Review

Functional genomics in farm animals – Microarray analysis

Christian Bendixen *, Jakob Hedegaard, Per Horn

Department of Genetics and Biotechnology, Danish Institute of Agricultural Sciences, P.O. Box 50, DK-8830 Tjele, Denmark

Abstract

The rapid progression of farm animal genomics has introduced novel technologies capable of presenting global descriptions of biological systems at the level of gene and protein expression and protein interaction. To fully benefit from these developments, experimental designs have to be adapted to these new technologies, and important considerations must be made in the choice of technologies and methods of analysis to be used. This paper addresses practical issues in the use of microarray based methods for gene-expression analysis in farm animals, and provides an overview of different array-platforms as well as a presentation of methods and software for the analysis of array data. Experimental design and the selection of animals and samples for microarray studies in farm animals present novel challenges, which are often overlooked. In particular, the frequent use of half sibs and full sibs in animal studies increases the risk of falsely identifying genes as being differentially expressed, due to genetic linkage of the gene to a QTL or a major gene affecting the trait in question.
© 2005 Elsevier Ltd. All rights reserved.

Keywords: Microarray; Review; Farm animals; cDNA array; Oligo array; Array analysis

Contents

---

* Corresponding author. Tel.: +45 8999 1356; fax: +45 8999 1300.
E-mail address: Christian.Bendixen@agrsci.dk (C. Bendixen).

1. Introduction

Microarray technology facilitates quantitative assessment of gene expression levels for several thousand genes simultaneously. Patrick Brown's group at Stanford was the first to print arrays of PCR fragments amplified from cDNA libraries on a glass surface the size of a standard microscope slide using a robotic printing device (DeRisi et al., 1996). The technology remains essentially the same today and is referred to as cDNA microarrays. To conduct a microarray analysis of gene expression, RNA is purified from tissues or cells of interest and labeled with fluorescent dyes. After hybridization of the labeled RNA to the array, the slides are scanned and the fluorescent signal in each cDNA element on the slide provides a measure of the expression of the corresponding gene (Duggan, Bittner, Chen, Meltzer, & Trent, 1999). The different steps used in a microarray experiments are described in more detail in the following sections. In addition to PCR-fragments amplified from cDNA clones, it is possible to print arrays using long synthetic oligonucleotides, giving increased flexibility in design and the potential to increase the specificity of the hybridization (Chou, Hsia, Mooney, & Schnable, 2004; Hessner et al., 2004; Hornshoj, Stengaard, Panitz, & Bendixen, 2004). Typically, oligonucleotides of 50–70 bp in length are used for printing, each representing a unique sequence close to the 3′-end of a particular transcript. A special type of oligonucleotide arrays is the high-density array produced by the use of photolithographic technology to chemically synthesize short oligonucleotides directly on the surface of the chip, a technology pioneered by Affymetrix.

2. Farm animal microarray resources

To conduct microarray experiments on farm animals a number of both commercial (Table 1) and custom-made arrays are available (see Section 5). An increasing number of vendors offer to print arrays from customer-provided clone collections or to synthesize sets of oligonucleotides based on customer-provided sequence information. However, many research groups and institutions have developed their own microarray facilities. Many of these focused initially on tissue specific or otherwise specialized arrays, made by printing cDNAs fragments expressed in the tissue of interest (Bai et al., 2003; Nobis et al., 2003; van Hemert, Ebbelaar, Smits, & Rebel, 2003). Construction of genome wide cDNA arrays requires access to cDNA resources from many tissues and developmental stages in order to obtain adequate gene representation. Currently there is a critical lack of availability of genome-wide farm animal cDNA arrays. Strategies to collect cDNA libraries from many different laboratories are hampered by the use of different vector systems resulting in the need for different sets of vector primers. In addition, logistic problems in assembling appropriate cDNA-panels, amplifying and purifying PCR fragments have resulted in numerous groups to focus on the development and use of long oligonucleotide arrays, which is made feasible due to the increased availability of genomic sequences.

The decision to print homemade arrays or to use commercially available arrays requires careful consideration. Custom-made arrays require investments in expensive equipment, like robotic printing devices, and rely on the availability of cDNA libraries for the production of cDNA microarrays or on access to panels of synthetic oligonucleotides for the printing of oligonucleotide arrays. This option is most suitable for research groups and institutions with preexisting genome research facilities. However, use of oligonucleotide arrays provides the freedom to optimize array-design and to easily include new genes as more sequences become available. The use of commercial arrays requires investments in less equipment (vs. custom-made arrays), but the price per array tends to be higher. Hence, the use of in-house or commercial arrays is highly dependent

Table 1
Selected commercial available microarray resources for farm animals

Name	Array type	No. spots/probe	No. genes	Vendor[c]
GeneChip® Porcine Genome Array	High density oligo	23,937	20,201	Affymetrix
GeneChip® Chicken Genome Array[a]	High density oligo		28,000	Affymetrix
GeneChip® Bovine Genome Array	High density oligo	24,072	23,000	Affymetrix
Pig Genome Oligo Set ver. 1.0	Array ready oligos	10,665	10,665	Operon
Pig Genome Oligo Extension Set ver. 1.0	Array ready oligos	2632	2632	Operon
Pig Immune Array	cDNA microarray		2860	ARK genomics
Chicken Embryo Array	cDNA microarray		1152	ARK genomics
Chicken Immune Array	cDNA microarray		5000	ARK genomics
Chicken Neuroendocrine Array	cDNA microarray		4800	ARK genomics
Pig Oligo array[b]	Oligo microarray	13,297	13,297	ARK genomics

[a] This array also contains 689 probe sets for detecting 684 transcripts from 17 avian viruses.
[b] Qiagen Pig Genome Oligo Set Version 1.0 and the Pig Genome Oligo Extension Set Version 1.0 printed on one slide.
[c] Affymetrix (http://www.affymetrix.com), Operon (http://www.operon.com), ARK Genomics (http://www.ark-genomics.org).

upon the number of arrays needed and the flexibility required.

3. Experimental design and methods

Studying changes in gene-expression for single genes has become a general tool, and many research groups have studied differential expression of single genes. However, with the introduction of array-based methods, the experimental hypotheses have changed from addressing differential expression of single genes towards a global description of expression changes with no prior assumptions made concerning the genes involved. Optimizing the designs of these experiments is a critical need, and some relevant issues will be discussed below.

3.1. Selection of animals and samples

Selecting meaningful samples for a microarray experiment is of great importance for reducing the risk of acquiring misleading results, and incorrect sampling is a major pitfall in microarray experiments (Takemura et al., 2005). In studies involving farm animals, groups of half-sibs or full-sibs are often chosen in order to reduce the background of biological variation unrelated to the phenotype, treatment, or genotype in question. In studies involving the testing of differential expression for a single or only a few genes, the use of family-related animals has not previously been of concern. However, when conducting array-based experiments assaying essentially all of the genes in an organism, special care must be taken to avoid identifying genes that are differentially expressed due to their genetic linkage to a QTL or to a major gene influencing the trait under investigation. In particular, special care must be taken when studying expression differences between haplotypes of candidate genes or between QTL-alleles. In these cases, the study of half-sibs and full sibs in families, segregation of genes in the QTL region, which contain genetic variation that affects gene expression, will co-segregate with the gene or QTL of interest. Studies in humans and other organisms have shown that differences in gene expression account for a major part of the variation within and among species. Yan, Yuan, Velculescu, Vogelstein, and Kinzler (2002) found genetic variation in 6 out of 13 genes examined using an experimental approach that could confidently identify variation when the expression of the two alleles of a gene differed by more than 20%, demonstrating the possible magnitude of the problem associated with linked genetic variation affecting gene expression.

The actual selection of tissue samples requires choices be made regarding the type of sampling needed, e.g., needle-biopsies or post-slaughter sampling. The size of the sample affects the yield of RNA and consequently determines whether one uses direct labeling of RNA or various methods for amplification (see below). Tissue samples are almost invariably a heterogeneous collection of different cell types, resulting in the generation of average expression profiles, which may hide differences in expression profiles between individual cell types. Therefore, methods for dissecting complex tissue samples, e.g., by using high precision laser systems, has made it possible to study expression patterns of specific and individual cell types (Luzzi, Mahadevappa, Raja, Warrington, & Watson, 2003; Upson et al., 2004). Samples must be handled carefully to avoid degradation of the RNA molecules, thus freezing samples in liquid nitrogen or submersion in RNA-stabilizing reagents immediately after sampling is important. Purification of intact and pure RNA from samples is critical for high quality results using microarrays. A variety of different purification methods are available from companies like Qiagen (Qiagen, 2005), Ambion (Ambion, 2005) and Invitrogen (Invitrogen, 2005).

3.2. Strategies for labeling and hybridization

Both cDNA arrays as well as long and short oligonucleotide arrays are hybridized to labeled samples. The most commonly used labels are fluorescent tags and a number of different technologies have been developed for labeling extracted mRNAs. Yu et al. (2002) reviews the different labeling technologies. Unamplified labeling during reverse-transcription of RNA into cDNA requires around 20 μg of total RNA, which is often not available due to limited sample size. However, techniques for amplifying RNA allow the use of smaller sample sizes (a few nanograms) of total RNA (Dafforn et al., 2004; Wang, Hu, Hamilton, Coombes, & Zhang, 2003). The specificity of hybridization of cellular mRNAs to the specific array elements is controlled by factors such as time, temperature, ionic strength and washing procedures. Hence, optimizing hybridization conditions is an important step, for which a wide range of methods and procedures has been published (Yu et al., 2002). Using Affymetrix chips, all samples are labeled with one single dye, and differential expression is measured by hybridizing comparable samples to separate arrays (Lockhart et al., 1996). Most cDNA and long-oligo array systems are based on the labeling of comparable samples with different fluorescent dyes, followed by co-hybridization of the two samples on one single array target (DeRisi et al., 1996). This presents a great advantage for reducing the impact of technical variation, and allows very robust approaches to the subsequent data analysis, as discussed below. The choice of the two samples, however, requires careful consideration. One approach is to co-hybridize all samples with a common reference sample on individual arrays and

compare the samples indirectly via the reference sample (reference design; Konig, Baldessari, Pollet, Niehrs, & Eils, 2004; Park et al., 2004). Another approach is to compare the samples directly on the same array using a loop design (Vinciotti et al., 2005). Optimizing experimental designs has been intensively debated, and interesting considerations can be found in numerous reviews (Bolstad, Collin, Simpson, Irizarry, & Speed, 2004; Churchill, 2002; Yang & Speed, 2002).

Array hybridization can be performed manually by applying the samples to the arrays under cover slips followed by incubation at a specific temperature. Better reproducibility, and in some cases stronger signal, can be achieved using automatic or semi-automatic hybridization stations. Apart from decreasing the experimental variation of the hybridization process these stations facilitate agitation of the hybridization mixture with the potential effect of speeding up and improving the hybridization process. Hybridization stations are available from (Genomic Solutions, 2004; Tecan, 2005; Ventana, 2005). After hybridization unbound sample are washed off and the slides are dried.

3.3. Scanning and image analysis

Hybridized slides are scanned using either of two systems. A CCD-based system uses filtered white light to excite the dyes. Another more commonly used system involves confocal scanners equipped with lasers of dye-specific wavelengths. Affymetrix arrays require a dedicated scanner, while custom-made or commercial arrays printed on standard microscope slides can be scanned using a variety of scanners. Leading manufacturers of scanners include Perkin–Elmer (PE, 2004), Axon Instruments (MDC, 2005) and Agilent Technologies (Agilent, 2005). One 16-bit tif image is obtained from each channel from each array and these are transformed to data matrixes using image analysis software. These matrixes contain both spot and background intensities, and the spot intensity values are indicative of gene expression levels. Comparing the relative expression data will be discussed below. Most scanners are supplied with image analysis software, but alternative image analysis programs, based on a wide range of algorithms are widely available (Glasbey & Ghazal, 2003; Jain et al., 2002; Petrov & Shams, 2004; Yang, Buckley, Dudoit, & Speed, 2002; Yang, Buckley & Speed, 2001). Databases dedicated to the storage of information obtained from microarray experiments are available both commercially and as freeware. A widely used solution is BioArray Software Environment (BASE) (Saal et al., 2002), which is a free web-based database for storing both the massive amounts of data generated by microarray analyses and the information regarding the samples used. To facilitate interpretation of results, the Microarray Gene Expression Data Society (Anonymous, 2002; MGED, 2005)

has proposed a set of guidelines (MIAME: Minimum Information About a Microarray Experiment; Brazma et al., 2001) to follow when publishing microarray experiments. The public repositories, ArrayExpress at EBI (EBI, 2005) and GEO at NCBI (GEO, 2005) are designed to accept, hold and distribute MIAME compliant microarray data.

4. Analysis of microarray data

A wide selection of commercial and non-commercial software programs for array data analysis have recently been developed. Commercial packages are usually more user friendly, but more expensive and inflexible than the non-commercial products, where new analysis tools can be more easily implemented. The software packages "Significance Analysis of Microarrays" (SAM, 2002; Tusher, Tibshirani, & Chu, 2001) and "Bioconductor" (Bioconductor, 2004; Gentleman et al., 2004; R, 2004) are widely used non-commercial software programs. See (Dresen, Husing, Kruse, Boes, & Jockel, 2003; Liu, Yao, Fayz, Womble, & Krawetz, 2004) for overviews of available software.

4.1. Normalization

An important issue of microarray data analysis is how to normalize the dataset in order to remove systematic variation (Quackenbush, 2002). The source of these variations has been comprehensively reviewed by Schuchhardt et al. (2000). Initial plotting of raw data is useful for detecting unwanted systematic variations in the data. Box-plots for each array of raw intensities and the raw log-ratios (the log to the ratio between the two "dye"-channels is a measure of relative expression) can be used to detect and exclude arrays of low quality (Fig. 1). Image plots of the raw log-ratios can be used to detect spatial deviations that should be accounted for in the downstream analysis (Fig. 2). Generating *MA*-plots by plotting the log-ratio (*M*) against the average intensity (*A*) for each spot facilitates the detection of intensity dependent deviations of the log-ratios due to dye effects and intensity dependent variation in the data (Yang, Buckley, Dudoit, et al., 2002) (Fig. 3). Variations in microarray data detected by the initial plots must be removed (normalized) before valuable information can be extracted. Typically, normalization follows an initial estimation of background signal and can be performed either within or between arrays. Simple subtraction of background signal from foreground intensities does not always give the most precise estimate of the real signal intensities (Yang, Buckley, Dudoit, et al., 2002) and background handling has gained increased awareness as a major issue in microarray data analysis. Background handling can be optimized by suiting the best segmentation
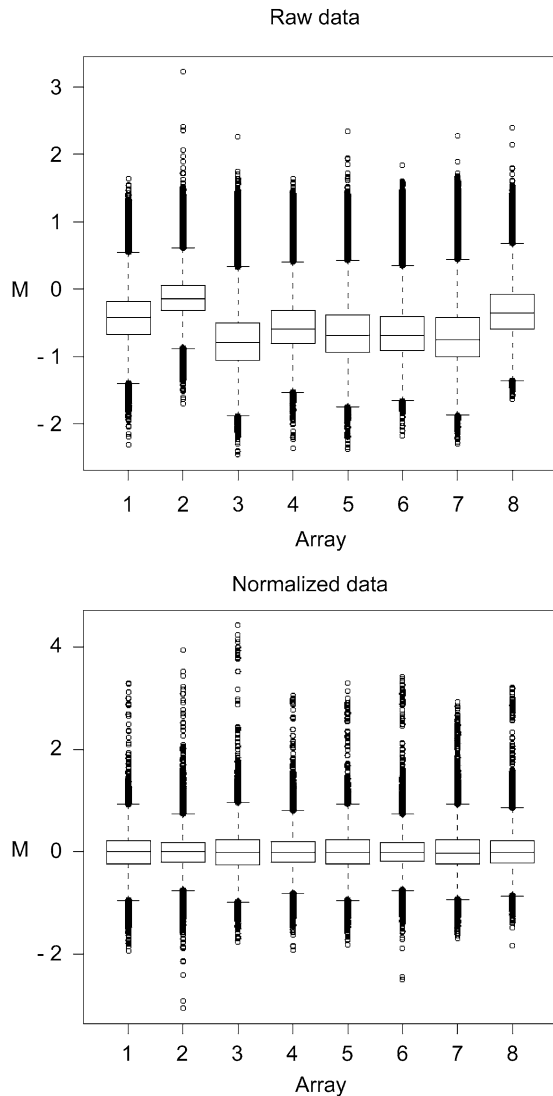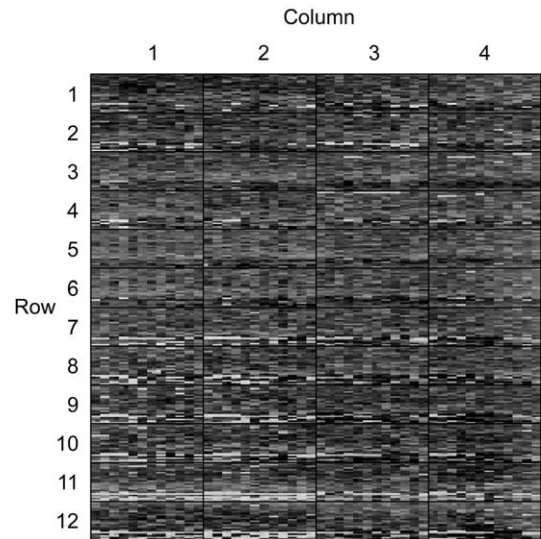
Fig. 2. Spatial image plot of raw, not background subtracted log ratios from array 6 (Fig. 1). Colours range from the most negative log-ratio (black) to the most positive log-ratio (white). Columns (1–4) and rows (1–12) indicate the 48 supergrids on the slide. The range of the log-ratios are seen to be unaffected by the spatial location on the slide indication that a global normalization approach could be used.

Fig. 1. Boxplots for eight arrays of raw, background subtracted log ratios (top) and of background subtracted loess normalized log ratios (bottom). The data is obtained from eight porcine cDNA arrays produced at DIAS which were hybridised with labelled porcine liver RNA and lung RNA. For individual arrays, it is seen that the raw log-ratios are quit narrow distributed around the median log-ratio. Comparing the boxplots of raw data between arrays reveal different distributions and a general tendency for the median log ratio to be below $M = 0$ indicating the need for normalization. Following background subtraction and loess normalization it is seen that the log ratios are centred on $M = 0$ and that the distribution for individual arrays is uniform.

method for the array (Ahmed, Vias, Iyer, Caldas, & Brenton, 2004) and spatial non-uniformity on the array is dealt with using a log-linear interpolation method to adjust lower intensities (Edwards, 2003). Normalization within arrays is generally based on the assumption that most genes are not differentially expressed or there is symmetry between up- and down-regulated genes. Most points in an *MA*-plot will hence fall along a horizontal line through $M = 0$. However, due to variation and

non-linearity in readouts from array scanners this is rarely the case and several strategies have been implemented to remove this deviation from normality. Curve-fitting strategies are the most well known approaches, where local regression is used to estimate a fit to the experimental curve followed by a re-centering of the data based on this fit. The lowess (Yang, Dudoit, et al., 2002) and centralization methods (Zien, Aigner, Zimmer, & Lengauer, 2001) belong to this type of technique. The spatial patterns of log-ratios that are sometimes observed in image plots or print-tip group *MA*-plots can be removed by regional smoothing by a local mean normalization (Colantuoni, Henry, Zeger, & Pevsner, 2002), a method that has been further developed to also remove intensity biases by a joint smoothing method (Cui, Kerr, & Churchill, 2003). Also print-tip loess normalization in combination with quality weights for individual spots has proven useful for removing spatial effects (Smyth & Speed, 2003). Normalization strategies based on scaling of total intensities or the use of housekeeping genes has been reviewed by Quackenbush (2001). The rationale behind normalizing between arrays is to adjust the signals from the different arrays to a comparable scale. This includes simple scaling between the arrays, assuming that the total intensity on each array should be the same (Bilban, Buehler, Head, Desoye, & Quaranta, 2002; Quackenbush, 2001), or that the coefficient of variation across the array is constant (Chen et al., 2002). ANOVA methods also have been used to adjust for overall effects of array and dye across genes (Kerr, Martin, & Churchill, 2000; Wolfinger et al.,
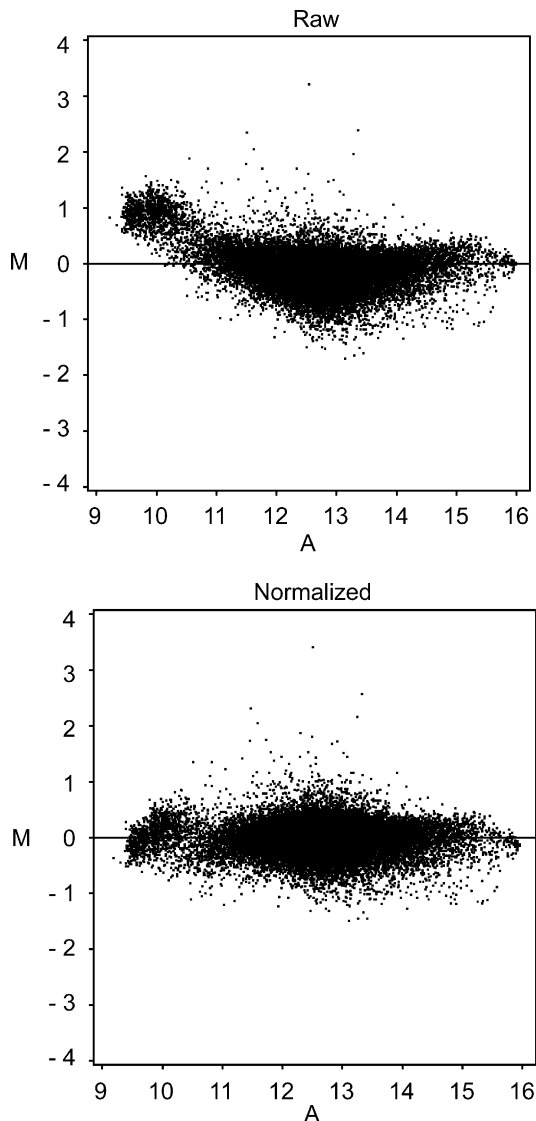
Fig. 3. *MA*-plots of raw, background subtracted log ratios (top) and loess normalized background subtracted log ratios (bottom) from array 6 (Fig. 1). In the top *MA*-plot, it is seen that the average intensity (*A*) of the log-ratios are distributed from just above 9 to 16 and that the log-ratios (*M*) are centred on $M = 0$, but show a clear deviation at low average intensities. This deviation has been corrected by applying loess normalization to the data as seen in the bottom *MA*-plot.

2001). Forcing the distribution of ratios between arrays to be the same is accomplished by the quantile normalization (Bolstad, Irizarry, Astrand, & Speed, 2003).

4.2. Analysis of significance and clustering

Normalized data are assumed to be corrected for systematic variations and the remaining variation is expected to reflect the biological mechanism(s) in question. Further analysis of data can be divided into supervised and non-supervised methods. Supervised methods are used to identify differentially expressed genes between groups of samples (Qu & Xu, 2004).

The grouping can, for example, be based on phenotypic observations or genotypes. Various statistical methods exist for selecting differentially expressed genes between groups. Speed and co-workers use a non-linear smoothing algorithm for the normalization of log ratios followed by permutation-based *t* statistics for testing the significance of each gene (Yang, Dudoit, et al., 2002) and others have demonstrated that ANOVA methods can be used to estimate changes in gene expression that are corrected for potential "noise" (Kerr et al., 2000). The framework of these two groups has been extended to account for correlations and multiple sources of variance in assessing gene significance via mixed models (Wolfinger et al., 2001). Traditionally, differentially expressed genes have been identified by means of their variations from a certain threshold. Typically, this threshold has been set to a factor of two (DeRisi, Iyer, & Brown, 1997). More recently, Yang, Chen, et al. (2002) define differential gene expression as greater than 2 SD from the mean, and they classify genes with fold changes greater than 1.5 as up or down regulated at a 95% confidence interval. Non-supervised methods for analysis of microarray data aim at identifying unknown relationships between samples and/or genes. A widely applied technique for this is hierarchical clustering with can be used in a one-way approach to identify clusters of either samples or genes or in a two-way approach where genes and samples are clustered simultaneously (Brown et al., 2000). Clustering the expression profiles of the genes can be used to identify possible co-varying and potentially functionally related genes. Other non-supervised method is k-means clustering (Tavazoie, Hughes, Campbell, Cho, & Church, 1999), self organizing maps (Toronen, Kolehmainen, Wong, & Castren, 1999) and principal component analysis (Raychaudhuri, Stuart, & Altman, 2000). For more detailed information regarding array analysis please refer to Speed (2003) and Parmigiani, Garret, Izizarry, and Zeger (2003).

5. Microarray studies in farm animals

Microarrays containing well-characterized genes have only recently become available for domestic animals such as cattle, chicken, and pigs, and the first reports on the development of microarrays for expression analysis in farm animals have been published. Many of those initial studies used tissue specific arrays and contain only a limited number of genes. The Center for Animal Functional Genomics, Michigan State University, USA, has constructed cDNA microarrays from a normalized bovine total leukocyte cDNA library (Yao, Burton, Saama, Sipkovsky, & Coussens, 2001) and from normalized porcine muscle (Yao, Coussens, Saama, Suchyta, & Ernst, 2002) and brain (Nobis et al., 2003) cDNA libraries, each representing less than 1000 genes. At the

Division of Animal and Veterinary Sciences, West Virginia University, USA, an oocyte microarray has been constructed and experiments performed to identify genes preferentially expressed in foetal ovary relative to somatic tissues (Yao et al., 2004). At the Department of Pathology, University of Guelph, Canada, Tao, Mallard, Karrow, and Bridle (2004) constructed a small-scale bovine immune-endocrine cDNA array consisting of 167 cDNA sequences and they used this array to demonstrate differential expression of cytokines and chemokines in a time-course study in response to ConA. Larger bovine liver and placenta arrays have been constructed at the Japanese National Institute of Agrobiological Sciences and they have been successfully applied in identifying differentially expressed genes during differentiation and development (Ushizawa et al., 2004). A porcine cDNA microarray comprising 5500 clones has been used to analyze differential transcript expression in phenotypically distinct muscle with the aim of identifying the genes involved in muscle phenotype determination (Bai et al., 2003). Lewin's group at the W.M. Keck Center for Comparative and Functional Genomics, University of Illinois, USA, has constructed a 3800 gene bovine microarray and applied this to profile transcripts expressed in spleen, placenta and brain (Band, Olmstead, Everts, Liu, & Lewin, 2002). Delaware Biotechnology Institute, University of Delaware, USA, has constructed a chicken array consisting of 13,007 genes selected from 363,838 ESTs representing 24 different adult or embryonic tissues (Burnside et al., 2005). CSIRO Livestock Industries developed a bovine cDNA microarray consisting of 9222 cattle probes from muscle and adipose tissues (Lehnert, Byrne, & Wang, 2004), and they used this array to contrast the gene expression profile in muscle tissue in response to various feeding regimes and between Japanese Black and Holstein cattle and for studying the mechanisms underlying in vitro adipogenesis in differentiating fibroblasts (Reverter et al., 2004). A collaboration in New Zealand has led to development of a 10,204 element bovine cDNA microarray, which has been used to study differential gene expression between lambs that are either genetically resistant or susceptible to larval nematodes (Diez-Tascon et al., 2005). At the Danish Institute of Agricultural Sciences a porcine cDNA array containing 26,878 cDNA clones, representing more than 22,000 genes, have been developed. The cDNAs were selected from clusters of an in-house collection of 810,124 porcine ESTs prepared from 98 cDNA libraries, representing a multitude of tissues and developmental stages. This repository was generated in a joint project between the Danish Institute of Agricultural Sciences, Danish Royal Veterinary and Agricultural University, and the Beijing Genomics Institute (Li, 2000).

It is clear from the previous discussion that the application of microarray technology to understand biological mechanisms in farm animals is at a very early stage. Few published data are currently available. An example of an ongoing project will be discussed in the following section.

5.1. Contrasting alleles of a major gene affecting meat quality

Naveau, Pommeret, and Lechaux (1985) showed segregation of a gene affecting a meat quality trait, the Napole technological yield (RTN). Further studies supported the hypothesis of a major locus with two alleles for this RTN trait, a recessive normal rn^+ allele and a dominant mutant allele RN^- (Le Roy, Naveau, Elsen, & Sellier, 1990). The RN^- phenotype is associated with elevated glycogen content in the sarcoplasmic as well as in the lysosomal compartments of glycolytic muscle cells (Estrade, Ayoub, Talmant, & Monin, 1994; Estrade, Vignon, Rock, & Monin, 1993). Meat from RN^- pigs have a low ultimate pH and a reduced water-holding capacity (WHC) resulting in a reduced yield of cured cooked ham (Enfalt, Lundstrom, Hansson, Johansen, & Nystrom, 1997; Estrade et al., 1993). The gene was mapped to chromosome 15 (Milan et al., 1995; Mariani et al., 1996), and Milan et al. (2000) reported that the RN^- phenotype is caused by an R225Q mutation in the γ3-subunit of AMP-activated protein kinase (AMPK). We have previously presented proteome studies that describe differentially expressed proteins and genes unleashed by the RN^- mutation (Hedegaard et al., 2004). Based on these observations we have hypothesized that the RN^- mutation induces a constitutive activation of the AMPK activity, leading to an increased glucose uptake in glycolytic muscle cells. A parallel microarray study is ongoing in our laboratory where we are using an experimental design aimed at describing expression differences between carriers of the RN^- mutation and wild-type pigs. RNA from *longissimus dorsi* muscle samples from 10 wild-type and 10 carriers of the R225Q mutation has been isolated and labelled. Each sample has been co-hybridized to a porcine cDNA array along with a reference sample consisting of equal quantities of all 20 RNA samples. Our preliminary data shows 30 differentially expressed genes, which are mostly related to glucose metabolism, post-translational modification of proteins, and transcriptional regulation (Horn et al., 2005).

6. Discussion

The RN^- experiment described above provides an example of an experimental design using half-sib progeny from a single heterozygous boar. Increasing the relatedness of the animals in the study is intended to decrease the overall biological variation unrelated to the

RN^- phenotype. However, the half-sib progeny also increases the risk of measuring transcriptional differences caused by other genes located in the same chromosomal region rather than the RN^- gene. This latter problem might not be restricted to studies where QTL-alleles or alleles of known genes are contrasted, but may also be observed in studies designed to contrast biological extremes in seemingly unrelated animals due to the large extent of linkage disequilibrium in farm animals (Farnir et al., 2000). In this case, expression studies led us towards a more detailed understanding of the biological changes associated with a major gene affecting meat quality. However, if our objective had been to identify the causative gene underlying the allelic differences between RN^- and wild-type animals, we would most likely have failed, since the genes identified as differentially expressed between the two genotypes were far from providing an unambiguous lead towards identifying AMPK as the causative gene.

7. Conclusions

With an increasing number of researchers embracing these new technologies, experience will be gathered to improve methods as well as experimental design providing new and exciting information leading to increased biological insight of farm animal genetics. In the mean time, researchers need to be diligent in their selection and interpretation of microarray data.

Acknowledgement

References

Agilent. (2005). Agilent Technologies, Inc., 395 Page Mill Rd., Palo Alto, California, CA 94306, USA. Available from <http://www.agilent.com>.

Ahmed, A. A., Vias, M., Iyer, N. G., Caldas, C., & Brenton, J. D. (2004). Microarray segmentation methods significantly influence data precision. *Nucleic Acids Research, 32*(5), e50.

Ambion. (2005). Ambion, Inc., 2130 Woodward, Austin, TX 78744-1832, USA. Available from <http://www.ambion.com>.

Anonymous (2002). A guide to microarray experiments – an open letter to the scientific journals. *Lancet, 360*(9338), 1019.

Bai, Q., McGillivray, C., da Costa, N., Dornan, S., Evans, G., Stear, M. J., et al. (2003). Development of a porcine skeletal muscle cDNA microarray: Analysis of differential transcript expression in phenotypically distinct muscles. *BMC Genomics, 4*(1), 8.

Band, M. R., Olmstead, C., Everts, R. E., Liu, Z. L., & Lewin, H. A. (2002). A 3800 gene microarray for cattle functional genomics: Comparison of gene expression in spleen, placenta, and brain. *Animal Biotechnology, 13*(1), 163–172.

Bilban, M., Buehler, L. K., Head, S., Desoye, G., & Quaranta, V. (2002). Normalizing DNA microarray data. *Current Issues in Molecular Biology, 4*(2), 57–64.

Bioconductor. (2004). Bioconductor – Open source software for bioinformatics. Available from <http://www.bioconductor.org>.

Bolstad, B. M., Collin, F., Simpson, K. M., Irizarry, R. A., & Speed, T. P. (2004). Experimental design and low-level analysis of microarray data. *International Review of Neurobiology, 60*, 25–58.

Bolstad, B. M., Irizarry, R. A., Astrand, M., & Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics, 19*(2), 185–193.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., et al. (2001). Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nature Genetics, 29*(4), 365–371.

Brown, M. P., Grundy, W. N., Lin, D., Cristianini, N., Sugnet, C. W., Furey, T. S., Ares, M. J., & Haussler, D. (2000). Knowledge-based analysis of microarray gene expression data by using support vector machines [In Process Citation]. *Proceedings of the National Academy of Science, USA, 97*(1), 262–267.

Burnside, J., Neiman, P., Tang, J., Basom, R., Talbot, R., Aronszajn, M., et al. (2005). Development of a cDNA array for chicken gene expression analysis. *BMC Genomics, 6*(1), 13.

Chen, Y., Kamat, V., Dougherty, E. R., Bittner, M. L., Meltzer, P. S., & Trent, J. M. (2002). Ratio statistics of gene expression levels and applications to microarray data analysis. *Bioinformatics, 18*(9), 1207–1215.

Chou, H. H., Hsia, A. P., Mooney, D. L., & Schnable, P. S. (2004). Picky: Oligo microarray design for large genomes. *Bioinformatics, 20*(17), 2893–2902.

Churchill, G. A. (2002). Fundamentals of experimental design for cDNA microarrays. *Nature Genetics, 32*(Suppl.), 490–495.

Colantuoni, C., Henry, G., Zeger, S., & Pevsner, J. (2002). Local mean normalization of microarray element signal intensities across an array surface: Quality control and correction of spatially systematic artifacts. *Biotechniques, 32*(6), 1316–1320.

Cui, X., Kerr, M. K., & Churchill, G. A. (2003). Transformations for cDNA microarray data. *Statistical Applications in Genetics and Molecular Biology, 2*(1), 1–20.

Dafforn, A., Chen, P., Deng, G., Herrler, M., Iglehart, D., Koritala, S., et al. (2004). Linear mRNA amplification from as little as 5 ng total RNA for global gene expression analysis. *Biotechniques, 37*(5), 854–857.

DeRisi, J. L., Iyer, V. R., & Brown, P. O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science, 278*(5338), 680–686.

DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., et al. (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer [see comments]. *Nature Genetics, 14*(4), 457–460.

Diez-Tascon, C., Keane, O. M., Wilson, T., Zadissa, A., Hyndman, D. L., Baird, D. B., et al. (2005). Microarray analysis of selection lines from outbred populations to identify genes involved with nematode parasite resistance in sheep. *Physiological Genomics, 21*(1), 59–69.

Dresen, I. M., Husing, J., Kruse, E., Boes, T., & Jockel, K. H. (2003). Software packages for quantitative microarray-based gene expression analysis. *Current Pharmaceutical Biotechnology, 4*(6), 417–437.

Duggan, D. J., Bittner, M., Chen, Y., Meltzer, P., & Trent, J. M. (1999). Expression profiling using cDNA microarrays. *Nature Genetics, 21*(Suppl. 1), 10–14.

EBI. (2005). ArrayExpress at the European Bioinformatics Institute (EBI). Available from <http://www.ebi.ac.uk/arrayexpress>.

Edwards, D. (2003). Non-linear normalization and background correction in one-channel cDNA microarray studies. *Bioinformatics, 19*(7), 825–833.

Enfalt, A. C., Lundstrom, K., Hansson, I., Johansen, S., & Nystrom, P. E. (1997). Comparison of noncarriers and heterozygous carriers of the RN‑ allele for carcass composition, muscle distribution and technological meat quality in Hampshire-sired pigs. *Livestock Production Science, 47*, 221–229.

Estrade, M., Ayoub, S., Talmant, A., & Monin, G. (1994). Enzyme activities of glycogen metabolism and mitochondrial characteristics in muscles of RN⁻ carrier pigs (Sus scrofa domesticus). *Comparative Biochemistry and Physiology. Biochemistry and Molecular Biology, 108*(3), 295–301.

Estrade, M., Vignon, X., Rock, E., & Monin, G. (1993). Glycogen hyperaccumulation in white muscle fibres of RN⁻ carrier pigs. A biochemical and ultrastructural study. *Comparative Biochemistry and Physiology. B: Comparative Biochemistry, 104*(2), 321–326.

Farnir, F., Coppieters, W., Arranz, J. J., Berzi, P., Cambisano, N., Grisart, B., et al. (2000). Extensive genome-wide linkage disequilibrium in cattle. *Genome Research, 10*(2), 220–227.

Genomic Solutions. (2004). Genomic Solutions, 4355 Varsity Dr., Ann Arbor, Michigan, MI 48108 USA. Available from <http://www.genomicsolutions.com>.

GEO. (2005). Gene Expression Omnibus (GEO) at the National Center for Biotechnology Information (NCBI). Available from <http://www.ncbi.nih.gov/geo>.

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., et al. (2004). Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biology, 5*(10), R80.

Glasbey, C. A., & Ghazal, P. (2003). Combinatorial image analysis of DNA microarray features. *Bioinformatics, 19*(2), 194.

Hedegaard, J., Horn, P., Lametsch, R., Sondergaard, M. H., Roepstorff, P., Bendixen, C., et al. (2004). UDP-Glucose pyrophosphorylase is upregulated in carriers of the porcine RN(−) mutation in the AMP-activated protein kinase. *Proteomics, 4*(8), 2448–2454.

Hessner, M. J., Singh, V. K., Wang, X., Khan, S., Tschannen, M. R., & Zahrt, T. C. (2004). Utilization of a labeled tracking oligonucleotide for visualization and quality control of spotted 70-mer arrays. *BMC Genomics, 5*(1), 12.

Horn, P., Sørensen, P., Hornshoj, H., Hedegaard, J., Bendixen, E., Panitz, F. & Bendixen, C. (2005). Differential gene expression profiles revealed by cDNA microarray analysis between wildtype and carriers of the porcine RN⁻ mutation in AMP-activated protein kinase, in preparation.

Hornshoj, H., Stengaard, H., Panitz, F., & Bendixen, C. (2004). SEPON, a Selection and Evaluation Pipeline for OligoNucleotides based on ESTs with a non-target Tm algorithm for reducing cross-hybridization in microarray gene expression experiments. *Bioinformatics, 20*(3), 428–429.

Invitrogen. (2005). Invitrogen Corporation, 1600 Faraday Avenue, PO Box 6482, Carlsbad, CA 92008, USA. Available from <http://www.invitrogen.com>.

Jain, A. N., Tokuyasu, T. A., Snijders, A. M., Segraves, R., Albertson, D. G., & Pinkel, D. (2002). Fully automatic quantification of microarray image data. *Genome Research, 12*(2), 325–332.

Kerr, M. K., Martin, M., & Churchill, G. A. (2000). Analysis of variance for gene expression microarray data. *Journal of Computational Biology, 7*(6), 819–837.

Konig, R., Baldessari, D., Pollet, N., Niehrs, C., & Eils, R. (2004). Reliability of gene expression ratios for cDNA microarrays in multiconditional experiments with a reference design. *Nucleic Acids Research, 32*(3), E29.

Le Roy, P., Naveau, J., Elsen, J. M., & Sellier, P. (1990). Evidence for a new major gene influencing meat quality in pigs. *Genetical Research, 55*(1), 33–40.

Lehnert, S. A., Byrne, K. A., & Wang, Y. H. (2004). Development and application of a bovine cDNA microarray for expression profiling of muscle and adipose tissue. *Australian Journal of Experimental Agriculture, 44*(11), 1127–1133.

Li, H. (2000). Gene sequencing. China, Denmark team up to tackle the pig. *Science, 290*(5493), 913–914.

Liu, D. K., Yao, B., Fayz, B., Womble, D. D., & Krawetz, S. A. (2004). Comparative evaluation of microarray analysis software. *Molecular Biotechnologys, 26*(3), 225–232.

Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., et al. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays [see comments]. *Nature Biotechnology, 14*(13), 1675–1680.

Luzzi, V., Mahadevappa, M., Raja, R., Warrington, J. A., & Watson, M. A. (2003). Accurate and reproducible gene expression profiles from laser capture microdissection, transcript amplification, and high density oligonucleotide microarray analysis. *The Journal of Molecular Diagnostics, 5*(1), 9–14.

Mariani, P., Lundstrom, K., Gustafsson, U., Enfalt, A. C., Juneja, R. K., & Andersson, L. (1996). A major locus (RN) affecting muscle glycogen content is located on pig chromosome 15. *Mammalian Genome, 7*(1), 52–54.

MDC. (2005). Axon Instruments, Molecular Devices Corporation, 1311 Orleans Drive, Sunnyvale, California, CA 94089-1136, USA. Available from <http://www.moleculardevices.com/transition>.

MGED. (2005). Microarray Gene Expression Data Society – MGED Society. Available from <http://www.mged.org>.

Milan, D., Jeon, J. T., Looft, C., Amarger, V., Robic, A., Thelander, M., et al. (2000). A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle. *Science, 288*(5469), 1248–1251.

Milan, D., Le Roy, P., Woloszyn, N., Caritez, J. C., Elsen, J. M., Sellier, P., et al. (1995). The RN locus for meat quality maps to pig chromosome 15. *Genetics, Selection, Evolution, 27*, 195–199.

Naveau, J., Pommeret, P., & Lechaux, P. (1985). Proposition d'une méthode de mesure du rendement technologique: 'la méthode NAPOLE'. *Techni-porc, 8*(6), 7–13.

Nobis, W., Ren, X., Suchyta, S. P., Suchyta, T. R., Zanella, A. J., & Coussens, P. M. (2003). Development of a porcine brain cDNA library, EST database, and microarray resource. *Physiological Genomics, 16*(1), 153–159.

Park, P. J., Cao, Y. A., Lee, S. Y., Kim, J. W., Chang, M. S., Hart, R., & Choi, S. (2004). Current issues for DNA microarrays: Platform comparison, double linear amplification, and universal RNA reference. *Journal of Biotechnology, 112*(3), 225–245.

Parmigiani, G., Garret, E. S., Izizarry, R. A., & Zeger, S. L. (2003). *Statistics for Biology and Health*. New York: Springer-Verlag New York, Inc..

PE. (2004). PerkinElmer, 45 William Street, Wellesley, Massachusetts, MA 02481-4078, USA. Available from <http://www.perkinelmer.com>.

Petrov, A., & Shams, S. (2004). Microarray image processing and quality control. *Journal of VLSI Signal Processing, 38*, 211–226.

Qu, Y., & Xu, S. (2004). Supervised cluster analysis for microarray data based on multivariate Gaussian mixture. *Bioinformatics, 20*(12), 1905–1913.

Quackenbush, J. (2001). Computational analysis of microarray data. *Nature Review Genetics, 2*(6), 418–427.

Quackenbush, J. (2002). Microarray data normalization and transformation. *Nature Genetics, 32*(Suppl.), 496–501.

Qiagen. (2005). QIAGEN Inc., 27220 Turnberry Lane, Valencia, CA 91355, USA. Available from <http://www1.qiagen.com>.

R. (2004). The R project for statistical computing. Available from <http://www.r-project.org>.

Raychaudhuri, S., Stuart, J. M., & Altman, R. B. (2000). Principal components analysis to summarize microarray experiments: Application to sporulation time series. *Pacific Symposium on Biocomputing*, 455–466.

Reverter, A., Wang, Y. H., Byrne, K. A., Tan, S. H., Harper, G. S., & Lehnert, S. A. (2004). Joint analysis of multiple cDNA microarray studies via multivariate mixed models applied to genetic improvement of beef cattle. *Journal of Animal Science, 82*(12), 3430–3439.

Saal, L. H., Troein, C., Vallon-Christersson, J., Gruvberger, S., Borg, A., & Peterson, C. (2002). BioArray Software Environment (BASE): A platform for comprehensive management and analysis of microarray data. *Genome Biology, 3*(8), SOFTWARE0003.

SAM. (2002). Significance analysis of microarrays (SAM). Available from <http://www-stat.stanford.edu/~tibs/SAM>.

Schuchhardt, J., Beule, D., Malik, A., Wolski, E., Eickhoff, H., Lehrach, H., et al. (2000). Normalization strategies for cDNA microarrays. *Nucleic Acids Research, 28*(10), E47.

Smyth, G. K., & Speed, T. (2003). Normalization of cDNA microarray data. *Methods, 31*(4), 265–273.

Speed, T. (2003). *Interdisciplinary statistics*. Florida: Chapman & Hall/ CRC CRC Press LLC.

Takemura, F., Inaba, N., Miyoshi, E., Furuya, T., Terasaki, H., Ando, S., et al. (2005). Optimization of liver biopsy RNA sampling and use of reference RNA for cDNA microarray analysis. *Analytical Biochemistry, 337*(2), 224–234.

Tao, W., Mallard, B., Karrow, N., & Bridle, B. (2004). Construction and application of a bovine immune-endocrine cDNA microarray. *Veterinary Immunology and Immunopathology, 101*(1–2), 1–17.

Tavazoie, S., Hughes, J. D., Campbell, M. J., Cho, R. J., & Church, G. M. (1999). Systematic determination of genetic network architecture. *Nature Genetics, 22*(3), 281–285.

Tecan. (2005). Tecan Schweiz AG, Seeatrasse 103, CH-8708 Männedorf, Switzerland. Available from <http://www.tecan.com>.

Toronen, P., Kolehmainen, M., Wong, G., & Castren, E. (1999). Analysis of gene expression data using self-organizing maps. *FEBS Letters, 451*(2), 142–146.

Tusher, V. G., Tibshirani, R., & Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Science, USA, 98*(9), 5116–5121.

Upson, J. J., Stoyanova, R., Cooper, H. S., Patriotis, C., Ross, E. A., Boman, B., et al. (2004). Optimized procedures for microarray analysis of histological specimens processed by laser capture microdissection. *Journal of Cellular Physiology, 201*(3), 366–373.

Ushizawa, K., Herath, C. B., Kaneyama, K., Shiojima, S., Hirasawa, A., Takahashi, T., et al. (2004). cDNA microarray analysis of bovine embryo gene expression profiles during the pre-implantation period. *Reproductive Biology and Endocrinology, 2*(1), 77.

van Hemert, S., Ebbelaar, B. H., Smits, M. A., & Rebel, J. M. (2003). Generation of EST and microarray resources for functional genomic studies on chicken intestinal health. *Animal Biotechnology, 14*(2), 133–143.

Ventana. (2005). Ventana – Molecular Discovery Systems, 1910 Innovation Park Dr., Tucson, Arizona, AZ 85737, USA. Available from <http://www.ventanadiscovery.com>.

Vinciotti, V., Khanin, R., D'Alimonte, D., Liu, X., Cattini, N., Hotchkiss, G., et al. (2005). An experimental evaluation of a loop versus a reference design for two-channel microarrays. *Bioinformatics, 21*(4), 492–501.

Wang, J., Hu, L., Hamilton, S. R., Coombes, K. R., & Zhang, W. (2003). RNA amplification strategies for cDNA microarray experiments. *Biotechniques, 34*(2), 394–400.

Wolfinger, R. D., Gibson, G., Wolfinger, E. D., Bennett, L., Hamadeh, H., Bushel, P., et al. (2001). Assessing gene significance from cDNA microarray expression data via mixed models. *Journal of Computational Biology, 8*(6), 625–637.

Yan, H., Yuan, W., Velculescu, V. E., Vogelstein, B., & Kinzler, K. W. (2002). Allelic variation in human gene expression. *Science, 297*(5584), 1143.

Yang, Y. H., Buckley, M., Dudoit, S., & Speed, T. (2002). Comparison of methods for image analysis on cDNA microarray data. *Journal of Computational and Graphical Statistics, 11*, 1–29.

Yang, Y. H., Buckley, M. J., & Speed, T. P. (2001). Analysis of cDNA microarray images. *Briefings in Bioinformatics, 2*(4), 341–349.

Yang, I. V., Chen, E., Hasseman, J. P., Liang, W., Frank, B. C., Wang, S., et al. (2002). Within the fold: Assessing differential expression measures and reproducibility in microarray assays. *Genome Biology, 3*(11).

Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J., et al. (2002). Normalization for cDNA microarray data: A robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Research, 30*(4), e15.

Yang, Y. H., & Speed, T. (2002). Design issues for cDNA microarray experiments. *Nature Review Genetics, 3*(8), 579–588.

Yao, J., Burton, J. L., Saama, P., Sipkovsky, S., & Coussens, P. M. (2001). Generation of EST and cDNA microarray resources for the study of bovine immunobiology. *Acta Veterinaria Scandinavica, 42*(3), 391–405.

Yao, J., Coussens, P. M., Saama, P., Suchyta, S., & Ernst, C. W. (2002). Generation of expressed sequence tags from a normalized porcine skeletal muscle cDNA library. *Anim Biotechnology, 13*(2), 211–222.

Yao, J., Ren, X., Ireland, J. J., Coussens, P. M., Smith, T. P., & Smith, G. W. (2004). Generation of a bovine oocyte cDNA library and microarray: Resources for identification of genes important for follicular development and early embryogenesis. *Physiological Genomics, 19*(1), 84–92.

Yu, J., Othman, M. I., Farjo, R., Zareparsi, S., MacNee, S. P., Yoshida, S., et al. (2002). Evaluation and optimization of procedures for target labeling and hybridization of cDNA microarrays. *Molecular Vision, 8*, 130–137.

Zien, A., Aigner, T., Zimmer, R., & Lengauer, T. (2001). Centralization: A new method for the normalization of gene expression data. *Bioinformatics, 17*(Suppl. 1), S323–S331.