

EXPRESSION QTL (eQTL) ANALYSIS FOR IDENTIFYING GENES CONTROLLING ECONOMICALLY IMPORTANT TRAITS IN PIGS

C.W. Ernst¹, J.P. Steibel¹, A.M. Ramos¹, R.J. Tempelman¹, F.F. Cardoso^{1,3}, G.J.M. Rosa²,
D.B. Edwards^{1,4} and R.O. Bates¹

¹Michigan State University, East Lansing, MI 48824

²University of Wisconsin, Madison, WI 53706

Introduction

Gene expression profiling using microarrays and genomic scans to identify chromosomal regions containing genes controlling important trait phenotypes have, until recently, been pursued as independent avenues of research. One direction has focused on variation in gene expression patterns, while the other has attempted to associate DNA sequence polymorphisms with phenotypic variation. However, within the past several years significant interest has turned towards integration of these two research areas (Jansen and Nap, 2001; Schadt et al., 2003b; Darvasi, 2003; Li and Burmeister, 2005; Drake et al., 2006; Haley and de Koning, 2006; Kadarmideen et al., 2006; de Koning et al., 2007) such that complex traits can now be studied in greater detail than is possible using either approach alone. Jansen and Nap (2001) proposed the term “genetical genomics” to describe the combination of expression profiling and marker-based fingerprinting of individuals in a segregating population. Such combination exploits all available statistical tools and augments the power of both approaches to unravel physiological processes and identify key genes. Determination of mRNA transcript abundances for individuals in a segregating population makes it possible to treat the expression of individual genes as quantitative traits, designated by Schadt et al. (2003a; 2003b) as expression QTL or eQTL. When the population has been screened for phenotypic traits of interest and trait QTL have been identified, the combined approach will help identify the gene(s) responsible for putative QTL. Thus, for each gene considered, QTL analysis will determine the regions of the genome influential for its expression and elucidate what portion of the variation in gene expression maps to the genes themselves (cis-acting factors), as opposed to other co-regulated genomic locations (trans-acting factors). Furthermore, unlike classic quantitative traits whose measures are far removed from the biological processes giving rise to them, the genetic linkages associated with transcript abundances afford a closer look at biochemical processes at the cellular level (Schadt et al., 2003b).

Applications of this approach are emerging rapidly and examples of recent reports in populations of budding yeast (Brem et al., 2002), *Drosophila melanogaster* (Wayne and McIntyre, 2002), *C. elegans* (Li et al., 2006), *Arabidopsis thaliana* (DeCook et al., 2006; West et al., 2007), barley (Potokina et al., 2007), maize (Consoli et al., 2002; Schadt et al., 2003a; Shi et al., 2007), rats (Lee et al., 2002; Okuda et al., 2002; Petretto et al., 2006), mice (Gu et al., 2002; Lemon et al., 2002; Schadt et al., 2003a, Tabakoff et al., 2003; Bystrykh et al., 2005; Lum et al., 2006), baboons (Cox et al., 2002) and humans (Schadt et al., 2003a; Monks et al., 2004; Morley et al., 2004) demonstrate the power of genetic-genomic integration strategies.

³Present address: Embrapa Pecuária Sul, Bagé, RS, Brazil

⁴Present address: University of Kentucky, Lexington, KY

Until recently, genetic approaches for dissecting complex traits have met with limited success. However, the integration of gene expression and genetics data will bridge the gap between knowledge of physiological mechanisms and genetic predisposition and accelerate discovery of candidate genes for use in genetic selection programs. The addition of gene expression data can be used to refine trait phenotypes, directly implicate pathways and genes comprising pathways associated with phenotypes, and identify key drivers of the pathways underlying phenotypes (Schadt et al., 2003b). This approach has tremendous potential for application in livestock populations. Several reviews of potential applications in livestock, including discussions of challenges to such studies, have been published (Haley and de Koning, 2006; Kadarmideen et al., 2006; de Koning et al., 2007). In addition, numerous reports have considered issues of design (Bueno Filho et al., 2006; Nettleton and Wang, 2006; Rosa et al., 2006; Wang and Nettleton, 2006; Cardoso et al., 2007) and analysis (Kendzioriski et al., 2006; Jia and Xu, 2007; Perez-Enciso et al., 2007 as examples), most of which can be applied broadly across species including to animal populations.

Application of the genetical genomics approach in livestock populations is illustrated in the diagram in Figure 1. Phenotypic records as well as DNA and RNA from target tissues are obtained from experimental resource populations. The DNA and phenotypic records are subjected to a classical QTL analysis, while the RNA is used for transcriptional profiling analyses. Individual mRNA abundances from the transcript profiles are then considered as quantitative phenotypes and they are incorporated into the genome scan. This integration will reveal cis- and trans-eQTL associated with the phenotypic traits of interest in the population, which will in turn facilitate not only identification of candidate genes at QTL, but also elucidation of biochemical pathways controlling phenotypic traits. Genes identified by this coordinated and synergistic approach could then be validated for incorporation into marker assisted selection strategies in commercial populations.

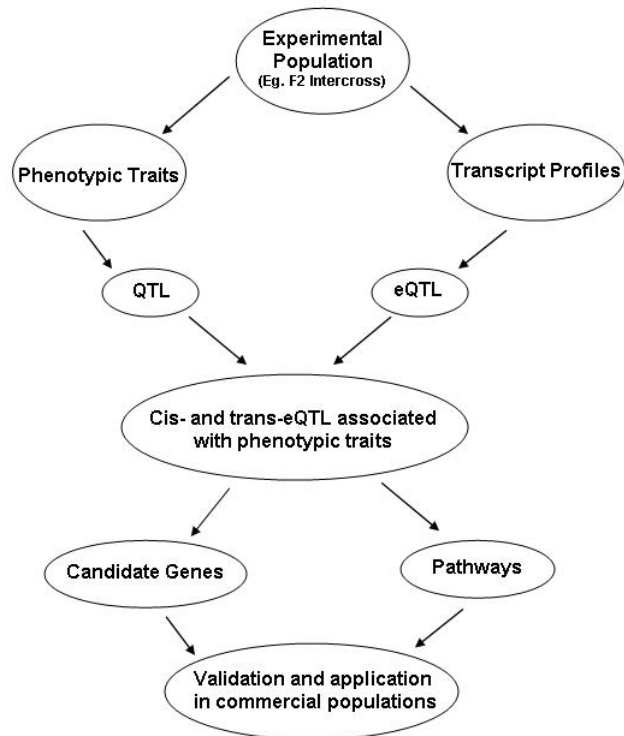


Figure 1. Genetical genomics in livestock populations.

A few studies are beginning to utilize this approach in livestock species. Liu et al. (2001) used a microarray containing 1,200 genes to detect gene expression differences between Marek's disease resistant and susceptible chickens. While this group was not able to collect enough gene expression information to model individual genes as

quantitative traits, they did genetically map 15 differentially expressed genes in order to integrate this information with previously identified Marek's disease QTL and the results revealed a strong positional candidate for one of the QTL. Further integration of gene expression and QTL studies in chickens have been proposed by Cogburn et al. (2003) and by de Koning et al. (2007). de Koning et al. (2007) proposed combining gene expression studies with fine mapping of functional trait QTL and they described a pilot study involving a body weight QTL on chicken chromosome 4. They utilized birds from the seventh generation of an advanced intercross and compared birds homozygous for markers flanking the QTL region inferred to be QQ or qq for the QTL (n=4 of each) using a cDNA microarray with over 12,000 features. After applying various normalization procedures they identified a consensus list of 45 genes at a false discovery rate (FDR) of 5%. Most of these genes were distributed across the genome, although 12 mapped to the QTL region and one was a valid functional candidate for future evaluation.

Objective

Our long term goals are to identify specific genes controlling economically important production efficiency and meat quality traits and to extend statistical tools to effectively incorporate this information into swine breeding programs. We are utilizing a three-generation pig resource population developed at Michigan State University (MSU) to not only identify classical QTL, but also eQTL in order to discover differentially expressed genes associated with economically important traits and determine the genomic regions with which they interact. Our specific objective for the study reported here was to conduct an expression QTL analysis of loin muscle tissue from F2 MSU Duroc x Pietrain resource population pigs exhibiting extremes for loin muscle area.

Methods

We have developed a Duroc x Pietrain resource population and conducted a primary genome scan utilizing 510 F2 offspring (of over 940 total F2s) for over 80 phenotypic traits (Edwards et al., 2008a,b). For transcriptional profiling, loin muscle tissue was collected at slaughter from 16 pigs per farrowing group (4 pigs from each of 4 litters; 11 farrowing groups). For 6 groups, 2 pigs per litter (1 barrow and 1 gilt) were selected that exhibited the largest loin muscle area (LMA) and two pigs per litter were selected that exhibited the smallest LMA adjusted to 105 kg as determined by B-mode ultrasound. For the remaining 5 groups, the 4 pigs from each litter were selected that exhibited extremes in tenth rib backfat thickness adjusted to 105 kg as determined by B-mode ultrasound.

For the present study, littermates exhibiting phenotypic extremes for LMA were paired within sex for transcriptional profiling using the Swine Protein-Annotated Oligonucleotide Microarray (<http://www.pigoligoarray.org>). This microarray was developed as a collaboration among several institutions and was supported in part by the U.S. Pig Genome Coordination Program. We have completed a pilot analysis with 60 microarrays. Data analysis methods were developed and tested through simulations using the MSU resource population QTL results (Edwards et al., 2008a,b). Median intensities for 20,736 expression traits were log-transformed and loess normalized (Yang et al., 2002), and all oligonucleotides were included in this preliminary analysis.

We used a modification of the first step of the Mixed Model Regression Mapping (MMRM) QTL mapping strategy proposed by Gilmour (2007). It consists of fitting all the markers in each linkage group as random effects with a linkage group specific variance, and testing for the significance of the corresponding variance components. The fundamental part of our random marker regression approach to select chromosomes is presented in the following model.

$$y = Dye + Sex + FG + \mathbf{Q}_l \mathbf{d}_l + Array + Litter + e, \quad [1]$$

where,

\mathbf{y} is the record (normalized intensity corresponding to a single animal),

$\mathbf{d}_k = [\mathbf{d}_{1k} \dots]$, are the random marker regression effects.

Dye , Sex and FG are the fixed effects of the fluorescence dye, sex and farrowing group, respectively.

$Litter$ is the random effect of litter, assumed n.i.i.d $N(0, \sigma_p^2)$,

$Array$ is the random effect of slide, assumed n.i.i.d $N(0, \sigma_A^2)$

\mathbf{Q}_l is a row vector of marker regression coefficients from linkage group l . The elements of \mathbf{Q}_l (q_{lm}) are obtained as the difference of the probabilities of breed of origin (A or B) at the position of each marker $m=1 \dots M_l$ in the linkage group: $q_{lm} = P(g_{lm} \equiv AA) - P(g_{lm} \equiv BB)$. The probabilities were obtained from an MCMC algorithm (Perez-Enciso et al., 2000) using the Qxpak software (Perez-Enciso and Misztal, 2004).

$\mathbf{d}_l \sim N(\mathbf{0}, \mathbf{I} \sigma_l^2)$, are the random marker regression effects.

$e \sim N(0, \mathbf{I} \sigma_e^2)$ vector of residual effects.

A set of L models (corresponding to L linkage groups) is fit, and the effect of the random markers are tested using a likelihood ratio test between model [1] and the null model (without the random marker effects). The procedure of Benjamini-Hochberg (Benjamini and Hochberg, 1995) was used to control FDR at the 5% level.

Results

We have so far performed the analysis for 12 chromosomes (SSC1-SSC8 and SSC15-SSC18). The histogram of p-values from this initial analysis reveals a substantial proportion of small p-values indicating differential expression explained by linkage groups (Figure 2). At FDR=0.05 ($\alpha=0.0066$), a total of 5,058 transcripts showed a significant effect for linkage group (at least one). Among those, 35% (1,769) of the transcripts showed association with only one linkage group and 47% (2,370) showed association with all 12 chromosomes tested (Figure 3). Figure 4 shows the distribution of transcripts with significant association to a single linkage group. SSC5 showed the

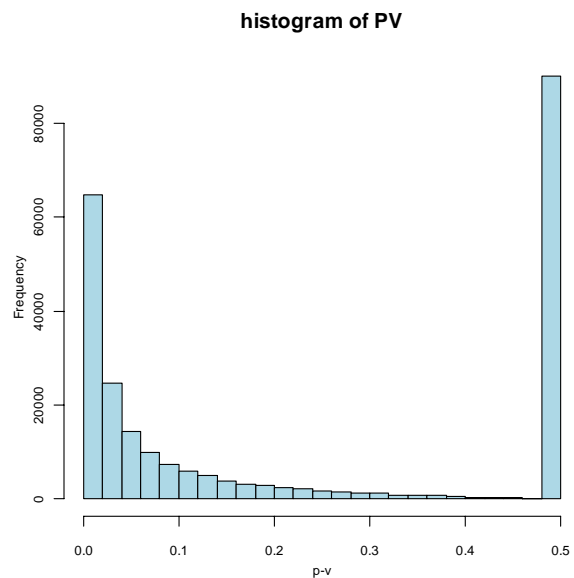


Figure 2. Histogram of p-values for transcript phenotypes.

largest number of significant transcripts.

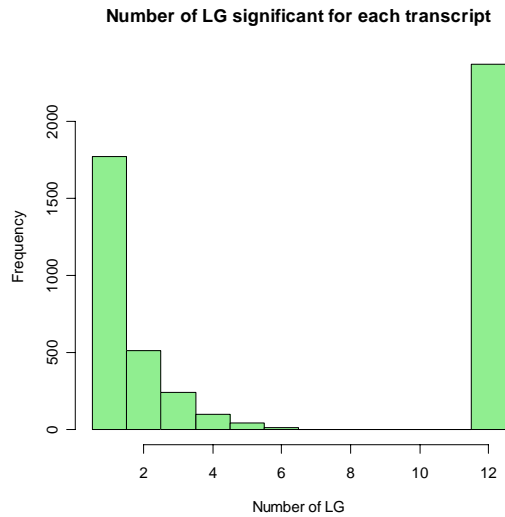


Figure 3. Number of significant linkage groups (chromosomes) for each transcript.

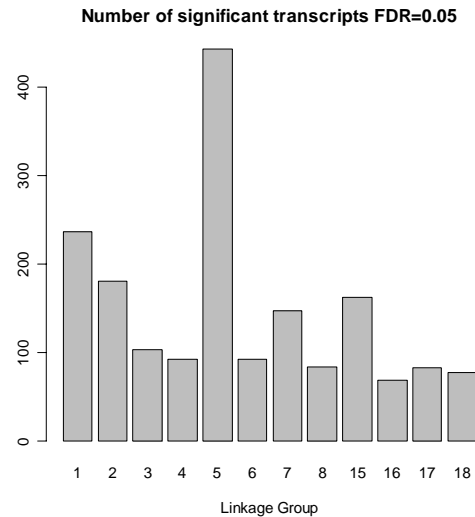


Figure 4. Distribution of transcripts with significant association to a single linkage group (chromosome).

Summary

Genetical genomics or eQTL analysis integrates genetic linkage analysis with gene expression evaluation in order to combine the power of recombination with functional analysis for identifying genes controlling economically important complex traits. Transcript abundance (expression) of genes within a tissue is affected by the millions of natural DNA variants that are segregating throughout a population. These transcript levels can be quantified using microarray technology and considered as expression phenotypes, which can subsequently be mapped using QTL analysis just as other phenotypes are mapped in order to identify genomic locations to which the expression traits are linked. Cis-acting variation may affect expression of a gene directly in the QTL region, while trans-acting variation may affect genes for transcription factors, signaling molecules, RNA processing or other processes.

Coordinating the map locations for QTL and genes with altered transcript abundances will lead to several outcomes as illustrated in Figure 5. Trait QTL mapping to the same position as a gene are considered cis-acting and the gene becomes a potential candidate gene for the QTL. Trait QTL affecting genes that map to other locations in the genome are considered trans-acting. The presence of “master regulators” may also emerge where a single QTL regulates the expression of dozens or hundreds of genes throughout the genome. In addition, situations may also exist where QTL affect expression of clusters of linked genes. Successful implementation of genetical genomics experiments are facilitated by availability of well annotated microarray resources and species genome sequences.

Initial eQTL analysis for our Duroc x Pietrain resource population is encouraging. Future efforts will include obtaining a complete set of loin muscle transcript profiles for the 176 available samples and completing the eQTL scan for the population. In addition,

we will be performing an eQTL analysis for backfat. Our strategy should reveal candidate genes for growth, carcass merit and pork quality phenotypes, as well as begin to elucidate gene pathways for these important traits.

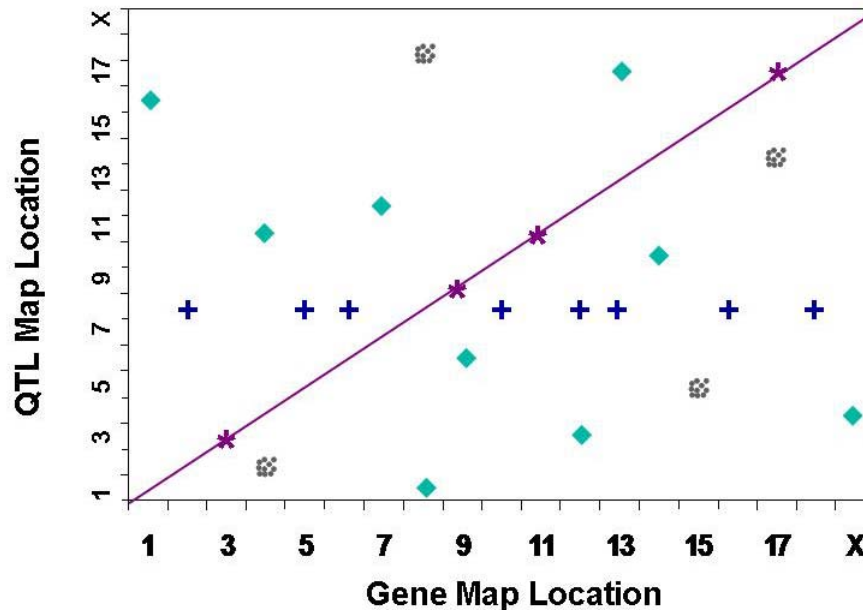


Figure 5. Expression QTL experimental outcomes. Genes identified by eQTL studies can be cis-acting (diagonal elements), trans-acting (off-diagonal elements designated by diamonds), multiple genes affected by the same QTL (off-diagonal elements designated by plus symbols) or clusters of linked genes co-regulated by a common QTL (off-diagonal clusters designated by gray dots).

Literature Cited

- Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. Royal Statistical Society Series B-Methodological* 57:289-300.
- Brem, R.B., G. Yvert, R. Clinton and L. Kruglyak. 2002. Genetic dissection of transcriptional regulation in budding yeast. *Science*. 296:752-755.
- Bueno Filho, J.S.S., S.G. Gilmour and G.J.M. Rosa. 2006. Design of microarray experiments for genetical genomics studies. *Genetics*. 174:945-957.
- Bystrykh, L., E. Weersing, B. Dontje, S. Sutton, M.T. Pletcher, T. Wiltshire, A.I. Su, E. Vellenga, J. Wang, K.F. Manly, L. Lu, E.J. Chesler, R. Alberts, R.C. Jansen, R.W. Williams, M.P. Cooke and G. de Haan. 2005. Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'. *Nature Genetics*. 37:225-210.
- Cardoso, F.F., J.P. Steibel, G.J.M. Rosa, C.W. Ernst, R.O. Bates and R.J. Tempelman. 2007. Assessment of different selective phenotyping strategies for genetical genomic studies with outbred F2 populations. *J. Anim. Sci.* 85(Suppl. 1):373.
- Cogburn, L.A., X. Wang, W. Carre, L. Rejto, T.E. Porter, S.E. Aggrey and J. Simon. 2003. Systems-wide chicken DNA microarrays, gene expression profiling, and discovery of functional genes. *Poult. Sci.* 82:939-951.

- Consoli, L., A. Lefevre, M. Zivy, D. de Vienne and C. Damerval. 2002. QTL analysis of proteome and transcriptome variations for dissecting the genetic architecture of complex traits in maize. *Plant Mol. Biol.* 48:575-581.
- Cox, L.A., S. Birnbaum and J.L. VandeBerg. 2002. Identification of candidate genes regulating HDL cholesterol using a chromosomal region expression array. *Genome Res.* 12:1693-1702.
- Darvasi, A. 2003. Genomics: Gene expression meets genetics. *Nature.* 422:269-270.
- DeCook, R., S. Lall, D. Nettleton and S.H. Howell. 2006. Genetic regulation of gene expression during shoot development in *Arabidopsis*. *Genetics.* 172:1155-1164.
- de Koning, D.J., C.P. Cabrera and C.S. Haley. 2007. Genetical genomics: Combining gene expression with marker genotypes in poultry. *Poultry Sci.* 86:1501-1509.
- Drake, T.A., E.E. Schadt and A.J. Lusis. 2006. Integrating genetic and gene expression data: application to cardiovascular and metabolic traits in mice. *Mamm. Genome* 17:466-479.
- Edwards, D.B. C. W. Ernst, R. J. Tempelman, G. J. M. Rosa, N. E. Raney, M. D. Hoge, and R. O. Bates. 2008. QTL mapping in an F2 Duroc x Pietrain resource population: I. Growth traits. *J Anim. Sci.* 86:241-253.
- Edwards, D.B. C. W. Ernst, N. E. Raney, M. E. Doumit, M. D. Hoge, and R. O. Bates. 2008. QTL mapping in an F2 Duroc x Pietrain resource population: II. Meat quality traits. *J Anim. Sci.* 86:254-266.
- Gilmour, A.R. 2007. Mixed model regression mapping for QTL detection in experimental crosses. *Computational Statistics and Data Analysis* 51:3749-3764.
- Gu, W., X. Li, K.H. Lau, B. Edderkaoui, L.R. Donahae, C.J. Rosen, W.G. Beamer, K.L. Shultz, A. Srivastava, S. Mohan and D.J. Baylink. 2002. Gene expression between a congenic strain that contains a quantitative trait locus of high bone density from CAST/EiJ and its wild-type strain C57BL/6J. *Funct. Integr. Genomics.* 1:375-386.
- Haley, C., and D.J. de Koning. 2006. Genetical genomics in livestock: potentials and pitfalls. *Anim. Genet.* 37(Suppl. 1):10-12.
- Jia, Z. and S. Xu. 2007. Mapping quantitative trait loci for expression abundance. *Genetics.* 176:611-623.
- Jansen, R.C. and J.P. Nap. 2001. Genetical genomics: the added value from segregation. *Trends Genet.* 17:388-391.
- Kadarmideen, H.N., P. von Rohr and L.L.G. Janss. 2006. From genetical genomics to systems genetics: potential applications in quantitative genomics and animal breeding. *Mamm. Genome.* 17:548-564.
- Kendzioriski, C.M., M. Chen, M. Yuan, H. Lan and A.D. Attie. 2006. Statistical methods for expression quantitative trait loci (eQTL) mapping. *Biometrics.* 62:19-27.
- Lee, S.J. and G.T. Cicila. 2002. Functional genomics in rat models of hypertension: using differential expression and congenic strains to identify and evaluate candidate genes. *Crit. Rev. Eukaryot. Gene Expr.* 12:297-316.
- Lemon, W.J., H. Bernert, H. Sun, Y. Wang and M. You. 2002. Identification of candidate lung cancer susceptibility genes in mouse using oligonucleotide arrays. *J. Med. Genet.* 39:644-655.
- Li, J. and M. Burmeister. 2005. Genetical genomics: combining genetics with gene expression analysis. *Human Molec. Genet.* 14:R163-R169.

- Li, Y., O.A. Alvarez, E.W. Gutteling, M. Tijsterman, J. Fu, J.A.G. Riksen, E. Hazendonk, P. Prins, R.H.A. Plasterk, R.C. Jansen, R. Breitling and J.E. Kammenga. 2006. Mapping determinants of gene expression plasticity by genetical genomics in *C. elegans*. *PLoS Genet.* 2:e222.
- Liu, H.C., H.H. Cheng, V. Tirunagaru, L. Sofer and J. Burnside. 2001. A strategy to identify positional candidate genes conferring Marek's disease resistance by integrating DNA microarrays and genetic mapping. *Anim. Genet.* 32:351-359.
- Lum, P.Y., Y. Chen, J. Zhu, J. Lamb, S. Melmed, S. Wang, T.A. Drake, A.J. Lusis and E.E. Schadt. 2006. Elucidating the murine brain transcriptional network in a segregating mouse population to identify core functional modules for obesity and diabetes. *J. Neurochem.* 97(Suppl. 1):50-62.
- Monks, S.A., A. Leonardson, H. Zhu, P. Cundiff, P. Pietrusiak, S. Edwards, J.W. Phillips, A. Sachs and E.E. Schadt. 2004. Genetic inheritance of gene expression in human cell lines. *Am. J. Hum. Genet.* 75:1094-105.
- Morley, M., C.M. Molony, T.M. Weber, J.L. Devlin, K.G. Ewens, R.S. Spielman and V.G. Cheung. 2004. Genetic analysis of genome-wide variation in human gene expression. *Nature.* 430:743-747.
- Nettleton, D. and D. Wang. 2006. Selective transcriptional profiling for trait-based eQTL mapping. *Anim. Genet.* 37(Suppl. 1):13-17.
- Okuda, T., T. Sumiya, K. Mizutani, N. Tago, T. Miyata, T. Tanabe, H. Kato, T. Katsuya, J. Higaki, T. Ogihara, Y. Tsujita, N. Iwai, K. Muzutani and H. Kato. 2002. Analyses of differential gene expression in genetic hypertensive rats by microarray. *Hypertens. Res.* 25:249-255.
- Perez-Enciso, M., and I. Misztal. 2004. Qxpak: A versatile mixed model application for genetical genomics and QTL analyses. *Bioinformatics* 20:2792-2798.
- Perez-Enciso, M., J.R. Quevedo and A. Bahamonde. 2007. Genetical genomics: use all data. *BMC Genomics.* 8:69.
- Perez-Enciso, M., L. Varona, and M. F. Rothschild. 2000. Computation of identity by descent probabilities conditional on DNA markers via a monte carlo markov chain method. *Genetics Selection Evolution* 32:467-482.
- Petretto, E., J. Mangion, N.J. Dickens, S.A. Cook, M.K. Kumaran, H. Lu, J. Fischer, H. Maatz, V. Kren, M. Pravenec, N. Hubner and T.J. Aitman. 2006. Heritability and tissue specificity of expression quantitative trait loci. *PLoS Genetics.* 2:e172.
- Potokina, E., A. Druka, Z. Luo, R. Wise, R. Waugh and M. Kearsey. 2007. Gene expression quantitative trait locus analysis of 16000 barley genes reveals a complex pattern of genome-wide transcriptional regulation. *Plant J.* doi:10.1111/j.1365-313x.2007.03315.x.
- Rosa, G.J.M., N. de Leon and A.J.M. Rosa. 2006. Review of microarray experimental design strategies for genetical genomics studies. *Physiol. Genomics.* 28:15-23.
- Schadt, E.E., S.A. Monks, T.A. Drake, A.J. Lusis, N. Che, V. Colinayo, T.G. Ruff, S.B. Milligan, J.R. Lamb, G. Cavet, P.S. Linsley, M. Mao, R.B. Stoughton and S.H. Friend. 2003a. Genetics of gene expression surveyed in maize, mouse and man. *Nature.* 422:297-302.
- Schadt, E.E., S.A. Monks and S.H. Friend. 2003b. A new paradigm for drug discovery: integrating clinical, genetic, genomic and molecular phenotype data to identify drug targets. *Biochem. Soc. Trans.* 31:437-443.

- Shi, C., A. Uzarowska, M. Ouzunova, M. Landbeck, G. Wenzel and T. Lubberstedt. 2007. Identification of candidate genes associated with cell wall digestibility and eQTL (expression quantitative trait loci) analysis in a Flint x Flint maize recombinant inbred line population. *BMC Genomics*. 8:22
- Tabakoff, B., S.V. Bhave and P.L. Hoffman. 2003. Selective breeding, quantitative trait locus analysis, and gene arrays identify candidate genes for complex drug-related behaviors. *J. Neurosci*. 23:4491-4498.
- Wang, D. and D. Nettleton. 2006. Identifying genes associated with a quantitative trait or quantitative trait locus via selective transcriptional profiling. *Biometrics*. 62:504-514.
- Wayne, M.L. and L.M. McIntyre. 2002. Combining mapping and arraying: An approach to candidate gene identification. *Proc. Natl. Acad. Sci. U.S.A.* 99:14903-14906.
- West, M.A.L., K. Kim, D.J. Kliebenstein, H. van Leeuwen, R.W. Michelmore, R.W. Doerge and D.A. St. Clair. 2007. Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in *Arabidopsis*. *Genetics*. 175:1441-1450.
- Yang, Y.H., S. Dudoit, P. Luu, D.M. Lin, V. Peng, J. Ngai and T.P. Speed. 2002. Normalization for cDNA microarray data: A robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Research* 30:e15.