

Combining Gene Expression with Marker Genotypes in Poultry

D. J. de Koning, Ian Dunn, C. P. Cabrera, J.M. Pethick, Yuna Blum, Suzanne Rowe and C. S. Haley

The Roslin Institute and R(D)SVS, University of Edinburgh, Roslin Biocentre, Roslin, EH25 9PS, United Kingdom
E-mail DJ.deKoning@roslin.ed.ac.uk

ABSTRACT

Microarrays have been widely implemented across the life sciences although there is still debate on the most effective uses of such transcriptomics approaches. In genetical genomics, gene expression measurements are treated as quantitative traits and genome regions affecting expression levels are denoted as expression quantitative trait loci or eQTL. The detected eQTL can either represent a locus that lies close to the gene that is being controlled (*cis*-acting) or one or more loci that are unlinked to the gene that is being controlled (*trans*-acting). One powerful outcome of genetical genomics is the reconstruction of genetic pathways underlying complex trait variation. Among livestock species, poultry is well placed to embrace this technology with the availability of the chicken genome sequence, microarrays for various platforms as well as experimental populations in which QTL have been mapped. In the build-up towards full-blown eQTL studies, we present two approaches in which gene expression studies have been combined with QTL data: a) a study where a gene expression study between broilers and layers is used to explore positional candidate genes for a pair of epistatic QTL, b) a study of the effects of a marked QTL at the gene expression level. In the latter, we have identified a positional candidate gene for a functional body weight QTL on chicken chromosome 4 in breast tissue samples from chickens with contrasting QTL genotypes.

INTRODUCTION

Dissecting the genetic control of variation in complex traits and identifying underlying loci controlling such variation has proved to be very challenging. While quantitative trait locus (QTL) detection has been successful in identifying chromosomal regions associated with a wide range of complex traits in many different species [e.g. experimental crosses (Doerge 2002), livestock (Andersson 2001; Andersson & Georges 2004), humans (Flint & Mott 2001)], these regions are sufficiently large to contain hundreds if not thousands of potential candidate genes. Likewise in chicken, many QTL have been mapped in experimental as well as commercial populations (Hocking 2005a) but the underlying mutation remains unknown for all but a few QTL. Further fine mapping of these QTL to reduce the size of these regions and hence refine the list of potential candidate genes can be achieved by creating additional recombination events through selective breeding (Darvasi 1998a) or by exploiting historical recombinations (Cardon & Bell 2001)

An approach that has great promise to make a major contribution to the dissection of complex traits is genetical genomics; the combined study of gene-expression and marker genotypes in a segregating population (Jansen & Nap 2001c; Jansen 2003). Genetical genomics is aimed at detecting genomic loci that control variation in gene expression, so called expression QTL (eQTL) (to distinguish them from functional QTL that affect traits at the whole-organism level). The detected eQTL can either represent a locus that lies close to the gene that is being controlled (*cis*-acting) or one or more loci that are unlinked to the gene that is being controlled (*trans*-acting) (Jansen & Nap 2001b). A major promise of genetical genomics is that by examining the relationship

between transcript location, location of eQTL and pleiotropic effects of eQTL, it might be able to reconstruct genetic pathways that underlie phenotypic variation (Jansen & Nap 2001a). Additional information to reconstruct pathways comes from the correlations between and among gene expression measurements and functional traits (Hitzemann *et al.* 2003) and the epistatic interactions between eQTL and functional QTL (Carlborg *et al.* 2004). Therefore, genetical genomics can be exploited as an additional tool to dissect phenotypic variation into its underlying components and elucidate how these components interact. If successful, genetical genomics will enhance and accelerate the characterization of functional QTL, which remains an arduous task, even in model species. At present, several studies have demonstrated the feasibility of eQTL studies and some of these have successfully integrated eQTL and gene expression with data on traditional phenotypes. What all these studies have in common is that, in comparison to ‘traditional’ QTL studies of functional traits, the sizes of the experiments are modest to small (de Koning & Haley 2005a). Consequently, the power of the studies is low and many important QTL will not have been detected and interactions between QTL will have been missed. Thus the results to date have not been very successful at reconstructing genetic pathways or identifying genes underlying functional trait variation. Therefore, more powerful experiments addressing these issues are necessary to realise the full potential of eQTL mapping.

A general conclusion from the published eQTL studies is that the most convincing evidence for eQTL is for the *cis*-acting effects (de Koning & Haley 2005b) while the reconstruction of genetic networks would require the identification of a larger proportion of *trans*-acting eQTL, including those with moderate effects. In short, current eQTL studies miss many important loci and fail to reconstruct genetic pathways underlying functional variation. At the same time, efforts to find the gene(s) underlying functional QTL via fine mapping and/or gene expression studies would be more effective if they were better integrated.

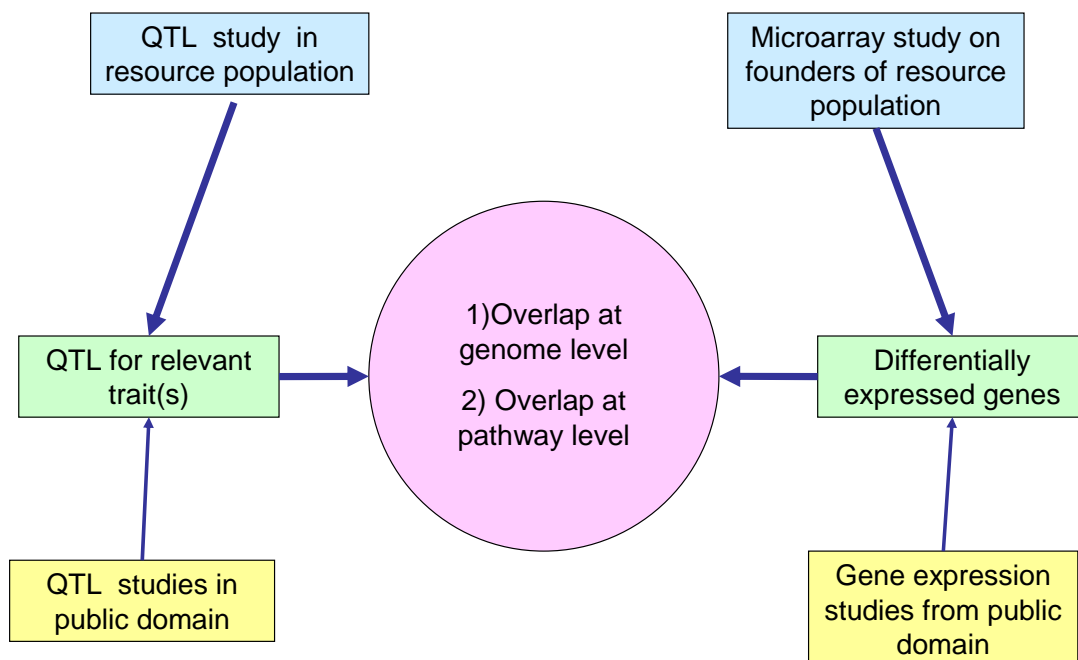


Figure1. Combining QTL studies with gene expression

Integrating QTL and Gene Expression studies

In a number of cases traditional QTL studies have been supplemented with microarray data in an attempt to move from a functional QTL to the underlying gene(s) (Wayne & McIntyre 2002). Below, we outline a case study where detection of functional QTL was followed up by a gene expression analysis. In this example, microarray experiments were carried out on the founder lines of the study. The underlying idea was that genes that were differentially expressed between the founder lines AND were located in the areas of the QTL that were found in the cross resulting from these lines, would be prime positional candidates for the functional genes underlying the QTL. The concept of such a combined approach is illustrated in Figure 1. In genetical genomics terms, this type of analysis explores whether the functional QTL is also a *cis*-acting eQTL. It would be much more difficult for such a study to determine the genetic basis of a QTL that had its functional effect through *trans*-acting regulation of expression of genes located outside the QTL region. This is because there are likely to be many differences in expression between lines for genes across the genome. This study provides no information on where in the genome the control of those expression differences lies and hence which of these genes are associated with the QTL region.

The first successful application of this approach in poultry dates back to 2001. Marek's disease (MD) is an infectious viral disease and a member of the herpes virus family. MD costs the poultry industry about 1 billion USD per annum. To study the genetic control of MD susceptibility, an experimental cross was established between a resistant and a susceptible inbred line of chicken (Vallejo *et al.* 1998b). F₂ offspring from this cross were experimentally challenged and genotyped, providing the data for a QTL analysis that resulted in seven QTL for susceptibility to MD (Vallejo *et al.* 1998a; Yonash *et al.* 1999). Subsequently, the founder lines of the F₂ cross were used for a microarray study to identify genes that were differentially expressed between the two lines following artificial infection. Fifteen of these genes were mapped onto the chicken genome and two of them mapped to a QTL region for Marek's resistance (Liu *et al.* 2001a). At the same time, protein interaction studies between a viral protein (SORF2) and a chicken splenic cDNA library revealed an interaction with the chicken growth hormone (GH) (Liu *et al.* 2001e). This led to the detection of a polymorphism in the GH gene that was associated with differences in the number of tumours between the susceptible and the resistant line (Liu *et al.* 2001d). The GH gene coincided with a QTL for resistance and also showed up as differentially expressed between the founder lines in the expression study (Liu *et al.* 2001c). More recently, the same group describe detection of lymphocyte antigen 6 complex (*LY6E*) as a putative Marek's disease resistance gene, again using the virus-host protein interaction screen (Liu *et al.* 2003). *LY6E* had been demonstrated earlier to be differentially expressed between resistant and susceptible chickens, but its location was not near a MD QTL (Liu *et al.* 2001b). Hence one could speculate that one of the MD QTL could act through *trans* acting control of the expression of this locus.

TOWARDS eQTL IN POULTRY

In many aspects, chickens are very well placed to be used in full blown genetical genomics studies. There is a large number of chicken QTL regions in the public domain (Hocking 2005b) and the species has the benefit of a full genome sequence (Hillier *et al.* 2004) and a SNP database (Wong *et al.* 2004). In terms of the gene expression tools, there are a number of tissue specific as well as general two-colour arrays (both spotted cDNA and long oligonucleotide array; <http://www.ark->

genomics.org/resources/chickens.php) as well as an Affymetrix™ chicken genome array (<http://www.affymetrix.com/products/arrays/specific/chicken.affx>). Large resource populations of chicken can be bred in a timely fashion or obtained from commercial lines. Populations for fine mapping, like advanced intercross lines (AIL) (Darvasi 1998b), are available in several labs (e.g. Wageningen University, Netherlands; Iowa State, USA and Roslin Institute, UK). An area for further development in the immediate future is the ongoing annotation of the chicken genome and other bioinformatics tools like pathway databases that incorporate chicken specific information. However, the most limiting factor in the uptake of genetical genomics for poultry and other livestock species is the budget required to run microarray studies on large numbers of animals. The recently proposed design of distant pairing (Fu & Jansen 2006) for genetical genomics looks promising in that it offers the possibility to array $2n$ individuals using n microarrays. This concept has recently been expanded to outbred F2 populations (Lam *et al.* 2008).

Combining QTL and microarray info for a pair of epistatic QTL.

In this particular experiment, the trait of interest is yellow yolkie follicle number in chickens, examining the genetics behind the phenotypic variation between broiler and layer chicken lines. Broilers have too many follicles, and therefore do not lay eggs as efficiently as layer chickens, which lay on average about 1 egg per day. In order to achieve an increase in the number of eggs produced by broiler chickens, their food intake must be heavily controlled, as it has been found that follicle numbers increase as a result of over-eating in these chickens, and this leads to a lower egg production. The pathway participating in this trait is thought to involve the pituitary gland located at the base of the brain, since it is central in reproduction. An epistatic pair of QTLs has been mapped for the number of follicles: one on chromosome 4 and another on chromosome 13.

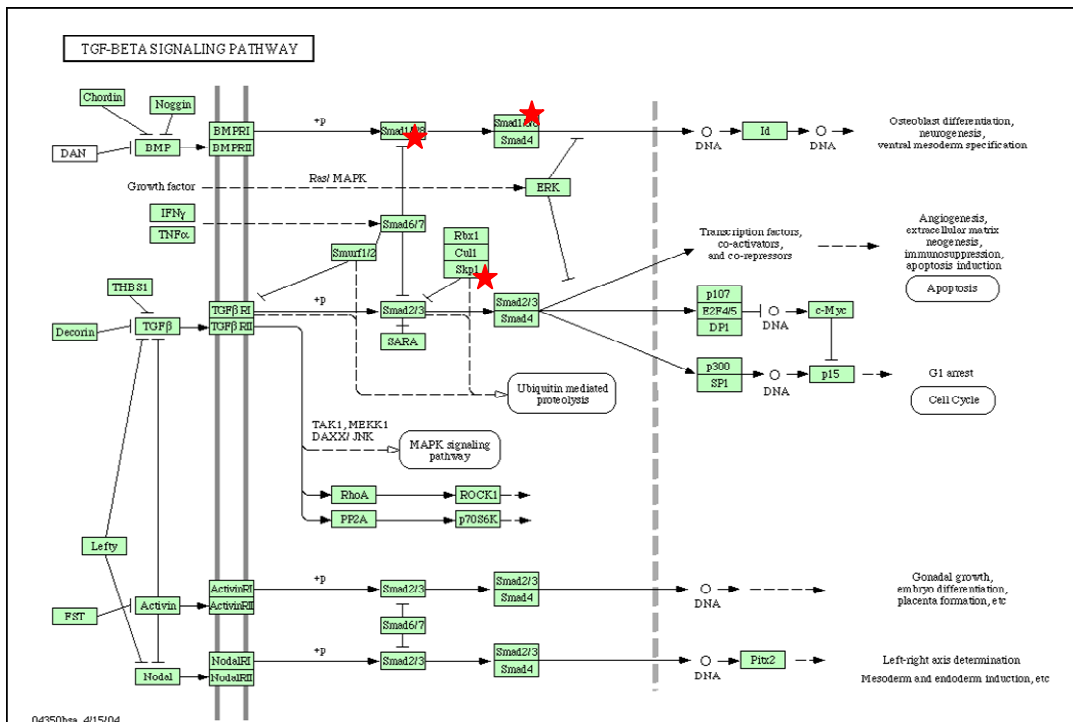


Figure 2. Enriched pathway with genes in GGA13 region indicated by red stars

The microarray study has identified 159 differentially expressed genes between the pituitary of broiler and layer chicken lines. The 20K oligonucleotide microarray is not

specific to a particular tissue, but contains oligonucleotide gene probes for most chicken genes. The pituitary was chosen on the basis of its central role in reproduction and growth for which the two lines are markedly different.

In order to compare whether any of the differentially expressed genes mapped to the QTL regions, they were all mapped against the chicken genome to obtain their genome location. Subsequently, the differentially expressed genes were mapped to functional pathways. It was then tested whether these pathways included genes in the QTL regions on GGA4 and GGA13. Especially enriched pathways with genes in both QTL regions were of particular interest, given the epistatic nature of the QTL. While some differentially expressed genes mapped to each QTL region, there was no pathway information available for these genes and they could not be studied further.

Furthermore, there were no enriched pathways that were common between the two regions. Figure 2 shows an example of an enriched pathway, with several genes in the QTL region on GGA13

Using Genetical genomics for a marked QTL

To illustrate potential of genetical genomics we describe a pilot study in chickens (Cabrera *et al.* 2006). The crucial part is the focused study of a particular putative QTL, in this case one affecting body weight segregating in an inter-cross of broilers and layers. Our objective was to identify candidate genes through the effect of the QTL at the gene expression level: what genes are affected, where do they map and in what kind of pathways are they involved?

We identified individuals that were homozygous for markers flanking a QTL region on chromosome 4 (GGA4) from the seventh generation of an advanced inter-cross between a single broiler and a single layer chicken. These were inferred to be either QQ (broiler allele) or qq (layer allele) for the QTL and matings were set up to provide birds with 'known' QTL genotypes. From the resulting offspring, QQ males and qq males were slaughtered at 21 days of age and a sample of the breast muscle was taken for RNA isolation and microarray studies. The microarrays design was a direct comparison of QQ versus qq for eight independent samples with a dye-swap (16 arrays used in total). The microarray was a chicken cDNA array with 12,877 functional features, spotted in duplicate (Ark-Genomics©, 2004). Using five alternative normalization procedures, we defined a consensus set of results consisting of 45 (895) differentially expressed genes when applying a false discovery rate (FDR) of 5% (20%) (Cabrera *et al.* 2006). This implies that out of 45 (895) results we expect less than 3 (180) false positive results. The genes that are differentially expressed seem evenly distributed over the genome and there appears to be no enrichment for affected genes in the QTL area on GGA4.

However, among the 12 differentially expressed genes (FDR < 20%) that map to the QTL region and should be considered positional candidate genes for the QTL. Among these, AADAT (FDR < 5%) is involved in lysine degradation, lysine biosynthesis and tryptophan metabolism, making it a promising candidate gene.

Current re-sequencing efforts are chasing SNPs that may be QTL specific and explain differences in alternative splicing. This demonstrates how a focussed study can aid the dissection of a QTL using limited resources.

CONCLUDING REMARKS

Although the existing eQTL studies demonstrate the utility of genetical genomics, they do not show its full potential because they miss many moderate effects and provide little opportunity to unravel genetic pathways due to a lack of *trans*-acting effects that would provide tangible links between eQTL and genes. We have demonstrated that with targeted studies, gene expression information can provided positional candidate genes

for QTL. However, this requires extensive bioinformatics effort and a healthy dose of good fortune.

A main shortfall when applying these principles to livestock populations like poultry is the lack of functional annotation. Impressive gene lists are reduced to small numbers of interesting genes with functional information, thus complicating the interpretation of the results. It is anticipated that next generation sequencing techniques might refine the application of gene expression studies to QTL regions and provide a more detailed characterization of the gene expression landscape.

Acknowledgements. The authors acknowledge financial support from the BBSRC through an ISPG to the Roslin Institute.

Literature Cited

- Andersson, L., 2001 Genetic dissection of phenotypic diversity in farm animals. *Nat.Rev.Genet.* **2**: 130-138.
- Andersson, L., and M. Georges, 2004 Domestic-animal genomics: deciphering the genetics of complex traits. *Nat.Rev.Genet.* **5**: 202-212.
- Cabrera C.P., Dunn I.C., Fell M., Wilson P.W., Burt D.W., Waddington D., Talbot R.T., Hocking P.M., Law A., Haley C.S., Knott S.A. & de Koning D.J. Application of genetical genomics to a marked QTL in poultry. Proceeding of 8th WCGALP in press. 2006. Brazil.
Ref Type: Conference Proceeding
- Cardon, L. R., and J. I. Bell, 2001 Association study designs for complex diseases. *Nat.Rev.Genet.* **2**: 91-99.
- Carlborg, O., P. M. Hocking, D. W. Burt, and C. S. Haley, 2004 Simultaneous mapping of epistatic QTL in chickens reveals clusters of QTL pairs with similar genetic effects on growth. *Genet.Res.* **83**: 197-209.
- Darvasi, A., 1998b Experimental strategies for the genetic dissection of complex traits in animal models. *Nat.Genet.* **18**: 19-24.
- Darvasi, A., 1998a Experimental strategies for the genetic dissection of complex traits in animal models. *Nat.Genet.* **18**: 19-24.
- de Koning, D. J., and C. S. Haley, 2005b Genetical genomics in humans and model organisms. *Trends Genet.* **21**: 377-381.
- de Koning, D. J., and C. S. Haley, 2005a Genetical genomics in humans and model organisms. *Trends Genet.* **21**: 377-381.
- Doerge, R. W., 2002 Mapping and analysis of quantitative trait loci in experimental populations. *Nat.Rev.Genet.* **3**: 43-52.
- Flint, J., and R. Mott, 2001 Finding the molecular basis of quantitative traits: successes and pitfalls. *Nat.Rev.Genet.* **2**: 437-445.
- Fu, J., and R. C. Jansen, 2006 Optimal design and analysis of genetic studies on gene expression. *Genetics* **172**: 1993-1999.
- Hillier, L. W., W. Miller, E. Birney, W. Warren, R. C. Hardison *et al.* 2004 Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**: 695-716.
- Hitzemann, R., B. Malmanger, C. Reed, M. Lawler, B. Hitzemann *et al.* 2003 A strategy for the integration of QTL, gene expression, and sequence analyses. *Mamm.Genome* **14**: 733-747.
- Hocking, P. M., 2005a Review of QTL results in chicken. *World's Poultry Science Journal* **61**: 215-226.
- Hocking, P. M., 2005b Review of QTL results in chicken. *World's Poultry Science Journal* **61**: 215-226.

- Jansen, R. C., 2003 Studying complex biological systems using multifactorial perturbation. *Nat.Rev.Genet.* **4**: 145-151.
- Jansen, R. C., and J. P. Nap, 2001a Genetical genomics: the added value from segregation. *Trends Genet.* **17**: 388-391.
- Jansen, R. C., and J. P. Nap, 2001b Genetical genomics: the added value from segregation. *Trends Genet.* **17**: 388-391.
- Jansen, R. C., and J. P. Nap, 2001c Genetical genomics: the added value from segregation. *Trends Genet.* **17**: 388-391.
- Lam, A. C., J. Y. Fu, R. C. Jansen, C. S. Haley, and D. J. de Koning, 2008 Optimal Design of Genetic Studies of Gene Expression With Two-Color Microarrays in Outbred Crosses. *Genetics* **180**: 1691-1698.
- Liu, H. C., H. H. Cheng, V. Tirunagaru, L. Sofer, and J. Burnside, 2001a A strategy to identify positional candidate genes conferring Marek's disease resistance by integrating DNA microarrays and genetic mapping. *Anim Genet.* **32**: 351-359.
- Liu, H. C., H. H. Cheng, V. Tirunagaru, L. Sofer, and J. Burnside, 2001c A strategy to identify positional candidate genes conferring Marek's disease resistance by integrating DNA microarrays and genetic mapping. *Anim Genet.* **32**: 351-359.
- Liu, H. C., H. H. Cheng, V. Tirunagaru, L. Sofer, and J. Burnside, 2001b A strategy to identify positional candidate genes conferring Marek's disease resistance by integrating DNA microarrays and genetic mapping. *Anim Genet.* **32**: 351-359.
- Liu, H. C., H. J. Kung, J. E. Fulton, R. W. Morgan, and H. H. Cheng, 2001e Growth hormone interacts with the Marek's disease virus SORF2 protein and is associated with disease resistance in chicken. *Proc.Natl.Acad.Sci.U.S.A* **98**: 9203-9208.
- Liu, H. C., H. J. Kung, J. E. Fulton, R. W. Morgan, and H. H. Cheng, 2001d Growth hormone interacts with the Marek's disease virus SORF2 protein and is associated with disease resistance in chicken. *Proc.Natl.Acad.Sci.U.S.A* **98**: 9203-9208.
- Liu, H. C., M. Niikura, J. E. Fulton, and H. H. Cheng, 2003 Identification of chicken lymphocyte antigen 6 complex, locus E (LY6E, alias SCA2) as a putative Marek's disease resistance gene via a virus-host protein interaction screen. *Cytogenet.Genome Res.* **102**: 304-308.
- Vallejo, R. L., L. D. Bacon, H. C. Liu, R. L. Witter, M. A. Groenen *et al.* 1998a Genetic mapping of quantitative trait loci affecting susceptibility to Marek's disease virus induced tumors in F2 intercross chickens. *Genetics* **148**: 349-360.
- Vallejo, R. L., L. D. Bacon, H. C. Liu, R. L. Witter, M. A. Groenen *et al.* 1998b Genetic mapping of quantitative trait loci affecting susceptibility to Marek's disease virus induced tumors in F2 intercross chickens. *Genetics* **148**: 349-360.
- 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.
- Wong, G. K., B. Liu, J. Wang, Y. Zhang, X. Yang *et al.* 2004 A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature* **432**: 717-722.
- Yonash, N., L. D. Bacon, R. L. Witter, and H. H. Cheng, 1999 High resolution mapping and identification of new quantitative trait loci (QTL) affecting susceptibility to Marek's disease. *Anim Genet.* **30**: 126-135.