Application of integrative genomics and systems biology to conventional and in vitro reproductive traits in cattle

Mazzoni, Gianluca; Pedersen, Hanne S.; de Oliveira Junior, Gerson A.; Alexandre, Pamela; Razza, Eduardo M.; Callesen, Henrik; Hyttel, Poul; Nogueira, Marcelo F. G.; Ferraz, Jose Bento S.; Kadarmideen, Haja

Published in: Proceedings of the 31st Annual Meeting of the Brazilian Embryo Technology Society (SBTE)

Link to article, DOI: 10.21451/1984-3143-AR993

Publication date: 2017

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Application of integrative genomics and systems biology to conventional and in vitro reproductive traits in cattle

Gianluca Mazzoni1,6, Hanne S. Pedersen2, Gerson A. de Oliveira Junior3, Pamela Alexandre3, Eduardo M. Razza4, Henrik Callesen2, Poul Hyttel1, Marcelo F.G. Nogueira5, Jose Bento S. Ferraz3, Haja N. Kadarmideen3, 4

1Department of Veterinary and Animal Sciences, University of Copenhagen, 1870 Frederiksberg C, Denmark.
2Department of Animal Science, Aarhus University, 8830 Tjele, Denmark.
3Research Center on Animal Breeding, Biotechnology and Transgenesis, FZEA, University of São Paulo, Pirassununga, SP, Brazil.
4Department of Pharmacology, Institute of Biosciences, São Paulo State University (UNESP), Botucatu, SP, Brazil.
5Department of Biological Sciences, School of Sciences and Languages, São Paulo State University (UNESP), Assis, SP, Brazil.
6Section of Systems Genomics, Department of Bio and Health Informatics, Technical University of Denmark, Lyngby, Denmark.

Abstract

Assisted reproductive technologies (ARTs) have a strong impact on breeding especially when coupled with genomic selection (GS). The routine implementation of in vitro production (IVP) and GS of embryos before embryo transfer (ET) in breeding companies is not yet possible. Improvement of oocyte donor and embryo recipient quality is needed to make realistic a commercialization of these procedures in the near future. A better understanding of both biological mechanisms and molecular markers associated to IVP-ET related traits is necessary to improve the prediction of donor and recipient cow quality for IVP procedures. The huge amount of data generated from high throughput technologies has a tremendous impact in the search for biomarkers of complex traits. This paper reviews integrative genomics and systems biology approaches as applied to both Bos indicus and Bos taurus cattle reproduction by both conventional and ARTs such as OPU-IVP. The integration of systems biology information across different biological layers generates a complete view of the different molecular networks that control complex traits and can provide a strong contribution to the understanding of traits related to ARTs.

Keywords: systems biology, IVP, reproduction, cattle, biomarkers, data integration.

Introduction

Breeding for more efficient animals is becoming of increasing importance, and new and faster breeding methods are needed. Assisted reproductive technologies (artificial insemination (AI), ovum pick-up (OPU), IVP and ET) have significantly contributed to animal breeding programs. Similarly, genomics has significantly increased both speed of genetic gain and selection accuracy (Kadarmideen et al., 2015). However, the greatest benefits of those tools can only be expected when they are combined, allowing animals to be selected accurately early in life and with a more precise estimation of their breeding value.

The combined GS-IVP-ET procedures are not routinely implemented yet. While it is possible to collect embryo biopsies for DNA genotyping without affecting the pregnancy rate of the embryos (Saadi et al., 2014), the IVP-ET efficiency is still relatively low and remains the bottleneck in this process. A possible way to increase the efficiency of IVP-ET procedures is identification and selection of high-quality oocyte donors and embryo recipients. In order to successfully predict donor and recipient cow quality in IVP procedures, a better understanding of the molecular mechanisms responsible for these traits is needed (Salilew-Wondim et al., 2010).

The literature provides a complex picture of the molecules associated with oocyte competence at different molecular level: proteomics, transcriptomics (Jiang et al., 2010; Gilbert et al., 2012; Nivet et al., 2013) and metabolomics (Matoba et al., 2014). Furthermore, numerous studies focused on the characterization of the expression profiles of granulosa cells from follicles at different developmental stages (Hatzirodos et al., 2014b; Girard et al., 2015) and physiological condition (Hatzirodos et al., 2014a). Research in recipient cows on transcriptomic profiles has usually been applied to the endometrial tissue, and this has increased the knowledge about molecular mechanisms responsible for good or inadequate uterine environment for pregnancy recognition and implantation (Salilew-Wondim et al., 2010; Forde and Lonergan, 2012; Ponsuksili et al., 2012; Minten et al., 2013; Killeen et al., 2014). In the same way endometrial gene expressions have been characterized both during the estrous cycle and early pregnancy before implantation (Bauersachs et al., 2005; Mitko et al., 2008; Bauersachs et al., 2009; Mansouri-Attia et al., 2009; Ponsuksili et al., 2012) as well as after implantation (Bauersachs et al., 2006; Binelli et al., 2015).

Considering the huge amount of data generated by use of modern high-throughput technologies at all levels of biological systems (e.g. genome-wide, transcriptome-wide, metabolome-wide or proteome-wide measurements), systems biology analysis is the most promising approach to provide a holistic view of important biological mechanisms and of molecular
markers associated with complex traits, such as those related to IVP-ET procedures. Systems biology is a holistic approach to analyze and decipher complex biological systems through computational and mathematical modeling of the whole set of molecular networks within an organism. Systems biology aims to decipher systems structures (the network of gene interactions and biochemical pathways), but also the dynamic changes of these network under different conditions (Kitano, 2002; Chuang et al., 2010; Kadarmideen et al., 2011; Kadarmideen, 2014). A huge part of the computational methods in systems biology is based on network approaches that identify group of genes or molecules with common behavior assuming that clustered molecules are functionally related (Brazhnik et al., 2002). As reviewed in (Mazzoni et al., 2015), several approaches are available to build gene interaction or association networks: Co-expression/regulatory patterns (Horvath, 2011), Bayesian networks (Friedman et al., 2000), Random Forest Tree approaches (Breiman, 2001) or Artificial Neural Networks (ANN) (McCulloch and Pitts, 1943; Cookson et al., 2009).

Integrative systems biology is based on analyses of networks of different types of molecules trying to integrate information at different biological levels (Dixon et al., 2007; Mazzoni et al., 2015; Suravajhala et al., 2016; Wang and Michoel, 2016). The analyses of a specific outcome (blastocyst rate and quality or endometrial receptive competence) with systems biology methods together with the integration of systems biology approaches at different biological levels is of vital importance for identification of biomarkers for incredibly complex traits like IVP and ET outcome.

While above we discussed about ARTs, it is very important to apply integrative systems biology approaches also to conventional cattle reproduction via GS and genomic breeding. Fundamental to this are Genome Wide Association Studies (GWAS). A very important proportion of the world beef production trade comes from tropical and subtropical regions. Bos indicus, from Brahman breed in Australia, and Nellore in Brazil, and their crosses are the most important breeds used in tropical areas. Candidate genes and biological information to be included in genomic selection of very young donors are essential to increase the productivity in tropical and sub-tropical areas (Simianer, 2016). The identification of candidate genes and metabolic pathways associated with reproductive traits such as antral follicular population and early pregnancy in Nellore heifers has become of great importance with the increasing use of ARTs technologies. In this context, genome wide association studies followed by functional analysis such as Medical Subject Heading (MeSH) terms enrichment can give a strong contribution.

In this mini-review we describe (i) the importance of the application of integrative systems biology approaches in the analysis of ARTs related traits, (ii) the two integrative systems biology experimental designs from the GIFT project consortium ([www.gift.ku.dk](http://www.gift.ku.dk)) aimed to identify biomarkers for the selection of superior oocyte donor and embryo recipient cows, and (iii) a GWAS for the identification of candidate genes and metabolic pathways for early puberty and reproductive traits in Nellore heifers.

**Integrative systems biology analyses and application to IVP related traits in cattle**

To correctly predict donor and recipient quality we need a list of biomarkers and biological processes to be included in the prediction methods for the selection of the best donor or recipient cows. This implies gathering enough information about specific molecules and biological processes associated to the IVP-ET performance of the animals.

The definition of a biomarker is strictly dependent on the field of study. The features of a biomarker are described by Austin Bradford Hill's guidelines (Aronson, 2005) and here adapted to the context of IVP related traits: strength (a strong association between donor and IVP-ET outcome), consistency (the association identified in the cow reference set should persist in cow populations from different farms and different physiological status of the cows), specificity (the biomarker should be associated with the specific IVP-ET related trait), plausibility (the biomarker should be part of meaningful biological mechanisms), coherence (the biomarker and the biological functions performed should be consistent with previous knowledge), analogy (previous findings for the same association make the biomarkers more valuable). Furthermore, in order to be routinely applied in IVP-ET procedures, the biomarker should be easily accessible and the costs of its measurement affordable. According to the guidelines the search for a biomarker is not an easy task. Until now, the search of biomarkers for IVP related traits have been based mainly on transcriptomic data for two reasons: first, it allows amplification of small samples (Orozco-Lucero and Sirard, 2014) and second, it provides information about the biology of a trait. Many studies have focused on functional enrichment and analysis of the differentially expressed genes to give a biological context of the biomarker. However, systems biology approaches like co-expression analysis were rarely adopted. The possibility to analyze transcriptomic data in a more holistic way opens for a better understanding of the biological mechanisms and allows the identification of key genes that would be impossible to identify with traditional “reductionist” approaches (Kitano, 2002; Chuang et al., 2010; Kadarmideen, 2014). For example, co-expression network can be used to identify regulatory genes that are key genes responsible for the control of the expression of a set of genes expected to be involved in the same biological process (Zhao et al., 2010). Thus, regulatory genes are candidate genes to be used for biomarker development.

The mRNA expression data are used as a surrogate for protein expression. However, it is known that the correlation between mRNA and protein levels is moderate due to complex regulation mechanisms.
occuring after transcription (Lu et al., 2007; Schwanhäusser et al., 2011). A perfect systems biology experiment could avoid this issue including use of all the biological levels and integration of the information to create a complete overview. The integration of different biological levels does not imply necessarily to measure all biological levels in a sample set. A huge amount of multi-omics data (transcriptomics, genomics, proteomics, metabolomics, interactomics etc.) is available in public databases. A possibility would be the integration of experimental data with data from publicly available repositories (Mazzoni et al., 2015) for example STRING v.10 database (Szklarczyk et al., 2015) and Gene Mania (Montojo et al., 2010).

Measurements of the expression pattern of a specific set of mRNA molecules from follicular cells or endometrial biopsies could be used directly to select the best cows but they cannot be implemented in routine breeding procedures (Ponsuksili et al., 2010). Furthermore, expression profiles are subject to continuous variation due to environmental and physiological status of the cows. Most of the molecules associated with IVP traits are not consistent (as required by the Austin Bradford Hill's guidelines). This means that when the cows are tested for prediction they should be in the same biological status and environmental condition as the reference set used to identify these biomarkers. Integration of the transcriptomic level with the genomic level could overcome this issue. The approach is called expression QTL (eQTL) mapping and allows for the identification of genomic variants (e.g. SNPs) that are correlated with the expression level of a specific transcript (Dixon et al., 2007; Cookson et al., 2009; Wang and Michoel, 2016). The eQTL mapping integrates variation at the level of RNA expression with the variation at the DNA level. The advantage of eQTL mapping is that it identifies DNA variants that can be used in breeding (Ponsuksili et al., 2010), for example to select for ARTs related traits.

Co-expression network analysis of granulosa cells in donor cows

In order to identify biomarkers to predict the quality of a donor cow and to perform an eQTL analysis, it is necessary to collect data at the single animal level. An integrative systems biology analysis at the single animal level is the ideal way to obtain a more complete understanding of the biological processes associated with oocyte competence in follicular cells and to identify genetic variants to be included in breeding procedures. On this idea, we based the experimental design of our transcriptomic donor cow analysis (Mazzoni et al., 2017). Briefly, all antral follicles present in each pair of ovaries were collected and the IVP performances for each animal were evaluated. The mural granulosa cells and a small amount of granulosa cells from the cumulus layer were collected as a byproduct of the follicle aspiration, and the total RNA was extracted and 24 samples sequenced. The RNA-Seq data were analyzed with a bioinformatics pipeline to quantify the expression of the entire set of known genes. Therefore we generated the average ovarian expression profile of granulosa cells and the respective IVP performances of the entire set of cumulus-oocyte-complexes.

We identified 51 differentially expressed genes associated to IVP performances and seven candidate genes associated with all IVP parameters analyzed (Mazzoni et al., 2017). In the same study, the functional enrichment of the differentially expressed genes and the comparison with previous findings in the literature confirmed the positive association between the IVP outcome and the presence of early atresia as previously observed (Moor and Trounson, 1977; Wurth and Kruip, 1992; De Wit et al., 2000; Feng et al., 2007; Heleil et al., 2010).

A systems biology analysis of the same dataset would provide more information about the biological mechanisms controlling atresia and consequently result in more plausible biomarkers. Thus, we are working on a co-expression network analysis of the same dataset using the Weighted Gene Co-expression Network Analysis (WGCNA) R package. WGCNA identifies groups of co-expressed genes called “modules” that are expected to be involved in the same biological process (Zhao et al., 2010). Furthermore, WGCNA provides the ability to select only the modules that are correlated with a trait of interest. Briefly, we identified modules associated with IVP performances, and these data were then integrated with information about protein-protein association provided by STRING v.10 (Szklarczyk et al., 2015). The functional enrichment based on STRING information was performed (i) to understand the biology behind the IVP performance and (ii) to select the biologically meaningful modules that are the most suitable for selection of new candidate genes. We analyzed the selected modules to identify central genes (hub genes) and regulatory genes integrating information from other sources like Ingenuity® Pathway Analysis.

The preliminary results confirmed that the systems biology is a good approach to study IVP related traits in donor cows. The atresia mechanism was confirmed to be positively correlated with IVP performances. Moreover, the analysis provided a more detailed description of the molecular mechanisms that link atresia with IVP performances which are new evidences to support the candidate genes.

Co-expression analysis of endometrial biopsies

Similarly to the donor cow analysis, we structured the experimental design of the recipient cow analysis for the embryo implantation to collect information for each single animal. Briefly, endometrial biopsies from the uterine horn ipsilateral to the corpus luteum were sampled from experimental cows on day 6-8 in the estrous cycle, and RNA was extracted and sequenced. On day 6-8 in the following cycle in vitro produced blastocysts were transferred to the animals and the pregnancy status was determined at slaughter on day 26-47. In the pregnant animals, embryo/fetuses were fixed and their quality evaluated according to.
external characteristics and histology.

While in the donor cow the IVP performances were measured as continuous traits, in the analysis of the recipient cows the trait was a factor with two outcomes (pregnant and not pregnant). Thus, it is possible to study the differential wiring between two biological networks (one for each phenotypic group). In other words, we can generate one network for the group of pregnant cows and one for the group of non-pregnant cows by using WGCNA. With statistical approaches it is then possible to identify genes (nodes) that are differentially wired comparing the two networks. Therefore, we can identify gene groups that must be co-expressed and activated together to generate an ideal receptive condition of the endometrium during ET. The functional enrichment and the integration with publicly available data could shed light on the mechanisms and the differences between receptive and non-receptive animals. Thus, we expect to identify novel biomarkers and information about biological mechanisms that can be used for prediction of good quality recipient cows.

**Expression QTL mapping in donor and recipient cows**

The eQTL mapping is an example of integrative systems biology analysis. An eQTL is a genomic region associated with transcript expression levels, and it can be close to the transcription start site (cis-eQTL) or acting on a larger distance or on another chromosome (trans-eQTL; [Mazzoni et al., 2015; Wang and Michoel, 2016]). Mapping of eQTL can be used to link genetic variants to a specific trait ([Buchner and Michoel, 2016](https://doi.org/10.1093/bib/bbq074)). Mapping eQTL can be used to study the differential wiring between two specific conditions of a target phenotype ([Kadarmideen, 2008](https://doi.org/10.1093/bib/bqq113)). In the two GIFT consortia projects previously described, both transcriptomic and genomic data were collected for eQTL mapping ([Westra and Franke, 2014](https://doi.org/10.1093/bib/bwp008)). Furthermore, eQTL approaches require smaller sample size to obtain good detection power ([Kadarmideen, 2008](https://doi.org/10.1093/bib/bqq113)).

In the context of oocyte donor or embryo recipient cow selection, eQTLs can be a useful breeding tool ([Ponsuksili et al., 2010](https://doi.org/10.1093/bib/byp183); [Nica and Dermitzakis, 2013](https://doi.org/10.1093/bib/bjt040)). Furthermore, eQTL approaches require smaller sample size to obtain good detection power ([Kadarmideen, 2008](https://doi.org/10.1093/bib/bqq113)). In the two GIFT consortium projects previously described, both transcriptomic and genomic data were collected for eQTL mapping. The hub genes and the regulatory genes identified with co-expression analyses will be included in eQTL analysis. The eQTLs together with biological information obtained from integrative systems biology studies could be adopted in animal selection processes through the inclusion in GS methodologies utilizing functional information, e.g. sgBLUP (systems genomic BLUP; [Kadarmideen, 2014](https://doi.org/10.1093/bib/byu127); and BLUPGA (BLUP approach given the Genetic Architecture; [Zhang et al., 2014](https://doi.org/10.1093/bib/byu127)).

**GWAS and functional enrichment analyses of early pregnancy of Nellore heifers**

Function enrichment analysis such as GWAS can give a strong contribution to the identification of candidate genes and metabolic pathways associated with conventional reproductive techniques. This is even more relevant if applied to Nellore cattle, one of the most important *Bos indicus* breeds for beef production in the world. We performed a GWAS followed by MeSH (Medical Subject Headings) enrichment to analyze pregnancy and antler follicular population in Nellore heifers (*Bos indicus*). Briefly, ovarian ultrasound (7.5 MHz transrectal linear transducer, Mindray M5Vet, China) was performed to count visible follicles on Nellore heifers submitted to fixed-time artificial insemination. Nellore heifers (*n = 1,255*) from 3 different farms in Central-Western Brazil were exposed to breeding season at an average of around 16 month of age and were genotyped with GeneSeek GGP *Bos indicus* HD Technology (74,677 SNPs). After quality control with the software PREGSF90 (developed by Misztal I. et., al (2002), 64,753 SNP were included in the GWAS for both traits. The GWAS was performed using Gensel software ([Fernando and Garrick, 2012](https://doi.org/10.1093/bib/bfr071)) under Bayes B method. 

(Co)variance components study was performed with single step analysis ([Wang et al., 2012](https://doi.org/10.1093/bib/bbr023)) under Bayesian method, and heritability estimates were 0.28 ± 0.07 (heifer pregnancy) and 0.49 ± 0.09 (follicular population), with a genetic correlation of -0.21 ± 0.29, meaning that in this sample of Nellore heifers, the number of antral follicles and heifer pregnancy may have antagonism.

SNP markers, located in genomic windows of approximately 1Mb that explained more than one per cent of genetic variance, were included in the functional enrichment. The MeSH enrichment was performed with MESH ([Morota et al., 2015](https://doi.org/10.1093/bib/bwq007); [Tsuyuzaki et al., 2015](https://doi.org/10.1093/bib/bwq007)). The functional analysis revealed 74 terms related to heifer pregnancy. These terms were related to: i) the metabolic pathways of fucose, that is involved in fertilization and in particular in the interaction between spermatozoia and oviduct and fertilization, ii) to Munc18 that plays a role in pituitary hormone secretion and iii) hemoglobin that takes part in ovary vascularization and LH secretion.

The functional analysis of the follicular population revealed 48 terms related to number of follicles and highlighted the importance of the following molecules: i) Neureptide receptor and kisspeptin associated with GnRH expression ([Amstalden et al., 2014](https://doi.org/10.1093/bib/bvq053)); ii) Cathepsin B that affects oocyte quality and control heat stress ([Balboula et al., 2013](https://doi.org/10.1093/bib/bqs094)); and iii) palmitic acid associated with apoptosis in follicular cells and with reproductive problems ([Zeron et al., 2001](https://doi.org/10.1093/bib/bqk021)).

The MeSH term enrichment of the GWAS results proved to be a reliable approach to get insight into the biology of the reproductive traits. These findings confirmed some of the candidate genes and molecular pathways identified in previous studies and could contribute to the research of the ARTs related traits.

**Conclusion and future perspectives**

The huge amount of data generated from high
throughput technologies has had a tremendous impact in the research for biomarkers for complex traits like those related to IVP procedures. Similarly, genomic analysis coupled with functional enrichment could expand the knowledge about the biology of conventional reproductive traits in important breeds such as Nellore cows. A better understanding of these traits such as number of follicles or pregnancy outcome is of great importance for the improvement of ARTs. In this context systems biology analysis aimed to identify the emergent properties from the “omic” frame is the most promising tool.

The next step is the validation of the candidate genes in bigger reference populations to test their real predictive power. The candidate genes could be validated with q-PCR, while secreted proteins encoded by the candidate genes could be tested directly in vitro. Moreover, further analysis (proteomics and cytological/morphological studies) could be used to verify the correlation between atresia, gene expression and IVP performances. However, the integration of systems biology analysis, across different biological layers proved to be a good methodology to get a complete picture of IVP traits that are controlled at different molecular levels and to improve the prediction of donor and recipient cow quality for in vitro embryo production.

Acknowledgments

The authors thank the Danish Innovation Fund, which funded the GIFT (Genomic Improvement of Fertilization Traits) project (grant number 0603-00509B and project website: www.gift.dk) and the São Paulo Research Foundation (FAPESP ; grants #2012/50533-2 and 2014/07566-2) and CNPq.

References


Simianer H. 2016. Genomic and other revolutions: why
some technologies are quickly adopted and others are not. Anim. Front, 6:53-58.


