

SHORT COMMUNICATION

Open Access



# Genomic prediction of avian influenza infection outcome in layer chickens

Anna Wolc<sup>1,2\*</sup>, Wioleta Drobik-Czwarno<sup>3</sup>, Janet E. Fulton<sup>2</sup>, Jesus Arango<sup>2</sup>, Tomasz Jankowski<sup>2,4</sup> and Jack C. M. Dekkers<sup>1</sup>

## Abstract

Avian influenza (AI) is a devastating poultry disease that currently can be controlled only by liquidation of affected flocks. In spite of typically very high mortality rates, a group of survivors was identified and genotyped on a 600K single nucleotide polymorphism (SNP) chip to identify genetic differences between survivors, and age- and genetics-matched controls from unaffected flocks. In a previous analysis of this dataset, a heritable component was identified and several regions that are associated with outcome of the infection were localized but none with a large effect. For complex traits that are determined by many genes, genomic prediction models using all SNPs across the genome simultaneously are expected to optimally exploit genomic information. In this study, we evaluated the diagnostic value of genomic estimated breeding values for predicting AI infection outcome within and across two highly pathogenic avian influenza viral strains and two genetic lines of layer chickens using receiver operating curves. We show that genomic prediction based on the 600K SNP chip has the potential to predict disease outcome especially within the same strain of virus (area under receiver operating curve above 0.7), but did not predict well across genetic varieties (area under receiver operating curve of 0.43).

## Background

Avian influenza (AI) is a devastating disease and the current approach that is used to control it is based on isolation and extermination of affected flocks in order to stop the spread of the virus. Alternative approaches, including the use of vaccines, have been only 60% effective in chickens [1, 2] due to the high mutation rate of the virus and lack of cross-protection between viral strains. Moreover, most countries do not allow importation of vaccinated birds due to the inability to distinguish between antibodies originating from the vaccine versus infection. Even in the case of high mortality rates caused by highly pathogenic avian influenza (HPAI) infections, typically above 85% in unvaccinated birds [3], there is usually a limited number of birds that survive following infection. Factors that underlie individual survival are likely multifactorial, but probably include a strong genetic component. Susceptibility to infectious disease (including

influenza A virus) is highly heritable in humans [4–6] and there is evidence from inbred lines of chickens that this is also true for viral diseases in poultry, including AI [7–10]. A genetic approach using genotypes from the 600K Affymetrix single nucleotide polymorphism (SNP) chip was undertaken by Drobik-Czwarno et al. [11] to identify differences between survivors of two HPAI outbreaks (H5N2 in the US and H7N3 in Mexico) and their genetics- and age-matched controls from unaffected flocks. Heritability of survival to HPAI was estimated to be between 0.18 and 0.24, which indicated that almost 20% of the differences in survival could be attributed to genetics [11]. In addition, several of the genomic regions identified were associated with survival but none with a major effect, indicating a complex polygenic nature of resistance to the disease.

An alternative to searching for causal mutations, is to use high-density SNP genotypes for genetic improvement of disease resistance by using all genetic markers across the genome to predict breeding values of selection candidates through the concept of genomic prediction and selection [12]. Genomic prediction involves the

\*Correspondence: awolc@iastate.edu

<sup>1</sup> Department of Animal Science, Iowa State University, 806 Stange Road, 239E Kildee Hall, Ames, IA 50010, USA

Full list of author information is available at the end of the article



estimation of the effects of all genetic markers on the phenotype in a training dataset, followed by the use of these estimates to predict the genetic or breeding value of selection candidates based only on their SNP genotypes across the genome. If there is a significant genetic component to resistance, this prediction and selection method can be effective even if no clear signals are obtained from genome-wide association studies (GWAS).

In clinical epidemiology, the receiver operating characteristic (ROC) curve is often used to evaluate diagnostic methods in terms of their ability to distinguish between healthy and sick subjects. This method has several advantages, including independence of prevalence of disease or choice of the decision criterion [13]. In this study, we used ROC curves to evaluate the diagnostic value of genomic breeding values for predicting AI infection outcome within and across different viral strains and genetic lines of layer chickens.

## Methods

Three sets of samples were used for this study (see [11] for a more detailed description): Hy-Line US commercial samples (205 survivors and 397 controls) from the 2015 H5N2 outbreak, Hy-Line Mexico commercial samples (480 survivors and 176 controls) from a 2012 H7N3 outbreak, and non-Hy-Line US commercial samples (47 survivors and 45 controls) from the same 2015 H5N2 outbreak in the US. All birds were from White Leghorn varieties. Blood samples from HPAI survivors and their age- and genetics-matched unaffected controls were collected on FTA Elute Microcards (GE Healthcare, Piscataway, NJ). Controls were selected from contemporary Midwest flocks that were not affected by HPAI because it was impossible to collect samples from dead birds on the affected farms for biosecurity and practical reasons. Because mortality in the H5N2 outbreak was higher than 99%, it was assumed that age-matched random birds from the same genetic varieties would be susceptible. DNA was extracted and genotyped on the 600K Affymetrix Axiom chicken SNP panel [14]. After quality control, 420,458 SNPs were retained for analysis, including some that were fixed within some subsets of samples. In case-control studies, population stratification is always a concern because it can lead to spurious associations. However, using multidimensional scaling analysis, Drobik-Czwaro et al. [11] showed that there were no structural differences between survivors and controls within the analysed populations.

The phenotype of survivor/control (0/1) was analysed using the BayesB method by fitting all SNPs simultaneously, in GenSel [15], separately for each dataset. The proportion of SNPs assumed to have no effect on survival was set to 0.999 in order to keep the number of SNPs in

the model on the same scale as the number of available phenotypic observations. A mixed linear model with the fixed effect of an overall mean and random additive effects of SNPs was fitted. The length of the chain was 52,000 iterations with the first 2000 discarded as burn-in. Priors for variance components were 0.05 for genetic variance and 0.15 for residual variance, which is equivalent to a heritability of 0.25, with 4.2 degrees of freedom for residual variance to reflect the high level of uncertainty. Posterior estimates of SNP effects were used to compute genomic breeding values for survival to HPAI of genotyped birds that were not in the training data.

To verify the potential of genomic prediction for genetic improvement of AI resistance, random fivefold cross-validation was performed within Hy-Line US and Mexico commercial samples, as well as prediction of the Hy-Line US commercial samples from the Hy-Line Mexico commercial samples, to represent prediction across different virus strains, and prediction of non-Hy-Line US commercial samples from Hy-Line US commercial samples, to represent prediction across different genetic varieties of layer chickens. The usefulness of the genomic estimated breeding values (GEBV) for predicting disease outcome was summarized by ROC curves. Sensitivity (true positive rate) and specificity (true negative rate) were calculated across the whole range of thresholds to classify birds as survivors versus controls based on their GEBV, i.e., for a given GEBV threshold, sensitivity is the proportion of survivors among birds with GEBV above this threshold and specificity is the proportion of controls among birds with GEBV below the threshold. Sensitivity was plotted against (1-specificity) (= false positive rate) to create the ROC curves, together with a 45° line that represents random assignment of survivors and controls. Summary statistics of area under the curve (AUC) and significance test for AUC=0.5 (or breeding values being equivalent in predictive ability to random assignment) were performed using the program easyROC [16].

## Results and discussion

Windows explaining more than 1% of the genetic variance within each dataset are included in Table 1. A 1-Mb window on chromosome 1 at 126 Mb consistently explained the largest amount of genetic variance in Hy-Line Mexico commercial samples (27 to 32%). This is in agreement with the strongest association signal identified by Drobik-Czwaro et al. [11] on chromosome 1, using different GWAS on the complete dataset. In the Hy-Line US commercial samples, the windows explaining the largest proportion of genetic variance differed between the folds (Table 1) and the estimates of variance explained were smaller than in the Hy-Line Mexico commercial samples. This suggests a more polygenic determination

**Table 1** Location of the 1-Mb regions that explained more than 1% of genetic variance (%Var) for different datasets and the probability of that region having a nonzero effect ( $P > 0$ )

Data set	Chromosome	Position (Mb)	Number of SNPs	%Var	$P > 0^b$
Mexico Hy-Line1 <sup>a</sup>	1	126	306	32.1	1.0
Mexico Hy-Line2	1	126	306	28.1	1.0
Mexico Hy-Line2	29	0	5583	1.0	1.0
Mexico Hy-Line3	1	126	306	27.1	1.0
Mexico Hy-Line3	12	12	608	2.9	0.7
Mexico Hy-Line3	29	0	5583	1.2	1.0
Mexico Hy-Line4	1	126	306	28.8	1.0
Mexico Hy-Line4	4	69	340	1.4	0.6
Mexico Hy-Line5	1	126	306	26.0	1.0
US Hy-Line1	1	167	438	3.0	0.5
US Hy-Line1	20	3	615	1.1	0.5
US Hy-Line1	13	11	561	1.0	0.5
US Hy-Line1	1	166	352	1.0	0.4
US Hy-Line2	15	1	659	1.7	0.6
US Hy-Line2	15	0	422	1.1	0.5
US Hy-Line2	29	0	5583	1.0	1.0
US Hy-Line3	1	32	283	2.0	0.4
US Hy-Line3	15	1	659	1.9	0.6
US Hy-Line3	15	0	422	1.1	0.5
US Hy-Line4	1	71	364	1.0	0.4
US Hy-Line4	29	0	5583	1.0	1.0
US Hy-Line5	4	84	433	3.4	0.4
US Hy-Line5	1	167	438	2.3	0.5
US Hy-Line5	9	16	629	1.2	0.5
US Hy-Line All	7	28	497	3.0	0.4
US Hy-Line All	1	32	283	1.7	0.4
US Hy-Line All	9	16	629	1.1	0.5
US Hy-Line All	1	167	438	1.1	0.4

<sup>a</sup> The number at the end of the dataset name refers to the fold number of the fivefold cross-validation

<sup>b</sup>  $P > 0$  was calculated as the proportion of MCMC iterations in which at least one SNP from that window was fitted in the model with nonzero effect

of survival of infection with the H7N3 virus than with H5N2. The use of birds from unaffected flocks as controls instead of dead birds from the respective affected flocks could have reduced power of the analysis and decreased the accuracy of genomic prediction. However, because mortality was very high (>99%), the risk of including

some potential survivors in the control group is expected to be limited.

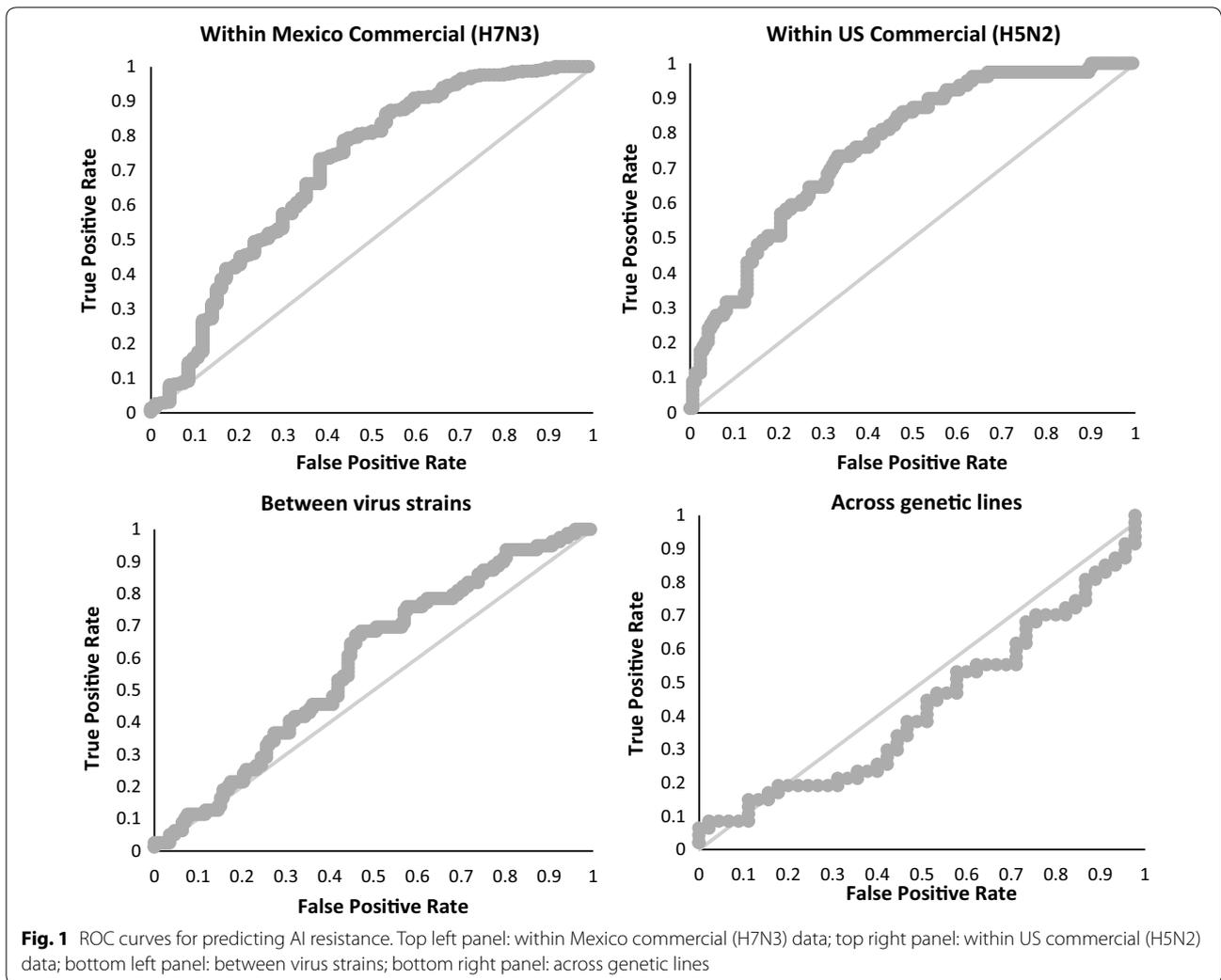
The ROC curves are in Fig. 1 and AUC results are in Table 2.

The ROC curves show significant improvement (Table 2) in accuracy of prediction compared to random assignment for all tested scenarios, except when predicting across different genetic varieties of chickens. The best diagnostic performance of GEBV was observed when training and validation were within the same virus strain and the same genetic variety (Table 2). In spite of differences in the windows that explained the largest proportion of genetic variance between the different virus strains, prediction from Mexico to US Hy-Line samples was significantly better than random assignment (AUC=0.58,  $P=0.02$ ). These results show that genomic predictions based on the 600K SNP chip have potential to predict disease outcome, especially within the same strain of virus, but do not predict well across genetic varieties. Methods to implement this type of crossbred performance for pure line selection using genomic data was discussed by Ibánñez-Escriche et al. [17].

Poor performance of genomic prediction across layer lines for egg production was previously reported by Calus et al. [18] using genotypes from a 60K SNP chip and interpreted as evidence of low consistency of linkage disequilibrium between SNPs and QTL between chicken lines. This limitation is expected to be overcome by the use of high-density genomic data (up to the sequence level) but so far sequencing costs have been prohibitive for generating large-scale training populations based on sequence data in chickens and the gains in accuracy for within-line predictions using sequence data have been limited [19]. In addition, if the SNP chip density is higher, a larger number of samples is required to estimate genotype effects accurately, which with the extremely high mortality in HPAI outbreaks may make it very difficult to collect sufficient numbers of survivor samples. Thus, at this point, the use of sequence data appears more promising for the identification of causal mutations than a practical application for genomic prediction.

## Conclusion

Genomic predictions based on the 600K SNP chip have potential to predict avian influenza infection outcomes, especially within the same strain of virus but do not predict well across genetic varieties.



**Table 2** Summary statistics of the area under the ROC curve including confidence interval and statistical test for differences from the value of 0.5 expected under random classifier

Training-validation scenario	Area under the ROC curve				z	P value
	Mean	SE	Minimum	Maximum		
Within US Hy-Line	0.76	0.03	0.70	0.82	8.41	4.0E-17
Within Mexico	0.71	0.03	0.64	0.77	6.20	5.7E-10
Across virus strains	0.58	0.04	0.51	0.66	2.25	0.02
Across genetics	0.43	0.06	0.31	0.55	-1.14	0.25

**Authors' contributions**

AW conceived the study, contributed to the analysis and wrote the draft. WDC and TJ performed the analysis. JA, JF conducted the experiment and collected the data. JCMD contributed to the methods and discussion. All authors read and approved the final manuscript.

**Author details**

<sup>1</sup> Department of Animal Science, Iowa State University, 806 Stange Road, 239E Kildee Hall, Ames, IA 50010, USA. <sup>2</sup> Hy-Line International, 2583 240th Street, Dallas Center, IA 50063, USA. <sup>3</sup> Department of Animal Genetics and Breeding, Faculty of Animal Science, Warsaw University of Life Sciences, Ciszewskiego 8, 02-786 Warsaw, Poland. <sup>4</sup> Nutribiogen, Witkowska 15/1, 61-039 Poznan, Poland.

### Acknowledgements

Juan Carlos Casillas from Hy-Line de Mexico and numerous commercial egg production facilities within Iowa enabled sample collection.

### Competing interests

The authors declare that they have no competing interests.

### Availability of data and materials

The data that support the findings of this study are available from Hy-Line International but restrictions apply to the availability of these data, which were used under license for the current study, and thus are not publicly available. However, data are available from the authors upon reasonable request and with permission of Hy-Line International.

### Consent for publication

Not applicable.

### Ethics approval and consent to participate

The data were collected on commercial farms, with the USA locations complying with the UEP (United Egg Producers) certified program for animal well-being. Welfare policies were regulated by each company policy and following the national regulations. Avian Influenza survivors were humanely euthanized as required by law to prevent the spread of the disease.

### Funding

The study was supported by a Grant from the Iowa Egg Industry Center and funding from Hy-Line International and Lohmann Tierzucht.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 16 January 2018 Accepted: 24 April 2018

Published online: 02 May 2018

### References

- Greene JL. Update on the highly-pathogenic avian influenza outbreak of 2014–2015. Congressional Research Service 7-5700. 2015. <https://fas.org/sgp/crs/misc/R44114.pdf>. Accessed 12 Jan 2018.
- Escorcía M, Del Río JC, Marrufo D, Godoy FS, Comonfort-Díaz S, Gómez B, et al. Clinical observations and pathology of the reproductive tract associated with natural infection of an avian influenza (AI) virus in an AI vaccinated layer flock. 2017. <https://en.engormix.com/poultry-industry/articles/clinical-observations-pathology-reproductive-t39955.htm>. Accessed 12 Jan 2018.
- França MS, Brown JD. Influenza pathobiology and pathogenesis in avian species. *Curr Top Microbiol Immunol*. 2014;385:221–42.
- Sorensen TI, Nielsen GG, Andersen PK, Teasdale TW. Genetic and environmental influences on premature death in adult adoptees. *N Engl J Med*. 1988;318:727–32.
- Albright FS, Orlando P, Pavia AT, Jackson GG, Cannon Albright LA. Evidence for heritable predisposition to death due to influenza. *J Infect Dis*. 2008;197:18–24.
- Everitt AR, Clare S, Pertel T, John SP, Wash RS, Smith SE, et al. IFITM3 restricts the morbidity and mortality associated with influenza. *Nature*. 2012;484:519–23.
- Sironi L, Williams JL, Stella A, Minozzi G, Moreno A, Ramelli P, et al. Genomic study of the response of chicken to highly pathogenic avian influenza virus. *BMC Proc*. 2011;5:S25.
- Smith J, Smith N, Yu L, Paton IR, Gutowska MW, Forrest HL, et al. A comparative analysis of host responses to avian influenza infection in ducks and chickens highlights a role for the interferon-induced transmembrane proteins in viral resistance. *BMC Genomics*. 2015;16:574.
- Wang Y, Lupiani B, Reddy SM, Lamont SJ, Zhou H. RNA-seq analysis revealed novel genes and signaling pathway associated with disease resistance to avian influenza virus infection in chickens. *Poult Sci*. 2014;93:485–93.
- Zhang L, Li P, Liu R, Zheng M, Sun Y, Wu D, et al. The identification of loci for immune traits in chickens using a genome-wide association study. *PLoS One*. 2015;10:e0117269.
- Drobik-Czwarno W, Wolc A, Fulton JE, Arango J, Jankowski T, O'Sullivan NP, et al. Identifying the genetic basis for resistance to avian influenza in commercial egg layer chickens. *Animal*. 2017. <https://doi.org/10.1017/S1751731117002889> (in press).
- Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 2001;157:1819–29.
- Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Casp J Intern Med*. 2013;4:627–35.
- Kranis A, Gheyas AA, Boschiero C, Turner F, Yu L, et al. Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics*. 2013;14:59.
- Garrick DJ, Fernando RL. Implementing a QTL detection study (GWAS) using genomic prediction methodology. *Methods Mol Biol*. 2013;1019:275–98.
- Goksuluk D, Korkmaz S, Zararsiz G, Karaagaoglu AE. EasyROC: an interactive web-tool for ROC curve analysis using R language environment. *R J*. 2016;8:213–30.
- Ibáñez-Escriche N, Fernando RL, Toosi A, Dekkers JC. Genomic selection of purebreds for crossbred performance. *Genet Sel Evol*. 2009;41:12.
- Calus MPL, Huang H, Vereijken A, Visscher J, ten Napel J, Windig JJ. Genomic prediction based on data from three layer lines: a comparison between linear methods. *Genet Sel Evol*. 2014;46:57.
- Ni G, Caverio D, Fangmann A, Erbe M, Simianer H. Whole-genome sequence-based genomic prediction in laying chickens with different genomic relationship matrices to account for genetic architecture. *Genet Sel Evol*. 2017;49:8.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

