

Control of Pig Reproduction VIII

Society of Reproduction and Fertility Volume 66

Proceedings of the Eighth International Conference
on Pig Reproduction

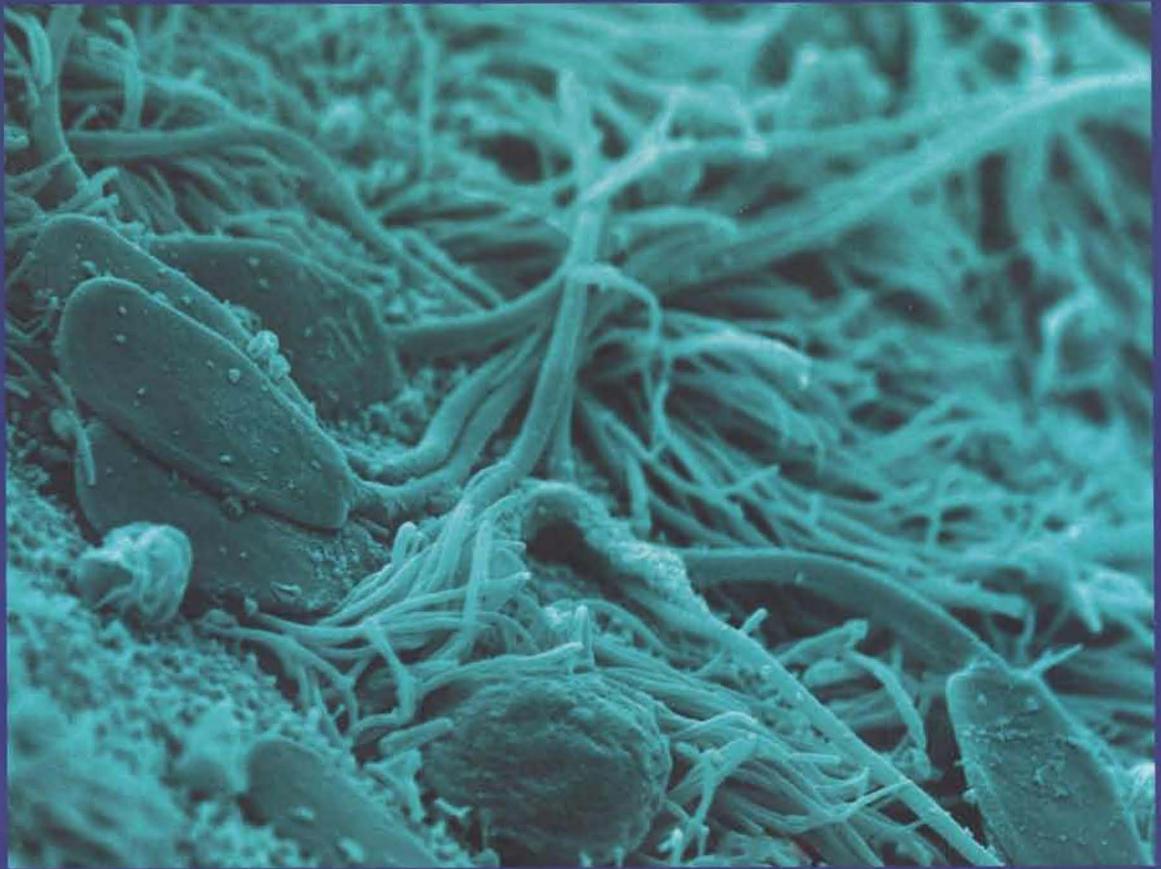
Alberta, Canada
June 2009

Edited by:
H. Rodriguez-Martinez, J.L. Vallet and A.J. Ziecik



Nottingham
University Press

Control of Pig Reproduction VIII



Edited by
H. Rodriguez-Martinez,
J.L. Vallet and A.J. Ziecik



Nottingham
University Press

Functional genomic approaches for the study of fetal/placental development in swine with special emphasis on imprinted genes

S.R. Bischoff^{1,3}, S. Tsai^{1,3}, N. Hardison^{1,4}, A.A. Motsinger-Reif⁴,
B.A. Freking², and J.A. Piedrahita^{1,3}

¹Department of Molecular Biomedical Sciences, College of Veterinary Medicine, North Carolina State University, Raleigh, NC, USA; ²USDA, ARS, U.S. Meat Animal Research Center, Clay Center, NE 68933-0166, USA; ³Center for Comparative Medicine and Translational Research, North Carolina State University, Raleigh, NC, USA; ⁴Bioinformatics Research Center, Department of Statistics, North Carolina State University, Raleigh, NC 27695-7566, USA

This chapter describes the application of functional genomic approaches to the study of imprinted genes in swine. While there are varied definitions of “functional genomics”, in general they focus on the application of DNA microarrays, single nucleotide polymorphism (SNP) arrays, and other high coverage genomic analyses, and their combination with downstream methods of gene modification such as silencing RNA (siRNA) and viral and non-viral transfection. Between the initial data acquisition and the actual genetic manipulation of the system lies bioinformatics, where massive amounts of data are analyzed to extract meaningful information. This area is in constant flux with an increased emphasis on detection of affected pathways and processes rather than generation of simple affected gene lists. We will expand on each of these points and describe how we have used these technologies for the study of imprinted genes in swine. First we will introduce the biological question to provide context for the discussion of the functional genomic approaches and the types of information they generate.

Part I. The biological question

While over 99% of genes in mammalian species are transcribed from both maternal and paternal alleles (bi-allelic expression), a small subset are transcribed from only one allele (mono-allelic expression). In some cases it is the maternal allele that is transcribed and in others the paternal allele. The choice of which allele is transcribed is dependent on markings placed in the chromosome during gametogenesis (Hajkova *et al.* 2002, Reik & Walter 2001). To date less than 100 imprinted genes have been identified, yet they have profound phenotypic effects, particularly in placental and fetal development and function (Angiolini *et al.* 2006). Yet, their role is not limited to fetal and placental development but can also affect other aspects of reproduction such as rearing behavior and lactation as will be described later. Our interest in these genes came about through the reports of abnormal placentation and fetal overgrowth of somatic-cell-nuclear-transfer-derived calves (Hill *et al.* 1999). The combined syndrome has been referred to as abnormal offspring syndrome (AOS) as well as large offspring syndrome

(LOS) (Farin *et al.* 2004 , Farin *et al.* 2006). Multiple laboratories working in this area reported epigenetic abnormalities in cloned cattle and mice, including dysregulation of imprinted genes (Dindot *et al.* 2004). It was through these original observations that we became interested in this complex and fascinating group of genes, and at the same time dismayed by the almost total lack of information of their function in swine.

Evolution of imprinted genes

Imprinted genes, which are defined as genes that display parent-of-origin, mono-allelic, expression, have only been found in placental mammals (Hore *et al.* 2007) and flowering plants (Huh *et al.* 2008) while non imprinted homologues have been found in reptiles, amphibians, fishes and the egg-laying monotremes (Edwards *et al.* 2007a). Yet, even if a small rudimentary placenta is present, such as that seen in marsupials, evidence for imprinting can be found. Thus, the placenta and imprinted genes appear to have co-evolved. This underlies the relevance of these genes to the formation and function of the placenta and in fetal development.

The parental-conflict hypothesis has emerged to explain the appearance of imprinting as a result of different evolutionary pressures influencing each parent in placental mammals. The hypothesis states that imprinting evolved to control energy flow between the mother and the developing fetus (Moore & Haig 1991) . The conflicting evolutionary outcomes are that the mother (and consequently her genome) is more successful by restricting nutrient flow to the fetus/offspring so that she does not commit too much of her energy resources to each fetus, leaving her more able to reproduce in large numbers. In contrast, the father (and his genome), is represented only in the fetus, and improves his success by extracting as much energy as possible from the mother to benefit each fetus/offspring. It is here where the "conflict" lies, and a careful balance between the two contrasting forces leads to a normal fetus. Unbalancing of these forces can lead to either a smaller than normal (small for gestational age or intrauterine growth restriction) or a larger fetus (large for gestation age or large offspring syndrome; Fig. 1).

The characteristics of uniparental conceptuses support components of the parental conflict hypothesis. Androgenotes (conceptuses derived from only the male) and gynogenotes (conceptuses derived from only the female) can be produced from either two male pronuclei or two female pronuclei (McGrath & Solter 1984). Parthenotes, which are a form of gynogenote, can be easily generated by activation of oocytes and inhibition of polar body extrusion by using cycloheximide (Tsai *et al.* 2006b), resulting in a diploid embryo carrying only maternally derived chromosomes. Although neither androgenotes nor gynogenotes can produce viable offspring, their characteristics are suggestive of the role of imprinted genes in energy distribution and placental development. Gynogenotes, with a double dose of maternally expressed genes, develop into small fetuses with small placentas, as would be expected from a reduction in energy delivery to the fetus. In contrast, androgenotes develop a very large placenta also supportive of the placental conflict hypothesis. However, they also lack a fetus suggesting that maternal imprints are an absolute requirement for fetal development. In addition, as will be discussed later, there is ample direct experimental evidence supporting both the parental conflict hypothesis and the role imprinted genes play in placental and fetal development, as well as in behaviors related to control of energy flow such as nurturing behavior and milk let down.

Can placentas exist without imprinting?

While the evidence from placental mammals and flowering plants strongly supports imprinting as competition for the flow of energy between the developing embryo and the energy source,

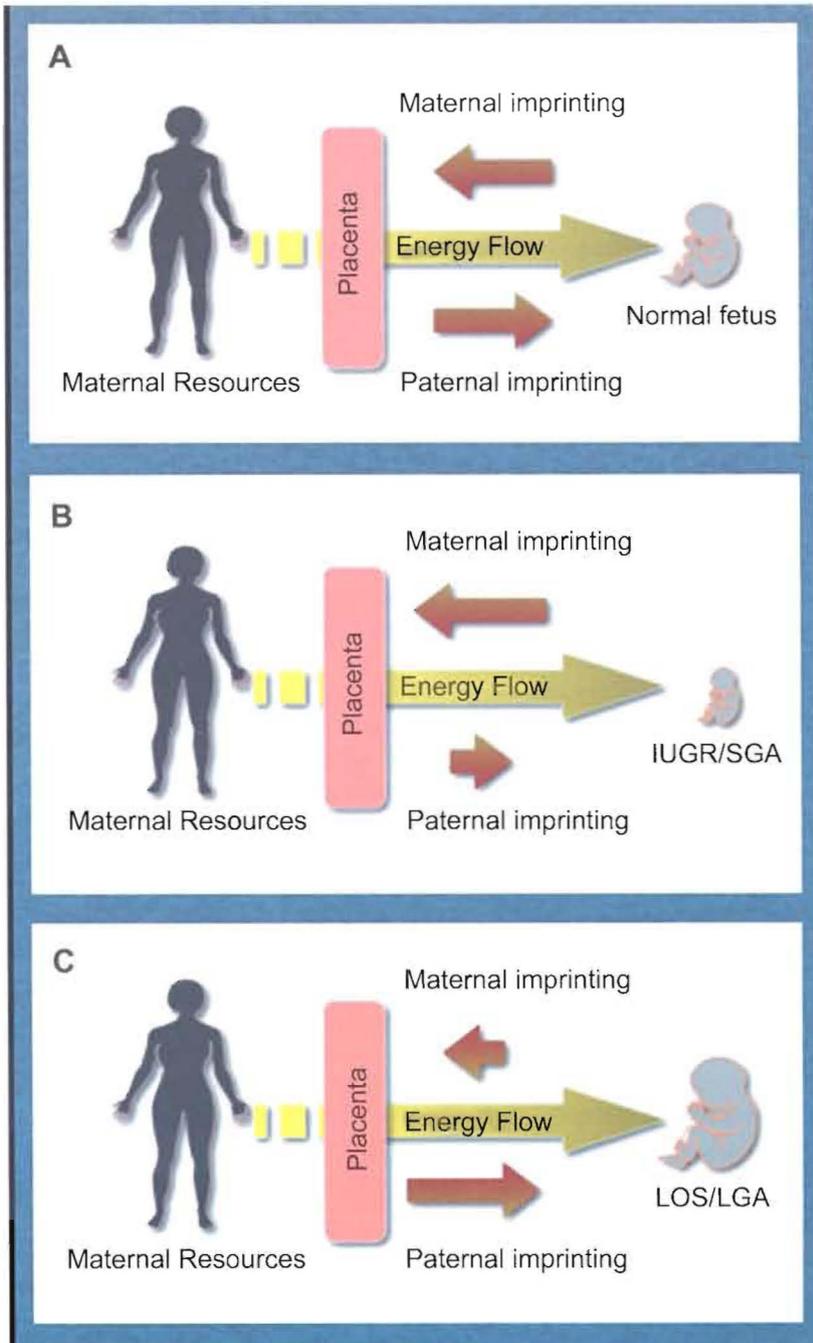


Fig. 1 Diagrammatic representation of the parental conflict hypothesis. A. A normal situation where the control of nutrients from the mother to the fetus is balanced leading to normal fetal growth. Notice maternal imprinting shifting the balance towards the mother while paternal imprinting shifts the flow towards the fetus. B. A case where the balance between maternal and paternal imprints is shifted towards the mother resulting in less nutrients reaching the fetus leading to intrauterine growth restriction/Small for gestational age (IUGR/SGA). C. The opposite case whereby nutrient flow is shifted from the mother to the fetus resulting in large offspring syndrome/large for gestational age (LOS/LGA).

the existence of placental fishes provides a paradox for the requirement of imprinted genes for placental development and function. At present, there has only been one report of analysis of imprinted genes in viviparous fishes and that single report indicates that the IGF2 gene is not imprinted in this species (Lawton et al. 2005). While that in itself is not sufficient evidence to classify these placental animals as a paradox, the recent observation that hammerhead sharks can reproduce by parthenogenesis is (Chapman et al. 2007). As mentioned above, uniparental offspring are embryonic lethal in all placental mammal species tested to date (Walsh et al. 1994, Hagemann et al. 1998, Zhu et al. 2003) and it is widely accepted that this lethality is due to the presence of imprinting. The observation of successful parthenogenesis in sharks, thus, suggest that imprinting may be absent in this species. If that is proven to be the case, how did this species evolve a placenta? Or is it that the fish placenta is functional and morphologically distinct from that of placental mammals?

Fetal nutrition within the fish species ranges from wholly dependent on deposited yolk during oogenesis (vitellogenesis), through an intermediate form of nutrition dependent on histotroph secretion from the uterus/oviduct, to nutrition dependent on yolk-sac based placentation (Hamlett 1989). In the scant literature in this area it is evident that the fish placenta is quite distinct from any known mammalian placenta in multiple aspects including the continued presence of an egg envelope throughout gestation (Heiden et al. 2005), the small area of actual attachment to the uterus/oviduct in relation to fetal size, and the reliance on the yolk sac as the major organ of nutrient exchange (Jones & Hamlett 2004, Reznick et al. 2007). This suggests that some aspects of placental development and function are independent of imprinting while others are more dosage sensitive and require the imprinting of one allele.

This is partially supported by observations in marsupials, with a rudimentary placenta, where imprinting has been observed in some genes such as IGF2 and PEG10 (Ager et al. 2007, Ager et al. 2008b) but not in others such as SNRPN, UBE3A, DIO3 (Rapkins et al. 2006), CDKN1C (Ager et al. 2008a) and DLK1 (Edwards & Ferguson-Smith 2007). A detailed analysis of marsupial placentation has been presented by Renfree et al. (2008) and elegantly describes how within the marsupial family different forms of placentation exist and that the more complex the placenta the greater the number of imprinted genes (Renfree et al. 2008). In summary, while species such as placental sharks suggest that in the absence of imprinting, a rudimentary form of placentation can exist, the preponderance of the evidence indicates that the emergence of complex placentation is associated with the presence of imprinted genes.

Experimental approaches to the study of imprinting in mammals.

The experimental evidence for the role of imprinted genes in placental and fetal development is derived from two general approaches, the analysis of uniparental animals and the direct observation of the effects of modification of imprinted genes by transgenesis. The uniparental models described above, in particular, have been very useful for broad and comprehensive analysis of imprinted genes between different mammalian species, as well as uncovering new imprinted genes (Barton et al. 1984, McGrath & Solter 1984, Surani et al. 1984, Cattanach & Kirk 1985, Dean et al. 2001, Zhu et al. 2003). A more focused approach is the analysis of the effects of transgenic manipulation of imprinted genes, usually by gene inactivation via homologous recombination. A few examples of placental defects resulting from manipulation of imprinted genes include larger placentas resulting from inactivation of the maternally expressed genes *Ascl2* (Guillemot et al. 1995), *Grb10* (Charalambous et al. 2003), *IGF2R* (Wang et al. 1994), *PhlA2* (Frank et al. 2002), and *p57 (Kip2)* (Takahashi et al. 2000). In contrast, inactivation of paternally expressed genes such as *Peg10* (Ono et al. 2006), *IGF2* (Sibley et al. 2004), *Peg3* (Li et al. 1999), and *Mest* (Lefebvre et al. 1998) result in smaller placentas.

The effect of Paternally Expressed Gene 3 (Peg3) deficiency is particularly intriguing as the phenotype illustrates the many roles imprinted genes can play with respect to energy utilization. The work of Curley *et al.* 2004 indicated that Peg3 deficiency influences the placenta (mentioned above), fetus and mother. At the fetal level, pups deficient in Peg3 have abnormal thermoregulation and suckling defects. Peg3 deficient mothers, in turn, have impaired maternal care, reduced feed intake during pregnancy and reduced milk-letdown. Peg3, therefore, can control energy flow at many levels, from food intake by the mother, to how much milk to provide the offspring, to how much the pup is able to extract from her during suckling (Curley *et al.* 2004). While the exact mechanism of action of Peg3 is unknown, the protein is expressed at high levels in the trophoblast layer of the mouse placenta and in the hypothalamus. As Peg3 is known to be involved in the control of apoptosis, it has been postulated that abnormal apoptosis leads to altered hypothalamic function affecting thermoregulation, maternal behavior and milk letdown, while defects in the placenta result in reduced fetal growth.

Combined, these observations indicate how complex the function of imprinted genes can be. Yet, in most cases the phenotype supports the parental conflict hypothesis with inactivation of maternally expressed genes leading to larger placentas, and inactivation of paternally-expressed genes leading to smaller placentas. While these results are important and support the role of imprinted genes in placental development they are limited to one species, a small fraction of the known imprinted genes, and perhaps with the exception of *Igf2*, *Igf2r*, and *Mash2/Ascl2* (Tanaka *et al.* 1997), the role of these genes in placental development is not known. Thus, it is the combination of the fascinating aspects of these genes, the extremely limited information of the function of imprinted genes in placenta of mice, and the absolute absence of information on their function in swine reproduction that encouraged us to embark in a comprehensive study of these genes in swine. To accomplish this goal, we used genomic approaches. In the next section we will describe the techniques we used to accomplish this goal and our positive and negative experiences with them.

Part II. Gene expression profiling methods

Gene expression profiling methods

The completion of the draft human genome sequence (Lander *et al.* 2001) demonstrated the feasibility of sequencing entire complex mammalian genomes, and marked the beginning of whole genome sequencing projects for many mammalian species. Prior to the availability of all of this genomic sequence, the classic approach in molecular genetics was the forward genetic screen. The goal of this approach was to find the genes responsible for a phenotype of interest, and indeed many interesting genes have been mapped in this way. However, the availability of whole genome and transcriptome sequences brings us to the unusual position of possessing information on the sequence of nearly all the genes in the genome, but understanding the function of a far smaller fraction. Simply put, we know where most genes *are*, but not what they *do*. So, how can we use the wealth of newly available genomic information to better understand gene function?

Microarray technology

Microarray technology provides a method of rapidly profiling gene expression genome-wide. There are two major microarray platforms currently available for gene expression profiling in swine: a commercial Affymetrix GeneChip Porcine Genome short oligonucleotide microarray

and a U.S. Pig Genome Coordination Program glass spotted long oligonucleotide microarray. The primary difference between these two platforms is that the Affymetrix platform is based on eleven 25-mer probes synthesized *in situ* on a solid support with a photolithographic mask, whereas the U.S. Pig Genome Coordination Program platform (Zhao et al. 2005) is based on traditionally synthesized oligonucleotides subsequently spotted onto a glass slide. In initial validation experiments, we directly compared the technical reproducibility and sensitivity of the two platforms, comparing the gene expression profiles of biparental and parthenogenetic fibroblast cell lines. From the same starting pool of total RNA, we found that the reproducibility of hybridization with the Affymetrix short oligonucleotide microarray was much higher than the U.S. Pig Genome Coordination Program microarray (Fig. 2). In probes shared across both platforms, we detected a greater number of differentially expressed genes using the Affymetrix platform. For the time being, the Affymetrix Porcine Genome Array is the most sensitive and reproducible platform for conducting gene expression profiling experiments with microarrays in swine (Tsai et al. 2006b). While the first generation of U.S. Pig Genome Coordination Program arrays suffered from printing defects, lower technical reproducibility, and lower gene coverage; the second generation of these arrays significantly improved coverage. However, due to their lower technical reproducibility, more than 2-3X the number of arrays are required to achieve equivalent statistical power to detect differential expression in contrast to Affymetrix Porcine GeneChip Arrays, so a cost/benefit analysis would still favor the Affymetrix platform for swine gene expression profiling.

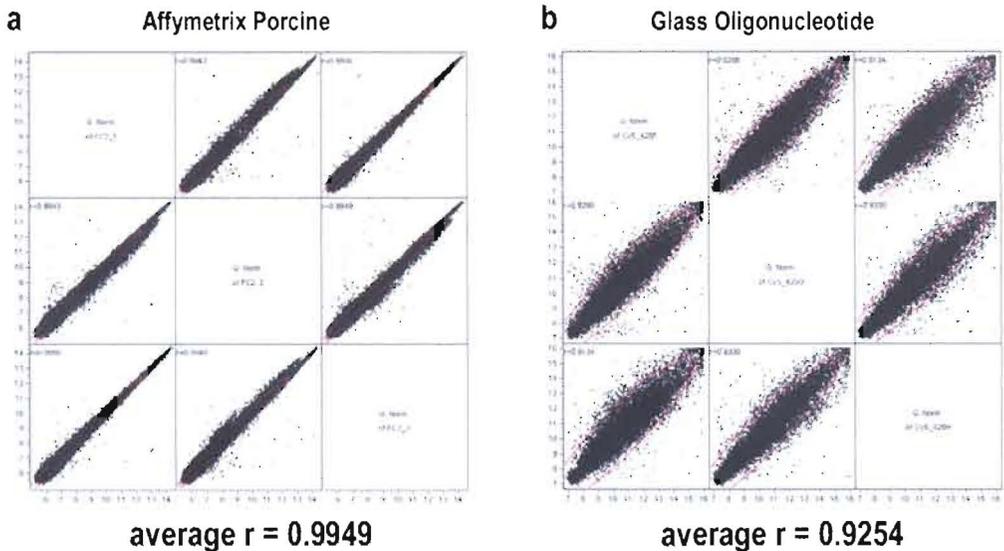


Fig. 2 Reproducibility of technical replicates in the Affymetrix and the glass array platforms. Pairwise scatterplots of control technical replicates of porcine fibroblast cell lines profiled on (a) Affymetrix Porcine and (b) U.S. Pig Genome Coordination Program long oligonucleotide glass microarrays. The lower reproducibility of the glass arrays reduces the ability of the arrays to detect statistically significant differences between experimental samples. This reduced accuracy can only be overcome by increasing the number of replicates in the glass array in comparison with the Affymetrix arrays.

There have been reports of using cross-species hybridization for the purposes of performing gene expression profiling in swine, primarily before the release of porcine specific arrays (Zhao *et al.* 2005). We compared cross-species hybridization of the same RNA described above onto Affymetrix Human U133 + 2.0 GeneChip Arrays. We found that this approach had the lowest power to detect differential expression, because of a high number of non-hybridizing probes. On average, 1-3 probes out of 11 in a probe set hybridized efficiently. Even after implementing various filtering algorithms, the sensitivity of detection was still significantly lower than using porcine specific microarrays.

Deep sequencing (RNASeq)

Microarrays provided a powerful tool for asking descriptive questions about gene expression genome wide. It is increasingly evident, however, that with rapidly evolving deep sequencing technologies microarrays will eventually be supplanted by direct sequencing of mammalian transcriptomes. In this approach, the complete transcriptome can be sequenced and matching transcripts counted rather than indirect quantitation based on hybridization intensities (Wang *et al.* 2009; (Wang *et al.* 2009; Wold & Myers 2008). This has several advantages in that: 1) there is a greater dynamic range in comparison to hybridization based technologies, single transcripts can be positively identified, 2) the technology does not rely on a *priori* knowledge of gene sequence, and 3) background from cross-hybridization is eliminated (Wang *et al.* 2009). Multiplex strategies have been developed to uniquely tag RNA samples with an unique error-correcting molecular barcode, so that the capacity of each sequencing run can be most efficiently utilized (Craig *et al.* 2008, Hamady *et al.* 2008). These strategies are based on the simple addition of 4-6 bp of sequence to the adapters that are used to create the libraries for resequencing; these unique identifiers allow the downstream determination of individual samples from a pool. The freedom from having to define which sequences to interrogate enables the possibility of novel transcript discovery. The sequence information provided allows the unambiguous identification of single transcripts, whereas detection by hybridization technologies is inevitably limited by background. Finally, the need for the complex nonlinear normalization strategies often employed in typical microarray experiments is lifted, as sequenced transcripts are simply mapped to a reference genome and counted. This may be useful in cases where one of the fundamental assumptions of most microarray normalization procedures, that the empirical distribution of transcripts is the same across samples, are on shaky ground, such as when comparing gene expression profiles across tissues or species. One limitation of RNAseq is that mapping the data produced is required by the existence of a reference genome; in swine ongoing sequencing efforts will increase the utility of RNAseq data in the coming years.

Annotation of microarrays

A determination of differential expression for a probe on a microarray or a sequence cluster from an RNAseq experiment is of limited utility without knowing what genes or transcriptional units they represent. Because of the limited annotation, with only about 20% of the probe sets present in the initial annotation of the Affymetrix Porcine Genome Array, we reannotated the probe sets against human cDNA and genomic DNA sequence (Tsai *et al.* 2006a). The approach we took was to extend the target sequence using sequence information available from The Institute for Genome Research (TIGR) swine gene index (currently Dana Farber gene index, website), and matching the extended sequence against other Ensembl human cDNA and genomic DNA sequences. This approach was successful in raising the percentage of annotated genes from

20 to 80% because the majority of the probes for this generation of microarrays are designed against the 3' untranslated region (UTR). In many cases, our annotation strategy extended the sequences beyond the 3' UTR. Our annotation provided a bit score (a measure of the likelihood of a correct sequence match) so that individual investigators can set their own threshold for acceptable annotation confidence. We have recently updated our annotation of the Affymetrix Porcine Genome microarray against bovine, mouse, and human. One additional unique feature of the annotation is our matching against the Affymetrix Human probe set IDs. This matching allows data generated from Affymetrix Porcine GeneChip arrays to be used with many of the pathway analysis solutions that are available for human gene expression profiling data.

Single nucleotide polymorphism (SNP) and single feature polymorphism (SFP) discovery

Affymetrix short oligonucleotide gene expression data can indicate single feature polymorphisms (SFP), as single nucleotide polymorphisms (SNP) that are close to the center of a 25-mer oligonucleotide probe can almost completely disrupt hybridization (Winzeler et al. 1998, Borevitz et al. 2003). The SFP are identified by disparate hybridizations among individual animals to one or more of the 11 targets for each mRNA on the array (Fig. 3) The exact SNP can then be identified by sequencing, however, SFP genotypes have also been directly used to generate high density haplotype maps for expression quantitative trait loci (eQTL) studies (West et al. 2006). Deep sequencing of the transcriptome can also contain not only information on the expression level of a transcript, but also allelic variations in the sequences obtained. The U.S. Pig Genome Coordination Program glass spotted long oligonucleotide microarrays cannot, however, be used for this purpose as a single SNP is insufficient to significantly disrupt the hybridization kinetics of a 70-mer probe.

We have demonstrated the feasibility of detecting SFP using Affymetrix short oligonucleotide arrays in swine (Bischoff et al. 2008). The basic idea behind the approach is to look for probes which have a substantially greater probe effect than expected, corresponding to the scenario where a SNP disrupts probe hybridization. Using this approach we detected 857 SFP between Chinese Meishan and European white composite breeds of swine, with a sensitivity of 0.65, specificity of 0.94, and a false discovery rate of 0.3. We have streamlined the method we used to determine the presence of an SFP into an easy to use downloadable procedure, *Click-N-SNP*, which will generate a list of putative SFP given raw Affymetrix data (.CEL files) and a simple experimental design (Bischoff et al. 2008). Similarly, given sufficient oversampling, it is possible to detect SNPs and indels in RNASeq data. Care, however, must be taken to use the appropriate statistical models to distinguish between true SNPs and sequencing error, given that the error rate in short-read sequencing platforms is relatively high.

One motivation behind obtaining genotype information from gene expression data, whether from microarrays or RNASeq, has been a method dubbed "genetic genomics" (Jansen & Nap 2001). The principle is that there is a heritable aspect to gene expression that may ultimately contribute to phenotypes of traits of interest. By merging information on allelic differences and gene expression, as well as the gene expression contribution to function, it may be possible to gain a better picture of the genes involved and their mode of regulation. Extracting this information from gene expression data, is a "free" source of this additional genetic information, qualified by the fact that SFP obtained from microarrays only localize the polymorphism to a 25 bp window, and therefore do not fully define the sequence variation. For higher density SNP genotyping applications, a 50k porcine SNP chip has been developed.

Finally, massively parallel targeted resequencing was recently demonstrated by coupling deep sequencing with solution hybrid selection with long oligonucleotides (Gnirke et al. 2009),

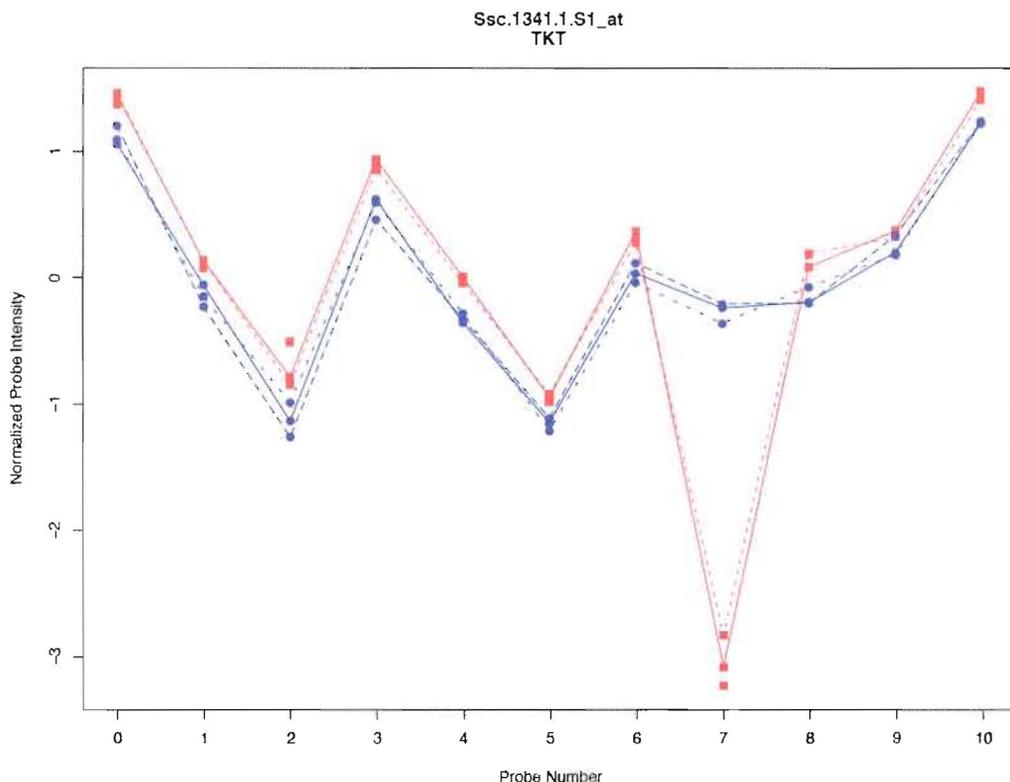


Fig. 3 Example of Single Feature Polymorphism (SFP) in swine detected between Chinese Meishan and European white composite breeds. Notice Probe 7 which exhibits hybridization intensities near background, while the remaining probes are 4-fold or more higher. This reduced hybridization was due to a SNP within that particular probe. Thus a probe-by-probe analysis can rapidly uncover a large number of potential SNPs within the population being used for the microarrays experiments.

should also be possible in swine. The principle of this approach is to synthesize a number of tiled probes against genomic region(s) of interest. These probes are around 200 bp in length, contain universal primer sequences for amplification, a T7 promoter for *in vitro transcription*, and are synthesized *in situ* on an Agilent custom microarray. After cleaving these probes off the microarray, the pool of probes is subjected to an *in vitro* transcription reaction containing biotinylated nucleotides to create a pond of biotinylated cRNA baits. This pond of biotinylated cRNAs hybridizes efficiently to sheared genomic DNA sequence. The captured, complementary genomic DNA sequence is then used as the input for resequencing library preparation. With the growing availability and near completion of a draft porcine genome sequence, an increasing proportion of the data generated via this approach will be mappable.

Part III. Moving beyond gene lists. Finding functional interactions from microarray data to study epigenetic asymmetry in porcine fetal tissues

In the following section we will enumerate a list of tools that facilitate analysis of gene expression datasets. Building on these descriptions, we draw on datasets generated by our laboratory

and apply the tools to systematically clarify functional relationships among expressed genes. The data are freely available at Gene Expression Omnibus (GEO) under accession number GSE10443.

Pattern discovery by clustering analysis

Cluster analysis software requires the ability to handle large input datasets (greater than 100,000 rows/columns) and should contain versatile microarray analysis features. Although a number exist, our laboratory has experience with the licensed package JMP Genomics (SAS, Cary, NC) and the freeware package R Bioconductor (<http://www.bioconductor.org>).

For pattern discovery a number of unsupervised methods exist to partition data into visual subsets by a common group of parameters or clusters (Allison et al. 2006 ; Kerr et al. 2008) which include hierarchical clustering, heat maps, k-means clustering and principal components analysis (PCA). For a concise summary of clustering methods useful to expression datasets, see D'Haeseleer (2005). Hierarchical clustering partitions data into groups of genes iteratively with each successive finer grouping being more similar. A central component of the method is depicting similarity by distance. Shorter branches represent more closely related items. In transcriptomic datasets, the distance metric is calculated from gene expression values. The output is often shown diagrammatically in a dendrogram, or branched-tree graph. An example of such output for imprinted gene expression in the placenta between day 30 control conceptuses and swine parthenote conceptuses is shown in Fig. 4).

DNA microarrays can also be depicted by two-dimensional graphical representations of gene expression values called heat maps, where color denotes expression intensity (i.e. red = low intensity, green = high intensity). A heat map can quickly show the level of expression of a gene across samples, time or treatment (Fig. 4) and is a rapid and simple way of looking at a large amount of data in a single figure. K-means analysis, a more complex method, can shuffle genes based on their geometric mean into a predicted number of clusters as defined by the hypothesis. The method is rapid, but requires a priori knowledge of how many clusters are expected, and is therefore biased.

Principal component analysis (PCA), another clustering method, reduces the complexity of a dataset by decomposing the variance into a limited number of dimensions or components using the mathematical tools of eigenvalues and covariance matrices. In this manner, the gene expression measurements can be visualized in a linear fashion to clarify how each array behaves in context with the other arrays in an experiment. The abscissa and ordinate axes represent the first and second principal components, respectively. The closer an expression array groups together indicates its similarity. Principal component analysis is a robust methodology for rapidly clustering data, and can provide a framework for quality control. An example of the use of PCA for this purpose is described in Fig. 5. The first three principal components were used as they explained 86%, 5%, and 5% of the total variation, respectively. Initial examination identified two discrepant arrays (LG2 and BG3). As 95% of the variation is contained within the concentration ellipse, both arrays were excluded from downstream analysis for technical reasons associated with RNA quality. Close examination of the array data then showed that the hybridization levels of both of these arrays were below that required to obtain a reliable signal. The arrays were then re-normalized excluding these two arrays to increase the accuracy of the data. It is interesting to note, that brain (BC, BG; blue) and placental (PC, PG; red) tissues grouped more closely than liver (LC, LG; green) or fibroblast (FC, FG; orange) day 30 swine fetal tissues and suggests brain and placental transcriptomes are more similar than fibroblast or liver transcriptomes. The main point of this example is that PCA can rapidly identify a hybridization/

technical issue in one or more of the arrays and prevent investigators from performing complex statistical analysis with data that are of low quality.

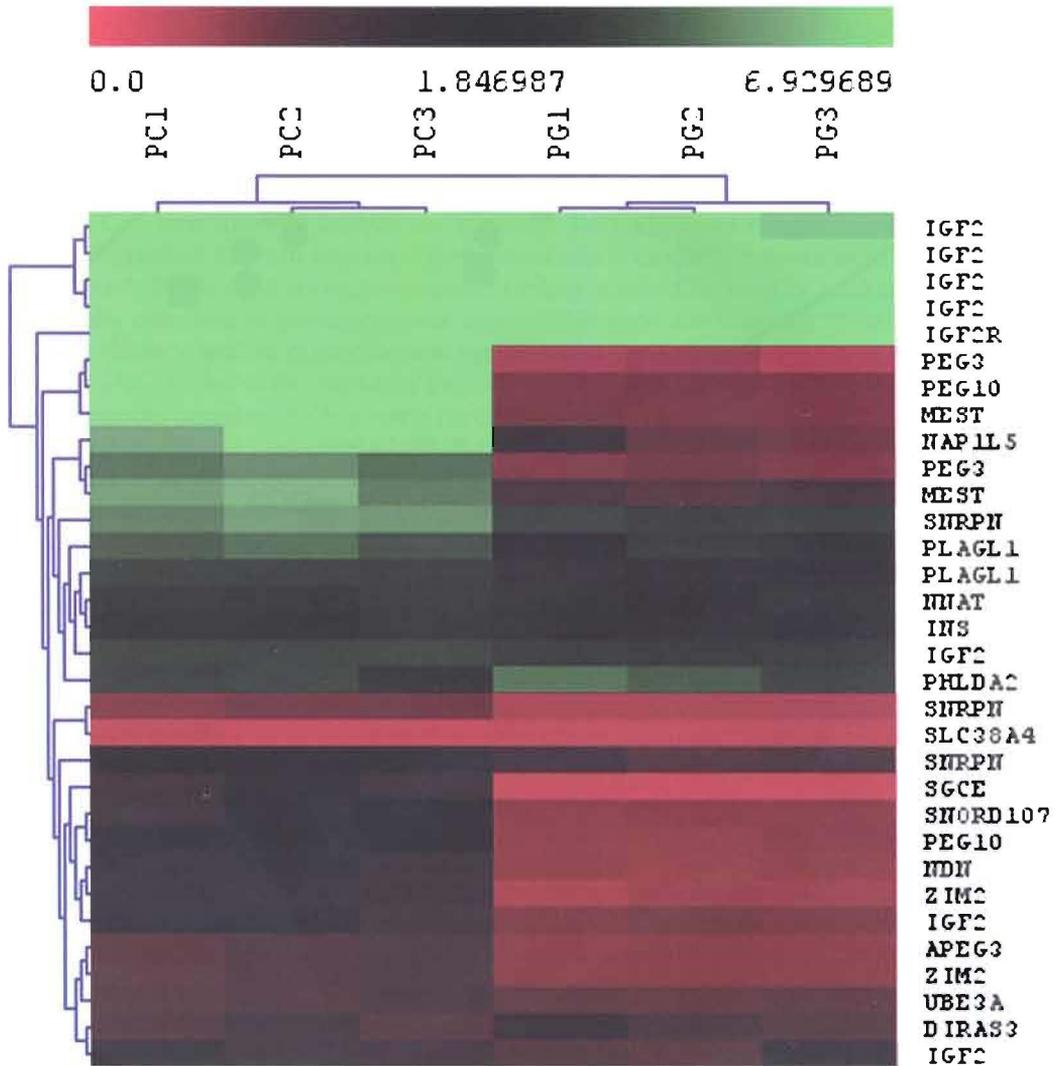


Fig. 4 Hierarchical Clustering and Heat Map of imprinted gene family in Day 30 fetuses. Microarrays containing placental RNA from biparental (PC) or parthenote (PG) D30 swine gestations were submitted to hierarchical clustering. Detransformed normalized values were used. Microarrays containing placental RNA from biparental controls or parthenote D30 swine gestations were profiled by one-color DNA short-oligonucleotide microarrays (Porcine GeneChip, Affmetrix). A gradient of red-black-green was used to denote expression intensity, where red denotes low expression, black shows median intensity, and green denotes high expression. Columns from left to right list samples of placental controls 1-3, placental parthenotes 1-3. Rows represent probe set intensities, and several genes contained multiple probesets which bind to the messenger RNA in different regions or alternatively spliced exons. Heat maps were generated with TM4 software using version MeV v4.3.02 (Saeed, Sharov et al. 2003) <http://www.tm4.org/mev.html>.

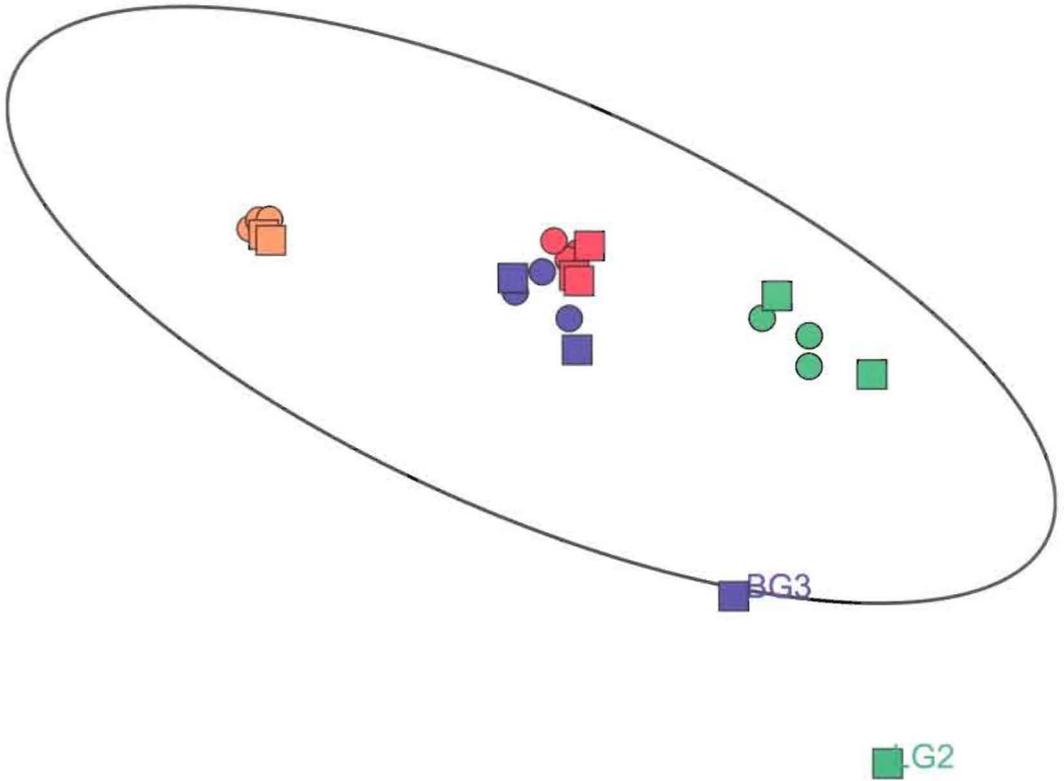


Fig. 5 Principal Component Analysis (PCA) identifies two discrepant arrays. Parthenogenic and biparental fetal tissues were hybridized to short-oligonucleotide arrays (PorcineGene Chip, Affymetrix, CA). The encircling ellipse explains 95% of the variation among samples. Circles denote parthenotes, squares represent controls. Brain = blue, fibroblast = orange, Liver = green, Placenta = red. Two arrays fell outside the concentration ellipse, BG3 and LG2, and were omitted from downstream analysis.

Enrichment analysis: functional annotation and pathway analysis

Classifying genes based on criteria such as biochemical function, genetic interaction and pathway, motif searching, and gene ontology is what broadly describes enrichment analysis. The overarching goal is to move from the daunting candidate gene lists and distill the dataset into meaningful biological processes that can be tested experimentally in the laboratory. A suite of over 68 enrichment analysis tools are now available and have recently been reviewed in Huang da *et al.* (2009). A full description of each is beyond the scope of this chapter, so we will focus on tools we have used and provide a summary of key findings. We refer the reader to Huang da *et al.* (2009) for a comparison of advantages, pitfalls, and operational classification of the current tools (e.g. statistical testing methods).

Gene ontology

Gene ontology (GO) provides a unique vocabulary that describes or annotates genes by molecular function, biological process and cellular distribution. Because the number and sophistication of GO-related programs has increased dramatically, a searching tool SerbGO (Mosquera & Sanchez-Pla 2008) is available on the web to identify which GO software application may be best for the

end-user's dataset. Alternatively, a summary of each application is available on the web at the Gene Ontology Consortium website: <http://www.geneontology.org/GO.tools.microarray.shtml>. Our group used SerbGO to identify significance analysis of function and expression (SAFE) (Barry *et al.* 2005, Gatti *et al.* 2009), Database for Annotation, Visualization and Integrated Discovery (DAVID) (Dennis *et al.* 2003), and Gene Set Enrichment Analysis (GSEA) (Subramanian *et al.* 2005) as appropriate tools for our purpose.

The porcine parthenogenetic conceptus develops as a small fetus and placenta which eventually dies at approximately day 32 of gestation (Fig. 6). To gain biological insight into pathways that differ during uniparental embryonic development compared to normal development, we used gene ontology (GO) descriptors to analyze our datasets. Two additional categories indicating parent of origin expression for each imprinted gene, were added to clarify paternal/maternal imprinting contributions. We used a permutation-based ranking method to identify functional categories differentially expressed in parthenogenetic fetuses. This approach is similar to SAFE (Barry *et al.* 2005) (significance analysis of function and expression) using SAS streamlined with JMP Genomics (SAS, Cary, NC) in lieu of Bioconductor (<http://www.bioconductor.org/>). Probe sets were assigned to GO categories based on the Affymetrix Porcine annotation by Tsai *et al.* (2006a). Initially, custom PHP code (<http://en.wikipedia.org/wiki/PHP>) reads in two files: one with the gene ID followed by any number of columns containing ranks for statistical tests that have been performed, and one containing the GO categories and the genes they contain. The algorithm calculates the rank sum for each GO category for each rank column, and subsequently permutes the gene labels with respect to their ranks. For each permutation, the GO category now contains a random set of genes and thus a random set of ranks. The permuted rank-sums are calculated, and a running total is kept of how frequently the permuted rank is less than or equal to the original rank. Dividing this by the total number of permutations provides the p-value estimate. The advantages of this permutation approach are to limit Type I errors for individual categories.

Summarized ranks of GO categories that were significant ($p < 0.05$) between parthenogenotes and biparental conceptuses for various tissues include paternally expressed imprinted genes, phosphatidylinositol binding, microtubule dynamics and lipid transporter activity. Not surprisingly, imprinted paternally expressed genes were ranked significant ($p < 0.006$) across all five datasets corresponding to each tissue indicating that parthenote profiling can reliably detect transcript dosage differences, regardless of tissue surveyed. Notably, there were marked differences in proliferation, biogenesis and biosynthesis pathways as predicted by the parent-conflict hypothesis. Consistent with our observations and others that parthenogenote conceptuses are developmentally delayed, various structural proteins ranked highly significant in most tissues. The artificial category "imprinted, maternally expressed" showed no significant difference, and may be related to a power-related problem to detect a theoretical 2:1 ratio upon comparison of maternally expressed genes between parthenotes and biparentals. As many biological pathways were affected that do not contain imprinted genes, we feel these data support conclusions that a gene network is present that extends beyond imprinted genes but is epistatically affected by them (Varrault *et al.* 2006). Thus, utilizing these methods it is possible to go from a list of genes, which in most cases is too large to properly address experimentally, to a more biologically relevant list of biological processes that are likely to be affected. This can greatly facilitate hypothesis generation as well as the design of physiological /biochemical experiments.

Pathway and interactome analysis

Additionally, to uncover new meaningful biological relationships it is often helpful to visualize gene signatures in the context of curated biochemical pathways as provided by resources such

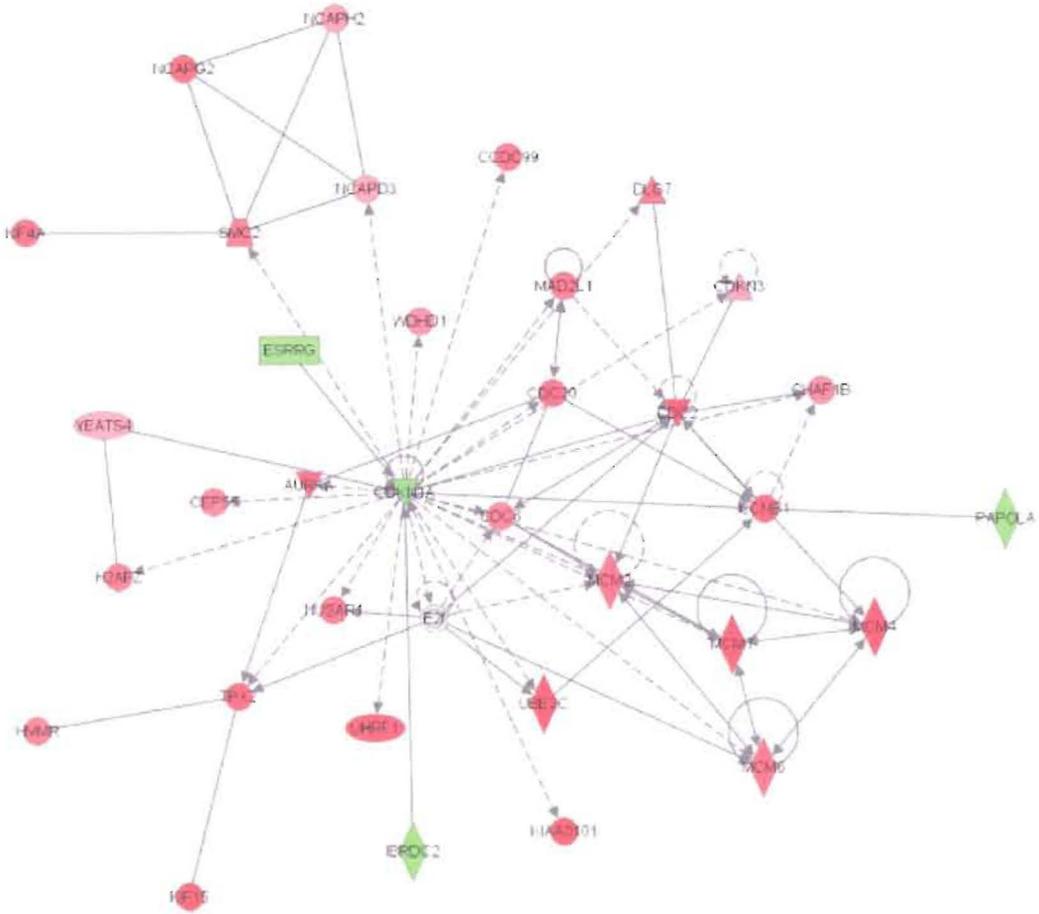


Fig. 6 Functional networks of genes dysregulated in swine parthenogenetic tissues. Ingenuity Pathway Analysis was used to map genes differentially expressed in swine parthenogenetic fetal tissues into functional pathways. Green represents up-regulation of genes with respect to the parthenote, while red represents down-regulation in the parthenote. The following panel depicts highest affected pathway in the combined analysis of all tissues. One of the central nodes or hub is CDKN1A. Up-regulation of CDKN1A results in cell cycle arrest at the G1/S checkpoint and induces apoptosis.

as KEGG (Kyoto Encyclopedia of Genes and Genomes) (Okuda et al. 2008), BioCarta (<http://www.biocarta.com/genes/allpathways.asp>), and Reactome (Matthews et al. 2009). For this process, open-source software such as Cytoscape (Cline et al. 2007, Yeung et al. 2008) and MadNET (Segota et al. 2008) are excellent programs that aid the investigator by mapping array expression datasets on canonical biochemical pathways. A handful of commercial applications like Ingenuity Pathways Analysis (IPA; IngenuitySystems, www.ingenuity.com) are also available. In general, one should choose the application that best suits the scientific question, helpful criteria include text mining options, curated pathway plug-ins, user-created network assembly, data visualization capabilities and flexibility of data import / output formats.

We used IPA to explore pathways altered in parthenogenetic swine conceptus tissues, similar to the approach taken by Jincho et al. (2008). An example of an interactome depicting

genes affected in the parthenote samples compared to normal conceptus samples is shown in Fig. 6. Pathway analysis indicated that cell cycle regulation, growth and proliferation, and cellular assembly pathways were among the most common and most affected pathways in each of the tissues profiled. Our functional analysis of parthenogenetic swine conceptus tissues is in agreement with the parent-conflict hypothesis, as one might expect growth pathways to be asymmetrically affected with a reduction in biogenesis, growth and proliferation pathways.

Finally, biological text mining has also become more attractive due to its ease of use and availability. For the exploration of alternative transcript isoforms, AceView (Thierry-Mieg & Thierry-Mieg 2006) is particularly handy. Recently, the community authorship or wiki concept inspired WikiGenes (Hoffmann 2008) (<http://www.wikigenes.org/>), WikiPathways (<http://www.wikipathways.org>) and regulatory networks based on biomedical discipline (i.e. pathways defining stem cell pluripotency using the PluriNET network (Muller *et al.* 2008 ; <http://www.openstemcellwiki.org/>). The main advantage of wiki concept is dynamic, collaborative forum for scientists to engage in sharing data and publishing ideas.

microRNA (miRNA) and target mRNA

MicroRNAs are known to critically regulate many developmental processes by translational inhibition or destabilizing target mRNAs and are often evolutionarily conserved (Grun *et al.* 2005 , Chen & Rajewsky 2006). Comprehensive arrays are available containing a large number of known mouse and human miRNAs. In species such as swine, exploration of miRNAs is difficult due to the absence of full sequence information that would permit identification of conserved miRNAs and thus the use of these cross-species platforms, although recent reports suggest that the degree of microRNA conservation is such that other species platforms can be used to globally examine swine miRNAs (Huang *et al.* 2008). An alternate approach is to utilize microarray data as a way to predict which microRNAs are affected. This approach is facilitated by the existence of novel bioinformatic tools such as gene set enrichment analysis (GSEA).

Gene set enrichment analysis is a robust method which utilizes gene sets (Molecular Signature Database, MSigDB; <http://www.broad.mit.edu/gsea/msigdb>) to analyze microarray data. The GSEA-P software distinguishes whether genes in known biochemical pathways or coexpression patterns tend to be at the top or bottom of the ranked genome-wide expression dataset or randomly distributed. An enrichment score is provided, which is a value of statistical significance after correction for multiple testing. A full description of the method is available in (Subramanian *et al.* 2005).

Gene set enrichment analysis of the top hundred differentially expressed genes between biparental and parthenote placentas were used as inputs to determine whether there were common microRNAs that are dysregulated. This GSEA approach was able to identify five microRNAs predicted to be differentially expressed in the parthenote samples. Two of these microRNAs have been implicated as ligands for angiotensin receptor II type 1 (AGTR1), a gene responsible for angiogenesis (Sasaki *et al.* 2002), vasoconstriction, and increased pregnancy complication by preeclampsia (Wallukat *et al.* 1999). Gross morphological examination of swine parthenote placentas showed reduced number of blood vessels, and this observation is also supported by differential expression of AGTR1 in placental tissues ($p < 0.0009$) from our microarray datasets. At the time of writing, a single report of miRNAs surveyed in swine fetal tissues observed on day 33 and day 65 of gestation has been reported (Huang *et al.* 2008). While at this point we have not confirmed the differential expression of these miRNAs in our samples by Q-PCR, we have previously used this method to identify and confirm miRNAs affected in human intrauterine growth restriction (manuscript in preparation).

Conclusion

Swine are an attractive model to study fetal growth because their placental morphology is relatively simple—(diffuse, epitheliochorial, non-invasive)—and may provide clues to physiological defects of epigenetic gene dysregulation. The swine parthenote model is already yielding important insights into fetal growth retardation. Our collective analyses of these datasets will contribute a greater understanding of the role of epigenetic mechanisms critical to swine placental function and will hopefully aid our understanding of the formative interactions among fetus, placenta and mother, which are depicted and summarized in Fig. 7.

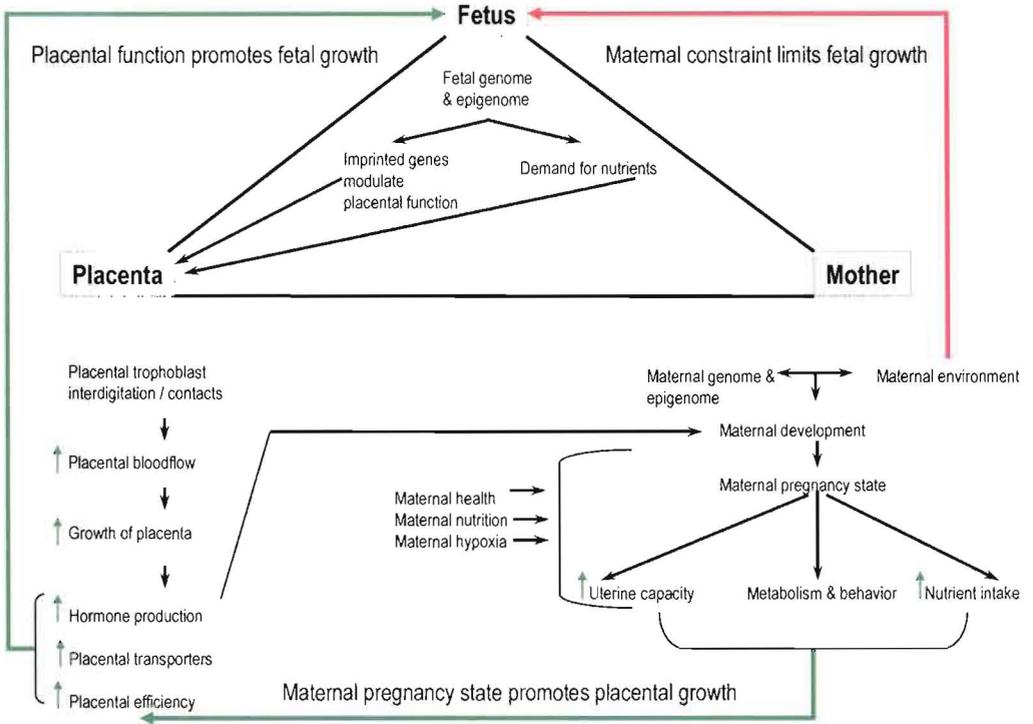


Fig. 7 Signaling between Fetus, Placenta and Mother. The diagram highlights interactions between fetus, placenta and mother in swine pregnancy and was modified from its original version as described in Murphy et al 2006 (Murphy, Smith et al. 2006). The placenta is the nexus between fetus and mother and its function in nutrient exchange is critical for fetal growth and pregnancy outcome as outlined by the various physiological crosstalk. For example, epithelial folds of chorionic trophoblasts create interdigitation and increase placental surface area, which ultimately promotes fetal blood flow, placental and fetal growth, and enhances transport of nutrients across the non-invasive swine placenta. Imprinted genes affect mammalian pregnancy outcome and functional studies by gene-targeting have described intrauterine growth restriction (IUGR) as one disease state by their perturbation. Knockout (KO) studies in mice have shown that the paternally expressed imprinted gene family, such as IGF2 and PEG10, results in placental hypotrophy, while KO conceptus of imprinted maternally expressed growth suppressor PHLDA2 results in placentomegaly. Maternal and fetal genotypes also affect conceptus size and placental efficiency, respectively (Biensen, Wilson et al. 1999). Growth retardation is not limited to placental insufficiency, as severe maternal caloric restriction results in preterm loss and postnatal runtling (Martin-Gronert and Ozanne 2007; Vuguin 2007).

Modern genomic approaches can greatly facilitate the study of physiological phenomena by providing a broad overview of the system, followed by the ability to focus on those pathways/systems that vary. Thus, while genomic analyses are not hypothesis driven, they greatly facilitate the development of hypotheses that have the most likelihood of yielding important biological information. We view genomic approaches as an initial unbiased screening step that can be followed up with more targeted functional experiments. They are not, by themselves typically conclusive, but are extremely useful for hypothesis generation. By comparison, many times candidate gene approaches suffer from too narrow a view of the biological system being studied, and fail to uncover novel interactions and pathways.

Acknowledgments

Portions of the work presented here were supported by a National Research Initiative Grant (2005-35604-15343) from the USDA Cooperative State Research, Education, and Extension Service to JP and BF, by grants HL 51587 and HD 048510 from the National Institutes of Health to JP, by a National Science Foundation IGERT fellowship to ST, a North Carolina State College of Veterinary Medicine doctoral fellowship to SB, and by a NIEHS training grant to NH. This work was performed as part of an initiative from the Center for Comparative Medicine and Translational Research (CCMTR) at the North Carolina State University College of Veterinary Medicine.

References

- Ager E, Suzuki S, Pask A, Shaw G, Ishino F & Renfree MB 2007 Insulin is imprinted in the placenta of the marsupial, *Macropus eugenii*. *Developmental Biology* **309** 317-328.
- Ager EI, Pask AJ, Gehring HM, Shaw G & Renfree MB 2008a Evolution of the CDKN1C-KCNQ1 imprinted domain. *BMC Evolutionary Biology* **8** 163.
- Ager EI, Pask AJ, Shaw G & Renfree MB 2008b Expression and protein localisation of IGF2 in the marsupial placenta. *BMC Developmental Biology* **8** 17.
- Allison DB, Cui X, Page GP & Sabripour M 2006 Microarray data analysis: from disarray to consolidation and consensus. *Nature Review Genetics* **7** 55-65.
- Angiolini E, Fowden A, Coan P, Sandovici I, Smith P, Dean W, Burton G, Tycko B, Reik W, Sibley C & Constancia M 2006 Regulation of placental efficiency for nutrient transport by imprinted genes. *Placenta* **27 Supplement A** 98-102.
- Barry WT, Nobel AB & Wright FA 2005 Significance analysis of functional categories in gene expression studies: a structured permutation approach. *Bioinformatics* **21** 1943-1949.
- Barton SC, Surani MA & Norris ML 1984 Role of paternal and maternal genomes in mouse development. *Nature* **311** 374-376.
- Bischoff SR, Tsai S, Hardison N, York A, Freking BA, Nonneman D, Rohrer G & Piedrahita JA 2008 Identification of SNPs and INDELS in swine transcribed sequences using short oligonucleotide microarrays. *BMC Genomics* **9** 252.
- Borevitz JO, Liang D, Plouffe D, Chang HS, Zhu T, Weigel D, Berry CC, Winzeler E & Chory J 2003 Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Research* **13** 513-523.
- Cattanach BM & Kirk M 1985 Differential activity of maternally and paternally derived chromosome regions in mice. *Nature* **315** 496-498.
- Chapman DD, Shivji MS, Louis E, Sommer J, Fletcher H & Prodohl PA 2007 Virgin birth in a hammerhead shark. *Biology Letters* **3** 425-427.
- Charalambous M, Smith FM, Bennett WR, Crew TE, Mackenzie F & Ward A 2003 Disruption of the imprinted *Grb10* gene leads to disproportionate overgrowth by an *Igf2*-independent mechanism. *Proceedings of the National Academy of Sciences* **100** 8292-8297.
- Chen K & Rajewsky N 2006 Deep conservation of microRNA-target relationships and 3'UTR motifs in vertebrates, flies, and nematodes. *Cold Spring Harbor Symposia On Quantitative Biology* **71** 149-156.
- Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, Vailaya A, Wang PL, Adler A, Conklin BR, Hood L, Kuiper M, Sander C, Schumleivich I, Schwikowski B, Warner GJ, Ideker T & Bader GD 2007 Integration of biological networks and gene expression data using Cytoscape. *Nature Protocols* **2** 2366-2382.
- Craig DW, Pearson JV, Szelinger S, Sekar A, Redman M, Corneveaux JJ, Pawlowski TL, Laub T, Nunn

- G, Stephan DA, Homer N & Huentelman MJ 2008 Identification of genetic variants using bar-coded multiplexed sequencing. *Nature Methods* **5** 887-893.
- Curley JP, Barton S, Surani A & Keverne EB 2004 Coadaptation in mother and infant regulated by a paternally expressed imprinted gene. *Proceedings of the Royal Society of Biological Sciences* **271** 1303-1309.
- D'Haeseleer P 2005 How does gene expression clustering work? *Nature Biotechnology* **23** 1499-1501.
- Dean W, Santos F, Stojkovic M, Zakhartchenko V, Walter J, Wolf E & Reik W 2001 Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos. *Proceedings of the National Academy of Sciences* **98** 13734-13738.
- 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** 3.
- Dindot SV, Farin PW, Farin CE, Romano J, Walker S, Long C & Piedrahita JA 2004 Epigenetic and Genomic Imprinting Analysis in Nuclear Transfer Derived Bos gaurus/Bos taurus Hybrid Fetuses. *Biology of Reproduction* **71** 470-478.
- Edwards CA & Ferguson-Smith AC 2007 Mechanisms regulating imprinted genes in clusters. *Current Opinion in Cell Biology* **19** 281-289.
- Edwards CA, Rens W, Clarke O, Mungall AJ, Hore T, Graves JA, Dunham J, Ferguson-Smith AC & Ferguson-Smith MA 2007 The evolution of imprinting: chromosomal mapping of orthologues of mammalian imprinted domains in monotreme and marsupial mammals. *BMC Evolutionary Biology* **7** 157.
- Farin CE, Farin PW & Piedrahita JA 2004 Development of fetuses from in vitro-produced and cloned bovine embryos. *Journal of Animal Science* **82** Supplement 53-62.
- Farin PW, Piedrahita JA & Farin CE 2006 Errors in development of fetuses and placentas from in vitro-produced bovine embryos. *Theriogenology* **65** 178-191.
- Frank D, Fortino W, Clark L, Musalo R, Wang W, Saxena A, Li CM, Reik W, Ludwig T & Tycko B 2002 Placental overgrowth in mice lacking the imprinted gene Ipl. *Proceedings of the National Academy of Sciences* **99** 7490-7495.
- Gatti DM, Sypa M, Rusyn I, Wright FA & Barry WT 2009 SAFEGUI: resampling-based tests of categorical significance in gene expression data made easy. *Bioinformatics* **25** 541-542.
- Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES & Nusbaum C 2009 Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology* **27** 182-189.
- Grun D, Wang YL, Langenberger D, Gunsalus KC & Rajewsky N 2005 MicroRNA target predictions across seven Drosophila species and comparison to mammalian targets. *PLoS Computational Biology* **1** 13.
- Guillemot F, Caspary T, Tilghman SM, Copeland NG, Gilbert DJ, Jenkins NA, Anderson DJ, Joyner AL, Rossant J & Nagy A 1995 Genomic imprinting of Mash2, a mouse gene required for trophoblast development. *Nature Genetics* **9** 235-242.
- Hamady M, Walker JJ, Harris JK, Gold NJ & Knight R 2008 Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nature Methods* **5** 235-237.
- Hagemann LJ, Peterson AJ, Weilert LL, Lee RS & Tervit HR 1998 In vitro and early in vivo development of sheep gynogenones and putative androgenones. *Molecular Reproduction and Development* **50** 154-162.
- Hajkova P, Erhardt S, Lane N, Haaf T, El-Maarri O, Reik W, Walter J & Surani MA 2002 Epigenetic reprogramming in mouse primordial germ cells. *Mechanisms of Development* **117** 15-23.
- Hamlett WC 1989 Evolution and Morphogenesis of the Placenta in Sharks. *Journal of Experimental Zoology* **252** 35-52.
- Heiden TC, Haines AN, Manire C, Lombardi J & Koob TJ 2003 Structure and permeability of the egg capsule of the bonnethead shark, *Sphyrna tiburo*. *Journal of Experimental Zoology* **303A** 577-589.
- Hill JR & Roussel AJ 1999 Clinical and pathologic features of cloned transgenic calves and fetuses 13 case studies. *Theriogenology* **51** 1451-1465.
- Hoffmann R 2008 A wiki for the life sciences where authorship matters. *Nature Genetics* **40** 1047-1051.
- Hore TA, Rapkins RW & Graves JA 2007 Construction and evolution of imprinted loci in mammals. *Trends in Genetics* **23** 440-448.
- Huang da W, Sherman BT & Lempicki RA 2009 Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37** 1-13.
- Huang TH, Zhu MJ, Li XY & Zhao SH 2008 Discovery of porcine microRNAs and profiling from skeletal muscle tissues during development. *PLoS ONE* **3** 3225.
- Huh JH, Bauer MJ, Hsieh TF & Fischer RL 2008 Cellular programming of plant gene imprinting. *Cell* **132** 735-744.
- Jansen RC & Nap JP 2001 Genetical genomics: the added value from segregation. *Trends in Genetics* **17** 388-391.
- Jincho Y, Sotomaru Y, Kawahara M, Ono Y, Ogawa H, Obata Y & Kono T 2008 Identification of genes aberrantly expressed in mouse embryonic stem cell-cloned blastocysts. *Biology of Reproduction* **78** 568-576.
- Jones CJ & Hamlett WC 2004 Structure and glycosylation of the term yolk sac placenta and uterine attachment site in the viviparous shark *Mustelus canis*. *Placenta* **25** 820-828.
- Kerr G, Ruskin HJ, Crane M & Doolan P 2008 Techniques for clustering gene expression data. *Computers in Biology and Medicine* **38** 283-293.
- Lawton BR, Sevigny L, Oberfell C, Reznick D, O'Neill RJ & O'Neill MJ 2005 Allelic expression of IGF2 in live-bearing, matrotrophic fishes. *Development Genes*

and *Evolution* 215 207-212.

- Lefebvre L, Viville S, Barton SC, Ishino F, Keverne EB & Surani MA 1998 Abnormal maternal behaviour and growth retardation associated with loss of the imprinted gene *Mest*. *Nature Genetics* 20 163-169.
- Li L, Keverne EB, Aparicio SA, Ishino F, Barton SC & Surani MA 1999 Regulation of maternal behavior and offspring growth by paternally expressed *Peg3*. *Science* 284 330-333.
- Mathews L, Gopinath G, Gillespie M, Caudy M, Croft D, de Bono B, Garapati P, Hemish J, Hermjakob H, Jassal B, Kanapin A, Lewis S, Mahajan S, May B, Schmidt E, Vastrik I, Wu G, Birney E, Stein L & D'Eustachio P 2009 Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Research* 37 619-622.
- McGrath J & Solter D 1984 Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* 37 179-183.
- Moore T & Haig D 1991 Genomic imprinting in mammalian development: a parental tug-of-war. *Trends in Genetics* 7 45-49.
- Mosquera JL & Sanchez-Pla A 2008 SerbGO: searching for the best GO tool. *Nucleic Acids Research* 36 71.
- Mouse Genome Sequencing Consortium, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexander S, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermizakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyraes E, Felsenfeld A, Fellwells GA, Flück P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Graves TA, Green ED, Gregory S, Guigó R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK, Karolchik D, Kasprzyk A, Kawai J, Keibler E, Kells C, Kent WJ, Kirby A, Kolbe DL, Korfi I, Kucherlapati RS, Kulbokas EJ, Kulp D, Landers T, Leger JP, Leonard S, Letunic I, Levine R, Li J, Li M, Lloyd C, Lucas S, Ma B, Maglott DR, Mardis ER, Mathews L, Mauceli E, Mayer JH, McCarthy M, McCombie WR, McLaren S, McLay K, McPherson JD, Meldrum J, Meredith B, Mesirov JP, Miller W, Miner TL, Mongin E, Montgomery KT, Morgan M, Mott R, Mullikin JC, Muzny DM, Nash WE, Nelson JO, Nhan MN, Nicol R, Ning Z, Nusbaum C, O'Connor MJ, Okazaki Y, Oliver K, Overton-Larty E, Pachter L, Parra G, Pepin KH, Peterson J, Pevzner P, Plumb R, Pohl CS, Poliakov A, Ponce TC, Ponting CP, Potter S, Quail M, Reymond A, Roe BA, Roskin KM, Rubin EM, Rust AG, Santos R, Sapojnikov V, Schultz B, Schultz J, Schwartz MS, Schwartz S, Scott C, Seaman S, Searle S, Sharpe T, Sheridan A, Shownkeen R, Sims S, Singer JB, Slater G, Smit A, Smith DR, Spencer B, Stabenau A, Stange-Thomann N, Sugnet C, Suyama M, Tesler G, Thompson J, Torrents D, Trevaskis E, Tromp J, Ucla C, Ureta-Vidal A, Vinson JP, Von Niederhausern AC, Wade CM, Wall M, Weber RJ, Weiss RB, Wendl MC, West AP, Wetterstrand K, Wheeler R, Whelan S, Wierzbowski J, Willey D, Williams S, Wilson RK, Winter E, Worley KC, Wyman D, Yang S, Yang SP, Zdobnov EM, Zody MC & Lander ES 2001 Initial sequencing and analysis of the human genome. *Nature* 409 860-921.
- Müller FJ, Laurent LC, Kostka D, Ulitsky I, Williams R, Lu C, Park IH, Rao MS, Shamir R, Schwartz PH, Schmidt NO & Loring JF 2008 Regulatory networks define phenotypic classes of human stem cell lines. *Nature* 455 401-405.
- Okuda S, Yamada T, Hamajima M, Itoh M, Katayama T, Bork P, Goto S & Kanehisa M 2008 KEGG Atlas mapping for global analysis of metabolic pathways. *Nucleic Acids Research* 36 423-426.
- Ono R, Nakamura K, Inoue K, Naruse M, Usami T, Wakisaka-Saito N, Hino T, Suzuki-Migishima R, Ogonuki N, Miki H, Kohda T, Ogura A, Yokoyama M, Kaneko-Ishino T & Ishino F 2006 Deletion of *Peg10*, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. *Nature Genetics* 38 101-106.
- Rapkins RW, Hore T, Smithwick M, Ager E, Pask AJ, Renfree MB, Kohn M, Hameister H, Nicholls RD, Deakin JE & Graves JA 2006 Recent assembly of an imprinted domain from non-imprinted components. *PLoS Genetics* 2 182.
- Reik W & Walter J 2001 Genomic imprinting: parental influence on the genome. *Nature Reviews Genetics* 2 21-32.
- Renfree MB, Ager EI, Shaw G & Pask AJ 2008 Genomic imprinting in marsupial placentation. *Reproduction* 136 523-531.
- Reznick D, Meredith R & Collette BB 2007 Independent evolution of complex life history adaptations in two families of fishes, live-bearing halfbeaks *zenarchopteridae*, *beloniformes* and *poeciliidae* *cyprinodontiformes*. *Evolution* 61 2570-2583.
- Sasaki K, Murohara T, Ikeda H, Sugaya T, Shimada T, Shintani S & Imaizumi T 2002 Evidence for the importance of angiotensin II type 1 receptor in ischemia-induced angiogenesis. *Journal of Clinical Investigation* 109 603-611.
- Segota I, Bartonicek N & Vlahovicek K 2008 MADNet: microarray database network web server. *Nucleic Acids Research* 36 332-335.
- Sibley CP, Coan PM, Ferguson-Smith AC, Dean W, Hughes J, Smith P, Reik W, Burton GJ, Fowden AL & Constância M 2004 Placental-specific insulin-like growth factor 2 *Igf2* regulates the diffusional exchange characteristics of the mouse placenta. *Proceedings of the National Academy of Sciences* 101 8204-8208.
- 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* **102** 15545-15550.
- Surani MA, Barton SC & Norris ML** 1984 Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* **308** 548-550.
- Takahashi K, Kobayashi T & Kanayama N** 2000 p57Kip2 regulates the proper development of labyrinthine and spongiotrophoblasts. *Molecular Human Reproduction* **8** 1019-1025.
- Tanaka M, Gertsenstein M, Rossant J & Nagy A** 1997 Mash2 acts cell autonomously in mouse spongiotrophoblast development. *Developmental Biology* 55-65.
- Thierry-Mieg D & Thierry-Mieg J** 2006 AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biology* **7** Supplement 1-14.
- Tsai S, Cassady JP, Freking BA, Nonneman DJ, Rohrer GA & Piedrahita JA** 2006a Annotation of the Affymetrix porcine genome microarray. *Animal Genetics* **37** 423-424.
- Tsai S, Mir B, Martin AC, Estrada JL, Bischoff SR, Hsieh WP, Cassady JP, Freking BA, Nonneman DJ, Rohrer GA & Piedrahita JA** 2006b Detection of transcriptional difference of porcine imprinted genes using different microarray platforms. *BMC Genomics* **7** 328.
- Varrault A, Gueydan C, Delalbre A, Bellmann A, Houssami S, Aknin C, Severac D, Chotard L, Kahli M, Le Digarcher A, Pavlidis P & Journot I** 2006 Zac1 regulates an imprinted gene network critically involved in the control of embryonic growth. *Developmental Cell* **11** 711-722.
- Wallukat G, Homuth V, Fischer T, Lindschau C, Horstkamp B, Jüpner A, Baur E, Nissen E, Vetter K, Neichel D, Dudenhausen JW, Haller H & Luft FC** 1999 Patients with preeclampsia develop agonistic autoantibodies against the angiotensin AT1 receptor. *Journal of Clinical Investigation* **103** 945-952.
- Walsh C, Glaser A, Fundele R, Ferguson-Smith A, Barton S, Surani MA & Ohlsson R** 1994 The non-viability of uniparental mouse conceptuses correlates with the loss of the products of imprinted genes. *Mechanisms of Development* **46** 55-62.
- Wang Z, Gerstein M & Snyder M** 2009 RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10** 57-63.
- Wang ZQ, Fung MR, Barlow DP & Wagner EF** 1994 Regulation of embryonic growth and lysosomal targeting by the imprinted Igf2/Mpr gene. *Nature* **372** 464-467.
- West MA, van Leeuwen H, Kozik A, Kliebenstein DJ, Doerge RW, St Clair DA & Michelmore RW** 2006 High-density haplotyping with microarray-based expression and single feature polymorphism markers in Arabidopsis. *Genome Research* **16** 787-795.
- Winzeler EA, Richards DR, Conway AR, Goldstein AL, Kalman S, McCullough MJ, McCusker JH, Stevens DA, Wodicka L, Lockhart DJ & Davis RW** 1998 Direct allelic variation scanning of the yeast genome. *Science* **281** 1194-1197.
- Wold B & Myers RM** 2008 Sequence census methods for functional genomics. *Nature Methods* **5** 19-21.
- Yeung N, Cline MS, Kuchinsky A, Smoot ME & Bader GD** 2008 Exploring biological networks with Cytoscape software. In *Current Protocols Bioinformatics*, Chapter 8, Unit 8.13. Eds Baxevanis AD, Petsko GA, Stein LD, Stormo GD, Yates JR & Davison DB. Hoboken: John Wiley & Sons, Inc.
- Zhao SH, Recknor J, Lunney JK, Nettleton D, Kuhar D, Orley S & Tuggle CK** 2005 Validation of a first-generation long-oligonucleotide microarray for transcriptional profiling in the pig. *Genomics* **86** 618-625.
- Zhu J, King T, Dobrinsky J, Harkness L, Ferrier T, Bosma W, Schreier LL, Guthrie HD, DeSousa P & Wilmot I** 2003 In vitro and in vivo developmental competence of ovulated and in vitro matured porcine oocytes activated by electrical activation. *Cloning & Stem Cells* **5** 355-365.