Proceedings, 10th World Congress of Genetics Applied to Livestock Production

Whole Transcriptome Sequencing in Reciprocal Crosses Suggests Parent-of-Origin Effects on Gene Expression in the Chicken Genome

Fernando Lopes Pinto*, Anna-Maja Molin†,*, Elizabeth R. Gilbert‡, Christa Honaker‡, Paul Siegel‡, Göran Andersson*, Leif Andersson†,* and *Dirk-Jan de Koning**
*Swedish University of Agricultural Sciences, Uppsala, Sweden; †Uppsala University, Uppsala, Sweden; ‡Virginia Tech, Blacksburg, VA, USA

ABSTRACT: The Virginia Tech body weight lines have undergone more than 50 generations of divergent selection for 56-day body weight. In this experiment we test for the preferential expression of alleles from certain lines (cis-acting effects) or parents (parent-of-origin effects) in six F1 individuals from reciprocal crosses of generation 54 parents. Using RNA samples extracted from liver, hypothalamus and breast muscle (*Pectoralis major*), we generated circa 250 million RNA sequencing reads per F1 individual as well as a 25-fold coverage DNA sequence of each of the parents. We identified 11338 line-specific SNPs in the RNA across the three tissues. Allelic imbalance was biased for the SNP allele corresponding to the reference genome with ~65% of the SNPs showing a significant imbalance at $P < 0.05$. The number of SNPs with parent-of-origin effects without allelic imbalance was between 500 and 650 for each tissue.
Keywords: Chicken; Transcriptome; Parent-of-origin effect; RNA

Introduction

The Virginia Tech selection lines have undergone more than 50 generations of divergent selection for body weight. Crosses based on these lines have been a valuable resource for QTL mapping (Jacobsson et al. 2005; Carlborg et al. 2006; Pettersson et al. 2011). More recently, these lines have been studied for the effects of selection at the DNA level (Johansson et al. 2010). Nevertheless, the genetic variants underlying the QTL have not yet been identified, except for the identification of a deletion disrupting the *SH3RF2* gene (Rubin et al. 2010). Through recent advances in next generation sequencing technologies we can study the effect of sequence variation on RNA expression in a detailed manner. Via whole transcriptome sequencing we can identify tissue-specific transcripts, splice variants, as well as allele-specific expression levels. With proper experimental design we can exploit allele-specific expression or allelic imbalance to study whether a certain allele is preferentially expressed because of its line origin (e.g. a local cis-effect) or through a parent-of-origin effect. In mice, approximately 1300 loci with parent-of-origin effects in the brain were identified using RNA sequencing (Gregg et al. 2010). A recent study suggested that the number of imprinted loci by Gregg et al. (2010) were overestimated and that rigorous analyses are necessary to reduce false positives (DeVeale, van der Kooy, and Babak 2012).

Parent-of-origin effects have been suggested for a number of QTL in experimental chicken crosses (Tuiskula-Haavisto et al. 2004) as well as commercial lines (Rowe et al. 2009). For a review of chicken QTL with parent-of-origin effects please see Abasht, Dekkers, and Lamont (2006). The most obvious cause of parent-of-origin effects is genomic imprinting where one allele is partially silenced depending on the parental origin of the allele. According to the tug-of-war theory of genomic imprinting (Moore and Haig 1991), imprinting should not be present in birds. An in-depth review on parent-of-origin QTL effects in chickens and their overlap with imprinted regions in mammals was provided by Tuiskula-Haavisto and Vilkki (2007). They argued that while chickens may not show genomic imprinting in the same way as mammals, several characteristics of imprinted genes are also found in the chicken genome (Tuiskula-Haavisto and Vilkki 2007). For example, Dünzinger et al. (2005) describe how chicken orthologues of mammalian imprinted genes are clustered on the macrochromosomes. Furthermore, these clusters replicate asynchronously in chickens, which is a requirement for genomic imprinting, adding to the argument that mechanisms that facilitate imprinting in mammals were already in place in the evolution of birds (Dünzinger et al. 2005).

In our study we used a combination of DNA and RNA sequencing on reciprocal crosses of the Virginia Tech selection lines to study allelic imbalance and parent-of-origin effects in a rigorous and heuristic approach.

Materials and Methods

Experimental design. Four birds from generation 54 of the high and low body weight selection lines (a male and female from each line) were reciprocally crossed. At 56 days of age, three birds (males) from each reciprocal cross were euthanized, and samples from the hypothalamus, liver, and breast muscle (*Pectoralis major*, abbreviated P_major) were collected and stored in RNA*later* (Life technologies). Blood samples of the parents were taken for DNA isolations.

RNA and DNA sequencing. Whole RNA was extracted from 18 samples (six animals and three tissues) and strand-specific libraries for sequencing were prepared by SciLifeLab Uppsala. RNA samples were barcoded and pooled and subsequently run across 12 lanes of a HiSeq2000 (Illumina) using 100 bp, paired-end reads. The DNA of the founder animals was sequenced on 4 lanes to

provide an approximate sequencing depth of 25x for each parent. The raw sequencing results were made available by SciLife Lab Uppsala as Fasta files.

Sequence analysis. 1.1 Tb of raw sequence data with phred scores > 30 was processed primarily using the Tuxedo suite (Trapnell et al. 2012)**.** Adapters were trimmed with trimmomatic and duplicate reads were removed with fastq-mcf. Subsequently the sequence reads were aligned with the reference transcriptome using Tophat2 and transcriptomes were assembled using Cufflinks. Transcript abundance between reciprocal crosses was estimated using Cuffdiff. For the identification of putative SNPs in the RNA sequences, the 18 separate transcriptomes (3 tissues in 6 birds) were analyzed individually using varscan. All putative SNPs were then compared against the corresponding DNA sequence of the parents. The DNA sequences were also trimmed with trimmomatic and aligned to the reference genome with Bowtie.

Data analysis. From the putative SNPs in the RNA data, corresponding sequences of the parents were extracted. Only SNPs for which all four parents were fully informative (e.g. homozygous for line-specific alleles) were retained for further analysis. SNP locations with less than 100 reads within a cross (across both alleles and three birds) were removed from the analysis to increase the statistical rigor of the analysis. For allelic imbalance we tested whether one allele (corresponding to a given line) was preferentially expressed over the other allele. For the parent-of-origin effect we tested whether the maternally inherited alleles were preferentially expressed over the paternally inherited alleles or vice versa. Both allelic imbalance and parent-of-origin effects were tested with standard Chi-squared goodness-of-fit tests. Both quantitative differences in RNA expression between crosses and allelic imbalance within crosses can severely bias the test for parent-of origin effects. To mitigate these biases we used an adjusted chi-squared test where we changed the expected values according to identified ratios in RNA counts between crosses as well as alternative alleles. Finally, we identified the subset of SNPs that have a parent-of-origin effect ($P < 0.05$) but no significant allelic imbalance ($P > 0.01$).

Results and Discussion

Differential expression. Comparing the transcript abundance between the reciprocal crosses showed 147, 756 and 277 differentially expressed transcripts for hypothalamus, liver and P_major, respectively ($P < 0.001$). Between four and six transcripts are only expressed in one of the crosses for each tissue. Interestingly, ENSGALG00000023606 is exclusively expressed in the cross between the high line sire and the low line dam across all three tissues. This transcript, also known as LOC769173, is a MOB-like protein phocein that may play a role in membrane trafficking, specifically in membrane budding reactions. A gene set enrichment analysis for the differentially expressed genes in the hypothalamus shows enrichment for terms related to hormone activity and neuropeptides. For liver the relevant terms were related to biosynthesis as well as iron-, heme-, vitamin- and cofactor binding. For the muscle the terms included extra cellular matrix, ECM-receptor interaction and cholesterol storage.

SNP detection. From the variant analyses across individual tissues for individual birds more than 300K candidate SNPs were identified within the RNA. From these, 11338 SNPs had more than 100 reads within a cross and were fully informative in the founder lines: the high line parents were homozygous for one allele and the low line parents were homozygous for the other allele.

Table 1. Overview of SNPs with allelic imbalance or parent-of-origin (POE) effects for different tissues and different thresholds

	Hypothalamus	Liver	P_major
Candidate SNPs	7340	4829	5029
Imbalance P < 0.05	5564	3703	3793
Imbalance P < 0.01	5070	3360	3486
Imbalance P < 0.001	4508	3027	3159
POE P<0.05	1471	1047	1656
POE P<0.01	1264	875	1611
POE P<0.001	1076	731	1577
POE* P < 0.05	656 Pat 426 Mat 230	523 Pat 341 Mat 182	540 Pat 496 Mat 44

Allelic imbalance and parent-of-origin effects. The results of the analyses for allelic imbalance and parent-of-origin effects are summarized in Table 1. It is clear that the effects of allelic imbalance are very prominent and that most SNPs with a parent-of-origin effect are also affected by allelic imbalance (Table 1). Around 2/3 of all of the candidate SNPs show significant allelic imbalance. Between 10% and 20% of the candidate SNPs show parent-of-origin effects (Table 1). While allelic imbalance or allele-specific expression can point to interesting mechanisms, there are concerns about the inferences made from RNA sequencing (Stevenson, Coolon, and Wittkopp 2013). Because the RNA reads are aligned to a single reference transcriptome, the alignment will be more successful for the reads that contain the reference allele compared to those that contain the alternative allele, thus resulting in an artificial bias toward the reference allele. In preliminary studies, observed that around 75% of the SNPs with allelic imbalance showed preferential expression of the reference allele, indicating that these results need to be explored further. A potential solution would be to create separate 'parental' transcriptomes and align against these multiple references (Stevenson, Coolon, and Wittkopp 2013). Because of this bias and the sensitivity of the parent-of-origin test for allelic imbalance, we focused on the SNPs with a significant parent-of-origin effect but without allelic imbalance (Table 1). This limited the number of SNPs with parent-of-origin effects to about 10% of the candidate SNPs (POE* in Table 1). The SNPs with parent-of-origin effects showed a clear bias towards preferential paternal expression (Table 1).

Furthermore, this bias is strongest for the P_major and least extreme for the hypothalamus. The large bias of paternally expressed SNPs in the muscle could suggest that even in birds paternal alleles may be preferentially expressed to enhance growth. The SNPs with parent-of-origin effects show limited overlap between tissues: Only five SNPs showed parent-of-origin effects across all tissues while a further 146 SNPs showed parent-of-origin effects in two tissues. The locations of SNPs with parent-of-origin effects along chromosomes 1 and 3 are plotted in Figures 1 and 2. Both chromosomes 1 and 3 have been discussed extensively in the context of parent-of-origin QTL and imprinting mechanisms (Dünzinger et al. 2005; Tuiskula-Haavisto and Vilkki 2007). Results presented in Figures 1 and 2 show that the effects are not randomly scattered across the chromosomes, but that some regions are clearly enriched for SNPs with parent-of-origin effects. Furthermore, there is somewhat more agreement across the three tissues than suggested by the limited overlap in actual significant SNPs between the three tissues.

Next Steps. From a large amount of RNA data we have identified for each of three different tissues a manageable number of candidate SNPs for further study. A putative parent-of-origin effect should be experimentally verified, for example using pyro-sequencing. It is unlikely that such a validation could be carried out for all of candidate SNPs in all three tissues. As part of the next step, a number of bioinformatics filters can be applied to identify the most promising candidate SNPs for validation. A straightforward filter would be to select all candidate SNPs that are located in regions that are orthologous to imprinted regions in mammalian genomes. Alternatively, one can select SNPs that are located in regions where QTL with parent-of-origin effects have been identified (Abasht, Dekkers, and Lamont 2006). These QTL are summarized in the animal QTL database (http://www.animalgenome.org/cgi-bin/QTLdb/GG/index) where they can be directly linked to the genome coordinates. One drawback is that the QTL regions in this database tend to span tens of centiMorgans and thus containing large numbers of candidate SNPs. The transcripts that contain the candidate SNPs should be further annotated, followed by gene set enrichment analyses to identify pathways that may be enriched for parent-of-origin effects. The overwhelming number of SNPs with allelic imbalance prevents further analyses to uncover potential cis-acting variants that might explain QTL that have been discovered in crosses between these selection lines. The most promising strategy is to re-align the RNA data against multiple reference assemblies (Stevenson, Coolon, and Wittkopp 2013). A further scrutiny of the SNPs with putative parent-of-origin effects will follow the recommendations of (DeVeale, van der Kooy, and Babak 2012) for a more stringent analysis of the RNA data.

POE*, SNPs with a parent-of-origin effect but no allelic imbalance; Pat, preferential paternal expression; Mat, preferential maternal expression.

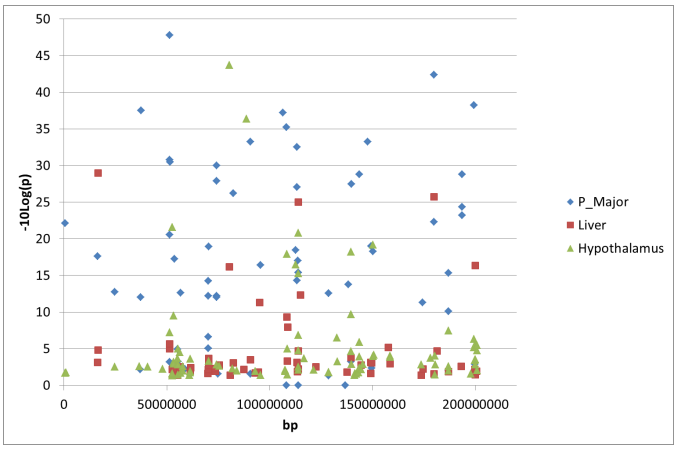

Figure 1. Parent-of-origin effects along chicken chromosome 1, plotted against the -10LOG(p) of the test for parent-of-origin effects.

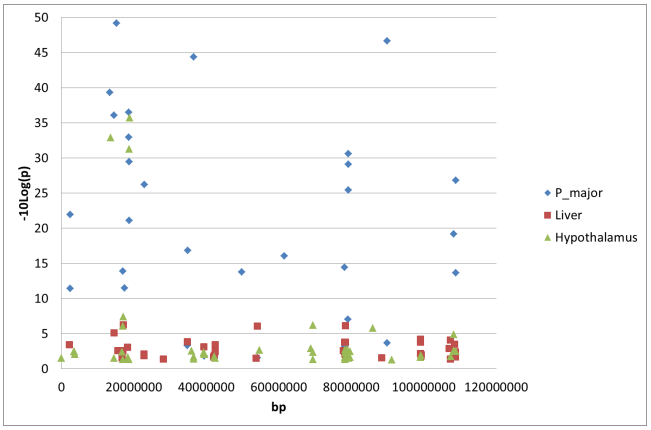

Figure 2. Parent-of-origin effects along chicken chromosome 3, plotted against the -10LOG(p) of the test for parent-of-origin effects.

Conclusion

Using reciprocal crosses between the Virginia Tech body weight lines, we have uncovered a considerable number of SNPs with possible parent-of-origin effects on mRNA expression effects. Further study of these candidate SNPs will show whether birds show processes that are similar to genomic imprinting in mammals. If this is the case, the theories about evolution of imprinting may require some revision.

Literature Cited

Abasht, B., J. C. M. Dekkers, and S. Lamont. 2006. *Poultry Science* 85: 2079–96.

Carlborg, Ö, L. Jacobsson, P. AAhgren, et al. 2006. *Nature Genetics* 38 (4): 418–20.

DeVeale, Brian, Derek van der Kooy, and Tomas Babak. 2012. *PLoS Genetics* 8 (3): e1002600.

Dünzinger, Ulrich, Indrajit Nanda, Michael Schmid et al. 2005. *Trends in Genetics* 21 (9): 488–92.

Gregg, C., J. Zhang, B. Weissbourd, S. et al. 2010. *Science* 329 (5992): 643.

Jacobsson, Lina, Hee-Bok Park, Per Wahlberg et al. 2005. *Genet. Res.* 86 (02): 115.

Johansson, Anna M., Mats E. Pettersson, Paul B. Siegel et al. 2010. *PLoS Genetics* 6 (11): e1001188.
Pettersson, Mats, Francois Besnier, Paul B. Siegel et al. *PLoS Genetics* 7 (7): e1002180.
Rowe, S. J, R. Pong-Wong, C. S Haley et al 2009. *Genet. Sel. Evol.* 41 (1): 1–11.
Rubin, Carl-Johan, Michael C. Zody, Jonas Eriksson et al. 2010. *Nature* 464 (7288): 587–91.
Stevenson, Kraig R, Joseph D Coolon, and Patricia J Wittkopp. 2013. *BMC Genomics* 14 (1): 536.
Trapnell, Cole, Adam Roberts, Loyal Goff et al. 2012. *Nature Protocols* 7 (3): 562–78.
Tuiskula-Haavisto, M., D. J De Koning, M. Honkatukia et al. 2004. *Genet. Res.* 84 (01): 57–66.
Tuiskula-Haavisto, M., and J. Vilkki. 2007. *Cytogenetic and Genome Research* 117 (1-4): 305–12.