Creating new knowledge for ruminant reproduction from rapidly expanding and evolving scientific databases

S Bauersachs, H Blum, S Krebs, T Fröhlich, GJ Arnold and E Wolf

Laboratory for Functional Genome Analysis (LAFUGA), Gene Center, Ludwig-Maximilians-Universität München, Feodor-Lynen-Str. 25, D-81377 Munich, Germany

Declining fertility is a major problem for the dairy industry. Recent developments of Omics-technologies facilitate a comprehensive analysis of molecular patters in gametes, embryos and tissues of the reproductive tract which may help to identify the reasons for impaired fertility. Large Omics-datasets require appropriate bioinformatics analysis in the context of rapidly expanding and evolving scientific databases. This overview summarizes the current status of ruminant genome projects, describes currently existing resources for ruminant genomics, transcriptomics and proteomics as well as databases and tools for the interpretation and exploitation of transcriptomics and proteomics datasets. Gene set enrichment analysis (GSEA) and transcription factor binding site (TFBS) analyses are strategies for the identification of regulatory genes. In general, the comprehensive analysis of molecular traits by Omics-technologies can enhance the interpretation of genome-wide association studies, providing insights into the biological pathways linking genotype and phenotype, and their modulation by endogenous and environmental factors.

Introduction

Reproductive success is a key component of economic production with ruminants, affecting both productivity and genetic progress. The decrease in dairy cattle fertility is a worldwide problem and a major cause of cow culling and economic losses. It is widely accepted that there is a strong association between high milk production and low fertility in dairy herds; however the reasons for the negative relationship between these traits remain to be resolved. On the one hand, metabolic problems that may be associated with high yield can influence molecular pathways controlling fertility at different levels and in various organs of the reproductive tract. On the other hand, preferential selection for production traits in the past may have led to genotypes in dairy cattle that are suboptimal for reproductive competence (Lucy 2001).

Reproductive success is determined by a cascade of biological processes: maturation and selection of gametes, fertilization, pre- and post-implantation embryonic development, fetal growth regulation, birth and early postnatal development of offspring. Among those, a reduction in fertilization and embryonic survival rates has been suggested as the most important component for decreasing reproductive efficiency in dairy cattle (Santos et al. 2004). Holistic and sensitive Omics-technologies characterizing the transcriptome, proteome, metabolome etc. of cells or tissues facilitate a comprehensive description of molecular patterns of gametes, embryos and

Corresponding author E-mail: ewolf@lmb.uni-muenchen.de

their maternal environment. For instance, our previous studies of bovine endometrium revealed characteristic transcriptome changes during the estrous cycle (Bauersachs et al. 2005; Mitko et al. 2008) and during early pregnancy (Bauersachs et al. 2006; Klein et al. 2006). Interestingly, we and others observed a different response of the endometrium to cloned vs. fertilized embryos (Bauersachs et al. 2009; Mansouri-Attia et al. 2009), suggesting the endometrium as a sensor for embryo quality and disturbed embryo-maternal communication in the peri-implantation period as a reason for structural and functional alterations of the placenta in clone pregnancies. Thus, changes of molecular patterns during development may point to genes or pathways that have an effect on reproductive success. Consequently, molecular patterns identified by Omics-technologies in organs and tissues that are relevant for reproduction can be viewed as "intermediate phenotypes" (Schadt 2009) of fertility, whose comprehensive description, interpretation and modeling may help to understand the genetic basis of cellular functions that are important for fertility (Figure 1).

This article provides an overview of the currently existing databases and bioinformatics tools that can be used to interpret Omics-data in the context of reproduction and other traits of ruminants.

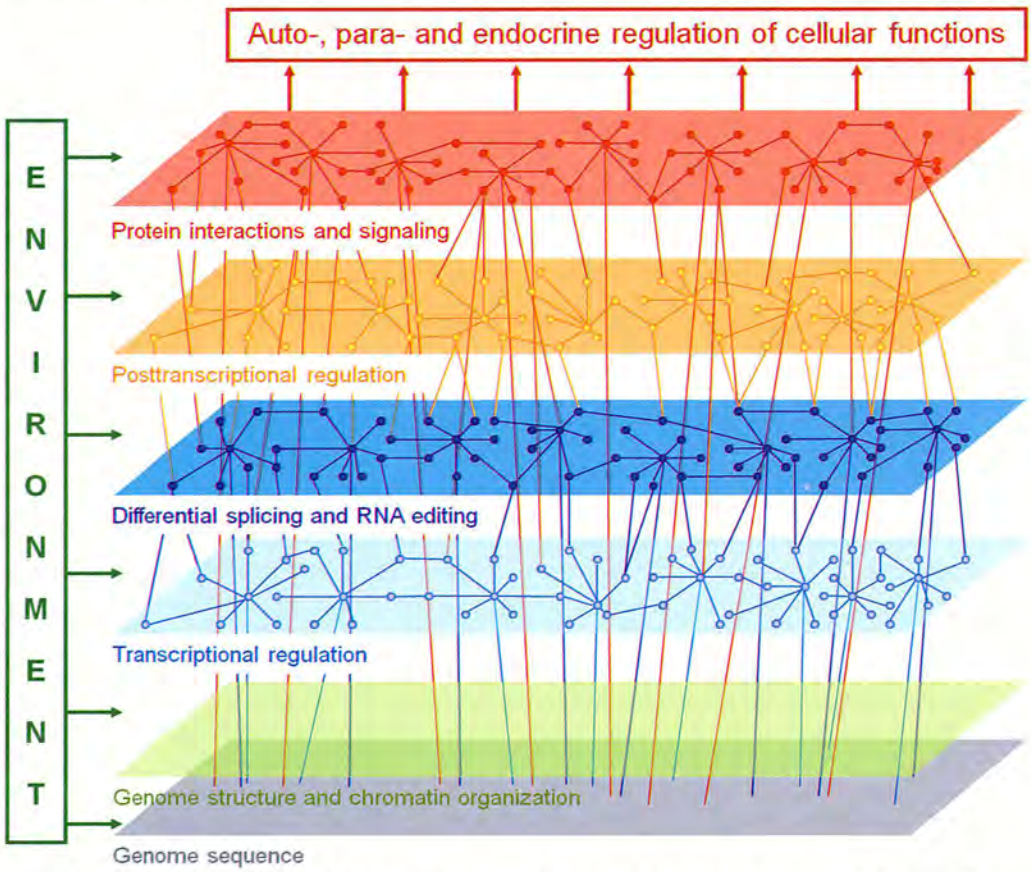

Fig. 1. Flow of genetic information via different classes of molecules producing molecular patterns and networks which affect cellular and organ functions. Integrating large-scale, high-dimensional molecular and physiological data holds promise for defining the molecular networks that respond to genetic and environmental perturbations of the physiological functions. The different layers of information provide a hierarchy of intermediate phenotypes, RNA being the most proximal non-DNA species of all molecular entities in the cell. Complex epigenomic mechanisms and interactions between the different classes of molecules modulate the flow of genetic information into biological functions.

Current status of ruminant genome projects

The largest current publicly available sequence database is GenBank located at the NCBI and contains nucleotide sequences of more than 300,000 organisms (Benson et al. 2009). GenBank is a collaborative effort between the International Nucleotide Sequence Database Collaboration (INSDC) of the NCBI, the DNA Data Bank of Japan (DDBJ) and the European Molecular Biology Laboratory (EMBL) Nucleotide Sequence Database (EMBL-Bank) at the European Bioinformatics Institute (EBI). Within this collaboration databases are synchronized by exchange and update of data. Founded in 1982, the content of GenBank has been growing exponentially, doubling approximately every 18 months (Figure 2). It currently comprises nearly $120*10^9$ nucleotides (GenBank release 177, April 2010). The generation of nucleotide sequence data was significantly accelerated by the development of powerful strategies for whole genome shotgun sequencing (WGS). The content of the WGS database grew faster than that of GenBank and exceeded it within two years (Figure 2). Among the twenty most sequenced organisms in GenBank Release 177 *Bos taurus* is the only ruminant species with a finished genome sequence, listed with roughly 5 gigabases of DNA/RNA sequences (Table 1) and 1.56 millions of expressed sequence tag (EST) records. The current assembly of the *Bos taurus* genome, Btau_4.0, is based on whole genome shotgun sequencing with 7× sequence coverage and comprises the sequence of all chromosomes except for the Y chromosome. The current release of GenBank contains no annotated genomes of other ruminants and only relatively small numbers of mRNA sequence data of other ruminants (e.g. mRNA entries: sheep: 3055, water buffalo: 695). This situation might be improved by the efforts of International Sequencing Consortia to sequence the Y chromosome of *Bos taurus* and the ovine genome (http://www.intlgenome.org/viewDatabase.cfm; Table 2).

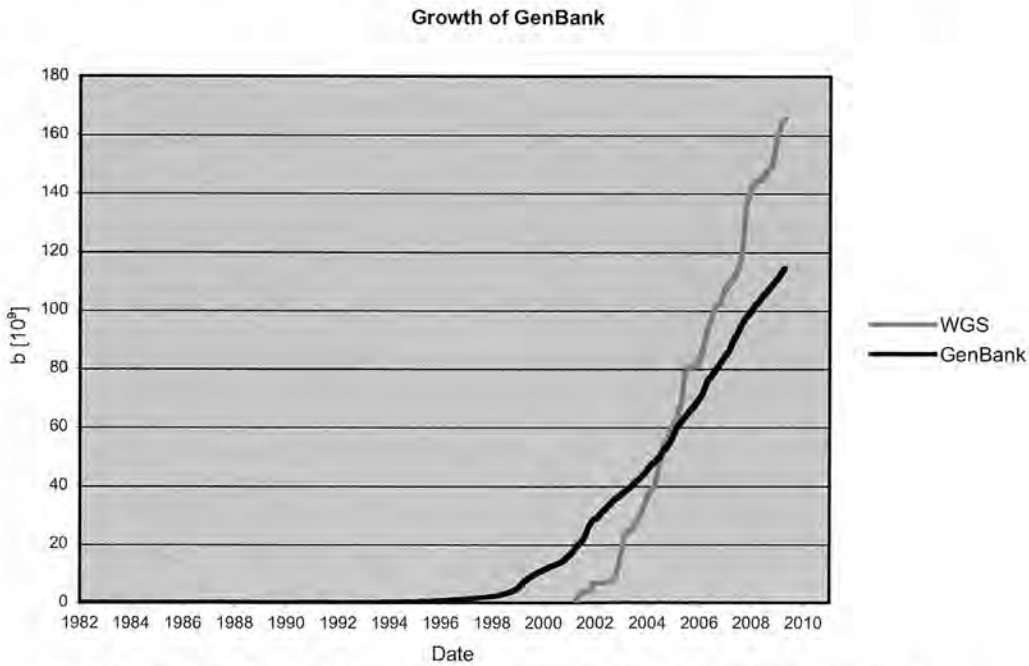

Fig. 2. Growth of GenBank. The number of bases in GenBank between 1982 and 2010 is blotted as black line, whereas the gray line displays the number of bases from whole genome shot gun sequencing projects. The data are derived from the distribution release note of GenBank release 177.

Table 1. Twenty most sequenced organisms in GenBank Release 177 (April 2010)

Entries	Bases*	Species
14661813	14675829480	*Homo sapiens*
7864719	8850187116	*Mus musculus*
2000118	6283974613	*Rattus norvegicus*
2179964	5355764557	*Bos taurus*
3890820	5034163602	*Zea mays*
3217695	4778268370	*Sus scrofa*
1697206	3052524947	*Danio rerio*
228209	1352880670	*Strongylocentrotus purpuratus*
1241806	1195864179	*Oryza sativa Japonica Group*
1753889	1185779622	*Nicotiana tabacum*
1423873	1146958794	*Xenopus (Silurana) tropicalis*
1205719	1043367600	*Drosophila melanogaster*
213748	999755722	*Pan troglodytes*
2286014	993961755	*Arabidopsis thaliana*
1435337	932443900	*Canis lupus familiaris*
655957	911105678	*Vitis vinifera*
808694	889675519	*Gallus gallus*
1840687	864995575	*Glycine max*
79871	816072773	*Macaca mulatta*
1216189	748184685	*Ciona intestinalis*

*: DNA/RNA, excluding chloroplast and mitochondrial sequences, metagenomic sequences, Whole Genome Shotgun sequences, and 'constructed' CON-division sequences. Ovine sequences are not included.

With the development of next-generation sequencing technologies the time-consuming process of generation of genome sequences has been dramatically shortened. Recently, a first draft sequence of the ovine genome sequence has been made available with a limited annotation that was produced by 454 sequencing and mapping the sheep sequences to the bovine genome sequence reordered according to the virtual sheep genome (www.sheephapmap.org/, www.livestockgenomics.csiro.au/). Furthermore, a low-coverage assembly of the alpaca (*Vicugna pacos*) is available that is actually not a ruminant but is similar to ruminants in its physiology and is thus generally classified as pseudo-ruminant. Recently, the Beijing Genomics Institute (BGI) has announced, that the genome of the Tibetan Antelope, a wild ruminant, has been finished (http://www.genomics.cn/en/search_show.php?type=show&id=530). Furthermore additional ruminant genomes might be sequenced within the Genome 10K project which plans whole-genome sequencing of 10.000 vertebrate species (http://www.genome10k.org/) (Genome 2009).

Current resources for ruminant genomics, transcriptomics and proteomics

The bovine reference genome is the first finished and fully annotated ruminant genome sequence (Elsik et al. 2009). Data from the bovine genome sequencing project as well as other

genome sequencing projects are available through a number of genome browsers (e.g. Ensembl and UCSC genome browser) and genome project databases (Table 2). Comparative genome alignments with other vertebrate genomes are available through these genome browsers, providing a source for comparative genome annotation. A specialized tool for visualizing the results of the mammalian genome comparative analysis is Evolution Highway that was used to decode the bovine genome evolutionary history (Elsik et al. 2009). In the context of genetic studies comparative genome annotation is of great value for positional cloning of bovine QTL. The Bovine Genome Database (http://bovinegenome.org) provides - in addition to the bovine genome sequence - up to date gene models and annotations and integration of physical and linkage maps with sequence, QTL, and SNP data. The investigation of variations in the genome is of major importance for animal breeding. For cattle and sheep there are large-scale ongoing SNP genotyping projects for genomic selection and HapMap development (bovinehapmap.org, sheephapmap.org). Publically available data in the context of QTL and GWAS analysis can be obtained for example from the Animal QTL database, the Bovine QTL viewer, and the USDA MARC website (see Table 2). Data derived from gene expression analyses can be found in public functional genomics data repositories such as Gene Expression Omnibus (GEO) and ArrayExpress (Table 2).

An important point for transcriptome and proteome analyses is a comprehensive gene annotation. Annotation data for protein-coding and non-coding transcripts and protein sequences can be downloaded, e.g. via Ensembl's BioMart. Genome annotation pipelines are using different strategies for gene annotation based on the alignment of sequences from public sequence databases, comparative alignment of Ensembl human and mouse proteins, and *ab initio* gene predictions. Another strategy to obtain full-length transcript and protein sequences is the generation of clusters of ESTs and full-length mRNA sequences derived from the same gene locus (UniGene, see Table 2). There are also attempts to find orthologous genes of several completely sequenced eukaryotic genomes (HomoloGene, see Table 2). In order to provide a defined functional description and classification of genes the Gene Ontology (GO) project has developed three structured controlled vocabularies (ontologies) that describe gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner. Genes are assigned based on data from the literature, on belonging to a known protein family, but also merely based on the presence of conserved protein domains. For ruminants only bovine genes are included in the GO classification. Likewise, only bovine genes are assigned to Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic, signaling and disease pathways.

With the discovery of small regulatory RNAs or so-called microRNAs (miRNAs) a database containing all known miRNA sequences was developed (mirBase, Table 2). In the current release (mirBase 15) there are 665 bovine and 4 ovine miRNAs. However, as many miRNAs are highly conserved between species annotations from other mammals can be used for ruminants where no annotation is available.

In addition to the information based on nucleic acid sequences and abundance, availability of data generated on the protein level is indispensable, since in all biological systems a broad spectrum of regulation phenomena occur on the protein level (e.g., secretion, activation of protein precursors by protein cleavage or phosphorylation, feedback inhibition, translational regulation, etc.). These crucial events are not mirrored by mRNA abundances, and their analysis is therefore addressable exclusively on the protein level. Moreover, the existence of mRNA in a tissue or cell type does not provide unequivocal evidence for the presence of the corresponding protein. As a consequence of substantial efforts in high resolution mass spectrometry (MS) of proteins and peptides, identification and quantification of thousands of proteins has become feasible and affordable during the last decade (for review, see Frohlich & Arnold 2006).

Table 2. Genomic resources for ruminant research

Resource	Content	Web site/Source
NCBI Genomic Biology	Links to genomic biology tools and resources for Bos taurus, Bubalus bubalis, Capra hircus, Ovis aries	www.ncbi.nlm.nih.gov/Genomes/
NCBI Entrez Genome Project database	Collection of complete and incomplete (in-progress) large-scale sequencing, assembly, annotation, and mapping projects for cellular organisms	www.ncbi.nlm.nih.gov/sites/entrez?db=genomeprj
Ensembl Genome Browser	Genome annotation for Bos taurus and Vicugna pacos	www.ensembl.org/Bos_taurus/Info/Index
UCSC Genome Browser	Genome annotation for Bos taurus	genome.ucsc.edu/cgi-bin/hgGateway?hgsid=158064562&clade=mammal&org=Cow&db=0
BioMart	Annotation tool	www.ensembl.org/biomart/martview
Entrez Gene	Searchable database of genes, from RefSeq genomes, and defined by sequence and/or located in the NCBI Map Viewer	www.ncbi.nlm.nih.gov/gene
HomoloGene	System for automated detection of homologs among the annotated genes of several completely sequenced eukaryotic genomes	www.ncbi.nlm.nih.gov/homologene
UniGene	UniGene entries are sets of transcript sequences that appear to come from the same transcription locus (gene or expressed pseudogene), together with information on protein similarities, gene expression, cDNA clone reagents, and genomic location	www.ncbi.nlm.nih.gov/sites/entrez?db=unigene
miRBase	Bovine and ovine micro RNAs	www.mirbase.org/cgi-bin/mirna_summary.pl?org=bta
Gene Expression Omnibus	Database for curated gene expression datasets	www.ncbi.nlm.nih.gov/geo
ArrayExpress Archive	Database of functional genomics experiments including gene expression	www.ebi.ac.uk/microarray-as/ae
Bos taurus genome project	Bovine Genome Database project	http://bovinegenome.org
Bovine HapMap project	Large-scale bovine single nucleotide polymorphism (SNP) genotyping for genomic selection and HapMap development	http://bovinehapmap.org
Animal Quantitative Trait Locus (QTL) database	Houses all publicly available QTL data on livestock animal species for easily locating and making comparisons within and between species	www.genome.iastate.edu/cgi-bin/QTLdb/index
Bovine QTL viewer	Contains all available public domain bovine QTL data for both dairy and beef traits	genomes.sapac.edu.au/bovineqtl
USDA MARC	Cattle and sheep genome maps	www.ars.usda.gov/main/docs.htm?docid=2340
Evolution Highway	Collaborative project designed to provide a visual means for simultaneously comparing genomes of multiple amniote species	http://evolutionhighway.ncsa.uiuc.edu
Ovis aries genome project	Home page of the International Sheep Genomics Consortium	www.sheephapmap.org
CSIRO livestock genomics web site	Access to data generated by genomics projects for major livestock species, major focus on the cattle and sheep genome mapping and sequencing projects, provides access to interactive genome maps of cattle and sheep and to the results of the Bacterial Artificial Chromosome (BAC) library contiging project for cattle	www.livestockgenomics.csiro.au

Due to the bioinformatics algorithms used for MS-based protein identification, comprehensive genomic databases of the organisms analyzed (or at least a closely related species) are indispensable prerequisites for protein identification. Hence, the availability of the bovine genome database (http://www.ncbi.nlm.nih.gov, Elsik et al. 2009) represents a milestone with respect to protein identifications in ruminants. Whole genome sequencing and annotation of sheep and other ruminants are currently in progress and will further stimulate the generation of data on the protein level.

Through various proteomic approaches performed during the last years, a tremendous amount of data has been generated, giving rise to the generation of databases focused on protein data. The most prominent of these is the "UniProt" database (***uni***versal ***prot***ein; www.uniprot.org), representing the largest collection of protein information from a broad variety of organisms and viruses. Uniprot combines data from Swiss-Prot, TrEMBL (Translated EMBL Nucleotide Sequence Data Library) (Boeckmann et al. 2003) and PIR (**P**rotein **I**nformation **R**esource, Wu et al. 2006) and is updated frequently.

Besides pure sequence data, UniProt provides information about post-translational modifications and functional aspects as well as relevant links to structural and gene ontology databases. A special feature of the Swiss-Prot part of UniProt is the manual annotation and review process performed by experts in the field, thereby providing a superior level of data reliability and relevance as compared to datasets automatically generated by unsupervised, computer based data mining.

To facilitate public access to the huge number of datasets generated in proteomic approaches, several databases are currently established containing raw data from mass spectrometry based protein identifications (peak lists, intensities, identified peptides etc.). Two prominent examples are the PRIDE database (**PR**oteomics **IDE**ntifications, Vizcaino et al. 2009) managed by the EBI at the EMBL and the "Peptide Atlas" (Deutsch et al. 2008) managed by the Seattle Proteome Center, containing data from 60 and 12 species (releases May 2010), respectively. Data from ruminants are contained only in PRIDE, currently comprising 66 different proteins from *Bos taurus* (PRIDE core version 2.8.0).

The benefit of these databases lies in the possibility to quickly gain information about the level of expression of proteins in individual cells, tissues or body fluids. Moreover, post-translational modifications identified by mass spectrometry can frequently be assigned to distinct amino acid positions of the protein. As a special advantage, the public availability of mass spectrometry raw data facilitates their re-analysis using rapidly evolving new algorithms.

The low number of ruminant data so far contained in PRIDE reflects both the lack of obligation to deposit raw data of proteomic experiments in public databases along with publication in scientific journals as well as the rather initial status of proteome research in farm animals. Along with ongoing new guidelines for proteomic data publication and exciting developments in protein analysis, e.g., mass spectrometry based quantification of proteins using SRM (**S**elected **R**eaction **M**onitoring) technology (Lange et al. 2008; Picotti et al. 2010), protein databases will provide a widely applicable source of information in the fields of basic research as well as veterinary medicine and animal reproduction.

Databases and tools to exploit transcriptomics and proteomics datasets for ruminants

As already mentioned the GO database provides a defined description of genes regarding the categories "Biological Process", "Molecular Function", and "Cellular Component". There are numerous tools for the analysis of GO terms associated with a list of differentially expressed transcripts or proteins (see Table 3). Many of these tools provide quantitatively enriched GO

terms associated with a gene list, i.e. GO terms for which significantly more associated genes were found than expected by chance. The processing of the results of such analyses can be very laborious due to the redundant structure of the GO categories. The "Functional Annotation Clustering" tool of the Database for Annotation, Visualization and Integrated Discovery (DAVID) (Dennis et al. 2003) reduces this problem by clustering enriched functional categories (from GO and other databases) that have overlapping gene contents.

Table 3. Tools for the bioinformatics analysis of Omics-data

Tool	Function	Web site/Source
Gene Ontology	provides a controlled vocabulary of terms for describing gene product characteristics and gene product annotation data	geneontology.org/
GO tools for analysis of gene expression/microarray datasets	overview of tools for Gene Ontology analysis of lists of differentially expressed genes/mRNAs or proteins	geneontology.org/GO.tools.microarray.shtml#goarray
Gene Set Enrichment Analysis (GSEA)	computational method that determines whether an a priori defined set of genes shows statistically significant, concordant differences between two biological states (e.g. phenotypes)	www.broadinstitute.org/gsea/
Database for Annotation, Visualization and Integrated Discovery (DAVID)	functional annotation, functional classification, gene ID conversion, gene name batch viewer	david.abcc.ncifcrf.gov/
CoPub text mining tool	text mining tool for detection of co-occuring biomedical concepts in abstracts from the Medline literature database significantly linked to a differential gene set	services.nbic.nl/cgi-bin/copub3/CoPub.pl
oPOSSUM	detection of over-represented transcription factor binding sites in the promoters of sets of genes	www.cisreg.ca/oPOSSUM/
KEGG pathway database	collection of manually drawn pathway maps	www.genome.jp/kegg/pathway.html
NCI/Nature Pathway Interaction Database	biomolecular interactions and cellular processes assembled into authoritative human signaling pathways	pid.nci.nih.gov/index.shtml
STRING	database and analysis tool for known and predicted protein-protein interactions	string.embl.de/newstring_cgi/show_input_page.pl?UserId=JE8TU8ELDuNa&sessionId=I1hDAbLNVFLZ
Cytoscape	open source bioinformatics software platform for visualizing molecular interaction networks and integrating these interactions with gene expression profiles and other state data	www.cytoscape.org/

A different strategy for the analysis of lists of differentially expressed genes or proteins is text-mining of PubMed abstracts with CoPub to analyze gene-gene co-citation and co-citation of genes with keywords (Table 3). This analysis identifies biological or disease-related keywords overrepresented within the differentially expressed genes.

Microarray datasets can also be characterized by comparison with gene sets derived from other gene expression studies or from defined functional categories by the use of "Gene set enrichment analysis" (GSEA) (Subramanian et al. 2005). GSEA compares a gene expression dataset with different collections of gene sets: positional gene sets, curated gene sets, motif gene sets, computational gene sets, and GO gene sets (for detailed explanation see www.broadinstitute.org/gsea/msigdb/index.jsp). The genes of an expression dataset are ranked according to differential expression with the most significantly up-regulated genes at the top and the most significantly down-regulated genes at the bottom. Based on the positions of the genes of the gene sets in the ranked gene expression dataset enrichment towards an end of the ranked list is calculated that indicates concordance of the gene set with the gene expression dataset. User-provided gene sets can also be used for comparison with the expression dataset. GSEA results can be helpful for example in concluding from regulatory mechanisms that are known for a given gene set or from gene sets which belong to defined functional categories or cellular pathways (Figure 3).

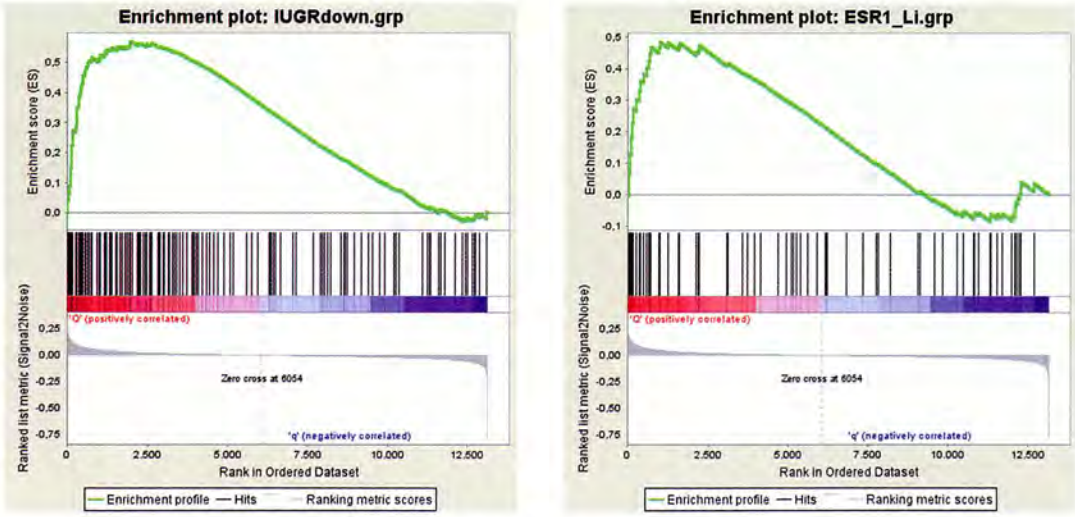

Fig. 3. Gene Set Enrichment Analysis. A set of 161 genes associated with human intrauterine growth retardation was tested for enrichment in a dataset of gene expression linked to the presence of a QTL haplotype for stillbirth risk (due to fetal overgrowth) in cattle (left). The enrichment result confirms that the differentially expressed QTL-associated genes are functionally linked to fetal growth. The same dataset was tested with a set of 67 genes reported to be estrogen-responsive in cattle (right). The enrichment result suggests involvement of estrogen-regulated genes and thus confirms *ESR1* as a functional candidate gene for the stillbirth QTL.

A further strategy to identify regulatory mechanisms underlying the observed gene expression changes is the identification of potential transcription factors by the analysis of regulatory motifs in the promoter regions of the differentially expressed genes, e.g. using oPOSSUM (Ho Sui et al. 2007). Unfortunately, this tool is available only for human, mouse and rat, i.e. the analysis in other species assumes conserved regulatory elements in the corresponding promoter regions.

Nevertheless, a comparative study of human and bovine transcription factor binding sites (TFBS) (Zadissa et al. 2007) encourages the use of human promoter databases for the inference of bovine gene regulation. In our recent study on local and systemic responses of the bovine mammary gland to experimental infection with a pathogenic *Escherichia coli* strain, TFBS analysis suggested NFkB and STAT1 as regulators of genes involved in immune response, inflammation, acute phase response and chemokine/cytokine signaling, which were differentially expressed locally in the infected quarter of the mammary gland (Mitterhuemer et al. 2010).

Finally, the potential interactions of the identified differentially expressed genes or proteins among themselves and with other genes or proteins could be of interest. There are also a number of different databases and tools for interaction analyses, for example STRING, a searchable database for known and predicted protein-protein interactions (Jensen et al. 2009).

Conclusions and perspectives

Holistic and sensitive Omics-technologies characterizing the transcriptome, proteome, metabolome and other molecular characteristics of cells or tissues facilitate the comprehensive description of molecular patterns of tissues that are associated with particular physiological or pathophysiological conditions. Importantly, dynamic changes of these patterns during development or disease may point to genes or pathways that have an effect on the trait under investigation. It is clear that the rapidly expanding scientific databases described in this article will help to dissect mechanisms of reproductive physiology and other important traits of ruminants at the molecular level.

In addition to the development and implementation of Omics-phenotypes, a refinement of physiological readouts is urgently required. Those can be obtained, for example, by the development of non-invasive longitudinal techniques such as remote/indirect sensing or imaging. Integrating large-scale, high-dimensional molecular and physiological data holds promise for defining the molecular networks that respond to genetic and environmental perturbations of physiological functions, including reproduction.

Acknowledgments

Our studies were/are supported by the Deutsche Forschungsgemeinschaft (FOR478, FOR1041), the Bundesministerium für Bildung und Forschung (FUGATO-projects FERTILINK, REMEDY and COMPENDIUM), and by the European Union (SABRE, ERIN, PluriSys).

References

Bauersachs S, Ulbrich SE, Gross K, Schmidt SEM, Meyer HHD, Einspanier R, Wenigerkind H, Vermehren M, Blum H, Sinowatz F & Wolf E 2005 Gene expression profiling of bovine endometrium during the oestrous cycle: detection of molecular pathways involved in functional changes. *Journal of Molecular Endocrinology* **34** 889-908.

Bauersachs S, Ulbrich SE, Gross K, Schmidt SEM, Meyer HHD, Wenigerkind H, Vermehren M, Sinowatz F, Blum H & Wolf E 2006 Embryo-induced transcriptome changes in bovine endometrium reveal species-specific and common molecular markers of uterine receptivity. *Reproduction* **132** 319-331.

Bauersachs S, Ulbrich SE, Zakhartchenko V, Minten M, Reichenbach M, Reichenbach H-D, Blum H, Spencer TE & Wolf E 2009 The endometrium responds differently to cloned versus fertilized embryos. *Proceedings of the National Academy of Sciences of the United States of America* **106** 5681-5686.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J & Sayers EW 2009 GenBank. *Nucleic Acids Research* **38** D46-51.

Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S & Schneider M 2003

The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research* **31** 365-370.

Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC & Lempicki RA 2003 DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biology* **4** P3.

Deutsch EW, Lam H & Aebersold R 2008 PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO Reports* **9** 429-434.

Elsik CG, Tellam RL, Worley KC, Gibbs RA, Muzny DM, Weinstock GM, Adelson DL, Eichler EE, Elnitski L, Guigo R, Hamernik DL, Kappes SM, Lewin HA, Lynn DJ, Nicholas FW, Reymond A, Rijnkels M, Skow LC, Zdobnov EM, Schook L, Womack J, Alioto T, Antonarakis SE, Astashyn A, Chapple CE, Chen HC, Chrast J, Camara F, Ermolaeva O, Henrichsen CN, Hlavina W, Kapustin Y, Kiryutin B, Kitts P, Kokocinski F, Landrum M, Maglott D, Pruitt K, Sapojnikov V, Searle SM, Solovyev V, Souvorov A, Ucla C, Wyss C, Anzola JM, Gerlach D, Elhaik E, Graur D, Reese JT, Edgar RC, McEwan JC, Payne GM, Raison JM, Junier T, Kriventseva EV, Eyras E, Plass M, Donthu R, Larkin DM, Reecy J, Yang MQ, Chen L, Cheng Z, Chitko-McKown CG, Liu GE, Matukumalli LK, Song J, Zhu B, Bradley DG, Brinkman FS, Lau LP, Whiteside MD, Walker A, Wheeler TT, Casey T, German JB, Lemay DG, Maqbool NJ, Molenaar AJ, Seo S, Stothard P, Baldwin CL, Baxter R, Brinkmeyer-Langford CL, Brown WC, Childers CP, Connelley T, Ellis SA, Fritz K, Glass EJ, Herzig CT, Iivanainen A, Lahmers KK, Bennett AK, Dickens CM, Gilbert JG, Hagen DE, Salih H, Aerts J, Caetano AR, Dalrymple B, Garcia JF, Gill CA, Hiendleder SG, Memili E, Spurlock D, Williams JL, Alexander L, Brownstein MJ, Guan L, Holt RA, Jones SJ, Marra MA, Moore R, Moore SS, Roberts A, Taniguchi M, Waterman RC, Chacko J, Chandrabose MM, Cree A, Dao MD, Dinh HH, Gabisi RA, Hines S, Hume J, Jhangiani SN, Joshi V, Kovar CL, Lewis LR, Liu YS, Lopez J, Morgan MB, Nguyen NB, Okwuonu GO, Ruiz SJ, Santibanez J, Wright RA, Buhay C, Ding Y, Dugan-Rocha S, Herdandez J, Holder M, Sabo A, Egan A, Goodell J, Wilczek-Boney K, Fowler GR, Hitchens ME, Lozado RJ, Moen C, Steffen D, Warren JT, Zhang J, Chiu R, Schein JE, Durbin KJ, Havlak P, Jiang H, Liu Y, Qin X, Ren Y, Shen Y, Song H, Bell SN, Davis C, Johnson AJ, Lee S, Nazareth LV, Patel BM, Pu LL, Vattathil S, Williams RL, Jr., Curry S, Hamilton C, Sodergren E, Wheeler DA, Barris W, Bennett GL, Eggen A, Green RD, Harhay GP, Hobbs M, Jann O, Keele JW, Kent MP, Lien S, McKay SD, McWilliam S, Ratnakumar A, Schnabel RD, Smith T, Snelling WM, Sonstegard TS, Stone RT, Sugimoto Y, Takasuga A, Taylor JF, Van Tassell CP, Macneil MD, Abatepaulo AR, Abbey CA, Ahola V, Almeida IG, Amadio AF, Anatriello E, Bahadue SM, Biase FH, Boldt CR, Carroll JA, Carvalho WA, Cervelatti EP, Chacko E, Chapin JE, Cheng Y, Choi J, Colley AJ, de Campos TA, De Donato M, Santos IK, de Oliveira CJ, Deobald H, Devinoy E, Donohue KE, Dovc P, Eberlein A, Fitzsimmons CJ, Franzin AM, Garcia GR, Genini S, Gladney CJ, Grant JR, Greaser ML, Green JA, Hadsell DL, Hakimov HA, Halgren R, Harrow JL, Hart EA, Hastings N, Hernandez M, Hu ZL, Ingham A, Iso-Touru T, Jamis C, Jensen K, Kapetis D, Kerr T, Khalil SS, Khatib H, Kolbehdari D, Kumar CG, Kumar D, Leach R, Lee JC, Li C, Logan KM, Malinverni R, Marques E, Martin WF, Martins NF, Maruyama SR, Mazza R, McLean KL, Medrano JF, Moreno BT, More DD, Muntean CT, Nandakumar HP, Nogueira MF, Olsaker I, Pant SD, Panzitta F, Pastor RC, Poli MA, Poslusny N, Rachagani S, Ranganathan S, Razpet A, Riggs PK, Rincon G, Rodriguez-Osorio N, Rodriguez-Zas SL, Romero NE, Rosenwald A, Sando L, Schmutz SM, Shen L, Sherman L, Southey BR, Lutzow YS, Sweedler JV, Tammen I, Telugu BP, Urbanski JM, Utsunomiya YT, Verschoor CP, Waardenberg AJ, Wang Z, Ward R, Weikard R, Welsh TH, Jr., White SN, Wilming LG, Wunderlich KR, Yang J & Zhao FQ 2009 The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* **324** 522-528.

Frohlich T & Arnold GJ 2006 Proteome research based on modern liquid chromatography-tandem mass spectrometry: separation, identification and quantification. *Journal of Neural Transmission* **113** 973-994.

Genome KCoS 2009 Genome 10K: a proposal to obtain whole-genome sequence for 10,000 vertebrate species. *Journal of Heredity* **100** 659-674.

Ho Sui SJ, Fulton DL, Arenillas DJ, Kwon AT & Wasserman WW 2007 oPOSSUM: integrated tools for analysis of regulatory motif over-representation. *Nucleic Acids Research* **35** W245-252.

Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P & von Mering C 2009 STRING 8–a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Research* **37** D412-416.

Klein C, Bauersachs S, Ulbrich SE, Einspanier R, Meyer HHD, Schmidt SEM, Reichenbach H-D, Vermehren M, Sinowatz F, Blum H & Wolf E 2006 Monozygotic twin model reveals novel embryo-induced transcriptome changes of bovine endometrium in the preattachment period. *Biology of Reproduction* **74** 253-264.

Lange V, Picotti P, Domon B & Aebersold R 2008 Selected reaction monitoring for quantitative proteomics: a tutorial. *Molecular Systems Biology* **4** 222.

Lucy MC 2001 Reproductive loss in high-producing dairy cattle: where will it end? *Journal of Dairy Science* **84** 1277-1293.

Mansouri-Attia N, Sandra O, Aubert J, Degrelle S, Everts RE, Giraud-Delville C, Heyman Y, Galio L, Hue I, Yang X, Tian XC, Lewin HA & Renard JP 2009 Endometrium as an early sensor of in vitro embryo manipulation technologies. *Proceedings of the National Academy of Sciences of the United States of America* **106** 5687-5692.

Mitko K, Ulbrich SE, Wenigerkind H, Sinowatz F, Blum H, Wolf E & Bauersachs S 2008 Dynamic changes in messenger RNA profiles of bovine endometrium

during the oestrous cycle. *Reproduction* **135** 225-240.

Mitterhuemer S, Petzl W, Krebs S, Mehne D, Klanner A, Wolf E, Zerbe H & Blum H 2010 Escherichia coli infection induces distinct local and systemic transcriptome responses in the mammary gland. *BMC Genomics* **11** 138.

Picotti P, Rinner O, Stallmach R, Dautel F, Farrah T, Domon B, Wenschuh H & Aebersold R 2010 High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nature Methods* **7** 43-46.

Santos JEP, Thatcher WW, Chebel RC, Cerri RLA & Galvao KN 2004 The effect of embryonic death rates in cattle on the efficacy of estrus synchronization programs. *Animal Reproduction Science* **82-83** 513-535.

Schadt EE 2009 Molecular networks as sensors and drivers of common human diseases. *Nature* **461** 218-223.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES & Mesirov JP 2005 Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102** 15545-15550.

Vizcaino JA, Cote R, Reisinger F, Foster JM, Mueller M, Rameseder J, Hermjakob H & Martens L 2009 A guide to the Proteomics Identifications Database proteomics data repository. *Proteomics* **9** 4276-4283.

Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Mazumder R, O'Donovan C, Redaschi N & Suzek B 2006 The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Research* **34** D187-191.

Zadissa A, McEwan JC & Brown CM 2007 Inference of transcriptional regulation using gene expression data from the bovine and human genomes. *BMC Genomics* **8** 265.