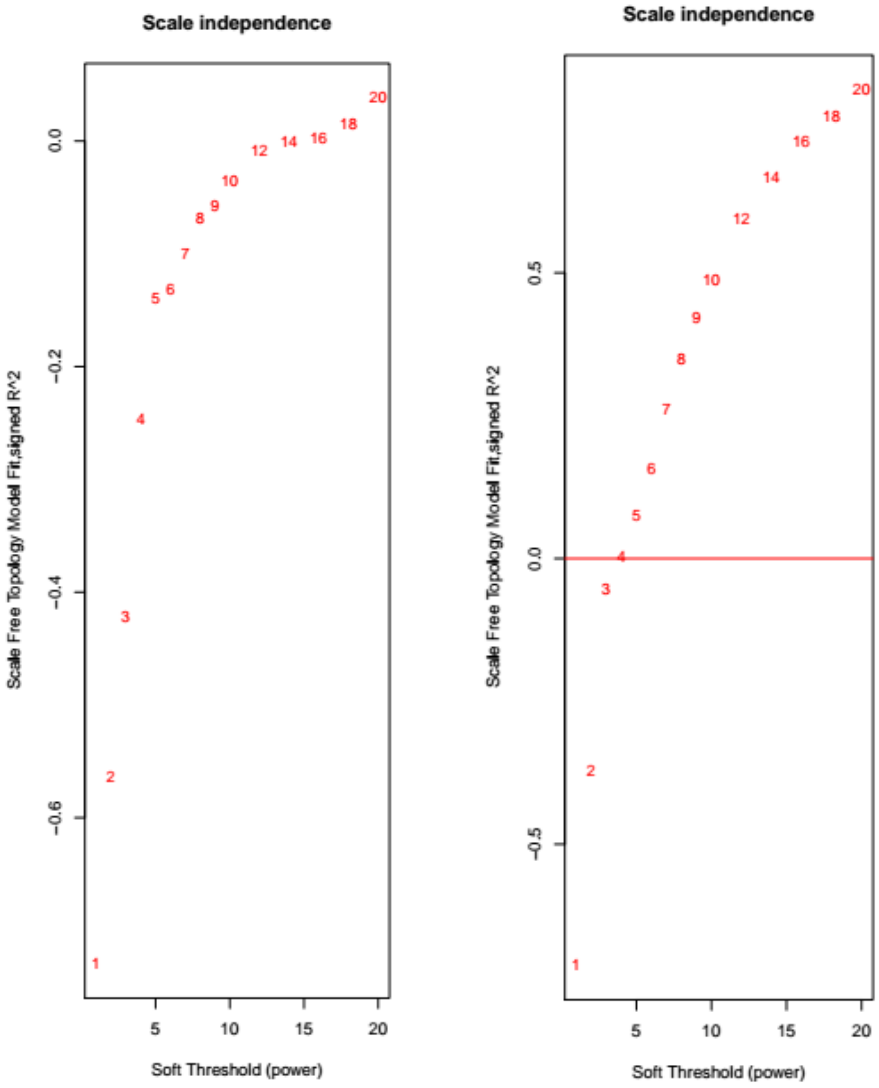


Figura 2. Gráficos de escala de topologia livre em BÚFALOS (A) e BOVINOS (B)

3.2. Construção da rede usando matriz de sobreposição topológica (TOM)

Foi construída a matriz de adjacência usando o *soft threshold* de 20 e a função *adjacency* do pacote R, essa matriz contêm as forças de co-expressão entre os genes, diz-se portanto, que é ponderada (*weighted matrix*).

O próximo passo, é usar esta matriz para calcular a medida de similaridade entre os genes (vértices do grafo). Essas medidas, são usadas para formar os clusters de genes, agrupando-os de acordo com a similaridade entre as forças das interações. Estes cálculos resultam em uma matriz de sobreposição topológica (TOM) que contêm as medidas de similaridade entre as conexões dos vértices do grafo. São executados pela função *TOMsimilarity* do WGCNA no programa R (Langfelder e Horvath, 2008).

Dessa forma, os genes que apresentaram forças de co-expressão similares foram agrupados mais próximos na representação do dendograma, também chamada de árvore de clusterização hierárquica de genes. As linhas de comando e as árvores estão abaixo.

```
softPower = 20
adjacency = adjacency(datExpr, power = softPower)
TOM = TOMsimilarity(adjacency)
dissTOM = 1-TOM
geneTree = hclust(as.dist(dissTOM), method = "average")
sizeGrWindow(12,9)
plot(geneTree, xlab="", sub="", main = "Gene clustering on TOM-based dissimilarity",
    labels = FALSE, hang = 0.04)
```

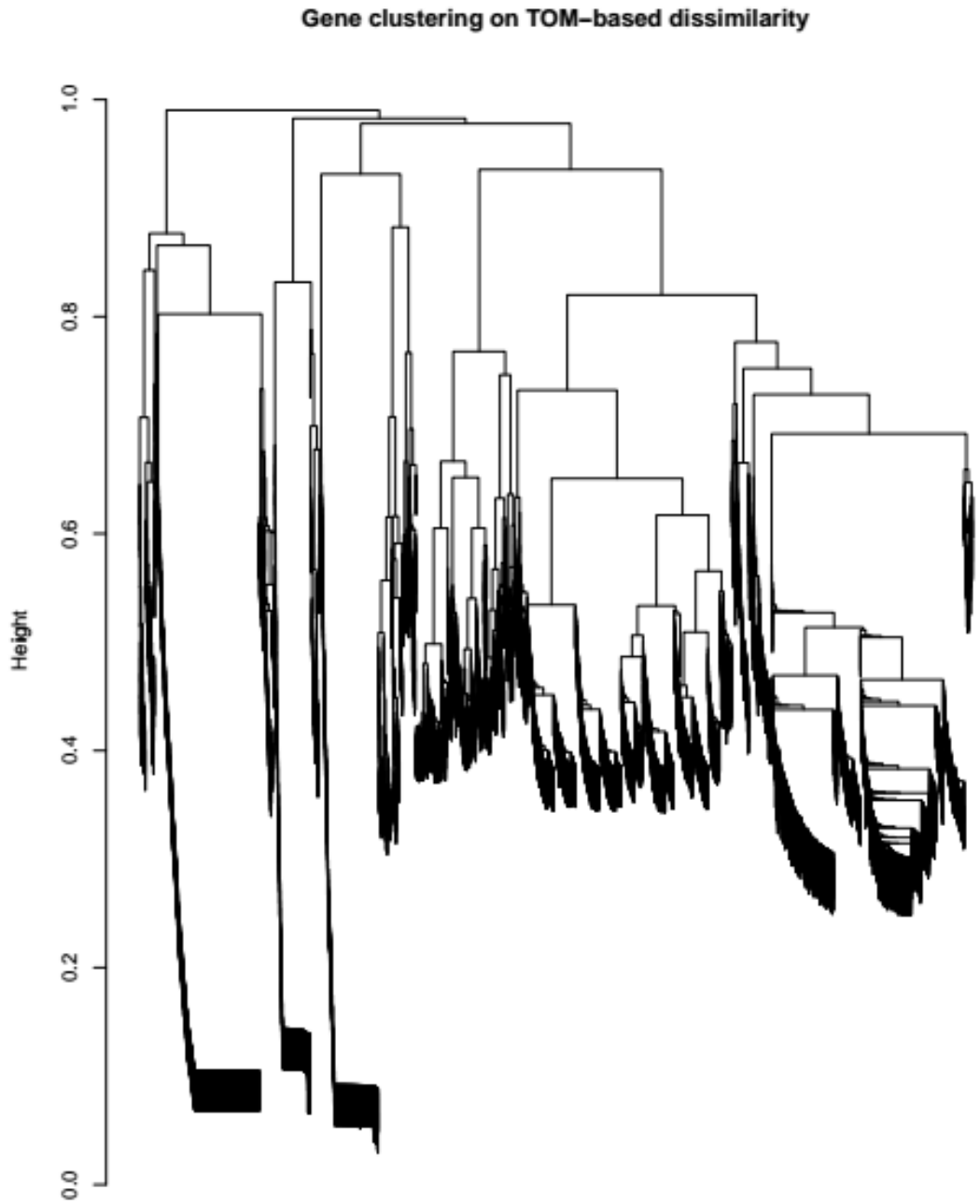


Figura 3. Dendograma ou árvore de genes co-expressos, com dissimilaridade baseada em método de sobreposição topológica (TOM), em BÚFALOS.

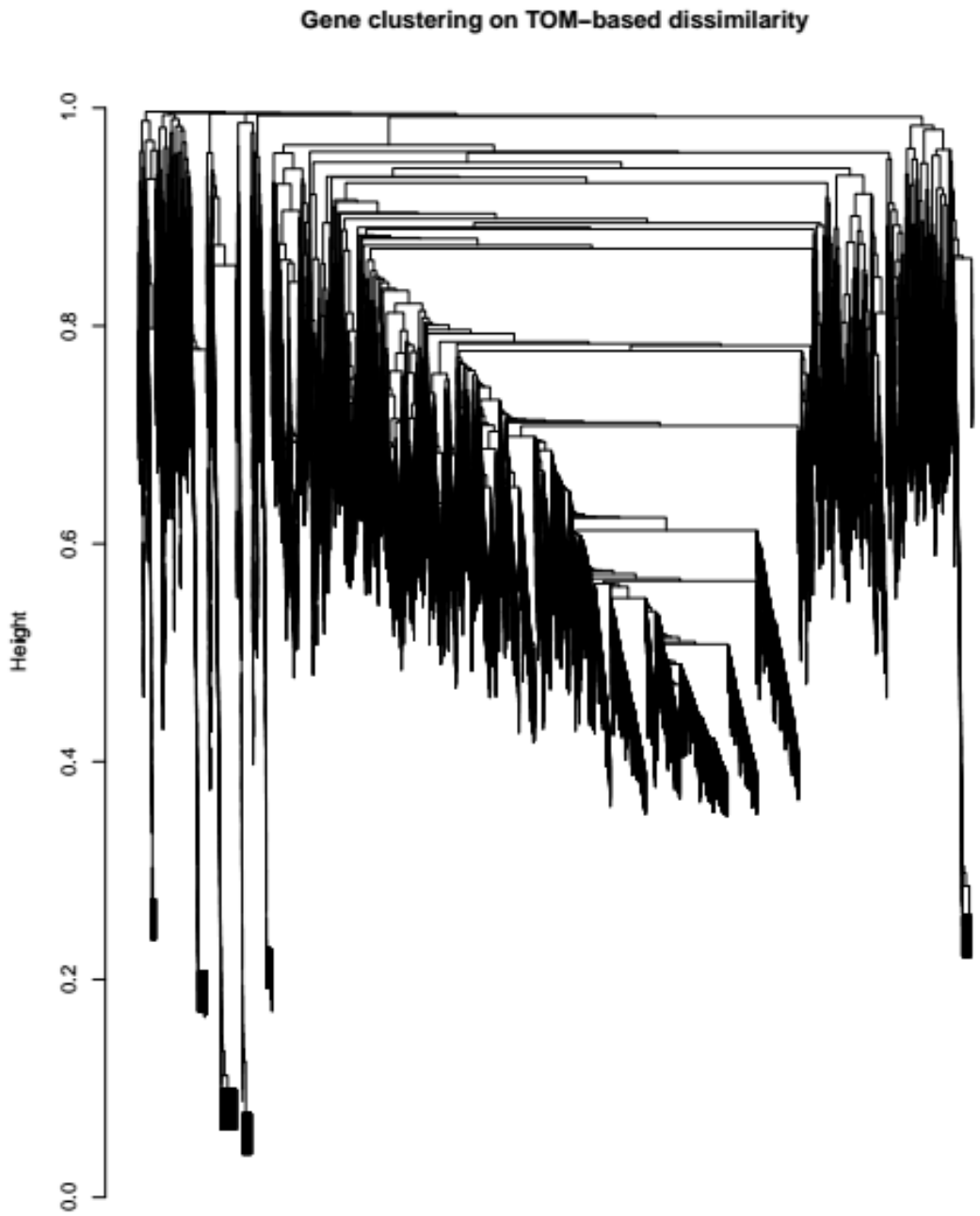


Figura 4. Dendograma ou árvore de genes co-expressos, com dissimilaridade baseada em método de sobreposição topológica (TOM), em BOVINOS.

3.3. Detecção dos módulos de genes co-expressos

Em seguida foi usado o algoritmo *Dynamic Hybrid Tree Cut* para definir os ramos da árvore de clusterização, ou seja, definir os módulos de genes co-expressos. Para, simplificar, os módulos com perfil de expressão muito similar (correlação de 0.75, valor *default*) foram unidos. Para quantificar esta similaridade de expressão entre os módulos, foram calculados os módulos *eigengenes,* definidos como o componente principal de cada módulo, ou a representação de seu perfil de expressão.

A seguir são mostradas as linhas de comando e os dendogramas ilustrando os módulos originalmente identificados e após a união dos módulos mais similares entre si.

```
minModuleSize = 30
dynamicMods = cutreeDynamic(dendro = geneTree, distM = dissTOM,
                        deepSplit = 2, pamRespectsDendro = FALSE,
                        minClusterSize = minModuleSize)
table(dynamicMods)
dynamicColors = labels2colors(dynamicMods)
table(dynamicColors)
sizeGrWindow(8,6)
plotDendroAndColors(geneTree, dynamicColors, "Dynamic Tree Cut",
                dendroLabels = FALSE, hang = 0.03,
                addGuide = TRUE, guideHang = 0.05,
                main = "Gene dendrogram and module colors")
MEList = moduleEigengenes(datExpr, colors = dynamicColors)
MEs = MEList$eigengenes
MEDiss = 1-cor(MEs)
METree = hclust(as.dist(MEDiss), method = "average")
sizeGrWindow(7, 6)
plot(METree, main = "Clustering of module eigengenes",
    xlab = "", sub = "")
MEDissThres = 0.25
abline(h=MEDissThres, col = "red")
merge = mergeCloseModules(datExpr, dynamicColors, cutHeight = MEDissThres, verbose = 3)
mergedColors = merge$colors
mergedMEs = merge$newMEs
sizeGrWindow(12, 9)
plotDendroAndColors(geneTree, cbind(dynamicColors, mergedColors),
                c("Dynamic Tree Cut", "Merged dynamic"),
                dendroLabels = FALSE, hang = 0.03,
                addGuide = TRUE, guideHang = 0.05)
bufalomoduleColors = mergedColors
colorOrder = c("grey", standardColors(50))
bufalomoduleLabels = match(bufalomoduleColors, colorOrder)-1
MEs = mergedMEs
save(MEs, bufalomoduleLabels, bufalomoduleColors, geneTree, file = "Bufalo-
networkConstruction-stepByStep.Rdata")
```

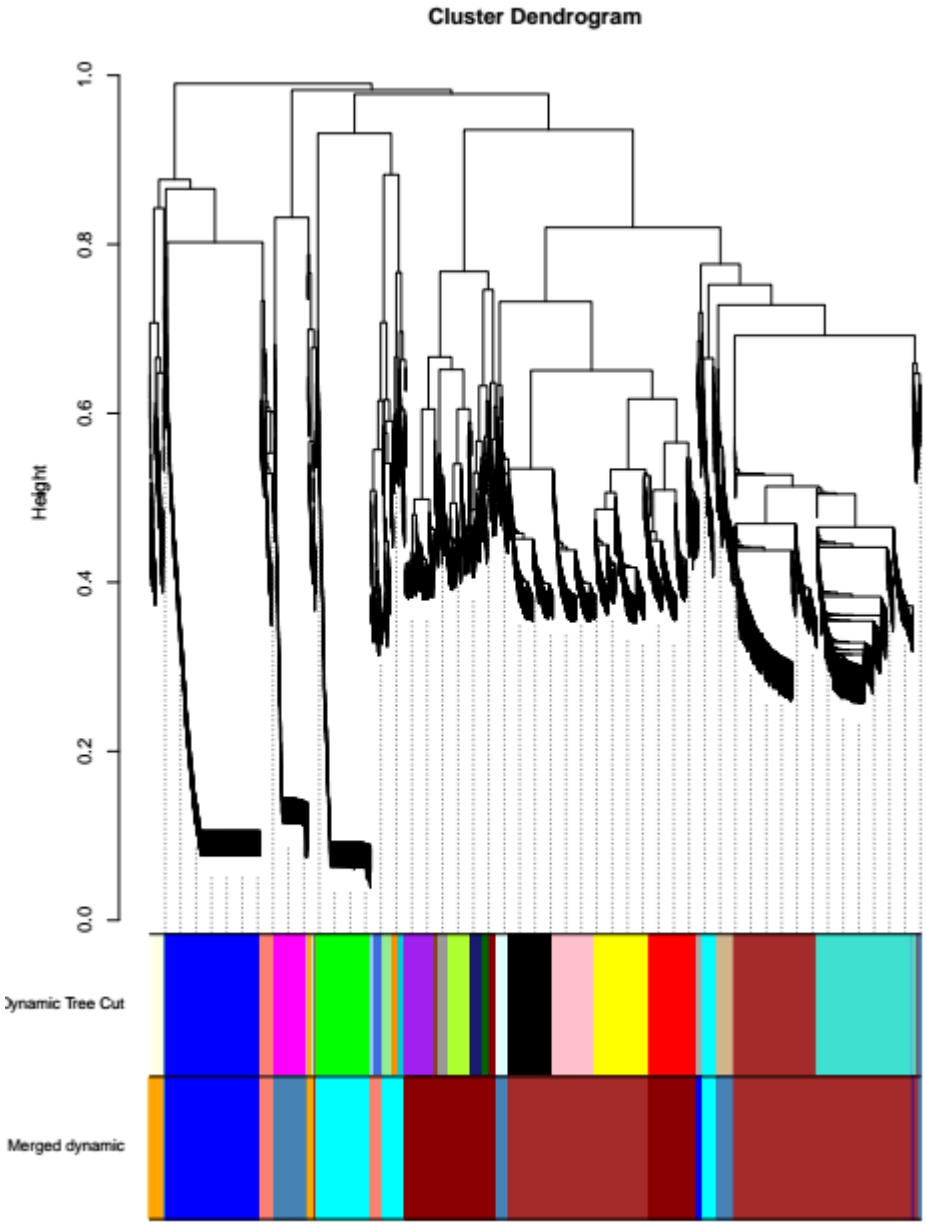


Figura 5. Dendograma de clusterização dos genes em oócitos e embriões de BÚFALOS, com dissimilaridade baseada no método de sobreposição topológica (TOM), seguido do código de cores dos módulos originalmente identificados, e dos códigos de cores dos módulos unidos com base em similaridade de perfil de expressão dos *eingengenes*.

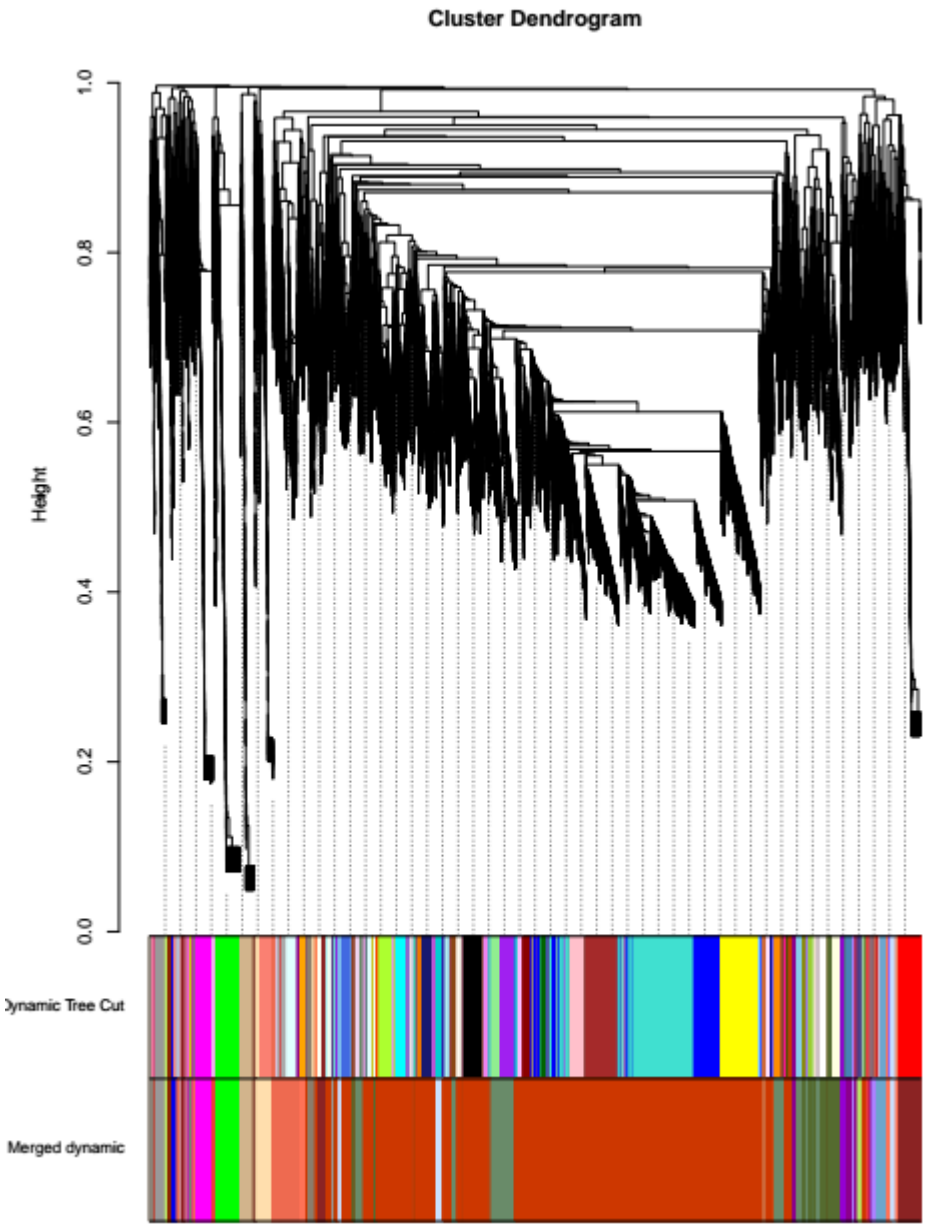


Figura 6. Dendograma de clusterização dos genes em oócitos e embriões de BOVINOS, com dissimilaridade baseada no método de sobreposição topológica (TOM), seguido do código de cores dos módulos originalmente identificados, e dos códigos de cores dos módulos unidos com base em similaridade de perfil de expressão dos *eingengenes*.

4. ANÁLISE DAS REDES DE CO-EXPRESSÃO

Foram seguidas as instruções do tutorial https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/Tutorials/FemaleLiver-03-relateModsToExt.pdf.

4.1. Identificação dos módulos de co-expressão estágio específicos

Foram identificados os módulos mais significativamente relacionados a oócitos e embriões. Para isso, foram utilizados os módulos *eigengenes* citados anteriormente, pois eles representam o perfil de expressão de cada módulo, sendo calculada a correlação dos *eigengenes* com os estágios. As linhas de comando são mostradas abaixo.

```
library(WGCNA)
options(stringsAsFactors = FALSE)
lnames = load(file="Bufalo_dataInput.RData")
lnames
lnames = load (file = "Bufalo-networkConstruction-stepByStep.RData")
lnames
nGenes = ncol(datExpr)
nSamples = nrow(datExpr)
MEs0 = moduleEigengenes(datExpr, BufalomoduleColors)$eigengenes
MEs = orderMEs(MEs0)
moduleTraitCor = cor(MEs, datTraits, use = "p")
moduleTraitPvalue = corPvalueStudent(moduleTraitCor, nSamples)
sizeGrWindow(10,6)
textMatrix = paste(signif(moduleTraitCor, 2), "\n(",
                   signif(moduleTraitPvalue, 1), ")", sep = "");
dim(textMatrix) = dim(moduleTraitCor)
par(mar = c(6, 8.5, 3, 3))
labeledHeatmap(Matrix = moduleTraitCor,
               xLabels = names(datTraits),
               yLabels = names(MEs),
               ySymbols = names(MEs),
               colorLabels = FALSE,
               colors = blueWhiteRed(50),
               textMatrix = textMatrix,
               setStdMargins = FALSE,
               cex.text = 0.5,
               zlim = c(-1,1),
               main = paste("Bufalo Module relationships"))
```

4.2. Identificação dos genes mais importantes (*hub genes*)

Foram identificados os genes com papel central e altamente conectados dentro dos módulos (conectividade intramodular), definidos como *hub genes*. Para isso, foi calculada a correlação entre os genes e o componente principal, ou módulo *eingengene*. Os *hub genes* foram sumarizados na tabela geneInfo.csv. Abaixo seguem as linhas de comando.

```
datTraits
embryo = as.data.frame(datTraits$embryo)
names(embryo) = "embryo"
modNames = substring(names(MEs), 3)
geneModuleMembership = as.data.frame(cor(datExpr, MEs, use = "p"))
MMPvalue = as.data.frame(corPvalueStudent(as.matrix(geneModuleMembership), nSamples))
names(geneModuleMembership) = paste("MM", modNames, sep="")
names(MMPvalue) = paste("p.MM", modNames, sep="")
geneTraitSignificance = as.data.frame(cor(datExpr, embryo, use = "p"))
GSPvalue = as.data.frame(corPvalueStudent(as.matrix(geneTraitSignificance), nSamples))
names(geneTraitSignificance) = paste("GS.", names(embryo), sep="")
names(GSPvalue) = paste("p.GS.", names(embryo), sep="")
module = "brown"
column = match(module, modNames)
bufalomoduleGenes = BufalomoduleColors==module
sizeGrWindow(7, 7)
par(mfrow = c(1,1))
verboseScatterplot(abs(geneModuleMembership[bufalomoduleGenes, column]),
                   abs(geneTraitSignificance[bufalomoduleGenes, 1]),
                   xlab = paste("Module Membership in", module, "module"),
                   ylab = "Gene significance for embryo",
                   main = paste("Module membership vs. gene significance\n"),
                   cex.main = 1.2, cex.lab = 1.2, cex.axis = 1.2, col= module)
names(datExpr)
names(datExpr)[BufalomoduleColors=="brown"]
annot = read.csv(file = "GeneAnnot.csv")
dim(annot)
names(annot)
probes = names(datExpr)
probes2annot = match(probes, annot$hgnc_symbol)
sum(is.na(probes2annot))
geneInfo0 = data.frame(hgnc_symbol = probes,
                       geneSymbol = annot$hgnc_symbol[probes2annot],
                       LocusLinkID = annot$entrezgene[probes2annot],
                       moduleColor = BufalomoduleColors,
                       geneTraitSignificance,
                       GSPvalue)
modOrder = order(-abs(cor(MEs, embryo, use = "p")))
for (mod in 1:ncol(geneModuleMembership))
{
  oldNames = names(geneInfo0)
  geneInfo0 = data.frame(geneInfo0, geneModuleMembership[, modOrder[mod]],
                         MMPvalue[, modOrder[mod]]);
```

```
  names(geneInfo0) = c(oldNames, paste("MM.", modNames[modOrder[mod]], sep=""),
                       paste("p.MM.", modNames[modOrder[mod]], sep=""))
}
geneOrder = order(geneInfo0$moduleColor, -abs(geneInfo0$GS.embryo))
geneInfo = geneInfo0[geneOrder, ]
write.csv(geneInfo, file="geneInfo.csv")
```

5. CONSERVAÇÃO DOS MÓDULOS DE CO-EXPRESSÃO DE BÚFALOS EM BOVINOS

Para acessar a preservação dos módulos de genes nas redes de co-expressão de búfalos e bovinos, foi utilizada a função *modulePreservation* conforme as instruções em https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/ModulePreservation/Tutorials/MiniTutorial-MouseLiver.pdf.

5.1. Obtenção do arquivo de entrada para a função de *modulePreservation*

A função *modulePreservation* compara se os módulos de genes em búfalos são preservados, ou seja, correspondentes aos módulos em bovinos. O primeiro arquivo de entrada consistiu na tabela com os dados de expressão normalizados de búfalo (*datExpr*) citada anteriormente (item 2), adicionada da coluna "*moduleColors*", que corresponde ao módulo em que o gene foi agrupado, e que é referido pela sua cor (*brown, cyan, blue etc*). Abaixo, foi ilustrado o primeiro arquivo de entrada (BufaloData.csv).

Gene	*Buf_oocyte1*	*Buf_oocyte2*	*Buf_embryo1*	*Buf_embryo2*	*moduleColors*
A2M	*4.37835926957161*	*4.37835926957161*	*7.81884349589515*	*6.90173345091496*	*brown*
A4GALT	*4.37835926957161*	*4.37835926957161*	*6.80729787364605*	*7.82882560532151*	*brown*
AAAS	*5.69584681756271*	*4.37835926957161*	*6.25252416898603*	*6.06300012493195*	*darkred*
AACS	*7.38124716583152*	*7.95568977621475*	*5.92884150990694*	*6.77159182795084*	*steelblue*
AAED1	*4.51896343396031*	*4.37835926957161*	*4.52927787524831*	*5.26784123881362*	*cyan*
AAGAB	*6.67538546064213*	*4.37835926957161*	*6.40164385497101*	*6.31621693148955*	*darkred*

O arquivo BufaloData.csv foi criado, com a união dos conteúdos de *datExpr* e *geneInfo* citado anteriormente (item 4.2) usando o pacote *dplyr* do programa R, e as linhas de comando descritas abaixo. Por fim, o segundo arquivo de entrada para o *modulePreservation* consistiu na tabela com os dados de expressão de bovinos (*datExpr*).

```
library(dplyr)
data <- read.csv(file="datExpBuf.csv")
data <- as.data.frame(data)
head(data)
mod <- read.csv(file="geneInfo.csv")
mod <- as.data.frame(mod)
head(mod)
inputPreservation <- data.frame(Buf_oocyte1=data$Buf_oocyte1, Buf_oocyte2=data$Buf_oocyte2,
Buf_embryo1=data$Buf_embryo1, Buf_embryo2=data$Buf_embryo2, Gene=data$gene,
geneSymbol=mod[match(data$gene, mod$geneSymbol), 3])
write.csv(inputPreservation, file="InputPreservation.csv")
```

5.2. Cálculo dos módulos preservados de búfalos em bovinos

Os arquivos de entrada foram carregados e checados para constatar se os nomes dos genes eram concordantes. Em seguida, cada conjunto de dados foi nomeado (búfalo e bovino) e a função *modulePreservation* foi executada para calcular a preservação dos módulos de búfalos em bovinos. As linhas de comando foram mostradas a seguir.

```
library(WGCNA)
options(stringsAsFactors = FALSE)
enableWGCNAThreads()
file = file("BufaloData.csv")
dat0=read.csv(file, header=TRUE)
names(dat0)
dim(dat0)
datSummary=dat0[,c(1,6)]
datExprBufalo <- t(dat0[,2:5])
no.samples <- dim(datExprBufalo)[[1]]
dim(datExprBufalo)
colnames(datExprBufalo) = datSummary$Gene
colorsBufalo = dat0$moduleColors
file = file("datExpBov.csv")
data = read.csv(file, header = TRUE)
dim(data)
colnames(data)
datExprBovine = t(data[, substring(colnames(data), 1, 4)=="Bov_"])
dim(datExprBovine)
colnames(datExprBovine) = data$"X"
setLabels = c("Bufalo", "Bovine")
multiExpr = list(Bufalo = list(data = datExprBufalo), Bovine = list(data = datExprBovine))
multiColor = list(Bufalo = colorsBufalo)
system.time( {
  mp = modulePreservation(multiExpr, multiColor,
                          referenceNetworks = 1,
                          nPermutations = 200,
                          randomSeed = 1,
```

```
                quickCor = 0,
                verbose = 3)
} )
save(mp, file = "modulePreservation.RData")
```

5.3. Geração do gráfico de preservação de módulos

O resultado foi visualizado na forma de gráfico ilustrando o valor *Z-summary* que indica a preservação dos módulos de búfalos em bovinos. Em geral, valor de *Z-summary* > 10 significa uma forte preservação, ou seja, os módulos de búfalo e bovinos são densamente conectados. O valor *Z-summary* entre 2 e 10 significa preservação moderada e *Z-summary* < 2 siginifica preservação fraca. A seguir, os comandos para gerar o gráfico.

```
ref = 1
test = 2
statsObs = cbind(mp$quality$observed[[ref]][[test]][, -1], mp$preservation$observed[[ref]][[test]][,
-1])
statsZ = cbind(mp$quality$Z[[ref]][[test]][, -1], mp$preservation$Z[[ref]][[test]][, -1])
print( cbind(statsObs[, c("medianRank.pres", "medianRank.qual")],
          signif(statsZ[, c("Zsummary.pres", "Zsummary.qual")], 2)) )
modColors = rownames(mp$preservation$observed[[ref]][[test]])
moduleSizes = mp$preservation$Z[[ref]][[test]][, 1]
plotMods = !(modColors %in% c("grey", "gold"))
text = modColors[plotMods]
plotData = cbind(mp$preservation$observed[[ref]][[test]][, 2], mp$preservation$Z[[ref]][[test]][, 2])
mains = c("Preservation Median rank", "Preservation Zsummary");
sizeGrWindow(10, 5)
par(mfrow = c(1,2))
par(mar = c(4.5,4.5,2.5,1))
for (p in 1:2)
{
  min = min(plotData[, p], na.rm = TRUE);
  max = max(plotData[, p], na.rm = TRUE);
  if (p==2)
  {
    if (min > -max/10) min = -max/10
    ylim = c(min - 0.1 * (max-min), max + 0.1 * (max-min))
  } else
    ylim = c(max + 0.1 * (max-min), min - 0.1 * (max-min))
  plot(moduleSizes[plotMods], plotData[plotMods, p], col = 1, bg = modColors[plotMods], pch = 21,
       main = mains[p],
       cex = 2.4,
```

```
      ylab = mains[p], xlab = "Module size", log = "x",
      ylim = ylim,
      xlim = c(10, 2000), cex.lab = 1.2, cex.axis = 1.2, cex.main =1.4)
  labelPoints(moduleSizes[plotMods], plotData[plotMods, p], text, cex = 1, offs = 0.08);
  if (p==2)
  {
    abline(h=0)
    abline(h=2, col = "blue", lty = 2)
    abline(h=10, col = "darkgreen", lty = 2)
  }
}
```

6. ANÁLISE DE ONTOLOGIA GÊNICA DOS MÓDULOS DE CO-EXPRESSÃO

Foi feita a análise de ontologia gênica dos módulos de co-expressão de búfalos. Para isso a lista de genes dos módulo foi exportada juntamente com o *LocusLinksID* (código identificador no Entrez) e submetidas aos pacotes *GO.db* (Carlson, 2017) e *AnnotationDBI* (Pages et al., 2017) do programa R. Foi usado o tutorial https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/Tutorials/FemaleLiver-05-Visualization.pdf.

```
library(WGCNA)
options(stringsAsFactors = FALSE)
lnames = load(file = "Bufalo_dataInput.RData")
lnames = load(file = "Bufalo-networkConstruction-stepByStep.RData")
annot = read.csv(file = "GeneAnnot.csv")
probes = names(datExpr)
probes2annot = match(probes, annot$hgnc_symbol)
allLLIDs = annot$entrezgene[probes2annot]
intModules = c("brown")
for (module in intModules)
{
  modGenes = (BufalomoduleColors==module)
  modLLIDs = allLLIDs[modGenes]
  fileName = paste("LocusLinkIDs-", module, ".txt", sep="")
  write.table(as.data.frame(modLLIDs), file = fileName,
              row.names = FALSE, col.names = FALSE)
}
fileName = paste("LocusLinkIDs-all.txt", sep="")
write.table(as.data.frame(allLLIDs), file = fileName,
            row.names = FALSE, col.names = FALSE)
GOenr = GOenrichmentAnalysis(BufalomoduleColors, allLLIDs, organism = "bovine", nBestP = 10)
tab = GOenr$bestPTerms[[4]]$enrichment
names(tab)
write.table(tab, file = "GOEnrichmentTable.csv", sep = ",", quote = TRUE, row.names = FALSE)
```

7. VISUALIZAÇÃO DOS HUB GENES NO MÓDULO BROWN

Para explorar a conexões entre os hub genes do módulo brown, foi gerada a visualização das conexões no programa Cytoscape (http://www.cytoscape.org/) (Shannon et al. 2003; Cline et al. 2007). Os arquivos de entrada no Cytoscape contêm os nodos e arestas das conexões do módulo, e foram exportados do WGCNA (*cytoscapeInput-edges1.txt* e *cytoscapeInput-nodes1.txt*) usando as linhas de comando a seguir.

```
library(WGCNA)
options(stringsAsFactors = FALSE)
enableWGCNAThreads()
lnames = load(file="Bufalo_dataInput.RData")
lnames = load (file = "Bufalo-networkConstruction-stepByStep.RData")
TOM = TOMsimilarityFromExpr(datExpr, power = 20)
save(TOM, file="BufTOM.RData")
modules = c("brown")
probes = names(datExpr)
inModule = is.finite(match(BufalomoduleColors, modules))
modProbes = probes[inModule]
modTOM = TOM[inModule, inModule]
dimnames(modTOM) = list(modProbes, modProbes)
cyt = exportNetworkToCytoscape(modTOM,
        edgeFile = paste("CytoscapeInput-edges1-", paste(modules, collapse="-"), ".txt", sep=""),
        nodeFile = paste("CytoscapeInput-nodes1-", paste(modules, collapse="-"), ".txt", sep=""),
                weighted = TRUE,
                threshold = 0.7,
                nodeNames = modProbes,
                altNodeNames = modProbes,
                nodeAttr = BufalomoduleColors[inModule])
```

Os arquivos .txt exportados foram abertos no cytoscape para visualização das 150 conexões mais fortes do módulo *brown*. No Cytoscape o arquivo apresenta os genes que estão conectados (*source* e *target*), a força de conexão entre eles (conectividade intramodular), o tipo de interação (pp, proteína-proteína) e a direção da conexão (sem direção, FALSE). A seguir consta o cabeçalho do arquivo ao ser aberto no Cytoscape.

source	target	value	interaction	directed
MB21D2	TCF20	0.715186958975012	pp	FALSE
LAMA1	RND1	0.71511516888714	pp	FALSE
ENPEP	MB21D2	0.715112141040459	pp	FALSE
RND1	TAF1A	0.715110592048089	pp	FALSE
KHDC3L	LAMA1	0.715097850713354	pp	FALSE
KHDC3L	TAF1A	0.715095693262126	pp	FALSE
DAB2	MB21D2	0.715090600729902	pp	FALSE

REFERÊNCIAS BIBLIOGRÁFICAS

Carlson M (2017). *GO.db: A set of annotation maps describing the entire Gene Ontology*.

R package version 3.4.1.

Cline, Melissa S et al. 2007. “Integration of Biological Networks and Gene Expression Data Using Cytoscape.” *Nature protocols* 2(10): 2366–82. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3685583/.

Graf, Alexander et al. 2014. “Fine Mapping of Genome Activation in Bovine Embryos by RNA Sequencing.” *Proceedings of the National Academy of Sciences of the United States of America* 111(11): 1–6. http://www.ncbi.nlm.nih.gov/pubmed/24591639.

Langfelder, Peter, and Steve Horvath. 2008. “WGCNA: An R Package for Weighted Correlation Network Analysis.” *BMC Bioinformatics* 9(1): 559. http://dx.doi.org/10.1186/1471-2105-9-559.

Langfelder, Peter, Rui Luo, Michael C Oldham, and Steve Horvath. 2011. “Is My Network Module Preserved and Reproducible?” *PLOS Computational Biology* 7(1): 1–29. https://doi.org/10.1371/journal.pcbi.1001057.

Pagès H, Carlson M, Falcon S and Li N (2017). *AnnotationDbi: Annotation Database Interface*. R package version 1.38.1.

Shannon, P et al. 2003. “Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks.” *Genome Research* 13. http://dx.doi.org/10.1101/gr.1239303.

Zhang, B, and S Horvath. 2005. “A General Framework for Weighted Gene Co-Expression Network Analysis.” *Stat Appl Genet Mol Biol* 4.

7. CONCLUSÕES GERAIS

Conclui-se que:

Os transcriptomas de oócitos maturados e blastocistos produzidos *in vitro* bubalinos, evidenciam em maior proporção a presença de genes relacionados a funções endógenas, e em menor proporção a funções especializadas durante a maturação oocitária e o desenvolvimento embrionário.

Os genes com maior variação de expressão em blastocistos foram relacionadas ao metabolismo de lipídios e implantação. E em oócitos foram relacionadas a fecundação e qualidade oocitária. Os resultados indicam que estes genes são importantes para o aspecto molecular de qualidade embrionária e oocitária em bubalinos.

Os transcriptomas de oócitos e blastocistos de búfalos e bovinos mostraram forte evidência de preservação indicando que seus funcionamentos são similares, exceto em termos de cinética de desenvolvimento e transporte de aminoácidos. Porém, os perfis de expressão dentro dos módulos de blastocistos mostrou maior dissimilaridade, indicando que mesmo dentro de módulos conservados há diferenças de expressão importantes entre búfalos e bovinos que justificam a elaboração de protocolos de PIVE específicos para búfalos.