Investigating the functional role of 1,012 candidate genes identified by a Genome Wide Association Study for body weight in broilers

Investigating the functional role of 1,012 candidate genes identified by a Genome Wide Association Study for body weight in broilers

E. Tarsani[1], A. Kranis[2,3], G. Maniatis[2] & A. Kominakis[1]

[1]Department of Animal Science, Agricultural University of Athens, Iera Odos 75, 11855, Athens, Greece

acom@aua.gr (Corresponding Author)

[2]Aviagen, EH28 8SZ, Midlothian, United Kingdom

[3] The Roslin Institute, University of Edinburgh, EH25 9RG, Midlothian, United Kingdom

Summary

A genome-wide association study (GWAS) was performed using 6,598 broilers and dense genome wide SNP data to identify QTLs and positional candidate genes for body weight at 35 days of age (BW35). A multi-locus mixed model analysis identified 12 genome-wide significant SNPs, dispersed on 9 autosomes and 1,012 positional candidate genes within a distance of 1Mb from the significant SNPs. Eight significant markers were located within genes and 17 genes were found to participate in skeletal system development. Candidate genes were found to participate in various pathways and biological processes related to growth. Current findings confirm previous results with regard to functional candidate genes, pathways and biological processes for body weight while proposing novel candidate genes for this trait.

Keywords: GWAS, body weight, broilers, candidate genes

Introduction

Body weight (BW) is an economically important trait for the broiler industry and presents considerable biological interest as it is a typical complex (polygenic) trait. To date, the ChickenQTLdb (https://goo.gl/j8B6Qe) has over 7,812 QTL/SNP associations of which 3,582 are related to growth traits and 166 to BW. Several GWASs have already been performed for growth traits (e.g. Xie et al., 2012). Despite the large number of findings by GWAS, the genetic architecture of BW in chicken remains elusive (Pettersson & Carlborg, 2010), since only a small number of positional candidate genes are confirmed as truly functionally relevant to the trait. In the present study, we first conducted a GWAS for BW at 35 days of age (BW35) using dense genome wide SNP data in a population of 6,598 broilers. We then employed various web-based tools to identify the most plausible functional candidate genes for the trait.

Material and methods

Data

A total of n=6,598 broilers (n=3,678 males and n=2,920 females) from a purebred commercial broiler line with records on body weight BW35 were made available by Aviagen. Animals were genotyped using an Affymetrix® Axiom® high-density genotyping array. After applying quality control filters (only autosomal polymorphic SNPs with call rate <0.99, MAF (minor allele frequency)<0.01 and linkage disequilibrium (LD) r^2 values greater than 0.99 within windows of 2 Mb inter-marker distance(s); and autosomal heterozygosity outside the 1.5 inter-quartile range) the final number of the SNPs used in this study was of 262,067 SNPs. Data filtering was performed using the SNP & Variation Suite v8.7.2 software (Golden Helix: http://www.goldenhelix.com). Phenotypic records for BW35 ranged from 1,130 to 2,630 g with an average of 1840.2 g (SD=194 g).

Statistical analysis

A multi-locus mixed (additive) model (Segura et al., 2012) and a stepwise regression procedure with forward selection and backward elimination was employed to identify genome-wide significant markers associated with the trait. The fixed effects part of the model included hatch week(36 classes), mating group (17 classes) and the sex (2 classes) as well as the SNP effects. The normalized genomic relationship matrix (GRM) was utilized for the estimation of the additive genetic effects. Statistically significant markers were selected at the optimal step according to the extended Bayesian Information Criterion (eBIC) and a FDR cutoff p-value of 0.05. This analysis was performed using SNP & Variation Suite v8.7.2 software.

Identification of QTL and positional candidate genes

We searched for reported QTL/associations as well as positional candidate genes within a distance of 1Mb around the significant SNPs in ChickenQTLdb (https://goo.gl/j8B6Qe) and the NCBI database (https://goo.gl/bxVS3d and https://goo.gl/laPQh), respectively.

Gene functional characterization and gene prioritization analysis

Functional enrichment analysis of positional candidate genes was done using the PANTHER database (Protein ANalysis THrough Evolutionary Relationships http://pantherdb.org/) (Mi et al., 2017). To detect the gene families among the candidate genes the ToppFun portal (Chen et al., 2009) was additionally employed. Pathway analysis was done using Cytoscape (http://www.cytoscape.org/) of ReactomeFIViz (Wu et al., 2014) via the Reactome pathway database. Positional candidate genes were submitted to Guilt By Association (GBA) based gene prioritization analysis (PA) based on their functional similarity to a training gene list of n=763 annotated genes extracted from NCBI data base using relevant search terms (body weight, body size, BMI) in human and mouse. This analysis was performed with ToppGene

portal (Chen et al., 2009). Finally, to perform network topological analysis (TA) of the candidate genes the NetworkAnalyst (http://www.networkanalyst.ca/) (Xia et al., 2014) portal was used.

Results

Significant SNPs, QTLs and positional candidate genes

Twelve SNPs dispersed across nine chromosomes were significant at the genome-wide level (FDR p-value<0.05, *Table 1*). A total number of 197 published QTL/associations related to growth traits and 1,012 positional candidate genes were identified to lie within the searched regions (*Table 1*). From the candidate genes, n=349 could not be identified as they involved orthologs that have not yet been identified (LOC genes). Eight out of twelve significant markers were located within genes (*Table 2*).

Functional enrichment and pathway analysis

Functional enrichment analyses of candidate genes showed 17 genes (*Table 2*) participating in the skeletal system development (GO:0001501, Bonferroni p-value=0.00398). Furthermore, 88 out of 1,012 genes were members of the S100 calcium binding proteins, the HOXL subclass homeoboxes and the type I Keratins gene families. Table 3 shows the candidate genes participating in the most significant pathways detected.

Gene prioritization and gene network analysis

A total number of 248 (out of 559) positional candidate genes were prioritized as most functionally relevant to the trait. The first 10 top ranked genes are presented in *Table 2*. Seven genes were always involved in six pathways (*Table 3*). The comparison of the PA and TA gene lists revealed 182 common genes, with 8 genes displaying the highest degree(s) of centrality (>50) (Table 2). A minimum network of the candidate genes is depicted in *Figure 1*.

Discussion

Current results confirm previous findings suggesting that GGA1 and GGA4 (Xu et al., 2013) as well as GGA10, GGA15, GGA22, GGA26 (Van Goor et al., 2015) and GGA27 (Lien et al., 2017) harbor QTLs/Associations related to BW. Five genes that participate in skeletal system development belong to the super-family of homeobox genes which play a fundamental role in embryonic development, cell proliferation and metabolic processes (Procino & Cillo, 2013). Additional genes in the same biological process were *MFGE8* that promotes obesity in mice (Khalifeh-Soltani et al., 2014), *SCUBE3* that affects fast muscle development in

zebrafish (Tu et al., 2014) and *PHOSPHO1* associated with body growth in piglets (Hu et al., 2016). Our findings also confirmed the importance of Wnt-signaling, MAPK and insulin signaling pathways for growth traits in the species (Xu et al., 2013) as well as the *TXK* and *PHOSPHO1* genes that are reported as significant for BW (Gu et al., 2011) and feed intake (Xu et al., 2016) in the species.
Pathway analysis highlighted the importance of *UBC, PSME3* and the *PSMD3* genes that participate in pathways relevant to BW (e.g. *PSMD3* in pigs, Wang et al., 2005). With regard to gene families, S100 proteins are regulators in several functions such as Ca^{2+} homeostasis, energy metabolism, proliferation and differentiation (Donato et al., 2013), while type I keratins are necessary for normal structure and tissue function (Schweizer et al., 2006). Regarding the eight common genes between PA and TA, *UBC* is significant for liver development in mice (Hallengren et al., 2013), while *SHC1* mediates the IGF-1 pathway and contributes to the activation of Ras/MAPK pathway leading to cell proliferation (Wagner et al., 2004). Furthermore, *DDX3X* participates in multiple functions including transcriptional and translational regulation as well as cell growth (Lai et al., 2010) and *KPNB1* plays an important role in embryonic development in mice (Miura et al., 2006). *SMAD4* regulates the balance between muscle atrophy and hypertrophy while, generally, common SMADs are coactivators and mediators of the signal transduction by TGF-beta (transforming growth factor) (Seong et al., 2007). The *TERF2* gene is involved in telomerase structure, conformation and tumor development (Benhamou et al., 2016), *SETDB1* participates in cell growth and tumor genesis (Ishimoto et al., 2016) and *RARA* affects the hippocampal development (Huang et al. 2008). Our findings being in agreement with previously published results point to several biological pathways affecting BW phenotype while at the same time supporting the hypothesis that its genetic architecture approximates the infinitesimal model.

References

Benhamou Y et al. 2016. Telomeric repeat-binding factor 2: a marker for survival and anti-EGFR efficacy in oral carcinoma. Oncotarget.;7(28):44236-44251. doi:10.18632/oncotarget.10005.

Chen J et al. 2009. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res.;37(Web Server issue):W305–W311. doi: 10.1093/nar/gkp427.

Donato R et al. 2013. Functions of S100 Proteins. Current molecular medicine. ;13(1):24-57.

Gu X et al. 2011. Genome-wide association study of body weight in chicken F2 resource population. PLoS One. 6: e21872-10.1371/journal.pone.0021872.

Hallengren J et al. 2013.Neuronal ubiquitin homeostasis. Cell biochemistry and biophysics. ;67(1):67-73. doi:10.1007/s12013-013-9634-4.

Hu J et al. 2016. Whole Blood Transcriptome Sequencing Reveals Gene Expression Differences between Dapulian and Landrace Piglets. BioMed Research International.;2016:7907980. doi:10.1155/2016/7907980.

Huang H et al. 2008. Changes in the expression and subcellular localization of RARalpha in the rat hippocampus during postnatal development. Brain Res. 1227, 26–33 10.1016/j.brainres.2008.06.073.

Ishimoto K et al. 2016. Ubiquitination of Lysine 867 of the Human SETDB1 Protein Upregulates Its Histone H3 Lysine 9 (H3K9) Methyltransferase Activity. Permyakov EA, ed. PLoS ONE. ;11(10):e0165766. doi:10.1371/journal.pone.0165766.

Khalifeh-Soltani A et al. 2014. Mfge8 promotes obesity by mediating the uptake of dietary fats and serum fatty acids. Nature medicine.;20(2):175-183. doi:10.1038/nm.3450.

Lai M-C et al. 2010. DDX3 Regulates Cell Growth through Translational Control of Cyclin E1 .Molecular and Cellular Biology. 30(22):5444-5453. doi:10.1128/MCB.00560-10.

Lien CY et al. 2017. Detection of QTL for traits related to adaptation to sub-optimal climatic conditions in chickens. Genetics Selection Evolution 49:39. https://doi.org/10.1186/s12711-017-0314-5.

Mi H et al.2017. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Research.;45(Database issue):D183-D189. doi:10.1093/nar/gkw1138.

Miura K et al. 2006.Impaired expression of importin/karyopherin beta1 leads to post-implantation lethality,Biochem. Biophys. Res. Commun., 341, pp. 132-138.

Pettersson M.E & Carlborg Ö, 2010. Dissecting the genetic architecture of complex traits and its impact on genetic improvement programs - lessons learnt from the Virginia chicken lines. Revista Brasileira de Zootecnia, 39(Suppl. spe), 256-260. https://dx.doi.org/10.1590/S1516-35982010001300028.

Procino A & Cillo C, 2013. The HOX genes network in metabolic diseases. Cell Biol Int, 37: 1145–1148. doi:10.1002/cbin.10145.

Schweizer J et al. 2006. New consensus nomenclature for mammalian keratins. The Journal of Cell Biology. ;174(2):169-174. doi:10.1083/jcb.200603161.

Segura V et al. 2012. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet. 44:825–830. doi: 10.1038/ng.2314.

Seong HA, et al. 2007. 3-Phosphoinositide-dependent PDK1 negatively regulates transforming growth factor-beta-induced signaling in a kinase-dependent manner through physical interaction with Smad proteins. J Biol Chem. ;282(16):12272-89. doi: 10.1074/jbc.M609279200.

Tu C-F et al. 2014. SCUBE3 (Signal Peptide-CUB-EGF Domain-containing Protein 3) Modulates Fibroblast Growth Factor Signaling during Fast Muscle Development. The Journal of Biological Chemistry.;289(27):18928-18942. doi:10.1074/jbc.M114.551929.

Van Goor A et al. 2015. Identification of quantitative trait loci for body temperature, body weight, breast yield, and digestibility in an advanced intercross line of chickens under heat stress. Genetics Selection Evolution 47:96. doi: 10.1186/s12711-015-0176-7.

Wagner K et al. 2004. The insulin-like growth factor-1 pathway mediator genes: SHC1 Met300Val shows a protective effect in breast cancer. Carcinogenesis, Volume 25, Issue 12: 2473–2478.

Wang Y et al.2005. Polymorphism detection of porcine PSMC3, PSMC6 and PSMD3 genes and their association with partial growth, carcass traits, meat quality and immune traits. Canadian journal of animal science.;85(4):475–80.

Wu G et al.2014. ReactomeFIViz: a Cytoscape app for pathway and network-based data analysis [version 2; referees: 2 approved]. F1000Research, 3:146 .doi: 10.12688/f1000research.4431.2.

Xia J et al. 2014. NetworkAnalyst - integrative approaches for protein–protein interaction network analysis and visual exploration. Nucleic Acids Research. ;42(Web Server issue):W167-W174. doi:10.1093/nar/gku443.

Xie L et al. 2012. Genome-wide association study identified a narrow chromosome 1 region associated with chicken growth traits. PLoS One. 2012, 7 (2): e30910-10.1371/journal.pone.0030910.

Xu Z et al. 2013. Overview of Genomic Insights into Chicken Growth Traits Based on Genome-Wide Association Study and microRNA Regulation. Current Genomics. ;14(2):137-146. doi:10.2174/1389202911314020006.

Xu Z et al. 2016. Combination analysis of genome-wide association and transcriptome sequencing of residual feed intake in quality chickens. BMC Genom 17:594. doi:10.1186/s12864-016-2861-5.

Table 1. Genome-wide significant SNPs for BW35. QTL/associations and positional candidate genes in flanking regions of 1Mb around the significant markers.

SNP ID	GGA	Position (bp)[1]	FDR p-value	Number of positional candidate genes	Number of QTL/ associations
rs13923872	1	112741685	0.0112	45	22
rs15608447	4	66885210	4.25E-09	38	37
rs312691174	4	29074989	0.00037	26	14
rs318199727	10	13536548	0.04111	44	13
rs318098582	11	18651449	0.00012	87	14
rs317945754	15	3557083	0.04594	29	21
rs316794400	22	4594855	6.07E-07	25	1
rs317288536	25	976833	8.05E-09	113	0
rs312758346	25	2412866	1.59E-05	198	0
rs317627533	26	4597439	2.12E-05	105	6
rs314452928	27	104022	0.0105	110	4
rs315329074	27	4528275	8.05E-16	192	65

[1]Positions are based on *Gallus gallus-5.0* genome assembly

Table 2. Genes participating in various descriptions according to analysis applied.

Description	Number of genes	Genes
SNPs within genes	7	*SLAIN2, ZC3H18, TMEM132D, F-KER, FCRL4, LEMD2, CACNB1*
Enrichment analysis: skeletal system development (GO:0001501)	17	*HOXB4, BGLAP, MFGE8, HOXB9, ACAN, HOXB3, HAPLN3, ADAMTS4, MEOX1, PHOSPHO1, CNTNAP1, SCUBE3, PRICKLE4, HOXB13, BCAN, HOXB5,HAPLN2*
Top ranked genes by PA	10	*SMAD4, CHRNB2, CDH1, NTRK1, RARA, STAT5B, SCARB1, NR1D1, SHC1, CYBB, PHB.*
Top prioritized genes by TA	16	*UBC, STAT3, SHC1, APP, ELAVL1, DDX3X, HNF4A, KPNB1, SMAD4, ERBB2, TERF2, SETDB1, RARA, SUMO2, CUL3, UBQLN4*

Common genes in PA and TA	8	*UBC, SHC1, DDX3X, KPNB1, SMAD4, TERF2, SETDB1, RARA*

Table 3. Prioritized genes involved in significant pathways.

Pathway	**FDR**	**Number of genes**	**Genes**
Signaling by NGF	2.37E-06	22	*NTRK1, SHC1, PHB, NGF, NRAS, LAMTOR2, NCSTN, PIP5K1A, UBC, PIP4K2B,THEM4, APH1A, FRS3, PSME3, PSMD7, PSMD4 ,PSMB4 ,AKAP13, PSMB3, PSMD3, RIT1 , ARHGEF11*
Signaling by EGFR	1.94E-03	15	*SHC1, PHB, NRAS , LAMTOR2, PIP5K1A, UBC, PIP4K2B, THEM4, FRS3, PSME3, PSMD7, PSMD4, PSMB4, PSMB3, PSMD3*
TCF dependent signaling in response to WNT	3.19E-03	7	*UBC, PSME3, PSMD7, PSMD4, PSMB4, PSMB3 ,PSMD3*
MAPK6/MAPK4 signaling	5.31E-03	9	*UBC, CCND3, IGF2BP1, PSME3, PSMD7, PSMD4 ,PSMB4, PSMB3, PSMD3*
Signaling by Type 1 Insulin-like Growth Factor 1 Receptor (IGF1R)	1.83E-02	13	*SHC1 ,PHB ,NRAS ,LAMTOR2, UBC, THEM4, FRS3, PSME3 , PSMD7, PSMD4 ,PSMB4 ,PSMB3 ,PSMD3*
Signaling by Insulin receptor	2.53E-02	14	*SHC1 ,PHB ,NRAS ,LAMTOR2, ATP6V0A1, UBC, THEM4, FRS3, PSME3, PSMD7 ,PSMD4 ,PSMB4 ,PSMB3, PSMD3*

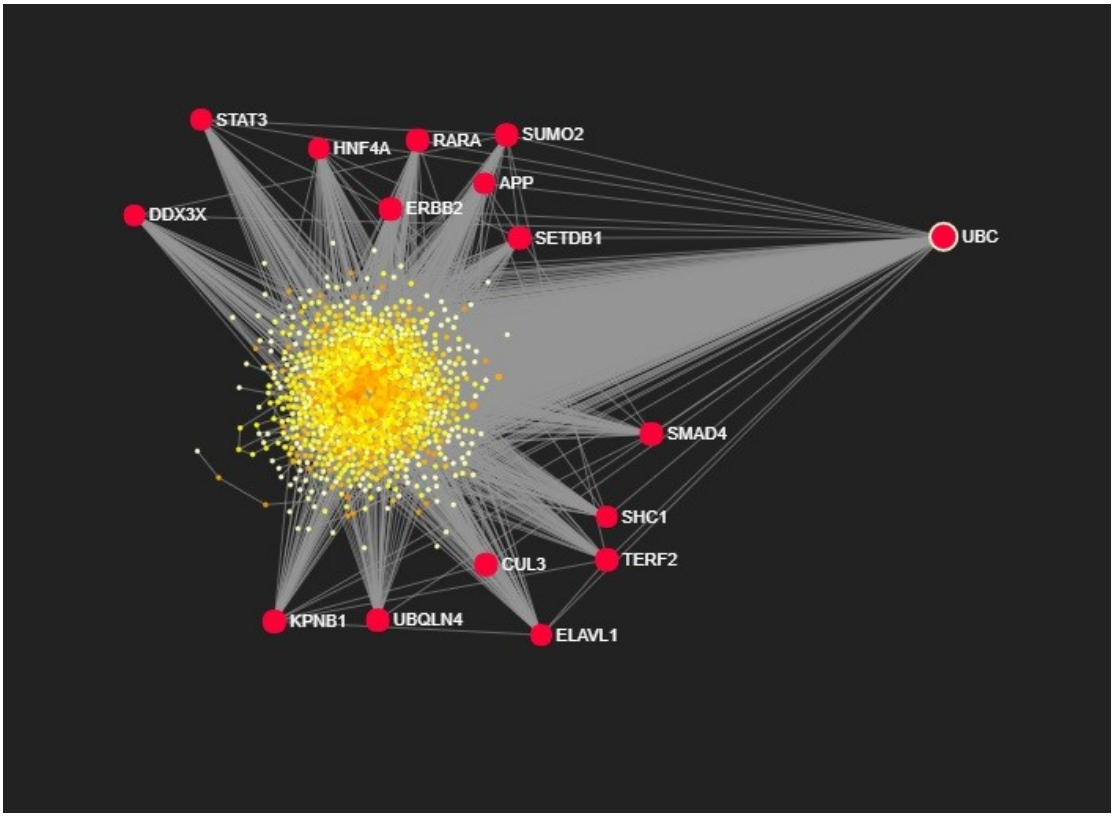

Figure 1. Depiction of a minimum gene network comprised of 1,163 nodes, 5,020 edges and 481 seed proteins. Genes with red color (see Table 2) represent the 16 top prioritized genes due to topological features i.e. degree of centrality >50. Orange, yellow and white colors represent genes with degree(s) of centrality <50.