

SNP prioritization in targeted sequencing data associated with humoral immune responses in chicken

Tomasz Suchocki ^{*,†}, Bartosz Czech,^{*} Aleksandra Dunislawska,[‡] Anna Slawinska,[‡] Natalia Derebecka,[§] Joanna Wesoly,[§] Maria Siwek,^{‡,1} and Joanna Szyda ^{*,†}

^{*}*Biostatistics Group, Department of Genetics, Wrocław University of Environmental and Life Sciences, Wrocław, Poland;* [†]*National Research Institute of Animal Production, Balice, Poland;* [‡]*Department of Animal Biotechnology and Genetics, UTP University of Science and Technology, Bydgoszcz 85-084, Poland; and* [§]*Laboratory of High Throughput Technologies, Faculty of Biology, Adam Mickiewicz University, Poznan, Poland*

ABSTRACT Our study aimed to identify single nucleotide polymorphisms (SNPs) with a significant impact on the innate immunity represented by antibody response against lipopolysaccharide (LPS) and lipoteichoic acid (LTA) and the adaptive immune response represented toward keyhole limpet hemocyanin (KLH) using the SNP prioritization method. Data set consisted of 288 F2 experimental individuals, created by crossing Green-legged Partridge-like and White Leghorn. The analyzed SNPs were located within 24 short genomic regions of GGA1, GGA2, GGA3, GGA4, GGA9, GGA10, GGA14, GGA18, and GGZ, pre-targeted based on literature references and database information. For the specific antibody response toward KLH at d 0 the most highly prioritized SNP for additive and dominance effects were located on GGA2 in the 3'UTR of MYD88. For the response at d 7, the most highly prioritized SNP pointed at the 3'UTR of MYD88, but potential causal additive variants were located within ADIPOQ and one

in PROCR. The highest priority for additive and dominance effects in the antibody response toward lipoteichoic acid at d 0 was attributed to the same SNP, located on GGA2 in the 3'UTR region of MYD88. Two SNPs among the top-10 for additive effect were located in the exon of NOCT. SNPs selected for their additive effect on antibody response toward lipopolysaccharide at d 0 marked 3 genes – NOCT, MYD88, and SNX8, while SNPs selected for their dominance effect marked – NOCT, ADIPOQ, and MYD88. The top-10 variants identified in our study were located in different functional parts of the genome. In the context of causality three groups can be distinguished: variants located in exons of protein coding genes (ADIPOQ, NOCT, PROCR, SNX8), variants within exons of non-coding transcripts, and variants located in genes' UTR regions. Variants from the first group influence protein structure and variants from both latter groups' exhibit regulatory roles on DNA (UTR) or RNA (lncRNA).

Key words: adaptive immunity, amplicon, innate immunity, mixed model

2021 Poultry Science 100:101433
<https://doi.org/10.1016/j.psj.2021.101433>

INTRODUCTION

Immune response is a complex quantitative trait and the knowledge on its genetic background in chickens is still limited (Biscarini et al., 2010; Berghof et al., 2018). The overall immunity is composed of innate and acquired parts. The innate immunity is the first line of the host organism defence against pathogens. The humoral part of the innate immunity is composed of several components, among them are the Natural

Antibodies (NABs) (Shishido et al., 2012). The NABs are defined as immunoglobulins, which are constitutively generated in the host in the absence of immunization or pathological condition (Baumgarth et al., 2005; Maddur et al., 2020). The major features of NABs are low antibody affinity, high avidity, and polyreactivity against antigens (Ochsenbein and Zinkernagel, 2000). Therefore, they are very effective as a first line of defense against pathogen invasion before the acquired antibodies are generated (Siwek and Knol, 2005). Some NABs have the ability to recognize the evolutionary conserved epitopes, which was used in this study to differentiate between the NABs populations. In the current study, the innate immunity was represented by NABs against lipopolysaccharide (LPS) and lipoteichoic acid (LTA). LPS is a molecule present on the outer membrane of the Gram-negative bacteria whereas LTA is an ingredient of

© 2021 The Authors. Published by Elsevier Inc. on behalf of Poultry Science Association Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Received January 13, 2021.

Accepted August 15, 2021.

¹Corresponding author: siwek@utp.edu.pl

the cell wall of Gram-positive bacteria (Siwek et al., 2006). The action of the innate immunity alone is in some cases insufficient to defend the host against pathogen invasion. Therefore, the adaptive immunity is required for successful elimination of the pathogen. The acquired immunity involves humoral and cellular immune responses (Erf, 2004). The adaptive immune response was represented by a specific antibody response towards keyhole limpet hemocyanin (KLH), which was determined in individuals specifically immunized with KLH antigen. This is a specific antigen obtained from the hemolymph of a certain sea mollusc species. The important characteristic of the KLH antigen is that birds never encountered this antigen before (Siwek et al., 2010).

The genetic background of the immune responses toward: LPS, LTA and KLH has been extensively analyzed in a unique experimental population composed of a cross between White Leghorn and Green-legged Partridge (WL × GP) (Siwek et al., 2010). These analyses covered all the steps of the classical quantitative trait *loci* (QTL) approach. The final step of each QTL analysis is a candidate gene selection. Identification of the candidate genes was performed in silico and the selected genes were verified in an association study based on a custom Illumina SNP genotyping panel (Siwek et al., 2015). One of the conclusions from this study was that the most significant single-nucleotide polymorphisms (SNPs) for immune responses toward KLH and LTA were located outside of the QTL regions originally proposed by the linkage analysis. Therefore, in the current study a different approach was undertaken. A selection of the candidate genes was based not only on the location near the QTL region, but also outside of the QTL region if the biological function matched the phenotypic trait of interest.

However, from the statistical perspective, performing an association study on tightly linked SNPs located within targeted regions poses problems in selecting specific polymorphisms (and genes) that most potently influence the variation of immune response measures. In such scenario, one can no longer rely on *P*-values resulting from testing SNP effects estimated in an association model. It is due to the location of the SNPs in genomic regions already known to be associated with trait variation, and because there is a very high intercorrelation between SNPs attributed to very high linkage disequilibrium. Causal mutations, due to their significant impact on the phenotypes, typically do not exhibit high genotype variation, and therefore they have low frequency of the negative allele in a population. As a consequence, it is more difficult to estimate a significant association for those SNPs as compared to non-causal SNPs which typically have higher frequency of the minor allele (Auer and Lettre, 2015). Testing of the SNP significance in the targeted regions sequenced with the high throughput technologies is also biased by the sequencing quality. In a typical NGS application, the quality of the sequencing is very nonuniform across regions. It impacts the number of successfully detected genotypes and their accuracy,

which may further be a confounding factor in significance testing. These drawbacks have been described in the literature, especially in the context of the human genomic data (Cantor et al., 2010). In this study, we propose a SNP prioritization approach based on the utilization of the variant sequencing information available in the variant calling format (VCF) file.

The goal of the current study was to investigate a new method of SNP prioritization and to identify genes, and possibly SNPs, located within those genes, that are potential candidates of being causal for the innate and specific immune response in chicken. In the context of the quantitative phenotypes, the causality is defined as a significant impact on a trait variation (Auer and Lettre, 2015).

MATERIALS AND METHODS

Animals and Phenotypes

Experimental dataset consisted of 288 F2 individuals, created by crossing Green-legged Partridge and White Leghorn. The population was described in details by Siwek et al. (2010). Birds were kept on a floor system on a farm at the University of Life Sciences in Lublin. The entire population was vaccinated according to the routine vaccination schedule, which incorporated a vaccine against *Salmonella*, Gumboro disease, bronchitis, bursa of Fabricius disease, and encephalomyelitis. Total antibody responses to KLH were measured in individual plasma samples obtained at 7 d after s.c. immunization with 1 mg of KLH (Cal Biochem-Novabiochem Co., La Jolla, CA) in 1 mL of PBS (pH 7.2) at 12 wk of age. Antibody titers to KLH of all birds were measured by an indirect ELISA as described by Sijben et al. (2000). The assay for 2 homotopes LPS and LTA was performed according to the methodology reported by Siwek et al. (2006). The titration plates were coated with 4 $\mu\text{g}/\text{mL}$ of LPS from *Escherichia coli* (Sigma-Aldrich GmbH) and 10 $\mu\text{g}/\text{mL}$ of LTA from *Staphylococcus aureus* (Sigma-Aldrich GmbH, St. Louis, MO). Serum samples were applied on the coated plates. Binding of Nab to LPS and LTA antigens was detected by 1:20,000 diluted rabbit anti-chicken antibody (IgG_{H+L}; Nordic, Tilburg, the Netherlands) conjugated to peroxidase (PO; RAch/IgG_{H+L}/PO). Substrate for enzymatic color reaction was 0.05% H₂O₂ with addition of tetramethylbenzidine. Absorbances were measured with a multiscan (Labsystems, Helsinki, Finland) at 450 nm. All of the phenotypes were expressed by titers as the log₂ values of the highest dilution giving a positive reaction as described by Siwek et al. (2003) for KLH and by Siwek et al. (2006) for LTA and LPS. KLH, LTA and LPS phenotypes were measured at d 0 without immunization (KLHd0, LTAd0, and LPSd0, respectively), to estimate the level of natural antibodies (innate humoral immune response). Additionally, the individuals were immunized with KLH and the second measurement of antibodies against KLH was taken at d 7 (KLHd7) to estimate the level of specific antibodies (adaptive

humoral immune response). Descriptive statistics of the phenotypic data were summarized in [Table 1](#).

SNP Genotyping

Selection of the candidate genes was performed based on literature references and data base information: the InnateDB (www.innatedb.com), the KEGG (www.kegg.jp), and the Reactome (reactome.org). All gene functions were assessed based on the NCBI Gene (www.ncbi.nlm.nih.gov) and the AmiGO (amigo.geneontology.org) databases to classify genes as immune-related. The primers for Long-Range PCR were designed according to the NEB (New England Biolabs, Ipswich, MA) guidelines for reactions using Q5 High-Fidelity DNA Polymerase. The list of the selected genes and their primers is presented in [Table 2](#). Primer3 was used for the primer design (bioinfo.ut.ee/primer3-0.4.0/). Amplicons' quality was verified based on band integrity (1% agarose gel) and the concentration (QUBIT4, ThermoFisher Scientific, measurement). The bands were assessed based on the size marker. Only PCR products with similar gel intensity and concentration were used in further analysis. The amplicons were cleaned up with magnetic beads and checked again on an agarose gel.

Amplicons' libraries were prepared and indexed with the use of Nextera XT DNA library preparation kit and Nextera XT Index Kit v2 (Illumina, San Diego, CA). Size of final library constructs was assessed by on-chip electrophoresis Agilent Bioanalyzer and Agilent High Sensitivity DNA Kit. Concentrations of libraries were measured with QUBIT 4 fluorometer and QUBIT dsDNA HS assay (Thermo Fisher Scientific, Waltham, MA). All libraries were diluted to 4 nM concentration, pooled and denatured. Final concentration of the libraries used for clustering was 10 pM, with 10% of PhiX control (Illumina) added in each run. For each individual, a panel of 18 amplicons spanning 18 genes on eight autosomes and the Z chromosome were sequenced in four paired-end runs (2 × 250 bp) with the use of Illumina MiSeq System and MiSeq Reagent Kits v2 (500 cycles). Over 70M of high quality, passing filter pair-end reads were generated from all runs (average 250K reads/per sample). Quality control of each FASTQ file was performed using FastQC software ([Andrews, 2010](#)). Data was then filtered to remove adapter sequences and the threshold of 30 was set in order to remove low-quality bases from the 3' and 5' ends using CutAdapt software ([Martin, 2011](#)). After quality control, reads were aligned to the GRCg6a reference genome (https://www.ncbi.nlm.nih.gov/assembly/GCF_000002315.5/) using BWA-MEM software ([Li et al., 2009](#)). Resulting BAM files were sorted and

indexed using SAMtools ([Li et al., 2009](#)). The processed files were calibrated using BaseRecalibrator package in GATK ([McKenna et al., 2010](#)) in order to adjust base quality scores. Variant calling for each individual was performed using HpyplotypeCaller and GenotypeGVCFs packages also from GATK software. The resulting multi-sample VCF file describing identified variants was further downstream processed. InDels and variants that did not overlap with amplicon locations were removed by VCFtools ([Danecek et al., 2011](#)), leaving the total number of 262 SNPs. SNP preselection criteria for downstream analyses were based on a minor allele frequency of at least 0.01 and technical quality of genotyping expressed by a minimum call rate of 80%. Variants that did not match flanked regions by PCR primers designed for this study were excluded. Occurrence of these SNPs was caused by alignment of a sequence to a domain that is shared by 2 genes of high homology in amplicon sequences. Genomic annotation of SNPs was assigned by the Variant Effect Predictor software ([McLaren et al., 2016](#)).

Estimation of Additive and Dominance Effects of SNPs

For the estimation of cumulative additive and dominance effects of animals the following mixed model was applied ([Vitezica et al., 2018](#)):

$$\mathbf{y} = \mu + \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_a\mathbf{a} + \mathbf{Z}_d\mathbf{d} + \boldsymbol{\varepsilon}, \quad (1)$$

where \mathbf{y} represents a vector of phenotypes (i.e., KLHd0, KLHd7, LTAd0, or LPSd0), μ is a general mean, $\boldsymbol{\beta}$ is a vector of fixed effects representing sex and batch with a corresponding design matrix \mathbf{X} , \mathbf{a} is a vector of random additive animal effects, \mathbf{d} is a vector of random dominance animal effects with corresponding design matrices \mathbf{Z}_a and \mathbf{Z}_d , $\boldsymbol{\varepsilon}$ is a vector of residuals with $\boldsymbol{\varepsilon} \sim N(0, \mathbf{I}\hat{\sigma}_\varepsilon^2)$, where \mathbf{I} is an identity matrix and $\hat{\sigma}_\varepsilon^2$ represents the residual variance. The covariance structure of \mathbf{a} and \mathbf{d} was assumed as $\mathbf{a} \sim N(0, \mathbf{G}_a\hat{\sigma}_a^2)$ and $\mathbf{d} \sim N(0, \mathbf{G}_d\hat{\sigma}_d^2)$ respectively with

$$\mathbf{G}_a = \frac{\mathbf{M}\mathbf{M}^T}{2 \sum_{i=1}^{N_{SNP}^*} p_i q_i}, \quad (2)$$

and

$$\mathbf{G}_d = \frac{\mathbf{W}\mathbf{W}^T}{4 \sum_{i=1}^{N_{SNP}^*} p_i^2 q_i^2}, \quad (3)$$

where $M_{ij} = \{2 - 2p_i, 1 - 2p_i, -2p_i\}$ and $W_{ij} = \{-2q_i^2, -2p_i q_i, -2p_i^2\}$ were respectively calculated for homozygous, heterozygous, and alternative homozygous genotypes for i^{th} SNP of j^{th} individual, p_i/q_i represent frequency of the reference/alternative allele for i^{th} SNP and N_{SNP}^* is a total number of SNPs. $\hat{\sigma}_a^2$, $\hat{\sigma}_d^2$ and $\hat{\sigma}_\varepsilon^2$ denote additive, dominance, and residual variances, respectively, which were assumed to be known as $\hat{\sigma}_a^2 = 0.3\hat{\sigma}_y^2$, $\hat{\sigma}_d^2 = 0.2\hat{\sigma}_y^2$ and $\hat{\sigma}_\varepsilon^2 = 0.5\hat{\sigma}_y^2$.

In the next step additive (vector \mathbf{snp}_a) and dominance (vector \mathbf{snp}_d) SNP effects were calculated based on

Table 1. Descriptive statistics of phenotypes.

Trait name	Mean	Standard deviation	Minimum	Maximum
KLHd0	3.81	1.55	1.3	9.6
KLHd7	11.55	1.90	5.5	16.7
LTAd0	5.67	1.79	1.0	12.0
LPSd0	2.96	1.28	1.1	8.5

Table 2. The list of selected genes.

Gene name	Gene ID	Chromosome	F primer	R primer
<i>PIK3R1</i>	427171	Z	GCATGAAAGGACGCACTGA	AAGGTACCATCGGCAGTGTC
<i>AP3B1</i>	427646	Z	GGCTCAGCAGGCAGTAAATC	TCCGAGAGACACCAGAATTG
<i>CD36</i>	417730	1	GGCCTCTTCTCTCGTGTGAC	ACATGTTTCAGTGGGCAACA
<i>MYD88</i>	420420	2	GGCCCTGACTGCTACTTGAG	TCTGACCTTGCAGATGTTGC
<i>TLR5</i>	554217	3	ACATCAACGGCTAATAATTGTCTT	TATAGTTTTGCAGCTCTATACCACT
<i>TAB2</i>	421622	3	TCTGTTTTCTCCAGCCTCGT	GCAGAAGTGGCTTCCCTGAAC
<i>CCRN4L (NOCT)</i>	404779	4	TTCCATGGGAAACAGCACCAG	TTATAACAGTCTGTCCAGGGTCTG
<i>MAP3K13</i>	424876	9	TGTTCTGTGAGCGAGAATGG	GAGGCAAGGTTTTGCAGAAAG
<i>ADIPOQ</i>	404536	9	CAGCTCTCCAGCTTGCTTCT	TCCATCTTTTTCCATCCTTGC
<i>PROCR</i>	424867	9	CACGCATCACCTACAGCACT	TGACTTTCACCCCTCTCCAC
<i>MAP2K1</i>	415549	10	TGTTGTGGTGAGAGCAGGAG	ACAAGGTTCCAAGGATGCAC
<i>SNX8</i>	416468	14	TTGAATTGCCAGTTGCTGAG	CCTGACCCCAAAAGACTGT
<i>E4F1</i>	416560	14	AAAGAAGGGCGGTATGTGTG	GCTCAGTGAAGACGGAAAG
<i>PDGFA</i>	374196	14	TCAGGCTCTTTTCGTGAGGT	CATTACGGAGCACATGGTTG
<i>SOCS1</i>	416630	14	TGGTGCAGCACTGCTAATTC	CTGCAAAGCAAGGAGGTTTC
<i>IL21R</i>	416586	14	ATGGGGCAGTTAGGTGAGTG	TGAGACTCGCTCTCAAAGCA
<i>IL20RB</i>	768437	14	ATTTCCATGGGTTTGGATGA	GCTAAGATGGCAGCAACACA
<i>PECAM1</i>	771243	18	CCAGGCTGCACAGATGAGTA	TGCATGTCTCTCTGTCTCTG

animal effects (\mathbf{a} , \mathbf{d}) using the back-solve method proposed by Liu et al. (2014, 2016) for additive effects:

$$\mathbf{snp}_a = \left(\left(\frac{1-k}{N_{SNP}^*} \right)^{-1} + \frac{1}{k} \mathbf{M} \mathbf{A}^{-1} \mathbf{M}^T \right)^{-1} \frac{1}{k} \mathbf{M} \mathbf{A}^{-1} \hat{\mathbf{a}}, \quad (4)$$

and for dominance effects

$$\mathbf{snp}_d = \left(\left(\frac{1-k}{N_{SNP}^*} \right)^{-1} + \frac{1}{k} \mathbf{W} \mathbf{A}^{-1} \mathbf{W}^T \right)^{-1} \frac{1}{k} \mathbf{W} \mathbf{A}^{-1} \hat{\mathbf{d}}, \quad (5)$$

where k is tuning parameter set in our analysis to 0.2 and \mathbf{A} represents the numerator relationship matrix among animals calculated based on their pedigree records. Other elements of Equations (4) and (5) are the same as in Equations (2) and (3).

For testing the significance of additive and dominance SNP effects we used the Wald test, which under H_0 follows the standard normal distribution.

SNP Prioritization

Based on SNP descriptive information available in the VCF file the prioritization of SNPs was carried out separately for additive and dominance effects, based on the following mixed linear model:

$$\log(p) = \mu + \beta_1 \log(\mathbf{AF}) + \beta_2 \frac{1}{\mathbf{BaseQRankSum}} + \beta_3 \mathbf{MQ} + \beta_4 \frac{1}{\mathbf{SOR}} + \mathbf{Zs} + \varepsilon^*, \quad (6)$$

where p is the nominal P -value of the Wald test for the additive or dominance effect of each SNP and each trait from, \mathbf{AF} is the frequency of an alternative allele, $\mathbf{BaseQRankSum}$ is the Z-score from Wilcoxon rank sum test of alternative vs. reference base qualities, \mathbf{MQ} is the root mean square of the mapping quality of reads across all animals, \mathbf{SOR} is the symmetric odds ratio to detect strand bias, $\mathbf{s} \sim N(0, \mathbf{L}\sigma_s^2)$ is a random vector of SNP priorities with the covariance matrix \mathbf{L} with R^2 LD measure between pairs of SNPs as the off-diagonal elements and

ones on the diagonal and σ_s^2 representing the variance of SNP associated P -values, ε^* is the residual term with $\varepsilon^* \sim N(0, \mathbf{I}\hat{\sigma}_{\varepsilon^*}^2)$, where \mathbf{I} is an identity matrix and $\hat{\sigma}_{\varepsilon^*}^2$ represents the residual variance.

In model (6) SNP priorities are expressed by their P -values corrected for the SNP allele frequency (\mathbf{AF}), sequencing quality ($\mathbf{BaseQRankSum}$, \mathbf{MQ} , \mathbf{SOR}) and the LD between SNPs. These estimated priorities (\mathbf{s}) were used to rank SNPs according to their importance on the variation of immune response traits.

RESULTS

Genomic Annotation of SNPs From Targeted Sequencing

The 114 SNPs located within 24 short genomic fragments (Figure 1) assigned to eight autosomes (GGA1, GGA2, GGA3, GGA4, GGA9, GGA10, GGA14, and GGA18) and the sex chromosome (GGZ) was used in the downstream analysis. The average segment length was 435 ± 492 bp ranging between 89 bp and 2,205 bp. Genomic annotation of SNPs showed that nearly 57% of SNPs ($n = 65$) were localized in the intronic regions while 20% ($n = 23$) of SNPs overlapped exons. The remaining SNPs were located in the UTR and downstream/upstream regions of genes. SNPs located in exons are represented by 30% ($n = 7$) of missense variants and 70% ($n = 16$) of synonymous variants. Of all SNPs, 79% SNPs ($n = 90$), 6% ($n = 7$), and 15% ($n = 17$) had potential modifier, moderate, and low impact on the effectiveness of translated proteins respectively. The distribution of SNPs and segments on chromosomes with descriptive statistics was presented in Table S1.

SNP Priority

The results of a mixed model analysis of all genotyped SNPs (Figure 2) in the form of nominal P -values from the Wald test on SNP effects estimated by Equations (4) and (5) enhanced by the sequencing quality and

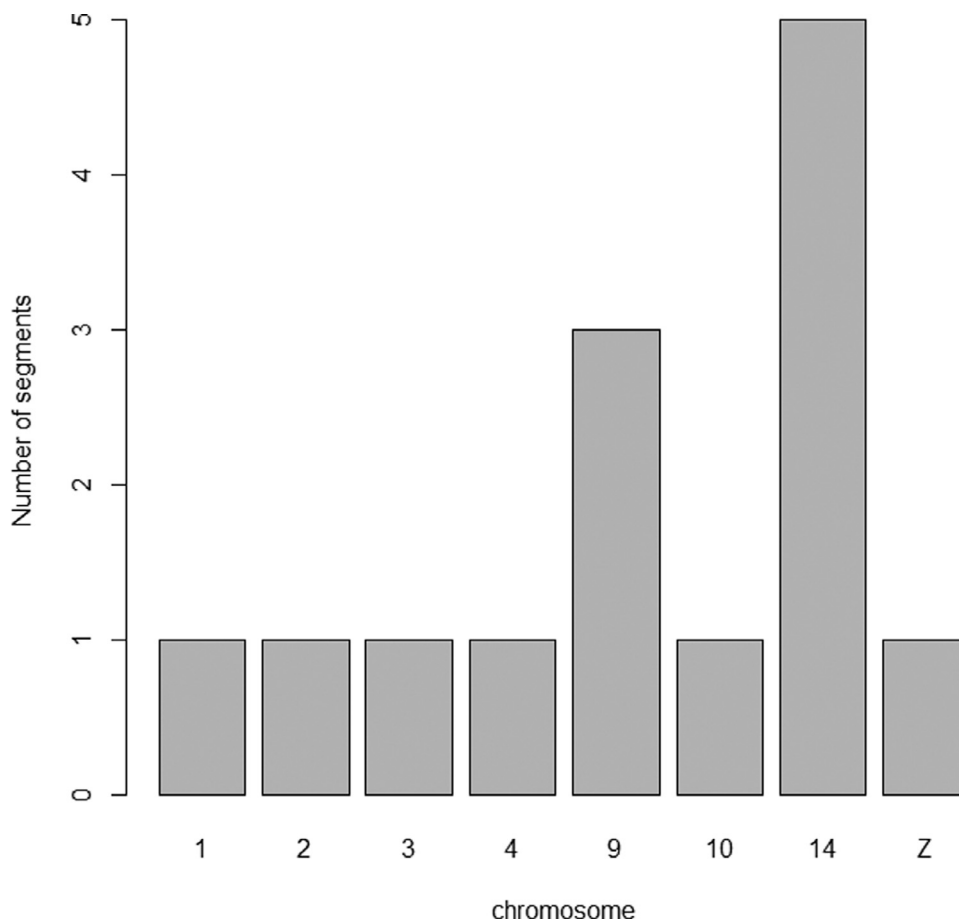


Figure 1. Number of segments with genetic markers per chromosome.

polymorphism information from the VCF files after SNP calling were utilized to construct a SNP priority list separately for additive and dominance effects on each trait. For the specific antibody response toward KLH at d 0 (KLHd0), the most highly prioritized SNP for additive and dominance effects was located on GGA2 in the 3'UTR of MYD88. For the additive effect, all the remaining nine polymorphisms from the top-10 list marked the ADIPOQ gene on GGA9, SNX8 on GGA14, MYD88 on GGA2, and NOCT on GGA4. One SNP was exonic and thus candidate for being a causal mutation. For the dominance effect, all the remaining nine polymorphisms from the top-10 list marked the same genes. Moreover, 2 SNPs were exonic and indicated potential for being a causal mutation. For the specific antibody response toward KLH at d 7 (KLHd7), there were 4 potential causal additive variants – 3 in ADIPOQ and one in PROCR on GGA9. The most highly prioritized SNP pointed at the 3'UTR of MYD88. The top-10 SNPs composition for dominance effects comprised polymorphisms located in 3'UTR regions of MYD88 (1st SNP), within ADIPOQ (including 2 exonic variants and 3 intronic variants), NOCT (one exonic and 2 intronic variants), and SNX8 (one intronic variant). The gene set related to the antibody response toward LTA at d 0 (LTAd0) comprised SNPs located in MYD88, NOCT, and ADIPOQ (additive), and MYD88, and NOCT (dominance). The highest priority for additive and dominance effects was attributed to the same SNP located

on GGA2 in the 3'UTR region of MYD88. Two SNPs among the top-10 for additive effect were located in the exon of NOCT (2 of them were also in the top 10 list for KLHd0), 4 in the intron of NOCT, and one more SNP in the 3'UTR region of MYD88. Eight SNPs among the top-10 for dominance effect were located in the NOCT (3 exonic and 5 intronic), and one more SNP in the 3'UTR region of MYD88. SNPs selected for their additive effect on antibody response toward LPS at d 0 (LPSd0) marked three genes – NOCT (with SNP ranked 4th located in an exon), MYD88 (one SNP located in the 3'UTR region), and SNX8 (with SNP ranked 7th located in an exon). SNPs selected for their dominance effect marked – NOCT (including four exonic SNPs ranked 4th, 5th, 7th, and 9th), ADIPOQ, and MYD88. The lists of top-10 ranked SNPs were presented in [Tables S2-S5](#).

Considering the SNPs with top-10 additive effects, there was no polymorphism common to all 4 traits. Three SNPs were common between KLHd0 and LTAd0 (rs14131328, rs316037040, rs14131330). The SNP rs14131328 was in the top 10 for KLHd0, LTAd0 and LPSd0 ([Figure 3A](#)). KLHd7 was represented by the unique set of SNPs (no intersection with the other traits). However, shifting to gene level, MYD88 was represented among top 10 additive SNPs in all 4 traits. The other genes represented by selected SNPs were NOCT and ADIPOQ present for three traits (NOCT in KLHd0, LTAd0, and LPSd0; ADIPOQ in KLHd0,

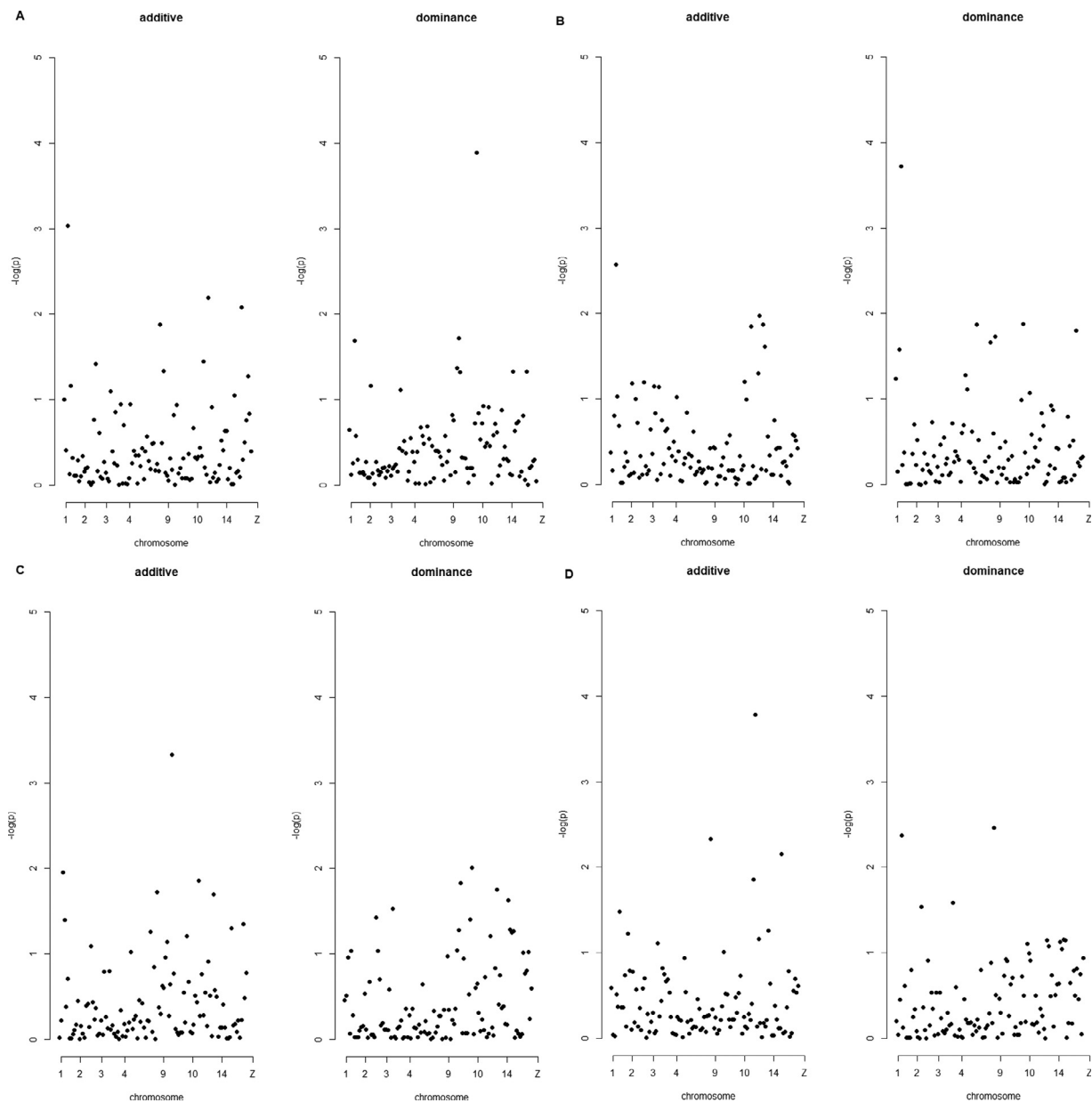


Figure 2. Manhattan plots of the amplicon association study.

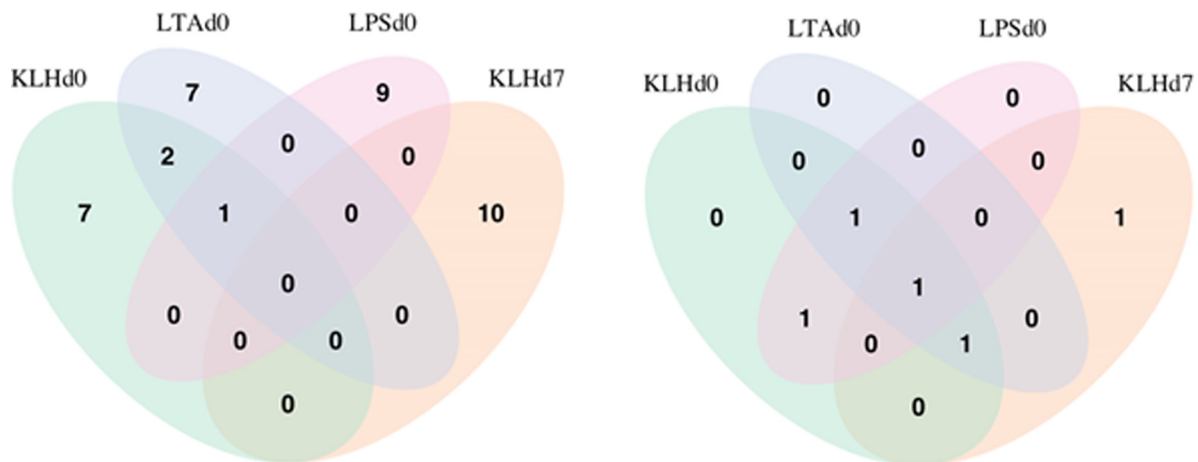


Figure 3. SNPs (A) and genes (B) represented among the top-10 additive effects.

KLHd7, and LTAd0), as well as SNX8 represented in the top 10 for KLHd0 and LPSd0 (Figure 3B for additive effects).

DISCUSSION

The analysis of the genetic background of the immune responses in the experimental population created by crossing White Leghorn layer with Green-legged Partridge-like dual-purpose native chicken goes back to 2010 (Siwek *et al.*, 2010; Sławińska *et al.*, 2011) (Figure 4). However, the QTL for innate immune responses toward LPS and LTA were initially detected in 2 independent populations of laying hens: a cross between 2 chicken

lines expressing different feather pecking behavior and a cross between 2 lines divergently selected toward SRBC (Sheep Red Blood Cells) response (Siwek *et al.*, 2006). Three QTL located on chicken chromosome (*Gallus gallus domesticus* [GGA]): GGA9, GGA18, and GGAZ were confirmed to be linked with NABs for LPS, and 2 QTL located on: GGA5 and GGA14 were validated as linked to NABs for LTA. Similar approach was undertaken for a primary antibody response toward KLH. The initial QTL for this antigen was detected in 2 independent populations of laying hens (Siwek *et al.*, 2003). The validation of this QTL was performed also in WL × GP population (Siwek *et al.*, 2010). A QTL for primary antibody response toward KLH located on

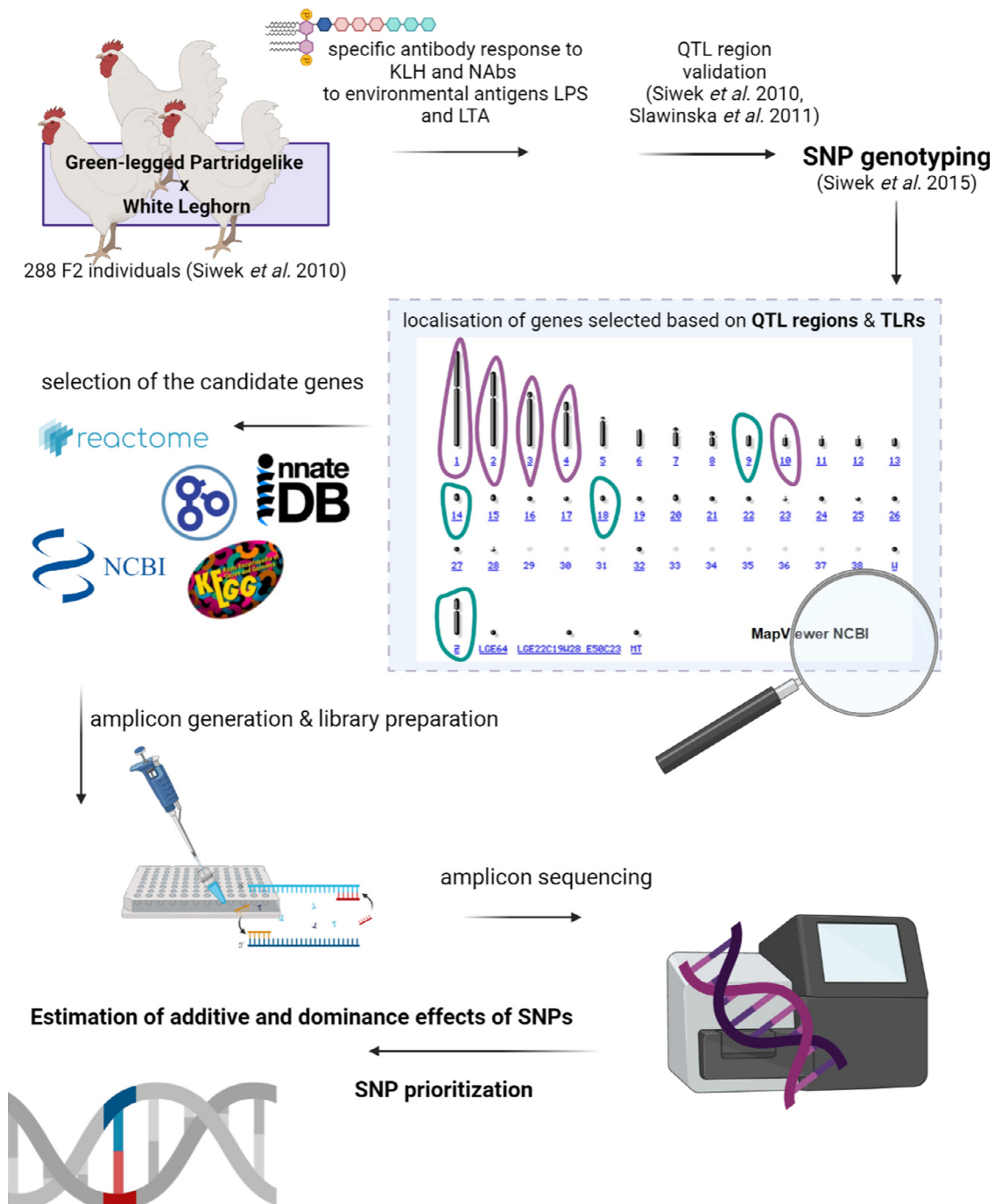


Figure 4. The overview of the different steps of the analysis. Created with BioRender.com.

GGA14 was further fine-mapped with a 2-QTL model (Siwek et al., 2012). The QTL detected for the innate immune responses toward LTA and LPS, and adaptive immune responses toward KLH in different experiments were narrowed down with 2 bioinformatics approaches: a meta-analysis and a combined analysis (Slawinska and Siwek, 2013). The information content used at each step of the analysis varied following the increasing genomic resolution. At first, the microsatellite markers pointed out to different chromosomal regions linked to each of the immune parameters (Siwek et al., 2003). An exception was GGA14 that harbored 2 QTL for LTA and KLH. The current experimental population was foreseen as a cross for a QTL validation study. Therefore, the selection of the microsatellite markers was focused on the previously identified QTL regions (Siwek et al., 2003, 2006). The availability of the SNP custom assays gave the opportunity to run an association analysis (Siwek et al., 2015). The major conclusion of the association study was that the most significant SNPs for the immune responses toward KLH and LTA are located outside of the QTL regions originally proposed by the linkage analysis. Therefore, in the search for causal mutations in candidate genes, a much higher marker resolution is necessary. Second, innate and adaptive immunity have some genes in common. The results of the current analysis confirmed this finding. Furthermore, 3 traits representing innate immunity (KLHd0, LTAd0, LPSd0) share a single variant located in MYD88 gene.

MYD88 participates in several KEGG pathways related to the immune responses, such as *MAPK signaling pathway* (gga04010), *Toll-like receptor signaling pathway* (gga04620), *NOD-like receptor signaling pathway* (gga04621), *Salmonella infection* (gga05132), *Influenza A* (gga05164), *Herpes simplex virus 1 infection* (gga05168). MYD88 encodes a protein which plays a central role in the innate and adaptive immune responses (Stelzer et al., 2016). The essential functions of this protein are related to signal transduction in the interleukin-1 and Toll like receptor signaling pathways, which regulate activation of proinflammatory genes (Stelzer et al., 2016).

Remaining genes pinpointed by the analysis in the current study are ADIPOQ, NOCT, SNX8, and PROCR. ADIPOQ encodes the adiponectin protein, whose role in the immune response and inflammation processes has been known for humans and mice (Fantuzzi, 2013) albeit not for poultry. Adiponectin is known for its anti-inflammatory activities such as inhibition of proinflammatory cytokines, induction of anti-inflammatory cytokines, downregulation of adhesion molecule expression, antagonism of toll like receptors (TLR) including their ligands for example, LPS (Fantuzzi, 2013). NOCT also known as Nocturnin is identified as a circadian clock regulated gene in *Xenopus laevis* (Stelzer et al., 2016). In vertebrates, this gene is known as Noct, Noc, CCR4L, CCRN4L or Ccr4c [22]. Niu et al. (2011) proved that Nocturnin is induced by LPS treatment in mouse embryonic fibroblasts. SNX8, sorting nexin 8 belongs to the family of proteins involved

in endocytosis, endosomal sorting, and signaling (Stelzer et al., 2016). The study performed by Wei et al. (2017) showed that SNX8 is a part of the IFN gamma noncanonical signaling pathway and participates in the host defense to microbial infection. PROCR, also known as Activated Protein C Receptor, is involved in direct cellular signaling function through inhibition of activation of NFkB expression of proinflammatory cytokines (IL-1, IL-6, TNF- α) (Cai et al., 2019). PROCR together with CD1 were identified as a distinct subfamily of MHC class-I-like genes. The function of the class-I-like genes ranges from specialized antigen presenting to the transport of IgG and lipid mobilization and catabolism (Maruoka et al., 2005).

SNP Effect

Three of the selected SNP variants showed pleiotropic effects. The most pronounced pleiotropy was attributed to rs14131328 located in the 3'UTR of MYD88, which was associated with KLHd0, LTAd0, and LPSd0. This single SNP has been identified as nonsynonymous mutation in the study on susceptibility to *Salmonella pullorum* in chickens (Liu et al., 2015). The 3 traits KLHd0, LTAd0, and LPSd0, represent innate immunity. This part of the immune response is involved in the first line of defense against microbial pathogens and plays a key role in bacterial infections. As it has been already described in the introduction, both LTA and LPS are present in the bacterial membranes. They belong to the pathogen-associated molecular patterns (PAMPs) which activate pattern recognition receptors (PRR) of the host organisms. One of the major classes of PRR is TLR.

There are 2 additional SNPs common to KLHd0 and LTAd0. The variant rs14131330 is located within 3'UTR of MYD88 gene. This SNP has been also identified as a nonsynonymous mutation in the study on susceptibility to *Salmonella pullorum* in chickens (Liu et al., 2015). The variant rs316037040 is located in ADIPOQ gene. The literature does not report previous phenotype association with this variant.

Considering inheritance modes, most of the selected SNPs exhibit either a strong additive or a strong dominance effect. Interestingly, most of the variants located within MYD88 (rs14131328, rs14131330, rs731606804, and rs14131330) and ADIPOQ (rs16662812, rs317362477, rs313075798, rs736460279, and rs316037040) revealed both, additive and dominance effects. This observation is important in view of the practical application of the results in breeding programs, since additive effects are independent on the commercial-cross design, while the dominance are.

Since all of the considered SNPs are located inside the genes representing genomic regions it is impossible to rely on *P*-values resulting from an association analysis anymore. This is because of 2 main factors: 1) huge intercorrelation between SNPs due to huge linkage disequilibrium (Biernacka and Cordell, 2007); 2) not high genotype variation due to low frequency of the negative allele of causal mutations (Auer and Lettre, 2015).

Because of this we used the novel SNP prioritization method using 2 sources of information: 1) *P*-values from an association analysis and 2) information coming out from variant calling file. Such statistical solution could improve the accuracy of detection association between phenotype and causal variant. Also, other authors suggested that prioritization methods are effective in such approaches (Hou and Zhao, 2013; Suravajhala and Benso, 2017).

From within the protein coding regions, 2 of the SNPs are good candidates for being causal mutations: for the adaptive response – a deleterious (SIFT = 0.02) variant rs741071044 located within PROCR, which changes valine to methionine and for the innate response – a tolerated (SIFT = 0.33) variant rs315673797 located within NOCT, which changes serine to asparagine. However, an interesting observation is that many of the top 10 SNPs were identified within regulatory sequences. It indicates the genetic determination of immune response traits seems to be shaped by the regulation of protein expression. Long non-coding RNAs are defined as non-protein coding transcripts longer than 200 nucleotides. Their role in immune response has been demonstrated by numerous studies (Heward and Lindsay, 2014). lncRNAs are expressed in tissue-specific or developmental stage-specific manner (Kwon et al., 2019). It was also reported that some lncRNAs regulate post-transcriptional gene regulation through binding to specific RBP (Kwon et al., 2019). Through the interaction with various molecules (protein, mRNA, miRNA) lncRNAs regulate transcription, splicing, RNA degradation and translation (Fitzgerald and Caffrey, 2014). There is also a demonstrated link between lncRNAs and 3'UTRs in the regulation of immune response (Schwerk and Savan, 2015). The 3'UTRs located downstream of the coding sequence are related to regulatory process of RNA stability, mRNA translation and localization. Any variation in 3'UTR both the length and the sequence may change the binding for miRNA and RNA binding proteins (RBP) leading to changes in the gene expression (Steri et al., 2018).

CONCLUSIONS

The major goal underlying our study was to identify causal mutations. Of course, in the case of complex traits, measured on a quantitative scale, no single causal mutation is expected to cause the response. Quantitative traits typically undergo a complex mode of inheritance determined by many genes with varying effect sizes. The top-10 variants identified in our study were located in different functional parts of the genome. In the context of causality, we want to highlight 3 groups: 1) variants located in exons of protein coding genes (ADIPOQ, NOCT, PROCR, SNX8) 2) variants within exons of non-coding transcripts, and 3) variants located in genes' UTR regions. Variants from the first group influence protein structure. While variants from both latter groups' exhibit regulatory roles on DNA (UTR) or RNA (lncRNA).

Summarizing, our study has demonstrated the important role of non-coding DNA variation in the determination of both innate and adaptive immune response in chicken, which complements coding genes with know impact on immune response.

ACKNOWLEDGMENTS

The research is co-financed by the: National Research Centre Grant number UMO/2014/13/B/NZ9/02123; UTP University of Science and Technology Grant number BN-8/2021 and The publication is co-financed under the Leading Research Groups support project from the subsidy increased for the period 2020–2025 in the amount of 2% of the subsidy referred to Art. 387 (3) of the Law of 20 July 2018 on Higher Education and Science, obtained in 2019.

DISCLOSURES

All authors declare no potential conflicts of interest.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.psj.2021.101433](https://doi.org/10.1016/j.psj.2021.101433).

REFERENCES

- Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data. Accessed November 2, 2020. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Auer, P. L., and G. Lettre. 2015. Rare variant association studies: considerations, challenges and opportunities. *Genome Med* 7:16.
- Baumgarth, N., J. W. Tung, and L. A. Herzenberg. 2005. Inherent specificities in natural antibodies: a key to immune defense against pathogen invasion. *Spring. Semin. Immuno.* 26:347–362.
- Berghof, T. V. L., M. H. P. W. Visker, J. A. J. Arts, H. K. Parmentier, J. J. van der Poel, A. L. J. Vereijken, and H. Bovenhuis. 2018. Genomic region containing toll-like receptor genes has a major impact on total IgM antibodies including KLH-binding IgM natural antibodies in chickens. *Front. Immunol.* 8:1879.
- Biernacka, J. M., and H. J. Cordell. 2007. Exploring causality via identification of SNPs or haplotypes responsible for a linkage signal. *Gen. Epidemiol.* 31:727–740.
- Biscarini, F., H. Bovenhuis, J. A. M. van Arendonk, H. K. Parmentier, A. P. Jungerius, and J. J. van der Poel. 2010. Across-line SNP association study of innate and adaptive immune response in laying hens. *Anim. Genet.* 41:26–38.
- Cai, X., I. Biswas, S. R. Panicker, H. Giri, and A. R. Rezaie. 2019. Activated protein C inhibits lipopolysaccharide-mediated acetylation and secretion of high-mobility group box 1 in endothelial cells. *J. Thromb. Haemost.* 17:803–817.
- Cantor, R. M., K. Lange, and J. S. Sinsheimer. 2010. Prioritizing GWAS Results: a review of statistical methods and recommendations for their application. *Am. J. Hum. Gen.* 86:6–22.
- Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, and R. Durbin. 2011. 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. *Bioinformatics.* 27:2156–2158.
- Erf, G. F. 2004. Cell-mediated immunity in poultry. *Poult. Sci.* 3:580–590.
- Fantuzzi, G. 2013. Adiponectin in inflammatory and immune-mediated diseases. *Cytokine.* 64:1–10.

- Fitzgerald, K. A., and D. R. Caffrey. 2014. Long noncoding RNAs in innate and adaptive immunity. *Curr. Opin. Immunol.* 26:140–146.
- Heward, J. A., and M. A. Lindsay. 2014. Long non-coding RNAs in the regulation of the immune response. *Trends Immunol.* 35:408–419.
- Hou, L., and H. Zhao. 2013. A review of post-GWAS prioritization approaches. *Front. Genet.* 4:280.
- Kwon, G. J., J. Henao-Mejia, and A. Williams. 2019. Editorial: long non-coding RNAs and immunity. *Front. Immunol.* 10:2378.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics.* 25:2078–2079.
- Liu, Z., M. E. Goddard, B. J. Hayes, F. Reinhardt, and R. Reents. 2016. Technical note: equivalent genomic models with a residual polygenic effect. *J. Dairy Sci.* 99:2016–2025.
- Liu, Z., M. E. Goddard, F. Reinhardt, and R. Reents. 2014. A single-step genomic model with direct estimation of marker effects. *J. Dairy Sci.* 97:5833–5850.
- Liu, X.-Q., F. Wang, J. Jin, Y.-G. Zhou, J.-S. Ran, Z.-Q. Feng, Y. Wang, and Y.-P. Liu. 2015. *MyD88* polymorphisms and association with susceptibility to salmonella pullorum. *BioMed Res. Internat.* 2015:1–7.
- Maddur, M. S., S. Lacroix-Desmazes, J. D. Dimitrov, M. D. Kazatchkine, J. Bayry, and S. V. Kaveri. 2020. Natural antibodies: from first-line defense against pathogens to perpetual immune homeostasis. *Clin. Rev. Aller. Immuno.* 58:213–228.
- Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17:10.
- Maruoka, T., H. Tanabe, M. Chiba, and M. Kasahara. 2005. Chicken CD1 genes are located in the MHC: CD1 and endothelial protein C receptor genes constitute a distinct subfamily of class-I-like genes that predates the emergence of mammals. *Immunogenetics.* 57:590–600.
- McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytzky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome. Res.* 20:1297–1303.
- McLaren, W., L. Gil, S. E. Hunt, H. S. Riat, G. R. Ritchie, A. Thormann, P. Flicek, and F. Cunningham. 2016. The ensembl variant effect predictor. *Genome Biol.* 17:122.
- Niu, S., D. L. Shingle, E. Garbarino-Pico, S. Kojima, M. Gilbert, and C. B. Green. 2011. The circadian deadenylase nocturnin is necessary for stabilization of the iNOS mRNA in mice. *PLoS One.* 6: e26954.
- Ochsenbein, A. F., and R. M. Zinkernagel. 2000. Natural antibodies and complement link innate and acquired immunity. *Immuno. Today.* 21:624–630.
- Schwerk, J., and R. Savan. 2015. Translating the untranslated region. *J. Immunol.* 195:2963–2971.
- Shishido, S., S. Varahan, K. Yuan, X. Li, and S. Fleming. 2012. Humoral innate immune response and disease. *Clin Immunol.* 144:142–158.
- Sijben, J. W., H. de Groot, M. G. B. Nieuwland, J. W. Schrama, and H. K. Parmentier. 2000. Dietary linoleic acid divergently affects immune responsiveness of growing layer hens. *Poult. Sci.* 79:1106–1115.
- Siwek, M., A. J. Buitenhuis, S. J. B. Cornelissen, M. G. B. Nieuwland, H. Bovenhuis, R. P. M. A. Crooijmans, M. a. M. Groenen, G. de Vries-Reilingh, H. K. Parmentier, and J. J. van der Poel. 2003. Detection of different quantitative trait loci for antibody responses to keyhole limpet hemocyanin and *Mycobacterium butyricum* in two unrelated populations of laying hens. *Poult. Sci.* 82:1845–1852.
- Siwek, M., B. Buitenhuis, S. Cornelissen, M. Nieuwland, E. F. Knol, R. Crooijmans, M. Groenen, H. Parmentier, and J. van der Poel. 2006. Detection of QTL for innate: non-specific antibody levels binding LPS and LTA in two independent populations of laying hens. *Dev. Comp. Immunol.* 30:659–666.
- Siwek, M., and Egbert. F. Knol. 2005. Genetic aspects of biological processes underlying the defense system in the neonate. *Folia Biol. (krakow).* 53:39–43.
- Siwek, M., A. Sławińska, M. Nieuwland, A. Witkowski, G. Zięba, G. Minozzi, E. F. Knol, and M. Bednarczyk. 2010. A quantitative trait locus for a primary antibody response to keyhole limpet hemocyanin on chicken chromosome 14—confirmation and candidate gene approach. *Poult. Sci.* 89:1850–1857.
- Siwek, M., A. Sławińska, M. Rydzanicz, J. Wesoly, M. Fraszczak, T. Suchocki, J. Skiba, K. Skiba, and J. Szyda. 2015. Identification of candidate genes and mutations in QTL regions for immune responses in chicken. *Anim. Genet.* 46:247–254.
- Siwek, M., J. Szyda, A. Sławińska, and M. Bednarczyk. 2012. Detection of two QTL on chicken chromosome 14 for keyhole lymphet heamocyanin. *J Appl. Genet.* 53:115–119.
- Sławińska, A., and M. Siwek. 2013. Meta - and combined - QTL analysis of different experiments on immune traits in chickens. *J Appl. Genet.* 54:483–487.
- Sławińska, A., A. Witkowski, M. Nieuwland, G. Minozzi, M. Bednarczyk, and M. Siwek. 2011. Quantitative trait loci associated with the humoral innate immune response in chickens were confirmed in a cross between green-legged Partridge-like and White Leghorn. *Poult. Sci.* 90:1909–1915.
- Stelzer, G., N. Rosen, I. Plaschkes, S. Zimmerman, M. Twik, S. Fishilevich, T. I. Stein, R. Nudel, I. Lieder, Y. Mazor, S. Kaplan, D. Dahary, D. Warshawsky, Y. Guan-Golan, A. Kohn, N. Rappaport, M. Safran, and D. Lancet. 2016. The GeneCards suite: from gene data mining to disease genome sequence analyses. *Curr. Prot. Bioinf.* 54:1.30.1–1.30.33.
- Steri, M., M. L. Idda, M. B. Whalen, and V. Orrù. 2018. Genetic variants in mRNA untranslated regions. *WIREs RNA* 9:e1474.
- Suravajhala, P., and A. Benso. 2017. Prioritizing single-nucleotide polymorphisms and variants associated with clinical mastitis. *AABC.* 10:57–64.
- Vitezica, Z. G., A. Reverter, W. Herring, and A. Legarra. 2018. Dominance and epistatic genetic variances for litter size in pigs using genomic models. *Genet. Sel. Evol.* 50:71.
- Wei, J., W. Guo, H. Lian, Q. Yang, H. Lin, S. Li, and H.-B. Shu. 2017. SNX8 mediates IFN γ -triggered noncanonical signaling pathway and host defense against *Listeria monocytogenes*. *Proc. Natl. Acad. Sci. U. S. A.* 114:13000–13005.