

Potential of functional genomics for reproductive medicine and biotechnology in domestic animals

Stefan Bauersachs and Eckhard Wolf

Institute of Molecular Animal Breeding and Biotechnology and Laboratory for Functional Genome Analysis (LAFUGA), Gene Center of the LMU Munich, Munich, Germany

Summary

The molecular mechanisms underlying maternal recognition and maintenance of pregnancy are highly complex and differ among mammalian species, including domestic animals. The advent of "omics" technologies facilitates systematic molecular analyses of the various steps in reproductive biology, providing the basis for the identification of key regulatory genes and pathways. Although full genome sequences for a number of domestic animals, like cattle, pig, and horse are already available in fairly high quality or are in progress, respectively, these sequence data will only provide a framework for functional studies of the genome. Microarrays are powerful tools to identify global messenger RNA expression changes associated with a variety of physiological and pathological processes in biology of reproduction. Recently, small regulatory RNAs, so-called microRNAs (miRNAs) have been shown to play important roles in the translational regulation of developmental and differentiation-related genes. The development of powerful tools for the analysis of miRNAs will facilitate future studies of their roles in biology of reproduction. Furthermore, the improvement of proteomics techniques will enable the direct analysis of the functional players – the proteins themselves. In cattle and sheep basic transcriptome and proteome analyses of the endometrium have been performed during the pre-attachment period and the oestrous cycle. These studies revealed hundreds of regulated genes reflecting the complexity of these reproductive processes. The results of such complex approaches will be the basis for the development of novel diagnostic tools and may provide fertility markers for selective breeding, and new strategies for improvement of female fertility.

Keywords: genomics, transcriptomics, bioinformatics, fertility, pregnancy, reproduction

Potential der funktionalen Genomanalyse für die Reproduktionsmedizin und -biotechnologie bei Haustieren

Die molekularen Mechanismen der maternalen Erkennung und Aufrechterhaltung einer Trächtigkeit sind sehr komplex und teilweise sehr unterschiedlich zwischen verschiedenen Säugerspezies und damit auch bei den Haustieren. Deshalb sind systematische Untersuchungsansätze von großer Bedeutung für ein umfassendes Verständnis dieser Vorgänge. Mit Hilfe solcher Ansätze können zentrale regulatorische Gene oder molekulare Pathways identifiziert werden, welche bei der maternalen Erkennung der Trächtigkeit, der Modulation des maternalen Immunsystems, der Sekretion von Nährstoffen für den frühen Embryo sowie für dessen Anheftung und Implantation eine Rolle spielen. Obwohl die Genomsequenzen für eine Reihe von Haustierspezies, wie beispielsweise Rind, Schwein und Pferd schon in relativ guter Qualität vorliegen bzw. deren Entzifferung weit fortgeschritten ist, stellen diese Sequenzdaten nur ein Grundgerüst für funktionale Studien zur Expression und Regulation des Genoms dar. In diesem Zusammenhang sind Microarray-Analysen besonders geeignet, globale Änderungen von mRNA-Spiegeln zu identifizieren, die mit verschiedenen physiologischen und pathophysiologischen Prozessen im Bereich der Reproduktion assoziiert sind. In den letzten Jahren wurde zudem gezeigt, dass kleine regulatorische RNAs, die so genannten microRNAs (miRNAs), auch in Säugern eine wichtige Rolle in der globalen Regulation von Genen spielen, welche Funktionen in der Entwicklungsbiologie und Gewebedifferenzierung haben. Die Regulation durch miRNAs erfolgt durch die Unterdrückung der mRNA-Translation. Dies stellt eine neue Dimension in der Kontrolle der Genexpression dar. Durch die Entwicklung leistungsfähiger Methoden zur Analyse der miRNAs können zukünftig deren potentielle Funktionen in reproduktionsbiologischen Prozessen untersucht werden. Eine Verbesserung der Technologien zur Proteomanalyse, wie beispielsweise hochsensitive Protein-Markierungsmethoden für die zweidimensionale Gelelektrophorese oder Massenspektrometrie-basierte Verfahren mit verbesserter Sensitivität, erlaubt es zudem in immer besserem Maße, die Proteine, welche die eigentlichen funktionalen Moleküle darstellen, quantitativ zu erfassen und zu identifizieren. Beim Rind und beim Schaf wurden bereits eine Reihe von grundlegenden Transkriptom- und auch einige wenige Proteomanalysen zu reproduktionsbiologischen Fragestellungen durchgeführt. Dabei wurde vor allem Endometrium in der Präimplantationsphase und auch in verschiedenen Phasen des Sexualzyklus untersucht. Diese Studien identifizierten Hunderte mRNAs, deren Konzentrationen sich während der frühen Trächtigkeit oder während des Zyklus ändern. Gegenwärtig sind wir dabei, mit solchen Experimenten auch beim Pferd zu beginnen. Die Ergebnisse dieser systematischen Untersuchungen liefern die Basis für die Entwicklung neuer diagnostischer Verfahren, stellen möglicherweise neue Parameter für die Selektion auf Fruchtbarkeit dar und dienen damit der Entwicklung neuer Strategien zur Verbesserung der weiblichen Fruchtbarkeit. Die Analyse qualitativ hochwertiger molekularer Datensätze von verschiedenen Spezies mit systembiologischen Ansätzen sollte langfristig zum Verständnis der Phylogenese reproduktionsbiologischer Strategien unserer Nutztiere beitragen.

Schlüsselwörter: Genomik, Transkriptomik, Bioinformatik, Reproduktion, Gravidität

With a delay of several years the spectrum of technologies required for functional genome analysis also entered research projects in domestic animal reproductive biology (Bazer and Spencer 2005, Hiendleder et al. 2005). The elucidation of the genome sequences for cattle, pig and horse is still in progress. Currently (state of May 2007), the bovine genome sequence

is most advanced with the preliminary assembly Btau_3.1 (www.ensembl.org/Bos_taurus/index.html) that has a total length of all contigs of 2.73 Gb. Considering the still existing gaps between contigs in scaffolds, the total span of the assembly is 2.87 Gb. For the horse there is a first preliminary assembly available (Equus1, pre.ensembl.org/Equus_ca-ballus/)

index.html). The total length of all contigs is 2.42 Gb. Taking the gaps between contigs in scaffolds into account, the total span of the assembly is 2.68 Gb, however, a full annotation is not yet available. For the pig, a preliminary assembly of the sequences of only four chromosomes has been performed (pre.ensembl.org/Sus_scrofa/index.html). Although in the near future full genome sequences of these three species will be available there is still a gap between the genomic sequences and the knowledge of complete mRNA transcripts. This is due to the problem of identifying the correct exon-intron structures of genes, which is particularly difficult for 5' and 3' untranslated regions. Mass sequencing of full-length cDNAs and ESTs would help to overcome this problem as it has been done in humans (Wiemann et al. 2001). Table 1 shows an overview of the current state of May 2007 content of the UniGene database for human, cattle, pig and horse. The UniGene database is an attempt to build sets of transcript sequences that appear to come from the same gene using procedures for automated sequence clustering. In addition to sequences of well-characterized genes, hundreds of thousands of novel expressed sequence tag (EST) sequences are included in this database (Wheeler et al. 2007). There are striking differences in the numbers of available sequences between the different species, with the lowest number being available for the horse.

With finishing of the genome sequences the work just now begins to explore the functions of the genes in the context of physiological and pathological processes. Among many different systematic approaches for characterization of gene functions, microarray analyses are mostly used to identify mRNAs, which differ in their concentration between two or more experimental samples, suggesting that the corresponding genes may be regulated in the context of the biological phenomenon under investigation. Such gene expression data are more and more deposited to gene expression/molecular abundance repositories like the NCBI Gene Expression Omnibus

(www.ncbi.nlm.nih.gov/geo/). With a search in the GEO database (May 2007) limited to "organism" 3,727 records were found for *Homo sapiens*, 72 for *Bos taurus*, 40 for *Sus scrofa*, and no record for *Equus caballus*. These figures indicate a clear positive correlation between the number of microarray experiments and the status of sequence data of the respective species (see Table 1). This correlation is due to the simple fact that sequence resources are needed for the production of microarrays either cDNA clones derived from cDNA libraries or at least in silico sequences for the deduction of oligonucleotides that can be synthetically produced on or spotted onto arrays. For cattle and pig oligonucleotide microarrays representing more than 15,000 genes are commercially available from Affymetrix. In addition, large cDNA arrays exist for these species (Suchyta et al. 2003). For the horse the first microarray was described in 2004 containing approx. 3,000 genes (Gu and Bertone 2004). So far there is no equine genome-wide microarray, which makes holistic transcriptome analyses more complicated than in cattle or pig.

Several years ago we faced the same problem in cattle. As we started to analyse mechanisms of embryo-maternal communication in the cow we decided to use a method that does not depend on known bovine cDNA sequences (Wolf et al. 2003). This technique is based on subtractive hybridisation (Diatchenko et al. 1996) to enrich cDNAs with different concentrations between two samples (e.g., endometrial samples from pregnant vs. non-pregnant cows). The enriched sequences are subsequently cloned in a subtracted library and screened for differential abundance by cDNA array hybridisation. A series of studies has been performed using this strategy to identify differentially expressed genes in the context of reproductive processes. Endometrium of day 18 pregnant vs. control animals has been analysed using a monozygotic twin model, where pregnancy was achieved by transfer of in vitro produced embryos (Klein et al. 2006) and in a second model, where cyclic heifers were

Table 1 Number of sequences and clusters in the UniGene database for *Homo sapiens*, *Bos taurus*, *Sus scrofa*, and *Equus caballus*. Anzahl an Sequenzen und Sequenz-Cluster in der UniGene-Datenbank für *Homo sapiens*, *Bos taurus*, *Sus scrofa*, und *Equus caballus*.

Species	<i>Homo sapiens</i>	<i>Bos taurus</i>	<i>Sus scrofa</i>	<i>Equus caballus</i>
Known genes are from	GenBank 01 Mar 2007	GenBank 25 Feb 2007	GenBank 27 Feb 2007	GenBank 24 Jan 2007
ESTs are from	dbEST through 01 Mar 2007	dbEST through 25 Feb 2007	dbEST through 27 Feb 2007	dbEST through 24 Jan 2007
mRNAs	172,323	20,531	5,060	804
Models (predicted from genome)	2,720	12,902	0	0
HTC (High-throughput cDNA sequences)	47,759	333	11,295	48
EST, 3'reads	1,680,320	222,187	150,941	4,623
EST, 5'reads	3,893,685	636,643	329,389	5,014
EST, other/unknown	898,026	319,140	42,433	17,612
Total sequences in clusters	6,694,833	1,211,736	539,118	28,101
Build Method:	Genome Based	Genome Based	Transcript Based	Transcript Based
UniGene Build #	201	83	28	1
Final Number of Clusters (sets):				
Sets total	124,179	36,421	37,938	5,499
Sets contain at least one mRNA	34,634	9,879	2,284	604
Sets contain at least one HTC sequence	13,085	309	6,951	37
Sets contain at least one EST	117,940	35,880	37,194	4,929
Sets contain both mRNAs and ESTs	29,000	9,346	1,592	34

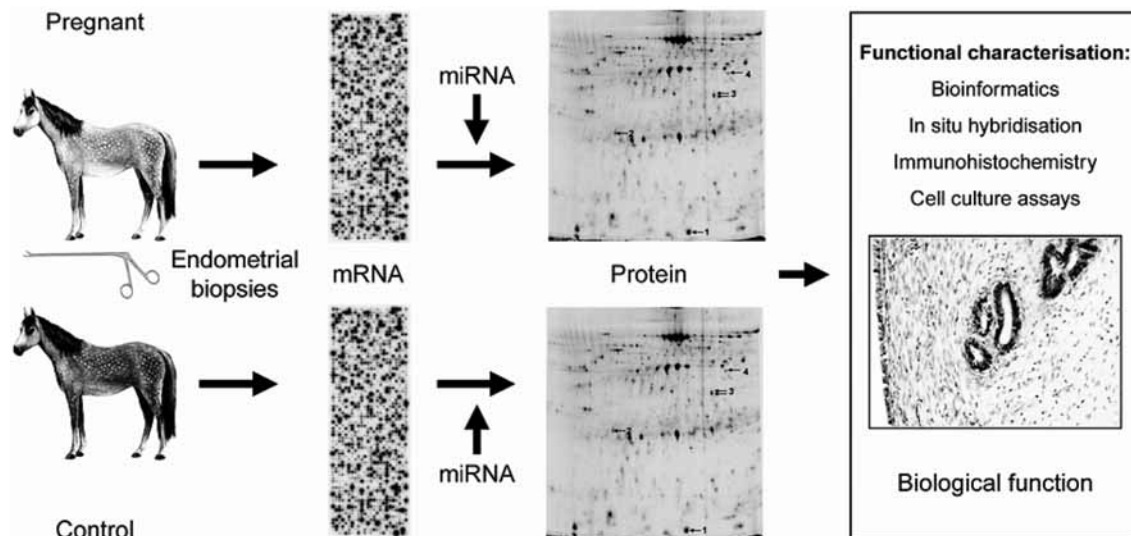


Fig 1 Strategy for functional genome analysis of early pregnancy in mares. Pregnancy recognition in the endometrium results in distinct changes of gene activation patterns. These changes are reflected by differences in mRNA concentrations between pregnant mares and corresponding controls, which can be identified using microarrays. Gene expression can also be regulated at the level of translational control. MicroRNAs play a pivotal role in this process. Differences at the protein level can be analysed with 2D-DIGE- or LC-MS-based approaches. Various bioinformatics tools can be used for data mining to identify key regulatory genes, which will be the subject of detailed functional characterisation.

Strategie zur funktionalen Genomanalyse der frühen Trächtigkeit bei der Stute. Die Erkennung der Trächtigkeit geht im Endometrium mit charakteristischen Veränderungen der Genaktivitätsmuster einher. Diese Veränderungen spiegeln sich in unterschiedlichen mRNA-Konzentrationen zwischen trächtigen und Kontrolltieren wider. Diese Unterschiede können mit Hilfe von Microarrays identifiziert werden. Die Genexpression kann auch auf der Ebene der Translation reguliert werden, wobei microRNAs eine zentrale Rolle spielen. Unterschiede auf Protein-Ebene können mit Hilfe der 2D-DIGE-Technologie oder mit LC-MS-Verfahren untersucht werden. Verschiedene bioinformatische Ansätze werden zum Daten-Mining eingesetzt, um zentrale regulatorische Gene zu identifizieren. Diese sind dann Gegenstand tiefer gehender funktionaler Charakterisierungen.

inseminated with semen or seminal plasma only and slaughtered at day 18 (Bauersachs et al. 2006). Furthermore, endometrial samples from heifers at day 0 (oestrus) vs. day 12 (dioestrus) (Bauersachs et al. 2005) have been compared. In the bovine oviduct epithelial cells have been analysed during the oestrous cycle and between the ipsilateral and the contralateral side (Bauersachs et al. 2003, Bauersachs et al. 2004). These studies revealed several hundred mRNAs, which show changes of their abundance during early pregnancy or during the oestrous cycle. A number of biological processes and molecular pathways have been identified which may play critical roles in early pregnancy in cattle. The mechanisms of recognition of pregnancy were also extensively investigated in the ovine model using candidate gene approaches but also functional genomics studies (Spencer et al. 2007).

In addition to the transcriptional control of genes the control of messenger RNA translation has been identified as an important regulatory mechanism of gene expression. Small RNAs of 21-23 nt in length, so-called microRNAs (miRNAs), have been shown to influence the expression of many target genes during developmental and differentiation processes, adding a new dimension of complexity in the control of gene expression (Bartel 2004, Kim and Nam 2006). It has been assumed that mammalian species have up to 1,000 different miRNAs. Several tools have been developed to analyse the expression of miRNAs, e.g. cloning for high-throughput sequencing, miRNA microarrays, and quantitative PCR adapted to miRNAs. With these tools it will be feasible to investigate possible functions of miRNAs in reproductive processes. First studies of miRNAs have been done in cattle where a considerable number of miRNAs was detected in several adult tissues and also in the embryo (Coutinho et al. 2007, Gu et al. 2007).

With the development of new sequencing technologies, e.g. the Genome Sequencer FLX from Roche/454, new possibilities emerged for transcriptome analyses especially for species where no genome-wide microarrays are available. These new sequencers allow up to 400,000 single reads per run, facilitating digital in-depth transcriptome analyses.

The major advantages of current transcriptomics techniques are their high sensitivity and their large dynamic range almost over the complete range of mRNA concentrations in cells and tissues. However, direct conclusions with regard to corresponding changes in protein abundance cannot be drawn from mRNA data. Furthermore, protein modifications important for regulation or signalling mechanisms cannot be predicted from the mRNA level (Hiendleder et al. 2005). Therefore, investigations of the proteome complementary to transcriptomics studies are essential for describing biochemical pathways and protein functions. But a major challenge is the complexity of the protein mixtures that must be analysed and their extremely heterogeneous physicochemical properties. In addition, proteins show an extraordinary variation in concentration, which complicates the detection of low-abundance proteins. In the last years a number of technical improvements and development of new technologies permitted holistic analyses of the proteome. A major advancement has been the application of liquid chromatography separation coupled with mass spectrometry (LC-MS) for qualitative and quantitative proteome analyses (Frohlich and Arnold 2006). In a bovine model for early pregnancy endometrial tissue samples collected during the pre-attachment stage have been analysed using the 2D difference gel electrophoresis (2D-DIGE) technique, an improvement of the classical 2D gel technique, resulting in the identification of four different pro-

teins with higher abundance in endometrium of pregnant animals (Berendt et al. 2005).

Since the primary results of gene expression analyses are mostly long lists of genes bioinformatics tools have to be applied for drawing meaningful conclusions. Similar to the differences between species in the sequence and expression databases the mapping to functional databases like Gene Ontology or molecular pathway databases also differs considerably due to the different status of gene annotation. Therefore it is necessary to use gene identifiers from the human orthologs for mapping to functional databases assuming similar functions of the same gene in different organisms. Once, key regulatory genes are identified they have to be further characterised. Figure 1 shows schematically a functional genomics strategy for the investigation of early pregnancy in the mare.

Due to the rapid progress of whole-genome sequencing projects and the development of novel high-throughput sequencing technologies, functional studies will be in the focus of future genome research projects. These require state-of-the-art technologies for holistic transcriptome and proteome studies, exactly defined animal models with a maximum of phenotypic information, and new bioinformatic approaches to facilitate a comprehensive description and eventually modelling of complex biological processes. In the context of reproduction, this integrated approach is expected to identify key genes or pathways, which may serve as diagnostic or therapeutic targets for sub- or infertility. Furthermore, mRNA or protein expression profiles of reproductive tissues will be evaluated as molecular fertility-related phenotypes and may provide the basis for the improvement of fertility by genetic selection.

Acknowledgement

Our functional genomics studies of bovine reproductive biology are funded by the DFG (FOR 478 "Mechanisms of embryo-maternal communication"; www.ematko.de) and by the BMBF (FUGATO–Fertilink).

References

Bartel D. P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281-297

Bauersachs S., Blum H., Mallok S., Wenigerkind H., Rief S., Prelle K. and Wolf E. (2003) Regulation of ipsilateral and contralateral bovine oviduct epithelial cell function in the postovulation period: a transcriptomics approach. *Biol. Reprod.* 68, 1170-1177

Bauersachs S., Rehfeld S., Ulbrich S. E., Mallok S., Prelle K., Wenigerkind H., Einspanier R., Blum H. and Wolf E. (2004) Monitoring gene expression changes in bovine oviduct epithelial cells during the oestrous cycle. *J. Mol. Endocrinol.* 32, 449-466

Bauersachs S., Ulbrich S. E., Gross K., Schmidt S. E., Meyer H. H., Einspanier R., Wenigerkind H., Vermehren M., Blum H., Sinowatz F. and Wolf E. (2005) Gene expression profiling of bovine endometrium during the oestrous cycle: detection of molecular pathways involved in functional changes. *J. Mol. Endocrinol.* 34, 889-908

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

Bazer F. W. and Spencer T. E. (2005) Reproductive biology in the era of genomics biology. *Theriogenology* 64, 442-456

Berendt F. J., Frohlich T., Schmidt S. E., Reichenbach H. D., Wolf E. and Arnold G. J. (2005) Holistic differential analysis of embryo-induced alterations in the proteome of bovine endometrium in the preattachment period. *Proteomics* 5, 2551-2560

Coutinho L. L., Matukumalli L. K., Sonstegard T. S., Van Tassell C. P., Gasbarre L. C., Capuco A. V. and Smith T. P. (2007) Discovery and profiling of bovine microRNAs from immune-related and embryonic tissues. *Physiol Genomics* 29, 35-43

Diatchenko L., Lau Y. F., Campbell A. P., Chenchik A., Moqadam F., Huang B., Lukyanov S., Lukyanov K., Gurskaya N., Sverdlov E. D. and Siebert P. D. (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. U.S.A.* 93, 6025-6030

Frohlich T. and Arnold G. J. (2006) Proteome research based on modern liquid chromatography—tandem mass spectrometry: separation, identification and quantification. *J. Neural. Transm.* 113, 973-994

Gu W. and Bertone A. L. (2004) Generation and performance of an equine-specific large-scale gene expression microarray. *Am. J. Vet. Res.* 65, 1664-1673

Gu Z., Eleswarapu S. and Jiang H. (2007) Identification and characterization of microRNAs from the bovine adipose tissue and mammary gland. *FEBS Lett* 581, 981-988

Hiendleder S., Bauersachs S., Boulesteix A., Blum H., Arnold G. J., Frohlich T. and Wolf E. (2005) Functional genomics: tools for improving farm animal health and welfare. *Rev. Sci. Tech.* 24, 355-377

Kim V. N. and Nam J. W. (2006) Genomics of microRNA. *Trends Genet.* 22, 165-173

Klein C., Bauersachs S., Ulbrich S. E., Einspanier R., Meyer H. H., Schmidt S. E., Reichenbach H. D., Vermehren M., Sinowatz F., Blum H. and Wolf E. (2006) Monozygotic twin model reveals novel embryo-induced transcriptome changes of bovine endometrium in the preattachment period. *Biol. Reprod.* 74, 253-264

Spencer T. E., Johnson G. A., Bazer F. W., Burghardt R. C. and Palmarini M. (2007) Pregnancy recognition and conceptus implantation in domestic ruminants: roles of progesterone, interferons and endogenous retroviruses. *Reprod Fertil Dev* 19, 65-78

Suchyta S. P., Sipkovsky S., Kruska R., Jeffers A., McNulty A., Coussens M. J., Tempelman R. J., Halgren R. G., Saama P. M., Bauman D. E., Boisclair Y. R., Burton J. L., Collier R. J., DePeters E. J., Ferris T. A., Lucy M. C., McGuire M. A., Medrano J. F., Overton T. R., Smith T. P., Smith G. W., Sonstegard T. S., Spain J. N., Spiers D. E., Yao J. and Coussens P. M. (2003) Development and testing of a high-density cDNA microarray resource for cattle. *Physiol. Genomics* 15, 158-164

Wheeler D. L., Barrett T., Benson D. A., Bryant S. H., Canese K., Chetvernin V., Church D. M., DiCuccio M., Edgar R., Federhen S., Geer L. Y., Kapustin Y., Khovayko O., Landsman D., Lipman D. J., Madden T. L., Maglott D. R., Ostell J., Miller V., Pruitt K. D., Schuler G. D., Sequeira E., Sherry S. T., Sirotkin K., Souvorov A., Starchenko G., Tatusov R. L., Tatusova T. A., Wagner L. and Yaschenko E. (2007) Database resources of the National Center for Biotechnology Information. *Nucleic. Acids Res.* 35, D5-12

Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S., Ansorge W., Bocher M., Blocker H., Bauersachs S., Blum H., Lauber J., Dusterhoft A., Beyer A., Kohrer K., Strack N., Mewes H. W., Ottenwalder B., Obermaier B., Tampe J., Heubner D., Wambutt R., Korn B., Klein M. and Poustka A. (2001) Toward a catalog of human genes and proteins: sequencing and analysis of 500 novel complete protein coding human cDNAs. *Genome Res.* 11, 422-435

Wolf E., Arnold G. J., Bauersachs S., Beier H. M., Blum H., Einspanier R., Frohlich T., Herrler A., Hiendleder S., Kolle S., Prelle K., Reichenbach H. D., Stojkovic M., Wenigerkind H. and Sinowatz F. (2003) Embryo-maternal communication in bovine - strategies for deciphering a complex cross-talk. *Reprod. Domest. Anim* 38, 276-289.

Dr. Stefan Bauersachs
Laboratory for Functional Genome Analysis (LAFUGA)
Gene Center of the Munich University
Feodor-Lynen-Str. 25, 81377 Munich, Germany
bsachs@lmb.uni-muenchen.de