Comprehensive Identification Of Specific Genes Controlling Complex Traits Through A Genome-Wide Screen For Cis-Acting Regulatory Elements - An Example Using Marek's Disease

S. MacEachern[1,2], W.M. Muir[3], S. Crosby[4], and H.H. Cheng[1]

Introduction

Marek's disease (MD), a T cell lymphoma induced by the highly oncogenic α-herpesvirus Marek's disease virus (MDV), is the main chronic infectious disease concern threatening the poultry industry. Enhancing genetic resistance to MD in commercial poultry is an attractive method to augment MD vaccines, which is currently the control method of choice. In order to optimally implement this control strategy through marker-assisted selection (MAS) and to gain biological information, it is necessary to identify specific genes that influence MD incidence.

Utilizing and integrating genomic QTL scans (Vallejo *et al*., 1998; Yonash *et al*., 1999; McElroy *et al.,* 2005; Heifetz *et al.,* 2009), functional genomic RNA profiling using microarrays (Liu *et al*., 2001a), and chicken-MDV protein-protein interactions using two-hybrid screens (Niikura *et al*., 2004), we have identified three MD resistance genes (growth hormone (Liu *et al.,* 2001b), SCA2 (Liu *et al*., 2003), and MHC class II β chain) and a number of other potential candidates. Despite this success, these genes can only account for a fraction of the total observed genetic variation.

To identify most, if not all, the remaining genes, we have implemented a genome-wide screen for allele specific expression (ASE) in response to MDV infection. ASE is a simple yet powerful approach, where the expression of each gene allele is compared **within** an RNA sample. When the expression of the alleles is not equal (i.e., allelic imbalance), then one can **unequivocally** declare ASE and the presence of a polymorphic cis-acting (genetic) element for that gene as LD is confined to the transcriptional unit. In brief, we developed a procedure that: (1) uses RNA seq on a limited number of samples to identify cSNPs and, when found, statistically analyzes for evidence of ASE, followed by (2) a comprehensive and more economical screen of all samples for ASE using Illumina GoldenGate assays to measure relative allele expression of each targeted cSNP. This method can be applied to all traits in all species though is simpler, more powerful to perform, and can detect trans-acting effects when inbred lines are available.

[1] USDA, ARS, ADOL, East Lansing, MI 48823, USA
[2] Currently at Cobb-Vantress Inc., Siloam Springs, AR 72761, USA
[3] Department of Animal Sciences, Purdue University, West Lafayette, IN 47907, USA
[4] Genome Sequencing Center, Washington U. in St. Louis, St. Louis, MO 63108, USA

Material and methods

Generation of samples. The highly inbred ADOL chicken lines 6 (MD resistant) and 7 (MD susceptible) were intermated in reciprocal crosses. Half of the progeny in each cross were challenged with MDV (2,000 pfu JM strain) at 2 weeks of age. At 1, 4, 7, 11, 13, and 15 days post-infection (dpi), 12 birds per treatment group were terminated; pure line birds were similarly treated with 14 birds per treatment group. RNA from spleen, thymus, and bursa was isolated and quantified.

RNA sequencing and analysis. Two pools containing equal amounts of RNA from 6 birds each per treatment group at 4 dpi were generated, then sequenced using an Illumina GA. Using the Maq algorithm (Li *et al*., 2008) to align the reads and a custom Python script, SNPs were detected, then filtered for sequence quality and alignment depth. Allelic imbalance was determined using a chi-square test ($p<0.05$).

Validation using Illumina GoldenGate assays and analysis. Based primarily on Illumina design scores, assays for 1536 selected cSNPs were developed and used on RNA samples from spleens of all F1 birds. In addition, artificial F1 pools generated by mixing equal amounts of RNA from each pure line were screened to query for the influence of trans-acting factors; 12 line 6x7 F1 and 12 line 7x6 F1 DNAs were also measured to normalize the assays. The dataset was analyzed using linear models.

Relative influence of cis- and trans-acting factors. True vs. artificial F1 expression ratios were plotted for each time point and condition for splenic RNAs. The total percentage of each SNP effect was estimated and the averages plotted over the course of infection.

Pathways analysis. Biological processes were screened by inputting genes exhibiting ASE in response to MDV infection into the Metacore (GeneGo) database.

Results and discussion

RNAs from parental lines or F1 progeny were obtained to evaluate potential influence of (1) maternal or epigenetic effects via reciprocal matings, (2) MDV infection, (3) time after infection, (4) different tissues, and (5) trans-acting factors via artificial F1 pools. As a result, for the splenic RNAs only, we had 288 F1 samples and 168 artificial F1 RNA pools.

For ASE assays (see Figure 1), it is necessary to have a cSNP for each to track each allele; the polymorphism does not have to be in an exon or nonsynonymous. Thus, to get a genome-wide and unbiased survey of all the expressed genes and an indication of ASE, replicate RNA pools from a single time point (4 dpi) were sequenced using a next generation sequencer (Illumina GA). This resulted in 11-13+ million mappable reads per treatment group with a total of 1.7+ Gb surveyed. The number of raw and filtered cSNPs is shown in Table 1. Statistical analysis of these cSNPs revealed that 5360 cSNPs in 3773 genes exhibited statistically significant allelic imbalance.

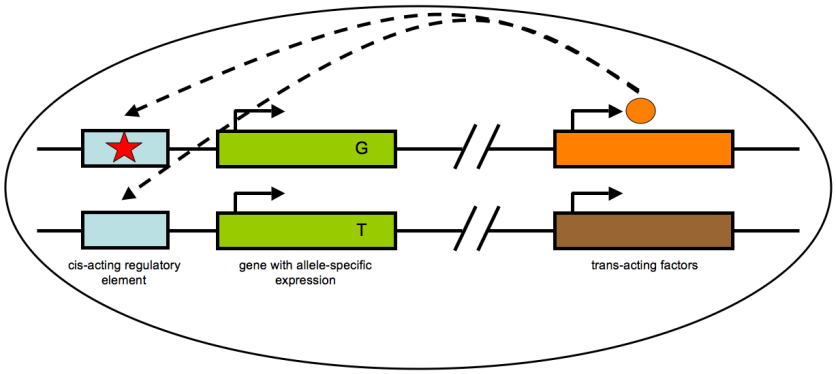

Figure 1: Detecting ASE due to cis-acting regulatory effects. Because the mutation in the cis element is specific to one chromosome (labeled with G SNP), it alone will be differentially expressed. Trans-acting factors do not influence ASE as they are present and can act equally on both alleles.

Table 1: Number of raw and filtered SNPs in RNA seq dataset

Cross[a]	6x7 U	6x7 I	7x6 U	7x6 I	Total
raw	375,732	315,361	378,586	361,386	670,031
filtered	11,574	12,472	11,939	10,849	22,655

[a]Lines 6 and 7 are MD resistant and susceptible, respectively. U and I stand for uninfected and infected, respectively

To economically validate and extend the results, 1536 SNPs exhibiting ASE were screened on all the RNA samples from spleens of all F1 birds using GoldenGate arrays. In addition, artificial F1 pools generated by mixing equal amounts of RNA from each pure line were screened to query for the influence of trans-acting factors; 12 RxS F1 and 12 SxR F1 DNAs were also measured to normalize the assays as both alleles are known to be equal. The initial statistical analysis of 1465 functioning assays for ASE effects in spleen is shown in Table 2.

Table 2: Factors influencing ASE

Direction of cross (CROSS)	481
Infection (INF)	1040
Days post infection (DPI)	1417
INF x DPI	1407
INF x CROSS	14
DPI x CROSS	1284
INF x DPI x CROSS	652
No effects	19

Besides finding genes with cis-regulatory element, the relative influence of trans-acting factors for each gene can be quantified by plotting the relative expression of each allele in true vs. artificial F1 RNAs as described by Wittkopp *et al*. (2004). Furthermore, the data can be further analyzed like microarrays to identify pathways that are enriched for ASE. Using

the Metacore (GeneGo) database, T cell activation, vesicle mediated transport, and cell cycling were identified

Thus, using our protocol of RNA seq followed by confirmation using Illumina GoldenGate arrays, we have identified a large number of genes exhibiting ASE in response to MDV infection. Furthermore, the direction of the cross also influences ASE, which suggests maternal or epigenetic influences. Undoubtedly, many of these genes contribute to genetic resistance to MD, which can be directly monitored in resource populations using the same cSNPs. Additional efforts are underway to integrate results from genomic DNA sequences of the two parental lines to identify possible causative mutations, and genomic sequences obtained from ChIP seq using MDV Meq, a bZIP transcription factor and the putative viral oncogene.

Conclusion

The described method, although demonstrated in inbred chicken lines, is applicable to all traits in any diploid species, and should prove to be a facile method to identify the majority of genes controlling any complex trait.

References

Heifetz, E.M., Fulton, J.E., O'Sullivan, N.P. *et al*. (2009) *BMC Genomics*, 10:20.

Li, H., Ruan, J., Durbin, R. (2008) *Genome Res*., 18:1851-1858.

Liu, H.-C., Cheng, H.H., Softer, L. *et al*. (2001a) *Anim. Genet*., 32:351-359.

Liu, H.-C., Kung, H.-J., Fulton, J.E. *et al*. (2001b) *PNAS*, 98:9203-9208.

Liu, H.C., Niikura, M., Fulton, J. *et al*. (2003) *Cytogen. Genome Res*., 102:304-308.

McElroy, J.P., Dekkers, J.C.M., Fulton, J.E. *et al*. (2005) *Poultry Sci*., 84:1678-1688.

Niikura, M., Liu, H.-C., Dodgson, J.B. *et al*. (2004) *Poultry Sci*., 83:1117-1123.

Vallejo, R.L., Bacon, L.D., Liu, H.-C. *et al*. (1998) *Genetics*, 148:349-360.

Wittkopp, P.J., Haerum, B.L., and Clark, A.G. (2004) *Nature*, 430:85-88.

Yonash, N., Bacon, L.D., Witter, R.L. *et al*. (1999) *Anim. Genet*., 30:126-135.