

Genetic variation in chicken interferon signalling pathway genes in research lines showing differential viral resistance

Joshua Mountford | Almas Gheyas | Lonneke Vervelde | Jacqueline Smith 

The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian, UK

Correspondence

Jacqueline Smith, The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush Campus, Midlothian EH25 9RG, UK.
Email: jacqueline.smith@roslin.ed.ac.uk

Funding information

This study was made possible through funding provided by the Biotechnology and Biological Sciences Research Council under project numbers BBS/E/D/10002071 (The function of genes and cellular phenotypes in animal systems) and BBS/E/D/20002174 (Host responses underlying immunity) and a National Capability Grant to the NARF (BB/J004219/1). LV was supported by funding from the European Union's Horizon 2020 Research and Innovation Programme under Delta-Flu grant agreement no. 727922

Abstract

Avian viruses of economic interest are a significant burden on the poultry industry, affecting production traits and resulting in mortality. Furthermore, the zoonosis of avian viruses risks pandemics developing in humans. Vaccination is the most common method of controlling viruses; however current vaccines often lack cross-protection against multiple strains of each virus. The mutagenicity of these viruses has also led to virulent strains emerging that can overcome the protection offered by vaccines. Breeding chickens with a more robust innate immune response may help in tackling current and emerging viruses. Understanding the genetic evolution of different lines will thus provide a useful tool in helping the host in the fight against pathogens. This study focuses on the interferon genes and their receptors in different chicken lines that are known to be more resistant or susceptible to particular avian viruses. Comparing genotypic differences in these core immune genes between the chicken lines may explain the phenotypic differences observed and aid the identification of causative variations. The relative resistance/susceptibility of each line to viruses of interest (Marek's disease virus, infectious bursal disease, infectious bronchitis virus and avian influenza virus) has previously been determined. Here we identify single nucleotide polymorphisms in interferons and downstream genes. Functional prediction tools were used to identify variants that may be affecting protein structure, mRNA secondary structure or transcription factor and micro-RNA binding sites. These variants were then considered in the context of the research lines and their distribution between phenotypes. We highlight 60 variants of interest in the interferon pathway genes that may account for susceptibility/resistance to viral pathogens.

KEYWORDS

chicken, disease resistance, genetic variation, inbred lines, interferon

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Animal Genetics* published by John Wiley & Sons Ltd on behalf of Stichting International Foundation for Animal Genetics.

INTRODUCTION

Study of the innate immune system is important as it is seen as a generalised first line of defence and a step in the initiation of immune responses to a pathogen. Induction of the interferon (IFN) signalling pathway is central to this response. The IFN system in chickens is of particular interest as it bears many similarities to the human system, but its simplified nature makes it an ideal model for studying this system. The interferon response in chickens is induced by pattern recognition receptors (PRRs) that react to pathogenic materials such as viral RNA, and this causes a signal cascade, resulting in upregulation of the type I (*IFN α* , *IFN β* , *IFN κ*), type II (*IFN γ*) and type III (*IFN λ*) interferon genes. Once produced, these interferons can induce an antiviral state by interacting with complementary receptor complexes as detailed in Figure 1 (Santhakumar et al., 2017a, Santhakumar et al., 2017b).

In vertebrate species, it was initially believed that type I and III interferons were redundant owing to the similarity in signal cascade and interferon stimulated genes (ISGs) affected. However, whilst type I receptors are expressed in most cells, the *IL-28R α* (*IFNLR1*) receptor appears to be preferentially expressed in specific cells such as epithelial cells. This suggests that the type I system works in a more systemic fashion compared with the cell-specific type III interferon system (Andreacos et al., 2019; Santhakumar et al., 2017b). Furthermore, inducing the expression of *IFNLR1* in chicken fibroblasts allows IFN λ to induce the type III antiviral response, further supporting the view that the distribution of this receptor determines the antiviral response in these cells (Reuter et al., 2014). The type II IFN system in vertebrate species is unique in comparison as its function is predominantly restricted to immune cells and IFN γ is principally produced by natural killer cells (Lee & Ashkar, 2018). Despite the overlap between the systems highlighted in Figure 1, it is evident the interferon systems play a distinct role in vertebrate species.

The economic impact of avian viruses on the poultry industry cannot be overstated. Infections can result in a reduction in egg production, the condemnation of consumable meat, immunosuppression leading to secondary infections and mortality. In the context of this study, we will focus on resistance to four economically important viruses: Marek's disease virus (MDV) (Boodhoo et al., 2016), infectious bursal disease virus (IBDV) (Dey et al., 2019), infectious bronchitis virus (IBV) (Legnardi et al., 2020) and avian influenza virus (AIV), which also poses a potential threat to human health (Lycett et al., 2019). Attempts to control these viruses by vaccination or alleviate the impacts of disease have been utilised over the years with varying degrees of success. Alongside the improvement of vaccines and/or vaccine adjuvants the identification of disease resistance genes can feed into conventional breeding programmes, but the selection of

birds with improved innate immune responsiveness may result in a broader effect than the selection of those with resistance to specific diseases (Kaiser, 2010; Swaggerty et al., 2009). Studies from MHC-congenic lines indicate that there are MHC-related and non-MHC related factors involved in viral disease resistance. Research using inbred/partially inbred chicken lines in response to viral disease has identified transcriptomic variations either in the interferon genes or in their up- or downstream effectors (Kaufman, 2018; Smith et al., 2011; Smith et al., 2015a; Stone, 1975). Given their role in the innate immune response and their variability in expression in response to viruses, it is possible that genetic variants in the interferon pathways between different chicken lines may contribute to resistant and susceptible phenotypes. Identifying variants causative for susceptibility or resistance could therefore allow for the development of breeding schemes to introduce viral resistance into commercial lines and also increase knowledge of virus–host interactions.

The goal of the current *in silico* study was to identify genetic variants from whole genome sequence data from eight chicken research lines and to annotate these variants, i.e. to determine if they are exonic, intronic, upstream or downstream of a gene. Functional prediction tools were used to determine the impact of these variants. Finally, using prior knowledge of the lines' susceptibility and resistance to IBV, IBDV, AIV and MDV, variants have been correlated to the disease response phenotype with the aim of identifying mutations potentially responsible for specific resistance or susceptibility.

MATERIALS AND METHODS

Determination of chicken line phenotype

The research lines examined in this study were lines WI, 15I, 7₂, 6₁, C.B12, N, 0 and P2a which have been previously described (Bacon et al., 2000; Mohd Isa et al., 2020; Stone, 1975). These are all White Leghorn lines inbred to varying degrees. The level of inbreeding is indicated in Table S1. Lines 6₁, 7₂, 15I and C are highly inbred (with inbreeding coefficients ranging from 0.72 to 0.85), whereas lines N, P2a and WI are only partially inbred (with inbreeding coefficients ranging from 0.37 to 0.59). Line 0, so called because it contains no avian leukosis virus subgroup E (ALVE) genes, was not developed as an inbred line, but does show a low level in this study. Data from Kaiser et al. (2008) was considered as a starting point for identifying line resistance and susceptibility phenotypes. However, these data were generated in the 1990s and contradictory findings have since been identified (Farhanah et al., 2018; Smith et al., 2015b). A broad search of the literature was completed in order to classify lines as either susceptible or resistant to each virus,

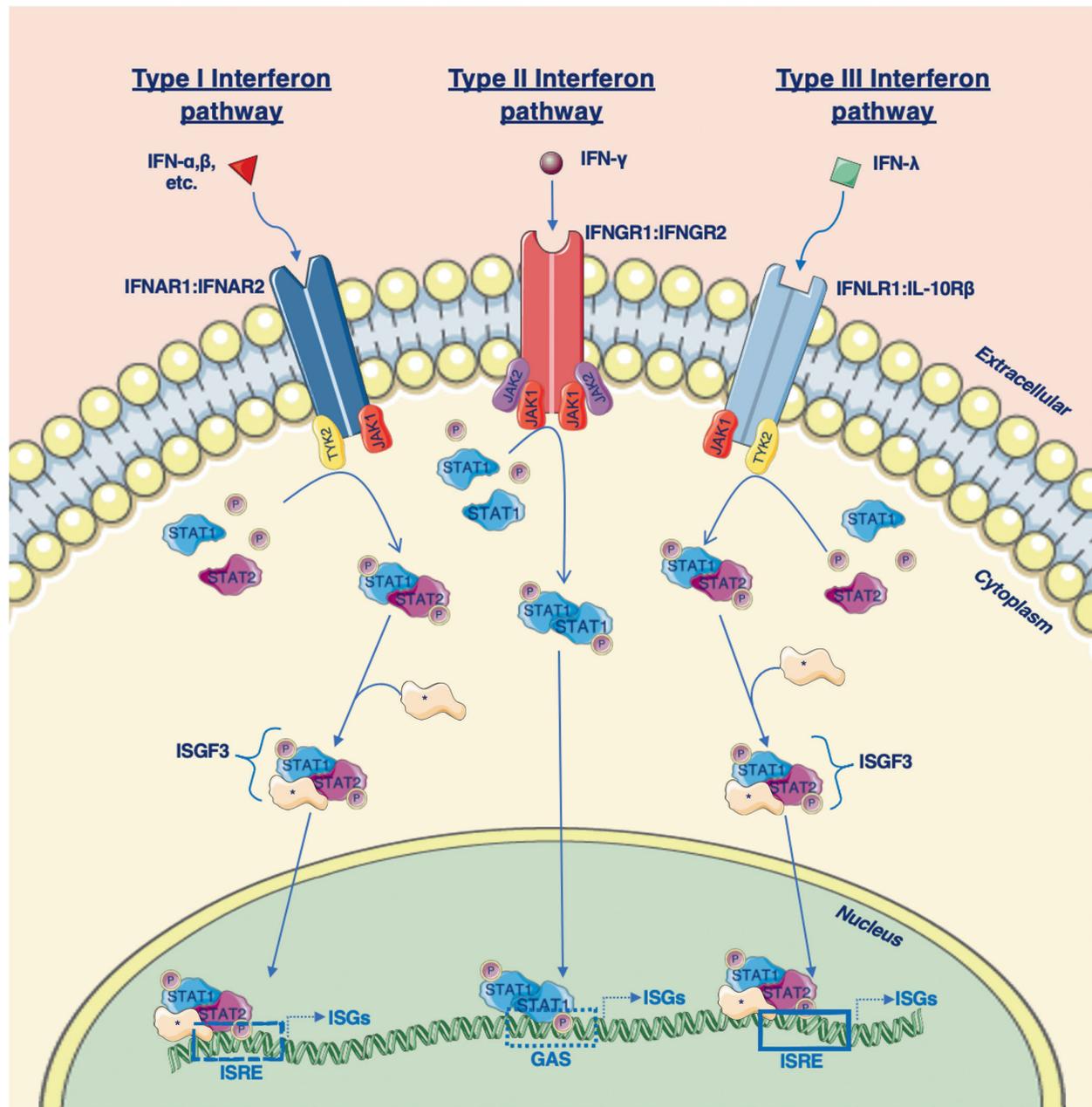


FIGURE 1 The current model for the interferon response pathway of the chicken. Type I, II and III interferon molecules act as cytokines in a paracrine and autocrine fashion by interacting with the IFNAR1:IFNAR2, IFNGR1:IFNGR2 and IFNLR1:IL-10R β dimeric receptor complexes respectively. Upon receptor binding, the Janus kinase (JAK) signal transducer and activation of transcription (STAT) pathway is activated, whereby phosphorylation of STAT1 and STAT2 causes them to dimerise. In mammals this dimer then goes on to form the complex interferon stimulated gene factor 3 (ISGF3 – in mammals this complex forms via interactions with IRF9, however a chicken orthologue of this protein has yet to be identified). The ISGF3 complex can then enter the nucleus and bind to interferon stimulated response element sequences, resulting in the upregulation of hundreds of ISGs and a specific antiviral state. In the case of the type II system, STAT1 and STAT2 form a homodimer that can then enter the nucleus and bind to gamma-activated sequences, resulting in upregulation of specific ISGs (Andreaskos et al., 2019; Santhakumar et al., 2017b). This image was created using modified images from Servier Medical Art, licensed under a Creative Commons Attribution 3.0 Generic License (<http://smart.servier.com>)

as much as was possible. Studies of IBV (Bumstead et al., 1989; Cook et al., 1990; Nakamura et al., 1991; Otsuki et al., 1990), IBDV (Asfor et al., 2021; Bumstead et al., 1993; Farhanah et al., 2018; Smith et al., 2015b), MDV (Burgess et al., 2001; Cole, 1968; Mohd Isa et al., 2020) and AIV (Ruiz-Hernandez et al., 2016) infections were referenced.

Identification of chicken interferon genes of interest

Ensembl genome browser release 103 (<https://www.ensembl.org/index.html>) was used to identify the interferon pathway genes of interest and their genomic co-ordinates. Genomic positions from this release

were used for interrogating sequence data from the research lines mapped to the GRCg6a chicken reference genome (GenBank assembly accession number GCA_000002315.5). Paralogues with >95% target identity were also identified and examined in this study (Gheyas et al., 2015).

Whole genome sequence data from the research lines

Genomic data from the research lines is the same as that used in the development of the 600 K chicken SNP genotyping array (Kranis et al., 2013). Briefly, 10 chickens from each line were supplied by the Institute of Animal Health (Compton, UK; now available through the National Avian Research Facility (NARF, Edinburgh, UK), with all chickens within lines being of the same gender (female in all lines except for male line N chickens). DNA was extracted from blood samples and pooled for sequencing on an Illumina GAIIX platform using a paired-end protocol where read length was 101 bases. These previously generated whole genome sequence data from each line were aligned to the GRCg6a chicken reference genome using the Burrows-Wheeler Aligner mem (BWA-mem) software package, version 0.7.15, using default settings (Li & Durbin, 2009).

Variant calling and annotation

The GENOME ANALYSIS TOOLKIT (GATK) version 3.8 and Picard 2.9.2 packages were used throughout the variant calling process following the best practices workflow for germline short variant discovery (GATK, n.d.). Picard's SortSam function was used to sort the mapping data into coordinate order and the MarkDuplicates function highlighted duplicate reads. Base Quality Score Recalibration of the mapped reads was then performed using the GATK. Over 20 million known single nucleotide polymorphisms (SNPs) in the chicken genome from Ensembl (release 92) were used in the BaseRecalibrator step as known variants to be masked from consideration as sequencing errors during the recalibration process. GATK's HaplotypeCaller function was used to call short variants. This resulted in a GVCF file for each line, and the files were then combined into one VCF file using the GenotypeGVCF function. The variant quality score recalibration (VQSR) approach was applied for filtering variants (GATK, n.d.). For this, over 1 million validated polymorphic SNPs (Kranis et al., 2013) were used as a truth dataset alongside the previous known dataset and the unfiltered variants from the research lines to create a trained Gaussian mixture model using VariantRecalibrator. The model calculated a variant quality score log-odds (VQSLOD) score for each SNP in the truth dataset and the research lines dataset. The ApplyRecalibration step of VQSR

(also called 'ApplyVQSR' in the latest version of the GATK) was used with a sensitivity filter of 99% to remove SNPs with a VQSLOD score lower than the top 99% of the truth set, as these are likely to be false positives. To further minimise the inclusion of false positives, SNPs with a VQSLOD score <20 were removed.

The ENSEMBL variant effect predictor (VEP) webtool (<https://www.ensembl.org/info/docs/tools/vep/index.html>) was used to annotate SNPs identified within and around the genes of interest. Variants were defined as upstream (if within 1 kb upstream of a transcription start site), downstream (within 1 kb downstream of the gene), 3' untranslated region (3'UTR), 5' untranslated region (5'UTR), intronic, synonymous or missense variants (McLaren et al., 2016).

Prediction of potential functional impacts of SNPs

To determine the potential impact of missense variants on protein function and amino acid sequences, corresponding variants were examined using the NORMAL-SMART, SIFT-SEQUENCE, PROVEAN PROTEIN and SNAP2 webtools. NORMAL-SMART (Simple Modular Architecture Research Tool; <http://smart.embl-heidelberg.de>) predicts domains within the protein sequence. Variants affecting particular protein domains could therefore be identified (Letunic et al., 2021). SIFT-SEQUENCE uses the SIFT (Sorting Intolerant From Tolerant) algorithm to predict whether an amino acid substitution, owing to a missense SNP, affects protein function based on sequence evolutionary conservation and the physical properties of each amino acid. Default settings were used for the prediction of the SIFT score for each missense SNP (Sim et al., 2012). Similarly, PROVEAN PROTEIN (Protein Variation Effect Analyser; http://provean.jcvi.org/seq_submit.php) is another method to investigate the impact of amino acid substitution on the biological function of a protein (Choi & Chan, 2015) and again the predictions were made using default settings. SNAP2 (Screening for Non-Acceptable Polymorphisms; <https://roslab.org/services/snap2web/>) uses a neural network-based classifier for distinguishing between functional and neutral variants within the protein sequence where the prediction algorithm is trained using more than 100000 experimentally annotated variants from multiple online databases (PMD, SWISS-PROT, OMIM) and HumVar (Hecht et al., 2015).

To determine the potential impact of variants on micro-RNA (miRNA) binding sites, 3'UTR variants were compared with miRNA binding sites identified via the miRDB (<http://www.mirdb.org>) online database (Chen & Wang, 2020). The potential impact of variants on transcription factor binding sites (TFBSs) was investigated by uploading 1 kb of upstream gene sequences into the MATCH 1.0 public webserver (<http://gene-regulation.com/pub/programs.html#match>).

TABLE 1 Resistant and susceptible phenotypes of research lines

Inbred chicken line	Virus resistance phenotype (Resistant/Moderate susceptibility/Susceptible)				
	MHC B haplotype	IBV (Bumstead et al., 1989; Cook et al., 1990; Nakamura et al., 1991; Otsuki et al., 1990)	IBDV (Asfor et al., 2021; Bumstead et al., 1993; Farhanah et al., 2018; Smith et al., 2015b)	MDV (Burgess et al., 2001; Cole, 1968; Mohd Isa et al., 2020)	LPAIV (Ruiz-Hernandez et al., 2016)
W1	B14	M	S	-	-
151	B15	S	R	S	-
7 ₂	B2	S	R	S	-
6 ₁	B2	M	M	R	-
C	B4 and B12	R	R	-	S
N	B21	R	M	R	-
0	B21	-	R	-	R
P2a	B19	-	-	S	-

LPAIV - Lowly Pathogenic Avian Influenza Virus; "R" denotes resistance, "M" denotes moderate susceptibility, "S" denotes susceptibility and "-" indicates no data identified. These data do not account for potential gender differences.

These lines were originally housed at the Institute for Animal Health, Compton, UK, but are now available at the NARF, Roslin, UK.

SNP positions were compared with known TFBSs and any changes recorded. Parameters used were 'high quality matrices only' and to 'minimise false positives' (Kel et al., 2001).

To determine the impact of variants on the mature mRNA secondary structure, SNPs identified as 3'UTR, missense, synonymous and 5'UTR were analysed using the RNASNP (<https://rth.dk/resources/rnasnp/>) and SNPFOLD (<http://ribosnitch-compute.bio.unc.edu/snpfold/SNPFold.html>) webtools alongside the mRNA sequence of the affected gene of interest. RNASNP assesses the effect of an SNP on a RNA secondary structure, as structural characteristics are essential for the correct functioning of RNA molecules (Sabarinathan et al., 2013). The RNASNP predictions were made using the 'Mode 2' function, which uses a local folding algorithm designed for long RNA molecules. SNPFOLD considers all possible RNA conformations for the input sequence and uses Pearson correlation coefficients to compare secondary structure change between wild-type and variant sequences (Halvorsen et al., 2010). The *p*-values indicating significance generated by these software tools were corrected for multiple tests using Bonferroni correction.

RESULTS

Phenotype definition of chicken lines

Previous observations of resistant and susceptible phenotypes in response to strains of IBV, IBDV, MDV and AIV for the research lines investigated in this study are described in Table 1. Resistance and susceptibility were determined based on variables such as viral titre, mortality, clinical symptoms and transmissibility. Issues encountered when considering such studies included the subjective nature of how resistance and susceptibility are defined and the lack of systematic study of these lines against the different viruses. These observations have also been made at different time points of infection, different chicken ages and while using different virus strains and infection doses or routes, adding further complexity to an already complex trait. The different MHC backgrounds of each line also need to be considered (Stone, 1975). The phenotypes defined were used to consider SNPs identified in the context of resistance and susceptibility depending on how they are distributed between lines.

Chicken interferon pathway genes

The genes of interest, the paralogues of these genes and their location in the GRCg6a reference genome are shown in Table 2. As the JAK-STAT pathway is intrinsic to ISG regulation, the *TYK2*, *JAK1*, *JAK2*, *STAT1*, *STAT2* and *IRF9* genes were examined along with the

TABLE 2 Genes of interest identified in GRCg6a

Gene	Gene/paralogue ID	Sequence location and direction
<i>IFNλ</i>	ENSGALG00000052146	Chromosome 7: 4610682-4612191 forward strand
	ENSGALG00000050947	Chromosome 7: 4572389-4573897 forward strand
	ENSGALG00000047344	Chromosome 7: 4591535-4593042 forward strand
<i>IFNα</i>	ENSGALG00000048874	Chromosome Z: 7395233-7395995 forward strand
	ENSGALG00000052209	Chromosome Z: 7410282-7411044 forward strand
	ENSGALG00000044725	Chromosome Z: 7399231-7399942 forward strand
	ENSGALG00000047630	Chromosome Z: 7381503-7382265 forward strand
	ENSGALG00000050924	Chromosome Z: 7401271-7402033 forward strand
	ENSGALG00000053752	Chromosome Z: 7387284-7388046 forward strand
	ENSGALG00000054396	Chromosome Z: 7385475-7386237 forward strand
	ENSGALG00000046996	Chromosome Z: 7377531-7378293 forward strand
	ENSGALG00000053207	Chromosome Z: 7414251-7415013 forward strand
	ENSGALG00000054368	Chromosome Z: 7391261-7392023 forward strand
	ENSGALG00000054104	Chromosome Z: 7421956-7422718 forward strand
	<i>IFNβ</i>	ENSGALG00000005759
<i>IFNκ</i>	ENSGALG00000015062	Chromosome Z: 34282011-34285224 reverse strand
<i>IFNγ</i>	ENSGALG00000009903	Chromosome 1: 35173604-3517772 reverse strand
<i>IFNARI</i>	ENSGALG00000030363	Chromosome 1: 106613562-106632348 forward strand
<i>IFNAR2</i>	ENSGALG00000041867	Chromosome 1: 106580169- 106592227 forward strand

TABLE 2 (Continued)

Gene	Gene/paralogue ID	Sequence location and direction
<i>IFNLRI</i>	ENSGALG00000004231	Chromosome 23: 5819774-5826643 reverse strand
<i>IL-10Rβ</i>	ENSGALG000000037989	Chromosome 1: 106593970- 106604799 forward strand
<i>IFNGRI</i>	ENSGALG00000013865	Chromosome 3: 54895190-54914161 forward strand
<i>IFNGR2</i>	ENSGALG000000032660	Chromosome 1: 106645075- 106653536 forward strand
<i>TYK2</i>	ENSGALG000000030599	Chromosome 30: 178603-197548 forward strand
<i>JAK1</i>	ENSGALG00000011031	Chromosome 8: 28552333-28603383 reverse strand
<i>JAK2</i>	ENSGALG00000015027	Chromosome Z: 27532087-27615380 forward strand
<i>STAT1</i>	ENSGALG00000007651	Chromosome 7: 7912708-7932691 reverse strand
<i>STAT2</i>	ENSGALG000000030661	Chromosome 33: 6934615-6940207 forward strand
<i>IRF9</i>	ENSGALG000000030291	Chromosome 20: 10063682-10066572 reverse strand

interferons and their receptors. Seventeen genes of interest were identified. *IFNα* and *IFNλ* were found to have 11 and three paralogues respectively, resulting in 29 genomic locations being explored. It should be noted that *IFNα*, *IFNβ*, *IFNκ* and *JAK2* are all located on the Z sex chromosome. In birds the female is the heterogametic sex, having ZW sex chromosomes, whereas the male has two copies of the Z chromosome. Incomplete dosage compensation in chickens results in higher expression of Z chromosome genes in males (Garcia-Morales et al., 2015).

SNPs across research lines identified in GRCg6a

Prior to VQSR, 12142027 SNPs were identified between the GRCg6a reference genome and the research lines. Over 81% of the SNPs (9862336) passed VQSR and 79% (9584464) had a VQSLOD score >20. Of these, 1760 were found to occur within the defined locations of the genes of interest, including 1 kb up- and downstream of the genes. The distribution of these SNPs after being defined

by VEP can be seen in Table 3. The SNPs are split into the following annotation categories: 81.4% (1432) intronic, 5.74% (101) upstream, 5.17% (91) downstream, 2.95% (52) 3'UTR, 2.33% (41) synonymous, 2.05% (36) missense and 0.40% (7) 5'UTR SNPs. Of the 29 genomic locations examined (Table 2), only 16 were found to contain one or more SNPs when research lines were compared with the GRCg6a reference.

Only one of the *IFN α* paralogues contained SNPs, all of which were upstream of the gene. *IFN λ* and its two paralogues were also not found to contain SNPs. It was considered that this may be due to these genes being highly conserved. However, this finding prompted examination of the sequence read depth of these genes in each line, and this identified low coverage for these regions. For example, the *IFN λ* paralogue (ENSGALG0000050947) with the highest coverage across all lines had an average read depth of only 1.55x. Compared with a longer gene with no paralogues, such as *IFNARI*, the average read depth across lines was found to be 15x. The issues of coverage observed are probably due to the presence of multiple similar paralogues in the genome and the short length of reads (101 bp) used in this study, as this will result in failure to distinguish unique genomic regions during read alignment. High GC content in particular regions will also mean problems with sequence coverage. Given this finding and the importance of both *IFN α* and *IFN λ* in these pathways it would be advisable to carry out long-read sequencing where reads can cover the entire gene region, as this would allow unique reads to be mapped and variants to be identified.

However, many more SNPs were detected in our current analysis using GRCg6a compared with the previous analysis that used the Galgal4 reference genome (Kranis et al., 2013). The comparison is shown in Appendix S2. The difference in the number reflects the better assembly and more complete build of GRCg6a, compared with Galgal4. For example, in Galgal4 no SNPs were detected from genes on chr30 or 33 but in GRCg6a we detected many SNPs (e.g. from genes *TYK2* and *STAT2*). Moreover, in our current analysis we have applied a more advanced variant calling approach (using GATK's VQSR method for filtration) compared with the hard filtration applied in the previous analysis.

Prediction of functional impacts of missense SNPs identified from genes of interest

The potential functional consequences of identified missense SNPs ($n = 36$) as predicted by SIFT, PROVEAN and SNAP2 scores can be found in Table 4. Those predicted to have a deleterious effect are shown in bold. A SIFT or PROVEAN score of <0.05 or -2.5 respectively, indicates a predicted deleterious change, whereas a SNAP2 score of greater than zero predicts an impact on protein function.

Thirteen of the amino acid substitutions were predicted to have an effect by at least one of the webtools, with five of these being predicted by two and the Y500D substitution in *JAK1* predicted to have an effect by all three. The locations of all 36 SNPs predicted protein domains impacted can be seen in Figures 2a–k.

Considering the different research lines, there are 42 instances where it could be inferred that a missense SNP contributes to resistance to one of the viruses, whereas 47 instances could be related to susceptibility. The arguments for these can be found in Table S3 (see the worksheet missenseSNPs). For example, the aforementioned Y500D substitution, visible in Figure 2i, occurs in lines C and W1 and affects the NORMAL-SMART-predicted SH2 domain of JAK1. It could be hypothesised that this SNP contributes to IBV resistance or AIV susceptibility in line C, whereas in W1 it could be linked to IBDV susceptibility. Of the 36 SNPs, 13 were found to be heterozygous. The remaining 23 SNPs were homozygous in at least one line. This finding could have further implications as allele-specific expression could potentially affect wild-type vs. mutant protein function.

SNPs potentially affecting mRNA secondary structure

Twenty-one of the 136 SNPs occurring in the mRNA sequence among the lines were predicted to affect mRNA secondary structure by at least one of the webtools (RNASNP and/or SNPFOLD with $p < 0.1$; Table 5). Only one SNP in the 3'UTR region of *IFN κ* (ChrZ:34282637) was indicated to significantly impact secondary structure by both webtools. This SNP occurs in the AIV resistant line 0, but is absent from the susceptible C line; therefore it could be perceived that this variation affects the *IFN κ* mRNA in a way that may contribute to AIV resistance. In the context of viral infection and how the SNPs distribute between the lines considered in this study, 31 instances of these could be related to viral resistance, whereas 21 instances could be related to susceptibility (Table S3; worksheet mRNAstructureAlteringSNPs).

SNPs in micro-RNA and transcription factor binding sites potentially contribute to resistance and susceptibility phenotypes

Three SNPs were found to impact either a TF or a miRNA binding site, as listed in Table 6. The gga-miR-1627-3p and gga-miR-196-1-3p binding sites in the 3'UTR of *IFNARI* are seen to be affected, whereas a V-MAF binding site upstream of *STAT1* is also disrupted. The miRNA gga-miR-196-1-3p is believed to have multiple roles including embryonic development and feather follicle development (Chen et al., 2017; Xu et al., 2013),

TABLE 3 Distribution of SNPs in genes of interest across lines

Distribution of SNPs in genes of interest across lines		Transcript ID	Total SNPs in gene (±1 kb)	Upstream							3'UTR	Intronic SNPs	Synonymous SNPs	Missense SNPs
Gene affected	Downstream			5'UTR	Downstream	5'UTR	3'UTR	Intronic SNPs	Synonymous SNPs	Missense SNPs				
<i>IFNγ</i>	ENSGGALT00000016105.4	53	6	13	0	2	32	0	0	0	0	0	0	
<i>IFNα</i>	ENSGGALT000000102661.1	8 ^a	8	0	0	0	0	0	0	0	0	0	0	
<i>IFNβ</i>	ENSGGALT00000039477.2	10 ^a	3	5	0	0	0	0	0	1	1	1	1	
<i>IFNκ</i>	ENSGGALT00000032834.4	9 ^a	1	1	0	2	4	1	1	1	1	0	0	
<i>IFNAR1</i>	ENSGGALT00000064880.2	208	9	13	1	15	160	4	4	4	6	6	6	
<i>IFNAR2</i>	ENSGGALT00000046822.2	185	20	16	1	6	133	1	1	1	8	8	8	
<i>IFNLR1</i>	ENSGGALT00000098112.1	16	1	1	2	1	7	1	1	1	3	3	3	
<i>IL-10Rβ</i>	ENSGGALT00000073701.2	169	18	12	0	7	126	2	2	2	4	4	4	
<i>IFNGR1</i>	ENSGGALT00000031776.5	163	9	8	1	7	136	0	0	0	2	2	2	
<i>IFNGR2</i>	ENSGGALT00000061931.2	85	5	13	0	11	53	2	2	2	1	1	1	
<i>TYK2</i>	ENSGGALT00000058693.2	23	0	1	0	0	15	1	1	1	6	6	6	
<i>JAK1</i>	ENSGGALT00000017969.6	626	11	5	2	1	591	14	14	14	2	2	2	
<i>JAK2</i>	ENSGGALT00000024235.6	143	2	1	0	0	136	4	4	4	0	0	0	
<i>STAT1</i>	ENSGGALT00000057169.2	21	3	0	0	0	17	0	0	0	1	1	1	
<i>STAT2</i>	ENSGGALT00000050242.2	12	0	2	0	0	7	3	3	3	0	0	0	
<i>IRF9</i>	ENSGGALT000000100907.1	29	5	0	0	0	15	7	7	7	2	2	2	
Total SNPs		1760	101	91	7	52	1432	41	41	41	36	36	36	
SNP % distribution		5.74	5.17	0.398	2.95	81.4	2.33	2.05	2.05	2.05	2.05	2.05	2.05	

The low number of identified SNPs may be a reflection of the low sequence coverage in these regions.

TABLE 4 Missense SNPs in genes of interest identified in chicken lines

SNP genome coordinate	Protein affected	Line(s) in which SNP is present	Alleles (ref/ SNP)	Amino acid substitution	SIFT score	PROVEAN score	SNAP2 score	Comments on SNP location
Chr1:106,613,646	IFNAR1	6 ₁ , 7 ₂ , 0	G/A	A4T	0 ^b	0.012	-84	
Chr1:106,613,656	IFNAR1	7 ₂	C/T	A7V	0 ^b	0.647	-25	
Chr1:106,621,892	IFNAR1	15I, 6 ₁ , 7 ₂ , C, N, P2a ^a , 0, W1	T/G	Y217D	0.29	0.363	32	Occurs in FN3 domain
Chr1:106,621,899	IFNAR1	N	A/G	E219G	0.44	-0.719	18	Occurs in FN3 domain
Chr1:106,623,894	IFNAR1	15I, 6 ₁ , 7 ₂ , C, N, 0, W1	G/A	S324N	1	1.099	-81	Occurs in tissue factor domain
Chr1:106,625,990	IFNAR1	0 ^a	C/T	A404V	0.76	0.729	-87	Occurs in interferon binding domain
Chr1:106,587,598	IFNAR2	C	C/T	P45S	0 ^b	0.759	-78	Occurs in tissue factor domain
Chr1:106,588,380	IFNAR2	15I, 7 ₂ , C, 0 ^a , W1	C/T	H137Y	1	0.037	-65	
Chr1:106,588,700	IFNAR2	15I, 7 ₂ , P2a ^a , 0 ^a , W1	A/G	N188D	0.7	-0.238	-89	Occurs in interferon binding domain
Chr1:106,589,401	IFNAR2	15I, P2a ^a , 0, W1	T/C	I214T	0.11	-0.435	-68	Occurs in interferon binding domain
Chr1:106,589,970	IFNAR2	15I, C, P2a ^a , 0 ^a , W1	G/T	G264V	0.15	0.584	-1	Occurs in transmembrane region
Chr1:106,590,778	IFNAR2	15I, C, P2a ^a , 0, W1	A/G	I296V	1	-0.005	-75	
Chr1:106,591,040	IFNAR2	15I, C, P2a ^a , 0 ^a , W1	G/A	G383D	0.69	1.243	-29	
Chr1:106,591,349	IFNAR2	15I, C, P2a ^a , 0, W1	C/T	T486M	^{aa0}	-0.991	-62	
Chr3:54,913,306	IFNGR1	15I, 6 ₁ , 7 ₂ , C, 0, W1	T/C	V395A	0.27	-0.16	-16	
Chr3:54,913,327	IFNGR1	15I, 6 ₁ , 7 ₂ , C, 0, W1	G/A	R402Q	0.48	1.272	-71	
Chr1:106,647,185	IFNGR2	6 ₁	T/C	C44R	0.62	3.391	-60	Occurs in tissue factor domain
Chr23:5,823,279	IFNLR1	N ^a	T/C	E151G	0 ^b	-1.341	-19	Occurs in interferon binding domain
Chr23:5,822,624	IFNLR1	C ^a	A/G	S199P	0.23	-1.046	14	Occurs in interferon binding domain
Chr23:5,822,588	IFNLR1	7 ₂ , C ^a , 0 ^a	A/G	S211P	0.06	-4.218	41	Occurs in interferon binding domain
Chr1:106,602,130	IL10RB	6 ₁	T/G	S248A	1	0.375	-66	
Chr1:106,603,837	IL10RB	6 ₁ , 7 ₂ , N, P2a ^a , 0 ^a	A/T	Q293H	0.04	-1.17	-95	
Chr1:106,603,910	IL10RB	7 ₂	A/G	R318G	0.38	-0.716	9	
Chr1:106,603,911	IL10RB	C, P2a ^a	G/A	R318K	1	-0.092	-78	

TABLE 4 (Continued)

SNP genome coordinate	Protein affected	Line(s) in which SNP is present	Alleles (ref/ SNP)	Amino acid substitution	SIFT score	PROVEAN score	SNAP2 score	Comments on SNP location
ChrZ:7,372,479	IFNβ	N ^a	T/C	S151P	0.36	-1.728	52	Occurs in IFNα/β/6 domain
Chr7:7,923,855	STAT1	151 ^a , 7 ₂ ^a , W1 ^a	A/C	V341G	0.03 ^b	-4.766	39	Occurs in STAT binding domain
Chr8:28,566,979	JAK1	N ^a , P2a ^a , 0	A/G	V332A	0.67	-0.305	-67	
Chr8:28,564,446	JAK1	C ^a , W1 ^a	A/C	Y500D	0.04	-3.757	61	Occurs in SH2 domain
Chr30:182,787	TYK2	7 ₂ ^a , W1 ^a	A/C	H215P	0.2	-2.732	63	Occurs in B41 domain
Chr30:183,686	TYK2	151 ^a , 6 ₁ ^a , 7 ₂ ^a , N, P2a, 0, W1	G/A	R299Q	0.44	-0.551	-17	Occurs in B41 domain
Chr30:186,858	TYK2	7 ₂ ^a , C ^a , 0 ^a	T/C	V550A	0.01	-2.65	-26	Occurs in SH2 domain
Chr30:190,505	TYK2	6 ₁ ^a , 7 ₂ ^a , P2a, 0 ^a	G/A	S875N	0.45	0.351	-96	Occurs in TyrKc domain
Chr30:190,528	TYK2	C ^a	T/C	S883P	0.25	-3.994	12	Occurs in TyrKc domain
Chr30:190,801	TYK2	6 ₁ ^a	A/C	D908A	0.34	-1.029	-53	
Chr20:10,064,956	IRF9	P2a ^a	T/C	H186R	0.51	-1.24	-38	
Chr20:10,064,260	IRF9	6 ₁ ^a , C ^a , 0 ^a	T/G	Y355S	0.15	-4.102	63	Occurs in IRF-3 domain

Scores in bold indicate a predicted significant effect.

^aSNP occurs in only one allele and is therefore heterozygous.

^bSIFT scores with low confidence in the value due to too few comparable sequences for the software to make a reliable prediction.

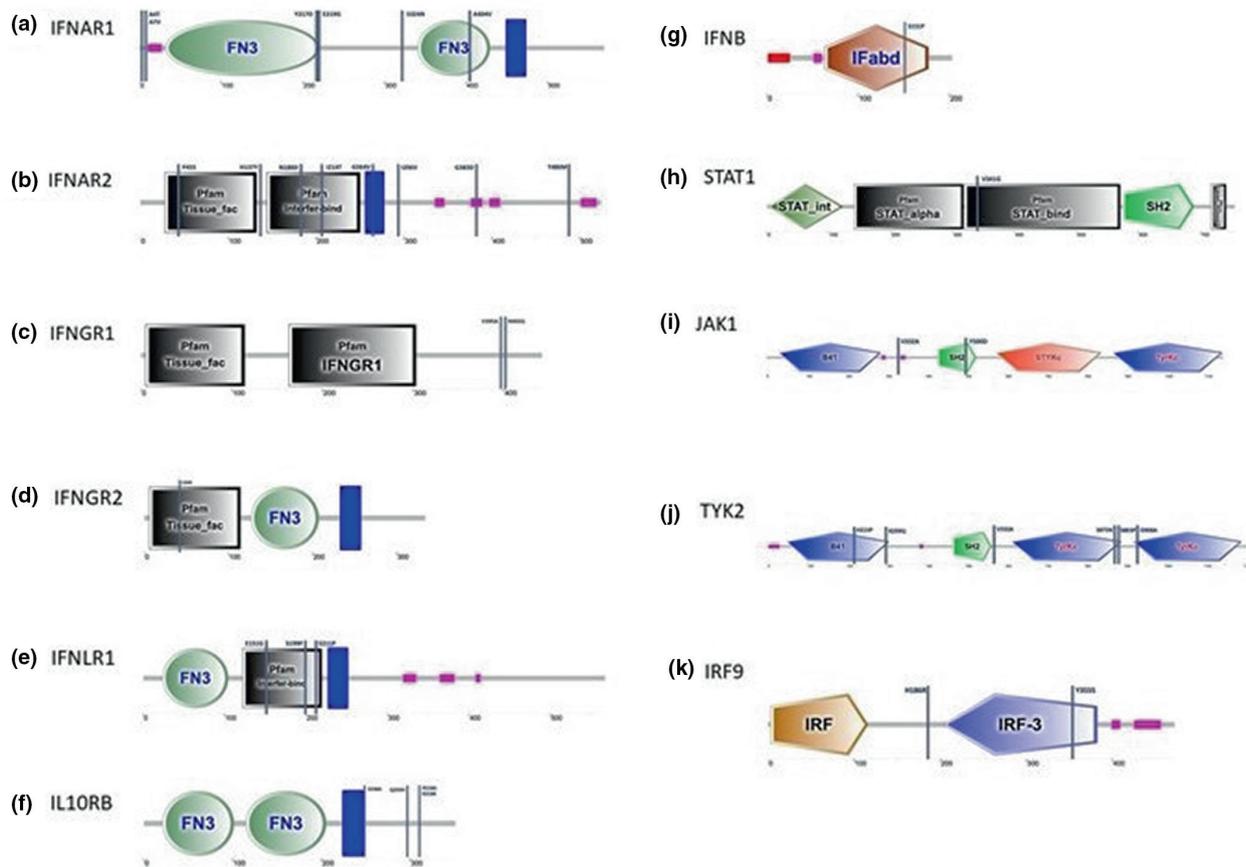


FIGURE 2 Location of missense SNPs with regards to predicted protein domains. (a) IFNAR1 – Y217, E219G and A404V can be found in fibronectin type 3 (FN3) domains. (b) IFNAR2 – P45S occurs in a tissue factor domain, N188D and I214T occur in an interferon binding domain and G264V occurs in the transmembrane region. (c) IFNGR1 – no SNPs were identified in the predicted domains. (d) IFNGR2 – C44R in tissue factor domain. (e) IFNLR1 – E151G, S199P and S211P occur in an interferon binding domain. (f) IL10Rβ – no SNPs were identified in predicted domains. (g) IFNβ – S151P SNP occurs in an IFNα/β/δ domain. (h) STAT1 – V341G SNP occurs in a STAT binding domain. (i) JAK1 – Y500D occurs in an Src Homolog 2 (SH2) domain. (j) TYK2 – H215P and R299Q occur in a band 4.1 (B41) domain and S875N, S883P and D908A occur in tyrosine kinase catalytic domains (TyrKc). (k) IRF9 – Y335S occurs in an interferon regulatory factor 3 (IRF3) domain. Note, chickens lack IRF3 but have an IRF7 orthologue. IFαβδ, Interferon alpha, beta and delta domain; STAT_int, STAT protein interaction domain; STAT_alpha, STAT protein all-alpha domain; STAT_bind, STAT protein DNA binding domain; B41, band 4.1 homologue domain (also known as ERM domain, erzin/radixin/moesin domain); STYKc, protein kinase of unclassified specificity; IRF, interferon regulatory factor domain. Pink indicates low complexity regions and dark blue blocks signify transmembrane regions

whereas the V-MAF TF is a viral oncogene (Nishizawa et al., 1989). In the context of viral infection and how the SNPs are distributed between the lines, there are five instances where viral resistance can be inferred, and two instances could be linked to susceptibility (Table S3, worksheet: TF-miRNA-bindingSNPs). For example, the SNP upstream of *STAT1* potentially affecting V-MAF binding occurs in the MDV resistant line 6₁ but is absent from the susceptible lines 7₂ and 15I.

DISCUSSION

Considering that the interferon pathways play a key role in immune responses, not only to viral infections, but also bacterial and parasitic infections, it is plausible that variations in the genes involved may contribute to differences in resistance and susceptibility. Here, we analysed

genome sequence data from eight chicken research lines and identified variants from interferon pathway genes. The impact of these SNPs was considered with regard to their potential functional effects and distribution amongst the lines. This process resulted in a total of 51 potentially relevant SNPs being identified. When considered in terms of the impact of the SNP and in the context of specific virus response, 78 instances could be related to resistance and 70 instances could be related to susceptibility. Table 7 highlights 19 SNPs of interest that correlate with resistance or susceptibility to at least two of the viruses within the scope of this study. Furthermore, prediction tools identified five of these 19 SNPs as having a significant impact on protein function or mRNA folding.

It is possible to hypothesise the mechanism by which some of these SNPs may act given their effect. Missense SNPs result in an amino acid change in the polypeptide

TABLE 5 SNPs of interest predicted to potentially perturb mRNA secondary structure

SNP genome coordinate	mRNA affected	Line(s) in which SNP is present	Alleles (ref/SNP)	mRNA base change	RNAseq P-value	SNPfold P-value	Region affected
Chr1:106613656	<i>IFNAR1</i>	7 ₂	C/T	C/U	0.4067	0.0078	Exonic
Chr1:106622867	<i>IFNAR1</i>	N, 0 ^a	C/T	C/U	0.6077	#0.016	Exonic
Chr1:106623894	<i>IFNAR1</i>	15I, 6 ₁ , 7 ₂ , C, N, 0, W1	G/A	G/A	0.0446	0.4078	Exonic
Chr1:106625990	<i>IFNAR1</i>	0 ^a	C/T	C/U	0.5257	0.0808	Exonic
Chr1:106631527	<i>IFNAR1</i>	P2a	G/A	G/A	0.8683	0.0086	3'UTR
Chr1:106632281	<i>IFNAR1</i>	15I, 6 ₁ , 7 ₂ , C, N, P2a ^a , 0 WL	T/C	U/C	0.9474	0.0093	3'UTR
Chr1:106589970	<i>IFNAR2</i>	15I, C, P2a ^a , 0 ^a , W1	G/T	G/U	0.424	0.046	Exonic
Chr1:106592036	<i>IFNAR2</i>	15I, C, 0, W1	T/G	U/G	0.5046	0.0275	3'UTR
Chr3:54913306	<i>IFNGR1</i>	15I, 6 ₁ , 7 ₂ , C, 0, W1	T/C	U/C	0.084	0.3036	Exonic
Chr1:106652516	<i>IFNGR2</i>	N, 0 ^a	G/A	G/A	0.9103	0.0198	3'UTR
Chr1:106653379	<i>IFNGR2</i>	N, 0 ^a	C/T	C/U	0.1912	0.0473	3'UTR
Chr23:5823279	<i>IFNLR1</i>	N ^a	T/C	A/G	0.0811	0.3326	Exonic
Chr1:106604195	<i>IL10RB</i>	7 ₂ ^a , C ^a	T/G	U/G	0.045	0.2181	3'UTR
Chr1:106604577	<i>IL10RB</i>	15I, 0 ^a , W1	G/A	G/A	0.0565	0.3225	3'UTR
ChrZ:34282637	<i>IFNK</i>	6 ₁ , 7 ₂ , N, P2a, 0, W1	C/T	G/A	#0.027	0.0968	3'UTR
Chr8:28553812	<i>JAK1</i>	15I, 7 ₂ , C, N ^a , P2a ^a	G/A	C/U	0.0978	0.591	Exonic
Chr8:28559109	<i>JAK1</i>	C	A/G	U/C	0.056	0.4255	Exonic
Chr8:28564450	<i>JAK1</i>	15I, 6 ₁ , 7 ₂ , C, N ^a , P2a ^a , W1	T/G	A/C	0.0972	0.4388	Exonic
Chr8:28566979	<i>JAK1</i>	N ^a , P2a ^a , 0	A/G	U/C	0.1547	0.0105	Exonic
Chr8:28567177	<i>JAK1</i>	P2a ^a	T/C	A/G	0.7025	0.0056	Exonic
Chr8:28576480	<i>JAK1</i>	N ^a , P2a ^a , 0	C/T	G/A	0.051	0.3395	Exonic

Allele changes are listed for the forward strand; *IFNLR1*, *IFNK* and *JAK1* are found on the reverse strand so the mRNA base changes observed in these genes reflect that of the template strand.

Scores in bold can be considered significant ($p > 0.1$), # – scores still significant after Bonferroni correction.

^aSNP occurs in only one allele and is therefore heterozygous.

TABLE 6 SNPs of interest predicted to affect TF or miRNA binding sites

SNP genome coordinate	Gene/mRNA affected	Line(s) in which SNP is present	Alleles (ref/SNP)	mRNA base change	Comments
Chr1:106631586	<i>IFNAR1</i>	0 ^a	C/T	C/U	Identified as gga-miR-1627-3p binding site
Chr1:106632281	<i>IFNAR1</i>	15I, 6 ₁ , 7 ₂ , C, N, P2a ^a , 0, W1	T/C	U/C	Identified as gga-miR-196-1-3p binding site
Chr7:7933106	<i>STAT1</i>	6 ₁ , C, P2a ^a , 0 ^a	C/T	N/A	Identified as V-MAF TF binding site

^aSNP occurs in only one allele and is therefore heterozygous.

sequence and this may have consequences on the final protein structure, possibly affecting features such as substrate binding efficiency, protein–protein complex interfaces or catalytic activity. The SNP at position Chr30:182787 where histidine is replaced by proline (H215P) lies in a predicted band 4.1 (B41) domain of TYK2. This SNP is present in the IBV- and MDV-susceptible line 7₂ and the IBDV-susceptible line W1 and may be a contributing factor to these phenotypes. For example, the change from the positively charged, polar histidine to the neutral, non-polar proline may result in a conformational change that negatively impacts the process of phosphorylation of the STAT proteins, resulting in reduced overall upregulation of ISGs. This change could potentially explain why line 7₂ developed

more progressive MD compared with lines 15L, 6₂ and N (Burgess et al., 2001), or why greater mortality was seen in line W1 chickens infected with IBDV compared with various other lines (Bumstead et al., 1993).

The effect of a SNP on protein function can be executed not just by alteration of the amino acid but also by affecting the mRNA secondary structure. Exonic (both synonymous and missense), 5'UTR and 3'UTR SNPs may affect a transcript's secondary structure by altering base pairing within the mRNA. Such structural alteration may affect mRNA stability. Altered secondary structure around the 5'UTR region may impact translation initiation efficiency by affecting the ability of the ribosome to bind to the mRNA, whereas secondary structure changes in the 3'UTR region may affect

TABLE 7 SNPs of interest correlating with resistance or susceptibility against more than one virus

SNP genome coordinate	Gene/mRNA affected	Phenotype (resistant/susceptible)	Virus affected (line with phenotype)	Comments
Chr1:106603837	<i>IL10RB</i>	Resistant	AIV (0) IBDV (7 ₂ , 15L) MDV (N, 6 ₁)	Significant SIFT score of 0.04 Heterozygous in line 0
Chr1:106621899	<i>IFNAR1</i>	Resistant	IBV (N) MDV (N)	Significant SNAP2 score of 18
Chr1:106622867	<i>IFNAR1</i>	Resistant	AIV (0) IBDV (0) IBV (N) MDV (N)	SNPfold P-value = 0.016 – significant after Bonferroni correction for $p < 0.1$ Heterozygous in line 0
Chr1:106625990	<i>IFNAR1</i>	Resistant	AIV (0) IBDV (0)	SNPfold P-value = 0.0808 – not significant after applying Bonferroni correction Heterozygous in line 0
Chr1:106652516	<i>IFNGR2</i>	Resistant	AIV (0) IBDV (0) IBV (N) MDV (N)	SNPfold P-value = 0.0198 – not significant after applying Bonferroni correction Heterozygous in line 0
Chr1:106653379	<i>IFNGR2</i>	Resistant	AIV (0) IBDV (0) IBV (N) MDV (N)	SNPfold P-value = 0.0473 – not significant after applying Bonferroni correction Heterozygous in line 0
Chr23:5823279	<i>IFNLR1</i>	Resistant	IBV (N) MDV (N)	SNPfold P-value = 0.0811 – not significant after applying Bonferroni correction Heterozygous in line N
Chr8:28566979	<i>JAK1</i>	Resistant	IBDV (0) AIV (0) IBV (N)	SNPfold P-value = 0.0105 – not significant after applying Bonferroni correction Heterozygous in line N
Chr8:28576480	<i>JAK1</i>	Resistant	AIV (0) IBDV (0) IBV (N)	RNAseq P-value = 0.051 – not significant after applying Bonferroni correction Heterozygous in line N
ChrZ:7372479	<i>IFNβ</i>	Resistant	IBV (N) MDV (N)	Significant SNAP2 score of 52 Heterozygous in line N
Chr1:106588380	<i>IFNAR2</i>	Susceptible	MDV (15L, 7 ₂) AIV (C)	No impact calculated by functional prediction tools
Chr1:106589970	<i>IFNAR2</i>	Susceptible	MDV (15L, P2a) AIV (C) IBDV (W1)	SNPfold p -value = 0.046 – not significant after applying Bonferroni correction Heterozygous in line P2a
Chr1:106590778	<i>IFNAR2</i>	Susceptible	MDV (15L, P2a) IBDV (W1)	No impact calculated by functional prediction tools Heterozygous in line P2a
Chr1:106591040	<i>IFNAR2</i>	Susceptible	MDV (15L, P2a) AIV (C) IBDV (W1)	No impact calculated by functional prediction tools Heterozygous in line P2a
Chr1:106591349	<i>IFNAR2</i>	Susceptible	MDV (15L, P2a) IBDV (W1)	No impact calculated by functional prediction tools Heterozygous in line P2a
Chr1:106592036	<i>IFNAR2</i>	Susceptible	MDV (15L) IBDV (W1)	SNPfold P-value = 0.0275 – not significant after applying Bonferroni correction
Chr30:182787	<i>TYK2</i>	Susceptible	IBV (7 ₂) MDV (7 ₂) IBDV (W1)	Significant PROVEAN score of -2.732 and significant SNAP2 score of 63 Heterozygous in lines 7 ₂ and W1
Chr7:7923855	<i>STAT1</i>	Susceptible	IBV (15L, 7 ₂) MDV (15L, 7 ₂) IBDV (W1)	Significant PROVEAN score of -4.766 and significant SNAP2 score of 39 Heterozygous in lines 15L, 7 ₂ and W1
Chr8:28564450	<i>JAK1</i>	Susceptible	AIV (C) IBDV (W1)	RNAseq P-value = 0.0972 – not significant after applying Bonferroni correction

the accessibility of miRNAs to binding sites, impacting their ability to silence translation in the mRNA (Haas et al., 2012; Mauger et al., 2019). In this study, an SNP at position Chr1:106652516 was identified that results in a guanine to adenine change in the 3'UTR region of the *IFNGR2* mRNA in the IBDV-resistant line N and was predicted to significantly affect the secondary structure. *IFNGR2* exhibits greater levels of expression after infection with virulent IBDV (vvIBDV; Farhanah et al., 2018), which would indicate the importance of this gene and the type II interferon pathway. It is possible that this SNP in line N may synergise with this upregulation by preventing miRNA binding owing to an altered secondary structure, leading to greater levels of translation and a greater type II pathway response compared with lines lacking this SNP, where miRNA binding and silencing may still occur.

SNPs can also directly affect miRNA binding sites in the 3'UTR of the mRNA by altering the recognition sequence, potentially impacting miRNA binding of the transcript and the resultant silencing effect of this interaction. Using miRDB, it was seen that potential miRNA binding sites for gga-miR-1627-3p and gga-miR-196-1-3p in the 3'UTR of *IFNAR1* were altered by SNPs at positions Chr1:106632281 and Chr1:106631586 respectively. *Gga-miR-1627-3p* was previously found to exhibit greater expression in macrophages with the MHC haplotype B19 compared with those with B2, suggesting that it may play a role in immune response (Irizarry et al., 2017). The SNP affecting the gga-miR-1627-3p binding site only occurs in line 0, which is perceived as resistant to AIV, so it is therefore plausible that this SNP could contribute to this resistant phenotype. Interferons were discovered owing to their ability to inhibit influenza virus (Isaacs & Lindenmann, 1957) and studies using type I interferons as both a pre-treatment and treatment *in vitro* and *in vivo* have identified a reduction in influenza virus replication (Jiang et al., 2011; Müller et al., 1994). Therefore, this SNP may reduce the silencing effect of gga-miR-1627-3p binding, allowing greater *IFNAR2* translation. Increased levels of *IFNAR2* at the cell surface could lead to a more responsive activation of the type I interferon pathway upon influenza infection, resulting in a greater antiviral response.

Another example of a variant affecting a regulatory binding site was observed with the SNP at position Chr7:7933106, just upstream of *STAT1*, where a base substitution is predicted in a V-MAF TF binding site. V-MAF is an oncogene that was first identified in avian musculoaponeurotic fibrosarcoma virus; however, a V-MAF protein homologue (C-MAF) exists in chickens, sharing 99.7% identity. It is likely that the chicken C-MAF will bind this same site, with binding affected in lines 6₁, C, P2a and 0, altering *STAT1* expression in these lines. In humans and mice, C-MAF has been shown to be involved in CD4 and CD8 T-cell subset differentiation, function and interleukin production (Ho et al., 1996;

Imbratta et al., 2020). Furthermore, research has indicated that *C-MAF* expression is induced in a STAT1-dependent fashion in CD4⁺CD25⁻ T cells. It is therefore possible that C-MAF binding at STAT1 could act as a self-upregulating positive feedback loop (Xu et al., 2009). *C-MAF* overexpression has been linked to T-cell lymphoma development in both mice and humans (Morito et al., 2006) and a key outcome in MDV pathogenesis is the development of T-cell lymphomas (Ross, 1999). Interestingly, the SNP within the potential MAF binding site occurs in line 6₁, which is known for its resistance to MDV, but is absent in its susceptible counterpart, line 7₂. Considering C-MAF's involvement in cancer development and potential regulation by STAT1, this SNP may prevent C-MAF binding upstream of *STAT1*. The lack of binding may mitigate carcinogenic effects that may be associated with upregulation of *STAT1*, leading to aberrant *C-MAF* expression after MDV infection, and may explain the reduced pathogenesis and mortality by MDV seen in line 6₁ (Mohd Isa et al., 2020).

Again, limitations on knowledge of TFBS and miRNA targets in chicken have to be considered when these predictions are made. Fifty-two 3'UTR variants and 101 upstream variants were identified in this study. Given the incomplete knowledge of miRNA and TF binding sites it is possible that more of these variants than are currently identified affect binding sites and potentially contribute towards resistant and susceptible phenotypes.

In the current study, functional variants were mainly predicted from exonic, upstream and UTR regions. However, such an approach fails to consider the potential impact of 81.4% of the SNPs (1432 intronic SNPs) in the studied regions. SNPs in intronic regions could potentially impact splice sites, giving a different mRNA splice variant or affecting splicing efficiency, resulting in varying protein isoforms or expression levels which may affect phenotypes such as disease resistance (Cooper, 2010). SNPs outside of genes can also affect expression by impacting distal regulatory regions; TF binding at these sites may exist spatially close to promoter regions and influence transcription initiation despite being distant in terms of genomic coordinate (Wang et al., 2019).

Also requiring further research is the impact of heterozygous SNPs. As shown in Table S1, the vast majority of identified SNPs (78–95%) are homozygous. However, the smaller proportion of heterozygous SNPs may represent the accumulation of new mutations over time, reflecting selective advantage (as being homozygous may have detrimental consequence). Allele-specific expression of genes is the result of differential expression occurring between the two parental alleles of one gene. This has implications in the context of this study as several SNPs were found to occur in only one allele and therefore may be upregulated or downregulated compared with the wild-type allele. For example, the aforementioned missense SNP at position Chr30:182787 which results in a H215P substitution in

the *TYK2* gene is present in the MDV-susceptible line 7₂ and absent from resistant line 6₁. The JAK/STAT pathway has been previously found to contain SNPs exhibiting allele specific expression in response to MDV (Perumbakkam et al., 2013). It is possible that SNPs occurring in alleles that experience greater expression upon infection may have a significant effect on resistant or susceptible phenotypes. Considering the findings above, MDV infection may result in greater H215P-TYK2 expression in line 7₂. This amino acid substitution was predicted to be deleterious by PROVEAN and to affect protein function by SNAP2. The increased levels of this affected TYK2 may result in less ISG induction, leading to greater viral replication and more severe infection.

The ultimate goal will be to breed for a more resilient bird that can tolerate and recover from challenges without having lost productivity at the end of its rearing or laying period. Previous work indicated that selection for production traits has resulted in weakened immunity in commercial lines, but also suggested that selection for immunity was not detrimental to growth (Van der Most et al., 2011). Furthermore, a Genome Wide Association Study (GWAS) of indigenous African chickens found no correlation between production traits and immunity traits (Psifidi et al., 2016). Therefore, breeding schemes to introduce variants that result in higher innate immune responsiveness may not negatively impact production. Another economic advantage with using breeding schemes to introduce resilience is that this process does not require gene editing and will therefore not result in any consumer concerns over genetic modification; however, it will be a much slower process.

To enable breeding for more resilient birds we have to define the mechanisms of resistance, as a variant conferring resistance to one virus could lead to susceptibility against another. The observation that some SNPs appear to confer resistance in one line and susceptibility in another may be due to the level of interferon response. For example, some viruses may require a more robust response, which a particular SNP might yield, whilst with other viruses this more robust response may give rise to the observed 'susceptibility'. For some viruses, a particular IFN may be detrimental and for others it will not because they have an ability to evade that response or to downregulate and block the induction of the IFN response. Hence breeding for resistance to virus A does not imply that it will also impart resistance to virus B. Our study indicated that the SNP at position Chr1:106603910 results in an R318K amino acid substitution in IL10Rβ. This SNP may contribute to IBDV resistance in line C whilst it may also contribute to MDV susceptibility in line 7₂. This SNP may enhance the type III interferon pathway's regulation of apoptosis (Li et al., 2008), which could therefore reduce bursal damage in IBDV-infected birds owing to aberrant apoptosis

(Cubas-Gaona et al., 2018), but may contribute to the development of lymphoma in MDV infected birds (Xu et al., 2011). This example indicates why unravelling the mechanism for resistance is of importance, as breeding for variants at such loci may introduce susceptibility to other viruses outside the scope of this study.

It is only possible to postulate the mechanism by which the SNPs identified in this study function. Further research will be required to validate the effects of these variants on phenotype. It is interesting to note that none of the examined genes are located on chromosome 16, so any role in resistance/susceptibility is not due to co-selection with the MHC during inbreeding of these lines. Furthermore, the current research is reductionist in considering only one SNP at a time; resistance and susceptibility are polygenic traits where effects may only be measurable in the presence of multiple SNPs working in concert with each other. The validation of candidate SNPs predicted in this work could be achieved by introducing variants *in vitro*, *in ovo* and *in vivo* and observing changes in antiviral responses, while also ensuring that responses to other pathogens are not negatively impacted. Our study further highlights the importance of the maintenance of rare, inbred chicken lines, given that they may already contain loci for resistance and once these have been identified a low-density SNP array could be developed to easily inform breeding schemes.

CONCLUSIONS

In the interferon pathway genes studied, 1760 SNPs were identified, of which 36 change an amino acid, 21 were predicted to affect mRNA secondary structure and three to affect TF/miRNA binding sites. Considering the research line phenotypes and the distribution of these SNPs, 78 instances could be related to resistance and 70 could be related to susceptibility to particular viruses. Further research will be required to validate these SNPs and examine the mechanisms involved in resistance or susceptibility. The current study has identified genomic variants in genes in the interferon pathways of different chicken research lines. This pathway is an integral component in innate immunity and can also contribute to the orchestration and development of the adaptive immune response. Not only will these findings help explain host innate immune responses and disease mechanisms used by these viruses and other micro-organisms, but it may also assist in the identification of targets for selective breeding, drug design or vaccine improvement. Eventually, this will pave the way for breeding programmes that result in more robust commercial chicken lines.

ACKNOWLEDGEMENT

For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any

Author Accepted Manuscript version arising from this submission.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

DATA AVAILABILITY STATEMENT

Genome sequence data are available in the European Nucleotide Archive under accession number PRJEB51681. All SNP information is provided in Table S4.

ORCID

Jacqueline Smith  <https://orcid.org/0000-0002-2813-7872>

REFERENCES

- Andreacos, E., Zaroni, I. & Galani, I.E. (2019) Lambda interferons come to light: dual function cytokines mediating antiviral immunity and damage control. *Current Opinion in Immunology*, 56, 67–75.
- Asfor, A.S., Nazki, S., Reddy, V.R.A.P., Campbell, E., Dulwich, K.L., Giotis, E.S. et al. (2021) Transcriptomic analysis of inbred chicken lines reveals infectious bursal disease severity is associated with greater bursal inflammation in vivo and more rapid induction of pro-inflammatory responses in primary bursal cells stimulated ex vivo. *Viruses*, 13(5), 933.
- Bacon, L.D., Hunt, H.D. & Cheng, H.H. (2000) A review of the development of chicken lines to resolve genes determining resistance to diseases. *Poultry Science*, 79(8), 1082–1093.
- Boodhoo, N., Gurung, A., Sharif, S. & Behboudi, S. (2016) Marek's disease in chickens: a review with focus on immunology. *Veterinary Research*, 47(1), 119.
- Bumstead, N., Huggins, M.B. & Cook, J.K. (1989) Genetic differences in susceptibility to a mixture of avian infectious bronchitis virus and *Escherichia coli*. *British Poultry Science*, 30(1), 39–48.
- Bumstead, N., Reece, R.L. & Cook, J.K. (1993) Genetic differences in susceptibility of chicken lines to infection with infectious bursal disease virus. *Poultry Science*, 72(3), 403–410.
- Burgess, S.C., Basaran, B.H. & Davison, T.F. (2001) Resistance to Marek's disease herpesvirus-induced lymphoma is multiphasic and dependent on host genotype. *Veterinary Pathology*, 38(2), 129–142.
- Chen, X., Ge, K., Wang, M., Zhang, C. & Geng, Z. (2017) Integrative analysis of the Pekin duck (*Anas anas*) MicroRNAome during feather follicle development. *BMC Developmental Biology*, 17(1), 12.
- Chen, Y. & Wang, X. (2020) miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Research*, 48(D1), D127–D131.
- Choi, Y. & Chan, A.P. (2015) PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, 31(16), 2745–2747.
- Cole, R.K. (1968) Studies on genetic resistance to Marek's disease. *Avian Diseases*, 12(1), 9–28.
- Cook, J., Otsuki, K., Huggins, M. & Bumstead, N. (1990) Investigations into resistance of chicken lines to infection with infectious bronchitis virus. *Advances in Experimental Medicine and Biology*, 276, 491–496.
- Cooper, D.N. (2010) Functional intronic polymorphisms: buried treasure awaiting discovery within our genes. *Human Genomics*, 4(5), 284–288.
- Cubas-Gaona, L.L., Diaz-Beneitez, E., Ciscar, M., Rodríguez, J.F. & Rodríguez, D. (2018) Exacerbated apoptosis of cells infected with infectious bursal disease virus upon exposure to interferon alpha. *Journal of Virology*, 92(11), e00364–e00318.
- Dey, S., Pathak, D.C., Ramamurthy, N., Maity, H.K. & Chellappa, M.M. (2019) Infectious bursal disease virus in chickens: prevalence, impact, and management strategies. *Veterinary Medicine*, 10, 85–97.
- Farhanah, M.I., Yasmin, A.R., Mat Isa, N., Hair-Bejo, M., Ideris, A., Powers, C. et al. (2018) Bursal transcriptome profiling of different inbred chicken lines reveals key differentially expressed genes at 3 days post-infection with very virulent infectious bursal disease virus. *Journal of General Virology*, 99(1), 21–35.
- Garcia-Morales, C., Nandi, S., Zhao, D., Sauter, K.A., Vervelde, L., McBride, D. et al. (2015) Cell-autonomous sex differences in gene expression in chicken bone marrow-derived macrophages. *Journal of Immunology*, 194(5), 2338–2344.
- GATK. Germline short variant discovery (SNPs + Indels). <https://gatk.broadinstitute.org/hc/en-us/articles/360035535932-Germline-short-variant-discovery-SNPs-Indels->. Accessed 16 October 2020.
- GATK. Variant Quality Score Recalibration (VQSR). <https://gatk.broadinstitute.org/hc/en-us/articles/360035531612-Variant-Quality-Score-Recalibration-VQSR->. Accessed 16 October 2020.
- Gheyas, A.A., Boschiero, C., Eory, L., Ralph, H., Kuo, R., Woolliams, J.A. et al. (2015) Functional classification of 15 million SNPs detected from diverse chicken populations. *DNA Research*, 22(3), 205–217.
- Haas, U., Sczakiel, G. & Laufer, S.D. (2012) MicroRNA-mediated regulation of gene expression is affected by disease-associated SNPs within the 3'-UTR via altered RNA structure. *RNA Biology*, 9(6), 924–937.
- Halvorsen, M., Martin, J.S., Broadaway, S. & Laederach, A. (2010) Disease-associated mutations that alter the RNA structural ensemble. *PLoS Genetics*, 6(8), e1001074.
- Hecht, M., Bromberg, Y. & Rost, B. (2015) Better prediction of functional effects for sequence variants. *BMC Genomics*, 16(Suppl 8), S1.
- Ho, I.C., Hodge, M.R., Rooney, J.W. & Glimcher, L.H. (1996) The proto-oncogene *c-maf* is responsible for tissue-specific expression of interleukin-4. *Cell*, 85(7), 973–983.
- Imbratta, C., Hussein, H., Andris, F. & Verdeil, G. (2020) C-MAF, a Swiss Army knife for tolerance in lymphocytes. *Frontiers in Immunology*, 11, 206.
- Irizarry, K.J.L., Downs, E., Bryden, R., Clark, J., Griggs, L., Kopulos, R. et al. (2017) RNA sequencing demonstrates large-scale temporal dysregulation of gene expression in stimulated macrophages derived from MHC-defined chicken haplotypes. *PLoS One*, 12(8), e0179391.
- Isaacs, A. & Lindenmann, J. (1957) Virus interference. I. the interferon. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 147(927), 258–267.
- Jiang, H., Yang, H. & Kapczynski, D.R. (2011) Chicken interferon alpha pretreatment reduces virus replication of pandemic H1N1 and H5N9 avian influenza viruses in lung cell cultures from different avian species. *Virology Journal*, 8, 447.
- Kaiser, P. (2010) Advances in avian immunology--prospects for disease control: a review. *Avian Pathology*, 39(5), 309–324.
- Kaiser, P., Howell, J., Fife, M., Sadeyen, J.R., Salmon, N., Rothwell, L. et al. (2008) Integrated immunogenomics in the chicken: deciphering the immune response to identify disease resistance genes. *Developmental Biology*, 132, 57–66.
- Kaufman, J. (2018) Generalists and specialists: a new view of how MHC class I molecules fight infectious pathogens. *Trends in Immunology*, 39(5), 367–379.
- Kel, A.E., Gössling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O.V., & Wingender, E. (2001) MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Research*, 31(13), 3576–3579.

- Kranis, A., Gheyas, A.A., Boschiero, C., Turner, F., Yu, L., Smith, S. et al. (2013) Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics*, 14(14), 59.
- Lee, A.J. & Ashkar, A.A. (2018) The dual nature of type I and type II interferons. *Frontiers in Immunology*, 9, 2061.
- Legnardi, M., Tucciarone, C.M., Franzo, G. & Cecchinato, M. (2020) Infectious bronchitis virus evolution, diagnosis and control. *Vet Sci.*, 7(2), 79.
- Letunic, I., Khedkar, S. & Bork, P. (2021) SMART: recent updates, new developments and status in 2020. *Nucleic Acids Research*, 49(D1), D458–D460.
- Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, 25(14), 1754–1760.
- Li, W., Lewis-Antes, A., Huang, J., Balan, M. & Kotenko, S.V. (2008) Regulation of apoptosis by type III interferons. *Cell Proliferation*, 41(6), 960–979.
- Lycett, S.J., Duchatel, F. & Digard, P. (2019) A brief history of bird flu. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 374(1775), 20180257.
- Mauger, D.M., Cabral, B.J., Presnyak, V., Su, S.V., Reid, D.W., Goodman, B. et al. (2019) mRNA structure regulates protein expression through changes in functional half-life. *Proceedings of the National Academy of Sciences of the United States of America*, 116(48), 24075–24083.
- McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A. et al. (2016) The Ensembl variant effect predictor. *Genome Biology*, 17(1), 122.
- Mohd Isa, F., Ahmed Al-Haj, N., Mat Isa, N., Ideris, A., Powers, C., Oladapo, O. et al. (2020) Differential expression of immune-related genes in the bursa of Fabricius of two inbred chicken lines following infection with very virulent infectious bursal disease virus. *Comparative Immunology, Microbiology and Infectious Diseases*, 68, 101399.
- Morito, N., Yoh, K., Fujioka, Y., Nakano, T., Shimohata, H., Hashimoto, Y. et al. (2006) Overexpression of c-Maf contributes to T-cell lymphoma in both mice and human. *Cancer Research*, 66(2), 812–819.
- Müller, U., Steinhoff, U., Reis, L.F., Hemmi, S., Pavlovic, J., Zinkernagel, R.M. et al. (1994) Functional role of type I and type II interferons in antiviral defense. *Science*, 264(5167), 1918–1921.
- Nakamura, K., Cook, J.K., Otsuki, K., Huggins, M.B. & Frazier, J.A. (1991) Comparative study of respiratory lesions in two chicken lines of different susceptibility infected with infectious bronchitis virus: histology, ultrastructure and immunohistochemistry. *Avian Pathology*, 20(2), 241–257.
- Nishizawa, M., Kataoka, K., Goto, N., Fujiwara, K.T. & Kawai, S. (1989) V-maf, a viral oncogene that encodes a 'leucine zipper' motif. *Proceedings of the National Academy of Sciences of the United States of America*, 86(20), 7711–7715.
- Otsuki, K., Huggins, M.B. & Cook, J.K. (1990) Comparison of the susceptibility to avian infectious bronchitis virus infection of two inbred lines of white leghorn chickens. *Avian Pathology*, 19(3), 467–475.
- Perumbakkam, S., Muir, W.M., Black-Pyrkosz, A., Okimoto, R. & Cheng, H.H. (2013) Comparison and contrast of genes and biological pathways responding to Marek's disease virus infection using allele-specific expression and differential expression in broiler and layer chickens. *BMC Genomics*, 14, 64.
- Psfidi, A., Banos, G., Matika, O., Desta, T.T., Bettridge, J., Hume, D.A. et al. (2016) Genome-wide association studies of immune, disease and production traits in indigenous chicken ecotypes. *Genetics, Selection, Evolution*, 48(1), 74.
- Reuter, A., Soubies, S., Härtle, S., Schusser, B., Kaspers, B., Staeheli, P. et al. (2014) Antiviral activity of lambda interferon in chickens. *Journal of Virology*, 88(5), 2835–2843.
- Ross, N.L. (1999) T-cell transformation by Marek's disease virus. *Trends in Microbiology*, 7(1), 22–29.
- Ruiz-Hernandez, R., Mwangi, W., Peroval, M., Sadeyen, J.R., Ascough, S., Balkissoon, D. et al. (2016) Host genetics determine susceptibility to avian influenza infection and transmission dynamics. *Scientific Reports*, 6, 26787.
- Sabarinathan R, Tafer H, Seemann SE, Hofacker IL, Stadler PF, Gorodkin J. The RNAsnp web server: predicting SNP effects on local RNA secondary structure. *Nucleic Acids Research*. 2013;41(Web Server issue):W475-9.
- Santhakumar, D., Iqbal, M., Nair, V. & Munir, M. (2017a) Chicken IFN kappa: a novel cytokine with antiviral activities. *Scientific Reports*, 7(1), 2719.
- Santhakumar, D., Rubbenstroth, D., Martinez-Sobrido, L. & Munir, M. (2017b) Avian interferons and their antiviral effectors. *Frontiers in Immunology*, 8, 49.
- Sim, N.L., Kumar, P., Hu, J., Henikoff, S., Schneider, G. & Ng, P.C. (2012) SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research*, 40(Web Server issue), W452-7.
- Smith, J., Sadeyen, J.R., Butter, C., Kaiser, P. & Burt, D.W. (2015b) Analysis of the early immune response to infection by infectious bursal disease virus in chickens differing in their resistance to the disease. *Journal of Virology*, 89(5), 2469–2482.
- Smith, J., Sadeyen, J.R., Cavanagh, D., Kaiser, P. & Burt, D.W. (2015a) The early immune response to infection of chickens with infectious bronchitis virus (IBV) in susceptible and resistant birds. *BMC Veterinary Research*, 11, 256.
- Smith, J., Sadeyen, J.R., Paton, I.R., Hocking, P.M., Salmon, N., Fife, M. et al. (2011) Systems analysis of immune responses in Marek's disease virus-infected chickens identifies a gene involved in susceptibility and highlights a possible novel pathogenicity mechanism. *Journal of Virology*, 85(21), 11146–11158.
- Stone, H.A. (1975) *Use of highly inbred chickens in research*. Washington, DC: USDA-ARS Technical Bulletin No. 1514.
- Swaggerty, C.L., Pevzner, I.Y., He, H., Genovese, K.J., Nisbet, D.J., Kaiser, P. et al. (2009) Selection of broilers with improved innate immune responsiveness to reduce on-farm infection by foodborne pathogens. *Foodborne Pathogens and Disease*, 6(7), 777–783.
- Van der Most, P.J., de Jong, B., Parmentier, H.K. & Verhulst, S. (2011) Trade-off between growth and immune function: a meta-analysis of selection experiments. *Functional Ecology*, 25(1), 74–80.
- Wang, Q., Jia, Y., Wang, Y., Jiang, Z., Zhou, X., Zhang, Z. et al. (2019) Evolution of cis- and trans-regulatory divergence in the chicken genome between two contrasting breeds analyzed using three tissue types at one-day-old. *BMC Genomics*, 20(1), 933.
- Xu, J., Yang, Y., Qiu, G., Lal, G., Wu, Z., Levy, D.E. et al. (2009) C-Maf regulates IL-10 expression during Th17 polarization. *Journal of Immunology*, 182(10), 6226–6236.
- Xu, S., Xue, C., Li, J., Bi, Y. & Cao, Y. (2011) Marek's disease virus type 1 microRNA miR-M3 suppresses cisplatin-induced apoptosis by targeting Smad2 of the transforming growth factor beta signal pathway. *Journal of Virology*, 85(1), 276–285.
- Xu, Z., Nie, Q. & Zhang, X. (2013) Overview of genomic insights into chicken growth traits based on genome-wide association study and microRNA regulation. *Current Genomics*, 14(2), 137–146.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Mountford, J., Gheyas, A., Vervelde, L. & Smith, J. (2022) Genetic variation in chicken interferon signalling pathway genes in research lines showing differential viral resistance. *Animal Genetics*, 53, 640–656. Available from: <https://doi.org/10.1111/age.13233>