ΓΕΩΠΟΝΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΤΩΝ ΖΩΩΝ
ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΖΩΙΚΗΣ ΠΑΡΑΓΩΓΗΣ
ΕΡΓΑΣΤΗΡΙΟ ΓΕΝΙΚΗΣ ΚΑΙ ΕΙΔΙΚΗΣ ΖΩΟΤΕΧΝΙΑΣ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Εντοπισμός γονιδιωματικών περιοχών και δικτύων γονιδίων που επηρεάζουν παραγωγικές και αναπαραγωγικές ιδιότητες σε πληθυσμούς κρεοπαραγωγικών ορνιθίων

ΕΙΡΗΝΗ Κ. ΤΑΡΣΑΝΗ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: ΑΝΤΩΝΙΟΣ ΚΟΜΙΝΑΚΗΣ

ΑΘΗΝΑ 2020

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Εντοπισμός γονιδιωματικών περιοχών και δικτύων γονιδίων που επηρεάζουν παραγωγικές και αναπαραγωγικές ιδιότητες σε πληθυσμούς κρεοπαραγωγικών ορνιθίων

Genome-wide association analysis and gene network analysis for (re)production traits in commercial broilers

ΕΙΡΗΝΗ Κ. ΤΑΡΣΑΝΗ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: ΑΝΤΩΝΙΟΣ ΚΟΜΙΝΑΚΗΣ

Τριμελής Επιτροπή:

Αντώνιος Κομινάκης (Αν. Καθ. ΓΠΑ)
Ανδρέας Κράνης (Ερευν. Β, Παν. Εδιμβούργου)
Αριάδνη Χάγερ (Επ. Καθ. ΓΠΑ)

Επταμελής εξεταστική επιτροπή:

Αντώνιος Κομινάκης (Αν. Καθ. ΓΠΑ)
Ανδρέας Κράνης (Ερευν. Β, Παν. Εδιμβούργου)
Αριάδνη Χάγερ (Επ. Καθ. ΓΠΑ)
Πηνελόπη Μπεμπέλη (Καθ. ΓΠΑ)
Δημήτριος Βλαχάκης (Επ. Καθ. ΓΠΑ)
Ευάγγελος Ζωίδης (Επ.Καθ. ΓΠΑ)
Γεώργιος Θεοδώρου (Επ.Καθ. ΓΠΑ)

Εντοπισμός γονιδιωματικών περιοχών και δικτύων γονιδίων που επηρεάζουν παραγωγικές και αναπαραγωγικές ιδιότητες σε πληθυσμούς κρεοπαραγωγικών ορνιθίων

Περίληψη

Σκοπός της παρούσας διδακτορικής διατριβής ήταν ο εντοπισμός γενετικών δεικτών και υποψηφίων γονιδίων που εμπλέκονται στο γενετικό έλεγχο δύο τυπικών πολυγονιδιακών ιδιοτήτων σε κρεοπαραγωγικά ορνίθια. Μία ιδιότητα σχετίζεται με την ανάπτυξη (σωματικό βάρος στις 35 ημέρες, ΣΒ) και η άλλη με την αναπαραγωγική ικανότητα (αριθμός αυγών ανά όρνιθα, ΑΩ). Για το σκοπό αυτό, διεξάχθηκαν γονιδιωματικής κλίμακας αναλύσεις συσχέτισης (GWAS) δεικτών-φαινοτυπικών τιμών και χρησιμοποιήθηκαν διάφορα εργαλεία Βιοπληροφορικής για τον εντοπισμό υποψηφίων λειτουργικών γονιδίων που επηρεάζουν έκαστη ή και τις δύο ιδιότητες, ταυτόχρονα.

Στο Κεφάλαιο 1 περιγράφεται η βέλτιστη στρατηγική δειγματοληψίας για τον εντοπισμό υπευθύνων γενετικών δεικτών και γονιδίων όταν λόγω περιορισμένων οικονομικών πόρων δεν είναι δυνατή η γονοτύπηση του συνόλου των ατόμων ενός πληθυσμού. Ειδικότερα, εξετάστηκε η αποτελεσματικότητα εναλλακτικών σεναρίων τα οποία περιελάμβαναν τυχαία υποσύνολα καθώς και υποσύνολα ακραίων φαινοτύπων μεγέθους 5% έως 50% ενός συνολικού πληθυσμού 6.700 ορνιθίων κρεοπαραγωγής. Το πιο αποτελεσματικό σενάριο δειγματοληψίας προσδιορίστηκε συγκρίνοντας τους στατιστικώς σημαντικούς δείκτες μεταξύ των υποσυνόλων και ολόκληρου του πληθυσμού. Στο βέλτιστο σενάριο, η αναζήτηση των πλέον πιθανών υποψηφίων δεικτών για την ιδιότητα βασίστηκε στο ποσοστό της φαινοτυπικής διακύμανσης που αυτοί εξηγούν χρησιμοποιώντας ένα κατώτερο όριο της τάξεως του 1%. Ως βέλτιστο σενάριο δειγματοληψίας αναδείχθηκε η χρησιμοποίηση του ημίσεος του συνολικού πληθυσμού το οποίο περιλαμβάνει ακραίες (υψηλές και χαμηλές) φαινοτυπικές τιμές για την ιδιότητα. Στο σενάριο αυτό, εντοπίστηκαν συνολικά πέντε δείκτες οι οποίοι εδράζουν εντός ή εγγύς έξι γονιδίων (*CACNB1, MYOM2, SLC20A1, ANXA4, FBXO32, SLAIN2*) και δέκα δημοσιευμένων ποσοτικών γονιδιακών τόπων (QTLs) σχετιζόμενων με την ανάπτυξη. Τα παρόντα ευρήματα έδειξαν ότι υπό συνθήκες περιορισμένων πόρων, η διεξαγωγή GWAS με το ήμισυ του πληθυσμού το οποίο περιλαμβάνει ακραίους φαινοτύπους για την ιδιότητα συνιστά μια αποτελεσματική στρατηγική για τον εντοπισμό υποψηφίων γονιδίων.

Στο Κεφάλαιο 2 διερευνάται η χρησιμότητα διαφόρων βιοπληροφορικών εργαλείων κατά την αναζήτηση λειτουργικών γονιδίων για ποσοτικές ιδιότητες όπως το ΣΒ. Αρχικά διεξήχθηκε ανάλυση συσχέτισης γενωμικών δεικτών-φαινοτυπικών τιμών για την ιδιότητα. Η ανάλυση αυτή ανέδειξε 12 στατιστικώς σημαντικούς δείκτες σχετιζόμενους με την ιδιότητα. Ακολούθησε αναζήτηση δημοσιευμένων ποσοτικών γονιδιακών τόπων (QTLs) και υποψηφίων γονιδίων εντός γενωμικών περιοχών εύρους 1 Mb γύρω από τους στατιστικώς σημαντικούς δείκτες. Η αναζήτηση αυτή ανέδειξε 1012 υποψήφια γονίδια θέσεως και 197 δημοσιευμένους QTLs σχετιζόμενους με την ανάπτυξη. Ακολούθησε λειτουργική ανάλυση (Functional Enrichment Analysis, FEA), ανάλυση ατραπών (Pathway Analysis, PA), τοπολογική ανάλυση δικτύου γονιδίων (Gene Network Analysis, GNA) και ανάλυση λειτουργικής προτεραιοποίησης (Gene functional Prioritization Analysis, GPA) με σκοπό την πρόβλεψη του λειτουργικού ρόλου των υποψηφίων γονιδίων. Η λειτουργική ανάλυση των υποψηφίων γονιδίων ανέδειξε 49 γονίδια ως εμπλεκόμενα σε αναπτυξιακές διαδικασίες ενώ 25 γονίδια βρέθηκαν να συμμετέχουν σε βιολογικές ατραπούς σχετιζόμενες με την ανάπτυξη. Η τοπολογική ανάλυση δικτύου γονιδίων και η ανάλυση λειτουργικής προτεραιοποίησης ανέδειξαν 14 κοινά γονίδια (*UBC, SMAD4, SHC1, NRAS, PSMD4, CDC6, PSMD7, RARA,*

PSMB4, CDH1, STAT5B, MED1, PSMD3, CDT1) ως πλέον σχετιζόμενα με την ιδιότητα. Η εφαρμογή των παραπάνω μεθόδων, είτε μεμονωμένα είτε συνδυαστικά, έδειξε ότι οι υπάρχουσες γνωσιακές βάσεις και τα διαθέσιμα εργαλεία μπορούν να συνεισφέρουν αποτελεσματικά στον εντοπισμό λειτουργικών γονιδίων για τις εξεταζόμενες ιδιότητες.

Στο Κεφάλαιο 3 επιχειρήθηκε ανίχνευση δομών[1] (modules) για τα υποψήφια γονίδια θέσεως και ακολούθως ανάλυση του λειτουργικού τους ρόλο με σκοπό τον εντοπισμό πλέον πιθανών γονιδίων για την ιδιότητα ΣΒ. Αρχικά, διενεργήθηκε γενωμική ανάλυση δεικτών-φαινοτυπικών τιμών. Ακολούθησε αναζήτηση δημοσιευμένων ποσοτικών γονιδιακών τόπων (QTLs) και γονιδίων θέσεως εντός γενωμικών περιοχών με υψηλά επίπεδα ανισορροπίας σύνδεσης ($D'>0.8$) γύρω από τους στατιστικώς σημαντικούς δείκτες. Η αναζήτηση αυτή ανέδειξε 645 υποψήφια γονίδια θέσεως γύρω από 11 στατιστικώς σημαντικούς δείκτες. Ακολούθησε εντοπισμός δομών εντός του δικτύου των υποψηφίων γονιδίων. Η ανάλυση αυτή ανέδειξε 5 δομές σχηματιζόμενες από 401 υποψήφια γονίδια. Η λειτουργική ανάλυση των γονιδίων που συνθέτουν τις παραπάνω δομές ανέδειξε 52 γονίδια ως συμμετέχοντα σε αναπτυξιακές διαδικασίες ενώ άλλα 14 γονίδια (*GABRG1, NGF, APOBEC2, STAT5B, STAT3, SMAD4, MED1, CACNB1, SLAIN2, LEMD2, ZC3H18, TMEM132D, FRYL, SGCB)* είχαν αποδεδειγμένα λειτουργικούς ρόλους σχετιζόμενους με την ανάπτυξη. Συνολικά, οι παραπάνω αναλύσεις ανέδειξαν 66 λειτουργικά γονίδια για την ιδιότητα, κάποια από τα οποία ήταν νέα και κάποια ήδη γνωστά.

Στο Κεφάλαιο 4 διεξήχθησαν αναλύσεις συσχέτισης γενωμικών δεικτών-φαινοτυπικών τιμών με σκοπό τον εντοπισμό δεικτών που ασκούν προσθετικές ή/και κυριαρχικές επιδράσεις στην αναπαραγωγική ικανότητα (ΑΩ) των ορνίθων κρεοπαραγωγής. Ακολούθησε ποσοτική διερεύνηση του τρόπου δράσης των δεικτών και εντοπισμός υποψηφίων γονιδίων για την ιδιότητα. Συνολικά, εντοπίσθηκαν 17 στατιστικώς σημαντικοί δείκτες σε επίπεδο χρωμοσώματος εκ των οποίων 7 σχετίστηκαν με προσθετικής φύσεως, 4 με κυριαρχικές και 6 και με τις δύο μορφές επιδράσεων (προσθετικές και κυριαρχικές). Ειδικότερα, οι 4 δείκτες κυριαρχίας σχετίστηκαν με φαινόμενα μερικής έως πλήρους κυριαρχίας. Η αναζήτηση υποψηφίων γονιδίων εντός γενωμικών περιοχών 50 kb γύρω από τους στατιστικώς σημαντικούς δείκτες κατέδειξε συνολικά 57 υποψήφια γονίδια θέσεως. Η ανάλυση λειτουργικής ενίσχυσης των υποψηφίων γονιδίων έδειξε ότι δύο από αυτά (*BHLHE40* και *CRTC1*) εμπλέκονται σε μηχανισμούς ανάδρασης του βιολογικού (cicardian) ρυθμού των πτηνών μέσω της φωτοπεριόδου. Επιπλέον, η ανάλυση λειτουργικής προτεραιοποίησης ανέδειξε επτά γονίδια (*GDF15, BHLHE40, JUND, GDF3, COMP, ELF3, CRTC1*) ως σχετιζόμενα με την αναπαραγωγή και δύο επιπλέον (*ITPR1, ELL*) με ποιοτικές ιδιότητες του αυγού. Συνολικά, τα ευρήματα του 4ου Κεφαλαίου υπογραμμίζουν τη σημασία και των μη προσθετικής φύσεως επιδράσεων στο γενετικό έλεγο της αναπαραγωγικής ικανότητας στις κρεοπαραγωγικές όρνιθες προτείνοντας παράλληλα νέα γονίδια για την ιδιότητα.

Στο Κεφάλαιο 5 αναζητήθηκαν οι υπεύθυνοι μηχανισμοί για την παρατηρούμενη αρνητική γενετική συσχέτιση (r_G=-0,17) μεταξύ του σωματικού βάρους (ΣΒ) και της ωοπαραγωγικής ικανότητας (ΑΩ) σε κρεοπαραγωγές όρνιθες. Αρχικά, εφαρμόστηκε διμεταβλητή ανάλυση συσχέτισης γενωμικών δεικτών - φαινοτυπικών τιμών για τις δύο ιδιότητες. Η ανάλυση αυτή ανέδειξε 51 σημαντικούς δείκτες εκ των οποίων οι 13 ήταν ανεξάρτητοι (περιοχές με χαμηλά επίπεδα ανισορροπίας σύνδεσης). Ακολούθησε αναζήτηση δημοσιευμένων ποσοτικών γονιδιακών τόπων (QTLs) και υποψηφίων γονιδίων θέσεως που περιλάμβαναν τους ανεξάρτητους δείκτες. Η αναζήτηση αυτή οδήγησε στον εντοπισμό 17 δημοσιευμένων ποσοτικών γονιδιακών τόπων (QTLs) και 14 υποψηφίων γονιδίων θέσεως. Ακολούθησε

[1] *Σ' ένα δίκτυο γονιδίων, ως δομή (gene module) ορίζεται ένα σύνολο γονιδίων του οποίου τα 'εσωτερικά' μέλη παρουσιάζουν υψηλότερο βαθμό συνδεσιμότητας (αριθμό αλληλεπιδράσεων) έναντι των 'εξωτερικών' μελών.*

εξέταση των βιολογικών διεργασιών ανά υποψήφιο γονίδιο, η οποία ανέδειξε δύο γονίδια *(ACVR1, CACNA1H)* ως συμμετέχοντα σε αναπτυξιακές και αναπαραγωγικές βιολογικές διεργασίες. Ειδικότερα, το γονίδιο *ACVR1* εμπλέκονταν άμεσα στην ανάπτυξη και στην αναπαραγωγή, ενώ το γονίδιο *CACNA1H* συμμετείχε άμεσα σε αναπαραγωγικές και έμμεσα σε αναπτυξιακές μεταβολικές λειτουργίες. Σύμφωνα με τη βιβλιογραφία, τα δύο προαναφερθέντα γονίδια παρουσίασαν τεκμηριωμένη εμπλοκή σε αναπτυξιακές και αναπαραγωγικές βιολογικές λειτουργίες μέσω της συμμετοχής τους σε διάφορες βιοχημικές ατραπούς. Τα ευρήματα του παρόντος Κεφαλαίου συνεισφέρουν στην βαθύτερη κατανόηση του γενετικού μηχανισμού που ευθύνεται για την αρνητική γενετική συσχέτιση μεταξύ της ανάπτυξης και της αναπαραγωγής στα ορνίθια κρεοπαραγωγής, προτείνοντας υποψήφιους γενωμικούς δείκτες και γονίδια με πλειοτροπική δράση.

Επιστημονικό πεδίο: Γονιδιωματική

Λέξεις κλειδιά: κρεοπαραγωγικά ορνίθια**,** σωματικό βάρος στις 35 ημέρες, αριθμός αυγών ανά όρνιθα, γονιδιωματικής κλίμακας αναλύσεις συσχέτισης, βέλτιστη στρατηγική δειγματοληψίας, εργαλεία Βιοπληροφορικής, επιδράσεις κυριαρχίας, πλειοτροπική δράση.

Genome-wide association analysis and gene network analysis for (re)production traits in commercial broilers

Abstract

Aim of the present PhD thesis was to identify genetic variants and plausible functional genes for two typical polygenic traits in broilers, one related to growth (body weight at 35 days of age, BW) and the other with reproduction efficiency (number of eggs per female broiler, EN). To this end, numerous genome-wide association studies (GWAS) were conducted to detect marker-trait associations and various post-GWAS *in silico* methods were applied in efforts to discover (novel) functional candidate genes for individual and joint traits.

Chapter 1 presents an optimal sampling strategy when conducting cost-effective GWASs. To this end, 19 GWASs for BW were conducted using random and extreme phenotype (continuous and dichotomized) samples with sizes ranging from 5% to 50% of a total population of 6700 broilers. The most efficient sampling scenario was identified by comparing genome-wide significant marker signals between sub-samples and the whole population. This comparison pointed out 50% extreme phenotype sampling as the optimal sampling scenario. In the optimal sampling scenario, putative genetic variants were selected using a threshold of 1% for the Proportion of Variance Explained associated with markers. This search strategy resulted in identification of a total number of five putative causal genetic variants. These variants resided in genomic regions harboring ten growth-related QTLs (e.g. breast muscle percentage, abdominal fat weight etc.) and six growth related genes (*CACNB1, MYOM2, SLC20A1, ANXA4, FBXO32, SLAIN2*). Chapter findings proposed the use of 50% extreme phenotype sampling as the optimal sampling strategy to detect causative genes when performing cost-effective GWASs.

Chapter 2 investigates the utility of various computational (*in silico*) techniques to propose functional candidate genes for BW. First, a GWAS for BW was conducted to detect significant marker-trait associations. This analysis pointed out twelve genome-wide significant SNPs across nine autosomes. The search for positional candidate genes in 1 Mb flanking regions around the significant markers revealed a total number of 1,012 candidate genes and 197 growth-related QTL/associations. Implementation of Functional Enrichment Analysis (FEA), Pathway Analysis (PA), Gene Network Analysis (GNA) and Gene functional Prioritization Analysis (GPA) to predict functional relevance of the candidate genes followed. FEA pointed out 49 candidate genes participating in system development while PA highlighted 25 member genes in growth-related pathways. GPA and GNA pointed out 14 common genes (*UBC, SMAD4, SHC1, NRAS, PSMD4, CDC6, PSMD7, RARA, PSMB4, CDH1, STAT5B, MED1, PSMD3* and *CDT1*) with highest functional relevance to the trait under study. Application of the preceding computational methods has, individually or jointly, demonstrated that extant knowledge and available tools can be useful to prioritize most likely candidates for traits under investigation.

In Chapter 3, the functional role of modular genes (genes organized into dense sub-networks) was explored in efforts to identify causal genes for BW. To this end, first a GWAS for BW was carried out. Then, genomic regions around the significant SNPs showing strong linkage disequilibrium levels ($D' > 0.8$) were searched to identify positional candidate genes and reported QTLs for the trait. This search revealed a total number of 645 positional candidate genes around 11 genome-wide significant markers. Community structure analysis was then conducted to detect densely interconnected nodes (modules) of the positional candidate genes. This analysis detected five modules formed by 401 candidate genes. Functional enrichment analysis of the modular genes showed 52 genes as participating in developmental processes. 14 more modular genes *(GABRG1, NGF, APOBEC2, STAT5B, STAT3, SMAD4, MED1,*

CACNB1, SLAIN2, LEMD2, ZC3H18, TMEM132D, FRYL and *SGCB)* had evidenced growth functional relevance. In the present Chapter, a total number of 66 functional candidate genes for BW were proposed, some of which were novel and some identified in previous studies. Chapter 4 aimed to provide the genetic variants impacting on egg number (EN) in female broilers. To this end, additive and dominant genetic models were applied to detect trait-marker associations. This analysis resulted in identification of a total number of 17 chromosome-wide significant markers. Of these, 7 were additive, 4 dominant and 6 additive plus dominant. Degree of dominance for the purely dominant markers ranged from partial to complete dominance. A total number of 57 positional candidate genes were identified within 50 kb flanking regions around the significant markers. Functional enrichment analysis of the positional candidate genes pinpointed two genes (*BHLHE40* and *CRTC1*) to be involved in the 'entrainment of circadian clock by photoperiod' biological process. Gene prioritization analysis of the positional candidate genes identified ten top ranked genes (*GDF15, BHLHE40, JUND, GDF3, COMP, ITPR1, ELF3, ELL, CRLF1* and *IFI30*). Seven prioritized genes (*GDF15, BHLHE40, JUND, GDF3, COMP, ELF3, CRTC1*) had documented functional relevance to reproduction, while two more prioritized genes (*ITPR1* and *ELL*) are reported to be related to egg quality in female chickens.
Chapter 5 aimed to shed light in the genetic mechanism underlying the antagonistic (negative) genetic correlation between growth (BW) and reproduction (EN) in broilers. First, a bivariate marker-traits association analysis was performed. This analysis resulted in detection of a total number of 51 genome-wide significant markers across 21 autosomes. Application of stepwise conditional-joint analyses pinpointed a total number of 13 independent markers. These independent markers were located within 17 reported QTLs and within or close proximity to 14 positional candidate genes. Examination of the Gene Ontology (GO) Biological Process (BP) profile per candidate gene highlighted two genes (*ACVR1* and *CACNA1H*) that were participating in relevant BPs to the traits under study. Of these genes, *ACVR1* was directly involved in both developmental and reproductive processes while *CACNA1H* was directly participating in reproduction and indirectly in growth via metabolic processes. Literature evidence of the functions of the two aforementioned genes (*ACVR1* and *CACNA1H*) enhanced their candidacy as pleiotropic genes for the examined traits and set them as typical gene exemplars of horizontal pleiotropy. Chapter findings provide novel insight in the nature of the antagonistic co-variation between growth and reproduction in female broilers.

Scientific field: Genomics

Keywords: broilers**,** body weight at 35 days of age, number of eggs per female broiler, genome-wide association studies, optimal sampling strategy, computational approaches, dominant genetic effects, pleiotropy.

Acknowledgements

Throughout the PhD, I acknowledge:

- Dr. Santiago Avendano (Aviagen Ltd) for financial support and data provision.
- Dr. A. Kranis and Dr. A. Hager for fruitful and helpful discussions in Genomics. Most and foremost, I would like to thank my supervisor A. Kominakis who shared his knowledge in statistical Genetics, guided me into the right direction whenever necessary and helped me to struggle.
- Dr. G. Maniatis (Aviagen Ltd) for data preparation.
- My family for their support.

Table of Contents

Περίληψη **3**

Abstract **6**

General Introduction **13**

Chapter 1 ..... **18**

Identification of Candidate Genes for Body Weight in Broilers Using Extreme-Phenotype Genome-Wide Association Study **18**

1.1. Abstract 18

1.2. Introduction 18

1.3. Material and methods 19

1.3.1. Animals and SNPs 19

1.3.2. Sampling scenarios 20

1.3.3. Marker-trait association analysis 20

1.3.4. Quantile-quantile plots and estimation of the genomic inflation factor 21

1.3.5. Proportion of variance explained by SNP per sampling case 21

1.3.6. Estimation of allelic effects based on PVE 21

1.3.7. Detection efficacy across the various sampling scenarios 21

1.3.8. Identification of putative causative genetic variants 21

1.3.9. Variant effect prediction of putative causative genetic variants 22

1.4. Results 22

1.4.1. Identification of significant SNPs in whole population sampling 22

1.4.2. Detection efficacy across the various sampling scenarios and estimated PVE 22

1.4.3. Identification and effect prediction of putative causative genetic variants 23

1.5. Discussion 23

1.5.1. Detection efficacy of EPS 23

1.5.2. Detection of causative genetic variants 23

1.6. Conclusions 24

1.7. References 25

Tables and Figures of Chapter 1 28

Chapter 2 ..... **43**

Elucidating the functional role of 1,012 candidate genes revealed by a Genome Wide Association Study for body weight in broilers **43**

2.1. Abstract 43

2.2. Introduction 43
2.3. Material and methods 44
2.3.1. Data 44
2.3.2. Statistical analysis 44
2.3.3. Identification of QTL and positional candidate genes 45
2.3.4. Functional enrichment and pathway analysis 45
2.3.5. Gene prioritization 46
2.4. Results 46
2.4.1. Significant SNPs and positional candidate genes 46
2.4.2. Detection of QTL/associations 46
2.4.3. Functional enrichment and pathway analysis 47
2.4.4. Gene prioritization and network analysis 47
2.5. Discussion 47
2.6. References 49
Chapter 3 58
Discovery and characterization of functional modules associated with body weight in broilers 58
3.1. Abstract 58
3.2. Introduction 58
3.3. Methods 59
3.3.1. Ethics Statement 59
3.3.2. Data and quality control 59
3.3.3. Statistical analysis 59
3.3.4. Detection of candidate genomic regions with strong LD 60
3.3.5. Identification of reported QTL and positional candidate genes 60
3.3.6. Detection of community structure and functional module characterization 60
3.4. Data Availability 61
3.5. Results 61
3.5.1. Significant SNPs and positional candidate genes 61
3.5.2. Reported QTL/associations 62
3.5.3. Detection of community structure 63
3.5.4. Functional enrichment analysis per module 63
3.5.5. Functional candidate genes 63
3.6. Discussion 63
3.7. References 65

Tables and Figures of Chapter 3 71
Chapter 4 82
Deciphering the mode of action and position of genetic variants impacting on egg number in broiler breeders 82
4.1. Abstract 82
4.2. Introduction 82
4.3. Material and methods 83
4.3.1. Data 83
4.3.2. Marker-trait association analysis 83
4.3.3. Quantile-quantile plots and estimation of the genomic inflation factor 84
4.3.4. Estimation of genomic heritability and proportion of variance explained 84
4.3.5. Identification of significant SNPs under multicollinearity conditions 84
4.3.6. Estimation of the degree of dominance 84
4.3.7. Detection, functional characterization and prioritization of positional candidate genes 85
4.4. Results 85
4.4.1. Significant SNPs and PVE 85
4.4.2. Estimation of the degree of dominance 86
4.4.3. Positional candidate genes 87
4.4.4. Functional enrichment analysis 87
4.4.5. Prioritized genes 87
4.5. Discussion 87
4.5.1. Mode of gene action 87
4.5.2. Functional candidate genes 88
4.6. Conclusions 89
4.7. References 90
Tables and Figures of Chapter 4 94
Chapter 5 103
Detection of pleiotropic loci with antagonistic effects on body weight and egg number in chickens 103
5.1. Abstract 103
5.2. Introduction 103
5.3. Material and methods 104
5.3.1. Data and quality control 104
5.3.2. Univariate and bivariate association analyses 104
5.3.3. Multiple-testing correction 106

5.3.4. Selection of independent SNPs 106
5.3.5. Effect prediction of the independent SNPs and identification of positional candidate genes and published QTLs 106
5.3.6. Functional profile of candidate genes and parent GO terms 106
5.4. Results 107
5.4.1. Significant SNPs obtained from univariate and bivariate analyses 107
5.4.2. Independent SNP signals 107
5.4.3. Effect prediction of the independent SNPs and identification of positional candidate genes and published QTLs 108
5.4.4. GO term profiling of candidate genes and GO slim categories 108
5.5. Discussion 108
5.6. References 110
Tables and Figures of Chapter 5 115
6. General discussion 125
Appendix – Supplementary Tables and Figures 132
Chapter 2 132
Chapter 3 191
Chapter 4 247
Chapter 5 257

General Introduction

1. Quantitative traits

In livestock, the most economically important traits are the quantitative (or complex or polygenic) traits. Apart from economic interest, these traits present also considerable interest from a biological point of view as they exhibit particular features: (i) phenotypic values follow continuous variation (often the normal distribution) (ii) are genetically controlled by an infinite number of loci and each locus has an infinitely small effect (Fisher's [1] infinitesimal model) and (iii) systematic environmental effects also contribute to phenotypic variation. Gradually, the infinitesimal model has been replaced by an finite genetic model, which states that quantitative traits are controlled by a few genes with large effects and many genes with small effects [2].

For many years, the 'black box' of complex traits (i.e. the exact number, genomic location and mode of action of causative genes) remained unrevealed. In an attempt to shed light to the 'black box', the interval mapping of Quantitative Trait Loci (QTL) approach [3] was initially employed. This approach was based on crosses between parental populations displaying maximally different phenotypes and limited number of DNA markers. Due to the small number of segregating polymorphic loci between the extreme populations, the aforementioned studies had limited potential for discovery of genetic associations. As a result, only large QTL intervals [4] and a limited number of trait-associated loci could be identified [5].

2. Genome Wide Association Studies

The advent of high-throughput genotyping technology in form of single nucleotide polymorphism (SNP) arrays has dramatically changed the landscape of gene discovery. Availability of thousands or millions of SNPs ensures the dense genome coverage and allows for efficient screening and detection of markers through genome-wide association studies (GWAS) [6]. Nevertheless, the sample sizes needed to identify SNPs that explain most of heritability (e.g. 80%) of polygenic traits are predicted to range from a few hundred thousand to several millions, depending on the underlying effect-size distributions of the traits [7]. In livestock populations, such sample sizes are unable to attain, at least for now, due to insufficient budgets to fully cover the expense of complete genotyping all phenotyped animals. In such cases, a useful cost-saving strategy is Selective Genotyping in which only a selected fraction of the animals with extreme phenotypic values for the trait under study, are genotyped [2,8]. Studies employing extreme phenotypes are called Extreme-Phenotype GWAS (XP-GWAS, [5]) and are considered to be particularly valuable for detecting genes responsible for quantitative variation in various species [12,10,13].

A typical GWAS includes the following steps [9]: (i) data pre-processing (i.e. applying quality control criteria at the marker and sample level), (ii) checking for population structure (e.g. via estimating the genomic inflation factor and visually examining the Quantile-Quantile (QQ) plot) and (iii) performing single-locus or multi-locus association analysis that involves regressing single SNP or multiple SNPs on a given trait, respectively. In GWA analysis, the additive, (over)dominant or/and recessive genetic model(s) is/are used to screen the genome and then identify SNP-trait associations [10]. For association analysis, linear mixed models [11,12] have been shown to be capable of correcting population structure, family relatedness and/or cryptic relatedness via the kinship matrix inferred from genome-wide markers (genomic relationship matrix). Particularly, multi-locus mixed models (MLMM [13]) involve a multi-dimensional genome scan, in which the effects of all markers are simultaneously estimated. Multi-locus GWAS methods are reported to improve the statistical power and

decrease the number of false positives when compared to single-locus mixed models [14] and thus seem to be a 'gold standard' to perform marker-trait association analyses.

3. Linkage Disequilibrium

GWAS rely on the principle of LD which describes the degree to which an allele of one SNP is correlated (or inherited) with an allele of another SNP. As a result, a significant SNP signal may pinpoint either a *direct* or an *indirect marker-trait association*. In the first case, a *functional SNP* that is influencing a biological system and ultimately affects the phenotype is detected. In the second scenario, the influential SNP is not directly typed, but instead a tag SNP in high LD with the influential SNP is typed and statistically associated to the phenotype. Because of these two possibilities, a significant SNP association from a GWAS should not be always assumed as the causal variant. To alleviate high SNP interdependency arising from LD between markers, *pruning*[1], *clumping*[2] or *joint and conditional association analysis*[3] are usually applied to obtain independent SNPs. Apart from avoiding false positives, relaxing SNPs dependency is also important when dealing with the large-scale testing problem as some multiple testing correction methods (e.g. the *Bonferroni correction* and the *Benjamini-Hochberg procedure)* assume that individual tests are independent of each other.

4. Search of causative genes

Intuitively, genes including lead SNPs and at the same time presenting functional relevance with the trait under study are considered ideal functional candidates. Yet, due to LD, the true causative genes may lie tens or hundreds kb up- and downstream from the significant SNPs. Because of this, the search for candidate genes around the lead SNPs comprises wider genomic distances that typically range from 50 to 1000 kb. Alternatively, this search can be performed within regions showing strong LD levels (e.g. D'>0.80) around the statistically significant SNP [15]. This search may result to tens or hundreds of positional candidate genes and assessing gene candidacy is an important step in efforts to produce a manageable subset of variants for further validation or exploration.

5. Gene functional prediction

When functional relevance of the candidate genes is unknown or limited, computational approaches such as gene set enrichment analysis [16], pathway analysis [17], gene network analysis [18] and gene prioritization analysis [19] can be employed to predict their functional relevance. By mapping genes to their associated biological annotations (such as gene ontology (GO) terms or pathway membership) and then comparing the distribution of the terms within a gene set of interest with the background distribution of these terms, enrichment analysis can identify terms which are statistically over-or under-represented within a given gene list [20]. However, it is widely thought that to understand and predict gene function, genes must be studied in the context of networks. Gene network analysis exploits several interactions [21] (e.g. protein-protein interaction networks (PPINs), interactions by RNA co-expression, literature-curated interactions and interactions derived from high-throughput experiments) and is based on the '*guilt-by-association'* (GBA [22]) principle. The GBA

---

[1] *discard pairs of SNPs that are in high or complete LD by removing one SNP from the correlated pair and keeping the one with the largest minor allele frequency*

[2] *sort the SNPs by importance (e.g. p-value) and keep only one representative SNP per region of LD*

[3] *selects the SNP with the lowest p value for conditioning the effect on neighboring loci based on the LD between the neighboring SNPs and the selected SNP*

principle states that genes with related function tend to be protein interaction partners or share features such as expression patterns. The GBA principle has been applied in gene prioritization analysis to predict novel genes for complex traits [23]. The latter relies on functional annotation similarity between the input genes and phenotypic keywords or known seed genes [24].
Genes in PPINs are organized into densely linked clusters i.e. communities or modules [25]. Modules present a structurally independent gene sub-network with more interior connections and consist of proteins which have the same or similar biological function(s) [26]. Modules could be further distinguished in protein complexes and in dynamic functional modules. Protein complexes are formed by several proteins which interact at the same place and time while dynamic functional modules are composed of few proteins participating in a specific cellular function not necessarily at the same place and time [27]. Moreover, functional modules consist of one or multiple protein complexes participating in a common biological process [28]. Modules do not emerge by chance and they can reveal interactions with biological importance within large PPINs [27]. For this reason, they can be used to efficiently cluster genes into functional groups and to predict protein functions [29].

6. Cross Phenotype Associations and Pleiotropy

GWAS may pinpoint the existence of genetic variants that are associated with multiple, sometimes seemingly distinct, quantitative traits. Such associations are termed cross-phenotype (CP) associations [30] and are potential evidence for pleiotropy. An important issue arising here is as how to distinct a *CP association* from true *pleiotropy*. The first occurs when a genetic locus is associated with more than one trait regardless of the underlying cause for the observed association [31], while the second arises when a genetic locus truly affects more than one trait and is one possible underlying cause for an observed CP association[30]. Detection of CP associations can be explored via multivariate or univariate statistical approaches in GWAS. While multivariate approaches [32] allow for direct identification of CP associations, in the context of univariate analyses, detection of CP associations relies on aggregating results of single traits analyses via meta-analysis techniques [33]. When searching for pleiotropic markers via GWAS it is important to obtain independent CP as a marker can be falsely associated with multiple phenotypes due to LD (spurious pleiotropy). Application of LD pruning [34] and/or conditional and joint analysis [35] of the SNP signals can effectively serve to this purpose.

7. Aim and structure of the thesis

The present doctoral thesis was designed to identify genomic markers and candidate genes attributable to the genetic control of two typical quantitative traits in *Gallus gallus* (chicken). One trait is associated with growth (body weight at 35 days, BW) and the other with reproductive efficiency (Egg Number per hen and year, EN). As the two traits display an antagonistic (negative) genetic correlation, a final goal of the current thesis was to identify genomic markers and possibly causal genes associated with antagonistic effects on the two traits. The thesis is comprised by 5 Chapters and is organized as follows.
Chapter 1 aimed to propose an optimal sampling strategy to performing cost effective GWAS for a typical complex trait such as BW at 35 days of age.
In ***Chapter 2***, various computational approaches (i.e. functional enrichment analysis, pathway analysis, GBA-based gene prioritization analysis and gene network analysis) were exploited to elucidate the functional role and prioritize candidate genes for BW.
In ***Chapter 3,*** an alternative approach for BW was followed. Here, it was investigated whether the trait is associated with *functional modules* and if so, they could be used to discover novel candidate gene for the trait under study.

Chapter 4 was designed to identify genetic variants associated with EN in female broilers, to describe the mode of their gene action (additive and/or dominant) and finally to provide a list of the implicated candidate genes for the trait.
Finally, aim of ***Chapter 5*** was to identify genetic variants and plausible candidate genes contributing to the negative genetic correlation observed between BW and EN in female broilers.

References

[1] Fisher RA. XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. Trans R Soc Edinburgh. Royal Society of Edinburgh Scotland Foundation; 1919;52: 399–433. doi:10.1017/S0080456800012163

[2] Hayes B, Goddard ME. The distribution of the effects of genes affecting quantitative traits in livestock. Genet Sel Evol. 2001;33: 209–229. doi:10.1186/1297-9686-33-3-209

[3] Thoday JM. Location of polygenes. Nature. 1961;191: 368–370. doi:10.1038/191368a0

[4] Flint J, Valdar W, Shifman S, Mott R. Strategies for mapping and cloning quantitative trait genes in rodents. Nat Rev Genet. Nature Publishing Group; 2005;6: 271–286. doi:10.1038/nrg1576

[5] Yang J, Jiang H, Yeh C-T, Yu J, Jeddeloh JA, Nettleton D, et al. Extreme-phenotype genome-wide association study (XP-GWAS): a method for identifying trait-associated variants by sequencing pools of individuals selected from a diversity panel. Plant J. John Wiley & Sons, Ltd (10.1111); 2015;84: 587–596. doi:10.1111/tpj.13029

[6] Gumpinger AC, Roqueiro D, Grimm DG, Borgwardt KM. Methods and tools in genome-wide association studies. Methods in Molecular Biology. Humana Press Inc.; 2018. pp. 93–136. doi:10.1007/978-1-4939-8618-7_5

[7] Zhang Y, Qi G, Park J-H, Chatterjee N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. Nat Genet. Nature Publishing Group; 2018;50: 1318–1326. doi:10.1038/s41588-018-0193-x

[8] Lander ES, Botstein D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics. Genetics Society of America; 1989;121: 185–99. Available: http://www.ncbi.nlm.nih.gov/pubmed/2563713

[9] Reed E, Nunez S, Kulp D, Qian J, Reilly MP, Foulkes AS. A guide to genome-wide association analysis and post-analytic interrogation. Stat Med. 2015;34: 3769–3792. doi:10.1002/sim.6605

[10] Wang MH, Cordell HJ, Van Steen K. Statistical methods for genome-wide association studies. Seminars in Cancer Biology. Academic Press; 2019. pp. 53–60. doi:10.1016/j.semcancer.2018.04.008

[11] Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S, Freimer NB, et al. Variance component model to account for sample structure in genome-wide association studies. Nat Genet. Nature Publishing Group; 2010;42: 348–354. doi:10.1038/ng.548

[12] Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of mixed-model association methods. Nat Genet. NIH Public Access; 2014;46: 100–6. doi:10.1038/ng.2876

[13] Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, et al. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet. Nature Publishing Group; 2012;44: 825–830. doi:10.1038/ng.2314

[14] Cui Y, Zhang F, Zhou Y. The Application of Multi-Locus GWAS for the Detection of Salt-Tolerance Loci in Rice. Front Plant Sci. 2018;9. doi:10.3389/fpls.2018.01464

[15] Yi G, Shen M, Yuan J, Sun C, Duan Z, Qu L, et al. Genome-wide association study dissects genetic architecture underlying longitudinal egg weights in chickens. BMC Genomics. BioMed Central; 2015;16: 746. doi:10.1186/s12864-015-1945-y

[16] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. National Academy of Sciences; 2005;102: 15545–50. doi:10.1073/pnas.0506580102

[17] García-Campos MA, Espinal-Enríquez J, Hernández-Lemus E. Pathway Analysis: State of the Art. Front Physiol. Frontiers; 2015;6: 383. doi:10.3389/fphys.2015.00383

[18] Leiserson MDM, Eldridge J V., Ramachandran S, Raphael BJ. Network analysis of GWAS

data. Current Opinion in Genetics and Development. 2013. pp. 602–610. doi:10.1016/j.gde.2013.09.003
[19] Bromberg Y. Chapter 15: Disease Gene Prioritization. Lewitter F, editor. PLoS Comput Biol. 2013;9: e1002902. doi:10.1371/journal.pcbi.1002902
[20] Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37: 1–13. doi:10.1093/nar/gkn923
[21] Gillis J, Pavlidis P. The Impact of Multifunctional Genes on "Guilt by Association" Analysis. Bader J, editor. PLoS One. 2011;6: e17258. doi:10.1371/journal.pone.0017258
[22] Walker MG, Volkmuth W, Sprinzak E, Hodgson D, Klingler T. Prediction of gene function by genome-scale expression analysis: prostate cancer-associated genes. Genome Res. 1999;9: 1198–203. Available: http://www.ncbi.nlm.nih.gov/pubmed/10613842
[23] Lee I, Blom UM, Wang PI, Shim JE, Marcotte EM. Prioritizing candidate disease genes by network-based boosting of genome-wide association data. Genome Res. Cold Spring Harbor Laboratory Press; 2011;21: 1109–21. doi:10.1101/gr.118992.110
[24] Moreau Y, Tranchevent LC. Computational tools for prioritizing candidate genes: Boosting disease gene discovery. Nature Reviews Genetics. 2012. pp. 523–536. doi:10.1038/nrg3253
[25] Liu G, Wang H, Chu H, Yu J, Zhou X. Functional diversity of topological modules in human protein-protein interaction networks. Sci Rep. Nature Publishing Group; 2017;7: 16199. doi:10.1038/s41598-017-16270-z
[26] Terentiev AA, Moldogazieva NT, Shaitan K V. Dynamic proteomics in modeling of the living cell. Protein-protein interactions. Biochemistry (Mosc). 2009;74: 1586–607. Available: http://www.ncbi.nlm.nih.gov/pubmed/20210711
[27] Spirin V, Mirny LA. Protein complexes and functional modules in molecular networks. Proc Natl Acad Sci U S A. National Academy of Sciences; 2003;100: 12123–8. doi:10.1073/pnas.2032324100
[28] Lu H, Shi B, Wu G, Zhang Y, Zhu X, Zhang Z, et al. Integrated analysis of multiple data sources reveals modular structure of biological networks. Biochem Biophys Res Commun. Academic Press; 2006;345: 302–309. doi:10.1016/J.BBRC.2006.04.088
[29] Sharan R, Ulitsky I, Shamir R. Network-based prediction of protein function. Mol Syst Biol. 2007;3: 88. doi:10.1038/msb4100129
[30] Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: Challenges and strategies. Nature Reviews Genetics. 2013. pp. 483–495. doi:10.1038/nrg3461
[31] Tyler AL, Crawford DC, Pendergrass SA. The detection and characterization of pleiotropy: discovery, progress, and promise. Brief Bioinform. 2016;17: 13–22. doi:10.1093/bib/bbv050
[32] Zhou X, Stephens M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat Methods. NIH Public Access; 2014;11: 407–9. doi:10.1038/nmeth.2848
[33] Li X, Zhu X. Cross-phenotype association analysis using summary statistics from GWAS. Methods in Molecular Biology. Humana Press Inc.; 2017. pp. 455–467. doi:10.1007/978-1-4939-7274-6_22
[34] Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, et al. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. Int J Methods Psychiatr Res. John Wiley and Sons Ltd; 2018;27. doi:10.1002/mpr.1608
[35] Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88: 76–82. doi:10.1016/j.ajhg.2010.11.011

Chapter 1

Identification of Candidate Genes for Body Weight in Broilers Using Extreme-Phenotype Genome-Wide Association Study[1]

1.1. Abstract

Traditionally, genome-wide association studies (GWAS) require maximum numbers of genotyped and phenotyped animals to efficiently detect marker-trait associations. Under financial constraints, alternative solutions should be envisaged such that of performing GWAS with fractioned samples of the population. In the present study, we investigated the potential of using random and extreme phenotype samples of a population including 6,700 broilers in detecting significant markers and candidate genes for a typical complex trait (body weight at 35 days). We also explored the utility of using continuous vs. dichotomized phenotypes to detect marker-trait associations. Present results revealed that extreme phenotype samples were superior to random samples while detection efficacy was higher on the continuous over the dichotomous phenotype scale. Furthermore, the use of 50% extreme phenotype samples resulted in detection of 8 out of the 10 markers identified in whole population sampling. Putative causative variants identified in 50% extreme phenotype samples resided in genomic regions harboring 10 growth-related QTLs (e.g. breast muscle percentage, abdominal fat weight etc.) and 6 growth related genes (*CACNB1, MYOM2, SLC20A1, ANXA4, FBXO32, SLAIN2).* Current findings proposed the use of 50% extreme phenotype sampling as the optimal sampling strategy when performing a cost-effective GWAS.

Keywords: Body Weight, Broilers, Extreme Phenotypes

1.2. Introduction

Quantitative Trait Loci (QTL) detection presumes the availability of both phenotypic trait values and marker genotypic data. In livestock populations where extensive individual performance recording takes place, collection and availability of phenotypic data on large numbers of animals is an ongoing situation. When costs are not of primary concern, all individuals with phenotypic data are genotyped and included for QTL analysis. However, this is seldom the case and under a limited budget it is necessary to make an effective allocation of genotyping costs. The latter could be extremely high for large sized populations and the high-throughput genotyping technology [1]. A useful genotyping cost-saving strategy is selective genotyping (SG) in which only a selected fraction of the phenotyped individuals, are genotyped [2,3].

The efficacy of SG to locate QTL has been extensively evaluated in simplified settings i.e. a single locus contributing to the phenotype. The basic experimental design was based on

[1] *Published in International Journal of Genetics and Genomics. Vol. 8, No. 1, 2020, pp. 29-40. doi: http://www.sciencepublishinggroup.com/journal/paperinfo?journalid=116&doi=10.11648/j.ijgg.20200801.14*

Current research was co-financed by Greece and the European Union (European Social Fund- ESF) through the Operational Programme «Human Resources Development, Education and Lifelong Learning 2014-2020» in the context of the project "Genome Wide Association Scanning and Gene Network Analysis of Body Weight in broilers" (MIS 5005869).

segregating populations arising from crosses (backcrossing or intercross) between parental populations displaying maximally different phenotypes. Following this approach, numerous studies have been carried out in an attempt to address issues related to the utility of extreme vs. random samples, the type of sampling strategy (one-tail vs. two-tail, symmetrical vs. asymmetrical sampling) and the optimum proportion(s) of selected samples [4-8].
Due to the small number of segregating polymorphic loci between the extreme populations, the aforementioned studies had limited potential for discovery of genetic associations. As a result, only large QTL intervals [9] and a limited number of trait-associated loci could be identified [10].
In modern GWAS, where diverse populations are used, the sample sizes needed to identify SNPs that explain most of heritability (e.g. 80%) of polygenic traits are predicted to range from a few hundred thousand to several millions, depending on the underlying effect-size distributions of the traits [11]. Samples of the above sizes are now in hand in human studies due to the successes of the large-scale consortia[11]. In livestock populations as well as in wild animal species, such sample sizes are unable to be realized, at least for now, due to insufficient budgets to fully cover the expense of complete genotyping all phenotyped animals.
Extreme-phenotype GWAS (XP-GWAS, [10]) are reported to be particularly valuable for detecting genes or alleles responsible for quantitative variation in species [12,10,13]. Furthermore, extreme phenotype sampling (EPS) is more effective in detecting (rare) variants when compared to random sampling (RS) [7,14,15] and EPS may deliver similar results when compared to a whole population GWAS [13].
So far, most of the studies carried out have aimed to compare the utility of RS and EPS when using comparable sample sizes while others have explored the statistical properties of using continuous vs. dichotomous phenotypes. Moreover, only a limited number of studies have focused on comparison of fractioned samples vs. whole population sampling and on detection of candidate genes for quantitative traits. Driven from the above, we have elaborated the present empirical study with the overarching aim to propose optimal sampling strategies when performing cost effective GWAS. The present report is organized as follows. First, we conducted a GWAS using all animals of a population (whole population sampling, PS) consisted of 6,700 broilers to identify significant SNPs associated with a typical complex trait (body weight at 35 days of age). Next, we selected random and extreme phenotype samples of progressively increasing sizes up to 50% of the whole population and identified SNP signals using continuous and dichotomized phenotypes. Finally, we compared SNP signals between sub-samples (extreme and random) and the whole population and examined whether sub-sampling may lead to the discovery of most plausible functional candidate genes for the trait.

1.3. Material and methods

1.3.1. Animals and SNPs

Genotypes from 6,700 broilers (3,718 males and 2,982 females) with corresponding records on BW at 35 days of age (average=2007.5 g, SD=222.8 g) were made available by Aviagen Ltd. The genotyping was conducted with the 600k Affymetrix HD SNP array [16] and included a total number of 547,904 autosomal SNPs dispersed on 28 chromosomes (GGA1-28). We applied the following quality control (QC) criteria at the marker level i.e. markers were excluded if: call rate<0.95, minor allele frequency (MAF)<0.05 and LD (r^2) values>0.99 for genomic distances up to 1 Mb. After application of QC, a final number of 215,555 SNPs remained for further analyses. Marker QC was carried out using the SNP & Variation Suite version 8.8.1 software (Golden Helix: http://www.goldenhelix.com).

1.3.2. Sampling scenarios

The first sampling scenario considered was RS. In this, random samples as high as 5% (RS_5%, n=335), 10% (RS_10%, n=670), 20% (RS_20%, n=1,340), 30% (RS_30%, n=2,010), 40% (RS_40%, n=2,680) and 50% (RS_50%, n=3,350) of the whole population were taken. Adjustment of BW records for three statistically significant ($p<0.05$) fixed effects: sex (n=2 classes), hatch (n=36 classes) and mating group (n=17 classes) followed, based on the least squares estimates of each class effect. We then performed EPS by taking fractions as high as 5% (EPS_5%), 10% (EPS_10%), 20% (EPS_20%), 30% (EPS_30%), 40% (EPS_40%) and 50% (EPS_50%) of the lower and upper tails of the adjusted phenotypic records. Only symmetrical sampling with equal fractions of the two extremes (low and high) was considered here. One more sampling scenario was also considered by dichotomizing the continuous extreme phenotypes and treating the two extremes as two groups representing a dichotomous phenotype (low and high). This scenario will be referred as the extreme phenotypes binary case (EPSB).

1.3.3. Marker-trait association analysis

An additive multi-locus mixed-model (MLMM) stepwise regression was applied with forward inclusion and backward elimination [17] to detect the significant markers associated with the trait. In the case of the whole and random sampling, the following mixed model was applied to the data:

$$y = X\beta + wa + Zu + e$$

where $\boldsymbol{y}$ is the n x 1 vector of the phenotypic values of BW for n broilers, X is the n x 55 matrix of fixed effects: sex (2 classes), hatch (36 classes), mating group (17 classes), $\boldsymbol{\beta}$ is the 55 x 1 vector of corresponding coefficients of fixed effects, $\boldsymbol{w}$ is the vector with elements 0 for the major homozygous genotype, 1 for the heterozygote genotype and 2 for the minor homozygous genotype (additive genetic model) , $\boldsymbol{\alpha}$ is the vector of the fixed effect for the minor allele of the candidate SNP to be tested for association, Z is the incidence matrix relating observations to the polygenic random effects, $\boldsymbol{u}$ is the vector of random polygenic effects, and $\boldsymbol{e}$ is the vector of random residuals. The random effects were assumed to be normally distributed with zero means and the following covariance structure:

$$Var\begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} G\sigma_u^2 & 0 \\ 0 & I\sigma_e^2 \end{bmatrix}$$

where σ_u^2 and σ_e^2 are the polygenic and error variance components, I is the nxn identity matrix, and G is the nxn genomic relationship matrix with elements of pairwise relationship coefficient using all the 215,555 SNPs. The genomic relationship coefficient between two individuals j and k, was estimated as follows:

$$\frac{1}{215{,}555}\sum_{i=1}^{215{,}555}\frac{(x_{ij}-2p_i)(x_{ik}-2p_i)}{2p_i(1-2p_i)}$$

where x_{ij} and x_{ik} are the numbers (0, 1 or 2) of the minor allele(s) for the i_{th} SNP of the j_{th} and k_{th} individuals, respectively, and p_i is the frequency of the minor allele [18]. Note that inclusion of the genomic relationship matrix in the model has been shown to correct for possible population structure and stratification in the data [19]. This analysis was carried out

with SNP & Variation Suite version 8.8.1 software (Golden Helix: http://www.goldenhelix.com).

During the analysis of the extreme phenotype samples, only the vector of SNP effects (***a***) was included in the model as the rest fixed effects (sex, hatch and mating group) had been appropriately accounted during trait adjustment (see 'Sampling scenarios'). Taken together, a total number of 19 analyses were carried out. All analyses were performed with SNP & Variation Suite (version 8.8.1) software (Golden Helix: http://www.goldenhelix.com). Each time, statistically significant markers were selected at the optimal step of the MLMM stepwise regression according to the extended Bayesian Information Criterion (eBIC, [20]). P-values of SNPs were corrected for multiple comparisons using the false-discovery rate (FDR) method [21] and significance was denoted using a FDR p-value less than 0.05.

1.3.4. Quantile-quantile plots and estimation of the genomic inflation factor

Quantile-quantile (Q–Q) plots were used to analyze the extent to which the observed distribution of the test statistic followed the expected (null) distribution. Q-Q plots along with the estimated genomic inflation factor lambda (λ) was used to assess potential systematic bias due to population structure or to the analytical approach [22].

1.3.5. Proportion of variance explained by SNP per sampling case

The Proportion of Variance Explained by a SNP *k* (PVE_k) was calculated as:

$$PVE_k = \frac{mrss_{h0} - mrss_k}{mrss_{h0}}$$

where $mrss_{h0}$ is the Mahalonobis root sum of squares (mrss) for the null hypothesis and $mrss_k$ is the same for marker *k*.

1.3.6. Estimation of allelic effects based on PVE

According to Falconer and Mackay [23], the PVE of a SNP is given by formula (1)

$$PVE = \frac{2p(1-p)\beta^2}{\sigma_p^2} \quad (1)$$

Where:
p is the MAF of the SNP and
σ_p^2 is the phenotypic variance of the trait
by solving formula (1) for the β term allows for the estimation of the SNP allelic effects (β) as follows:

$$\beta = \sqrt{\frac{PVE\,\sigma_p^2}{2p(1-p)}} \quad (2)$$

1.3.7. Detection efficacy across the various sampling scenarios

Detection efficacy (DE) of marker-trait associations across the various subsamples was explored by finding the maximum number of lead (i.e. significant) SNPs within 500 kb regions around SNPs detected in whole population sampling (PS).

1.3.8. Identification of putative causative genetic variants

In the most efficient sampling scenario, we used estimated PVE associated with lead SNPs to

infer their importance as causative genetic variants for the trait. Specifically, lead SNPs with PVE ≥ 2.0% were considered putative evidence of large genetic effects [24] while those with 2.0<PVE ≥ 1.0% were considered evidence for moderate genetic effects [24,25].

1.3.9. Variant effect prediction of putative causative genetic variants

Annotation of the putative causative SNPs was predicted using the Variant Effect Predictor tool (VEP, https://www.ensembl.org/Tools/VEP, [26]) and the latest *Gallus gallus* genome assembly (ver. GRCg6a (https://www.ncbi.nlm.nih.gov/assembly/GCF_000002315.6: accessed:21st April 2019) and NCBI Annotation release 104: https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Gallus_gallus/104/: accessed: 21st April 2019). VEP identified overlapping transcripts and predicted the effects that SNP alleles could have on genes, transcripts, protein sequence as well as regulatory regions. Apart from the aforementioned, the VEP tool was also used to infer associations of the queried variants with phenotypes via connections with Animal QTL database (Animal QTLdb) and Online Mendelian Inheritance in Animals (OMIA) database for the species.

1.4. Results

1.4.1. Identification of significant SNPs in whole population sampling

Figure 1 displays a Manhattan and Q-Q plot of SNP p-values in PS. As the Q-Q plot clearly shows, there is no evidence of any systematic bias due to population structure or analytical approach in our case. This can also be validated by the estimated value of lambda (λ=0.95). The Q-Q plot also shows that some SNPs depart from the expected probability indicating possible association with the trait. The significant (FDR p-value< 0.05) SNPs detected in PS are shown in Table 1 along with estimated PVE, respective MAF and allelic effects (β). PVE ranged from 1.44% (*rs315329074*) to 0.009% (*rs317777863*) while MAF ranged from 0.066 (*rs314844319) to* 0.469 (*rs15608447).* Highest PVE were attained for markers *rs15425131* (1.44%) and *rs315329074* (1.42%), albeit for different reasons. As Formula 1 implies, PVE is the product of MAF and β. In case of *rs15425131*, PVE is the result of low MAF (0.091) and highest β (3.86 g) while in the case of *rs315329074* is the product of higher MAF (0.171) with lower β (2.93 g). In general, highest PVE were associated with highest p-values on the $\log_{10}$ scale (Table 1).

1.4.2. Detection efficacy across the various sampling scenarios and estimated PVE

Figure 2 presents the genome-wide significant SNPs (n=49) and the corresponding sampling scenario(s) in which each SNP was found to be significantly associated with the trait. Specifically, 10 unique SNP signals were detected in PS plus 39 more in the rest sampling scenarios. As shown in more detail in Table 2, within sampling scenarios, the maximum number (n=13) of SNPs was identified in EPS_20%, followed by EPS_50% (n=11). DE i.e. number of SNP signals that were common or lied within 500kb distances from the PS SNPs are presented in Table 3. DE ranged from a minimum n=3 in RS_30% and EPSB_50% to a maximum n=8 in EPS_50% (Figure 3). A detailed view of SNPs detected across the sampling scenarios in relation to the position of PS SNPs on the same autosomes is provided in Figure 4. As Figure 4 displays, there were 3 SNPs detected within a distance of 436,398 bp (2,890,348-3,326,746 bp) on GGA25 with two markers (*rs317093585* and *rs313194380*) detected in EPS_50% and one (*rs312861757*) in PS. Of the three markers, *rs317093585* and *rs313194380* were distanced 29,091 bp apart and displayed moderate LD levels (D'=0.37), while markers *rs313194380* and *rs312861757* were distanced 407,307 bp and were in strong LD (D'=0.99).

DE was also found to be dependent on MAF as markers with moderate MAF such as *rs316794400* (MAF=0.17) and *rs315329074* (MAF=0.20) were detected even in small sized samples while markers with lower MAF such as *rs15425131 (*MAF=0.09) were detected only in large sized samples (50%) (Table 2). PVE associated with lead SNPs were higher in EPS than RS (Table 2) and estimated PVE in subsamples (random or extreme) were invariably higher than in PS due to the Beavis effect [27] or the winner's curse [28] as it is known in the biostatistics literature.

1.4.3. Identification and effect prediction of putative causative genetic variants

Tables 4 and 5 show the putative causative SNPs in the most efficient sampling scenario i.e. EPS_50%. Specifically, two SNPs i.e. *rs315329074* and *rs15425131* had PVE $\geq$ 2.0% (Table 4) while three markers (*rs14265664*, *rs316794400 and rs15608447*) had PVE<2.0% and PVE $\geq$ 1.0% (Table 4). *rs315329074* (PVE=3.2%, MAF=0.195) is an intron or downstream variant (at 3') of *CACNB1* gene where six growth-related QTLs (such as BW hatch, femur weight etc.) are reported. *rs15425131* (PVE=2.1%, MAF=0.12) is a synonymous variant within *MYOM2* gene where a comb weight QTL is reported. *rs14265664* (PVE=1.3%, MAF=0.10) underlies a region where a wattles weight QTL is reported. This intergenic variant is detected between genes *FBXO32* and *LOC112531900*. Of these, *FBXO32* is the nearest gene distanced only 9247 bp from the marker. *rs316794400* (PVE=1.3%, MAF=0.18) lies at 5' of *SLC20A1* gene and at 3' of the *ANXA4* gene in a region where 2 growth-related QTLs (breast muscle percentage, abdominal fat weight) are reported. Finally, *rs15608447* (PVE=1.2%, MAF=0.47) is an intron variant in *SLAIN2* gene (Table 5).

1.5. Discussion

1.5.1. Detection efficacy of EPS

A first interesting finding of the present study relates to the type of sampling and specifically the superiority of EPS vs. RS. This finding is not new and has been repeatedly validated in the relevant literature [7,14,15]. Yet, the most striking result obtained here was the remarkable efficiency of EPS in detecting marker-trait associations that reached a maximum value of 80% in the case of EPS_50%. This finding complies with results of a GWAS in *Larimichthys crocea* reporting that 40-60% EPS can deliver similar results as using whole population sampling PS [13]. Results on GGA25 have also demonstrated that GWAS-identified SNPs serve only as representatives for the SNPs in the same haplotype block and it is equally likely that SNP peaks may arise as a result of strong LD between the array-identified SNPs [29]. Apparently, this finding has important implications in terms of identifying true causative genetic variants and the underlying functional candidate genes.

A second important outcome deals with the utility of continuous vs. dichotomized phenotypes. Dichotomized phenotypes are favorable when accurate phenotyping is expensive, or phenotypes cannot be measured at a continuous scale and may offer additional advantages due to application of more powerful statistical methods [30,31]. However, there is also evidence (e.g. [7,15]) that this specific design can cause a loss of information and decrease the power. In concordance with the latter studies, present results have demonstrated that dichotomized extreme phenotypes did not offer any advantage over continuous phenotypes, at least in detection of causal SNPs.

1.5.2. Detection of causative genetic variants

Perhaps, the most intriguing task when performing a XP-GWAS is as how to screen and identify the putative causative SNPs. An obvious solution here is to select SNPs with highest PVE, or better, those surpassing a certain threshold (e.g. 1%). In doing so, one should be

wary of the fact that inferences on the true PVE of the causative variants are expected to be biased (inflated) due to the Beavis effect. The severity of the bias depends on sample size but also on the underlying distribution of the true PVE of all causative variants which is assumed to be exponentially or gamma distributed, with an abundance of low PVE loci and very few high PVE loci [32]. As King and Long [32] emphasized, when the vast majority of causative variants contribute 1% or less to the phenotype, the resulting bias is expected to be severe, even in large sized samples (e.g. 1000), because power declines with decreasing PVE.

Despite the aforementioned inherent limitations, in our case, the use of the PVE threshold has proved particularly useful in identifying true causative genetic variants for the trait under study. This may be fairly concluded by the fact that all 5 implicated markers with PVE ≥1.0% resided in genomic regions harboring a total number of 10 growth-related QTLs and 6 growth relevant genes. Among the implicated QTLs are breast muscle percentage, abdominal fat weight, body weight hatch, femur weight etc., just to mention some of the reported QTLs in the area. At the same time, the list with the candidate genes includes *CACNB1 (calcium voltage-gated channel auxiliary subunit beta 1)* that affects skeletal muscle development in mice [33], *MYOM2 (myomesin 2)* that encodes a fast-fibre isoform of myomesin called M-protein [34] that is mainly expressed in adult cardiac and fast-twitch fibers in skeletal muscles [35], *SLC20A1 (solute carrier family 20 member 1,* also known *as PiT1)* that is necessary for normal liver development [36], *ANXA4 (annexin A4)* that participates in epithelial cell proliferation [37], *FBXO32 (F-box protein 32,* also known as Atrogin 1 or MAFbx) a skeletal and cardiac muscle-specific F-box motif- containing protein associated with muscle atrophy [38] and *SLAIN2 (SLAIN motif family member 2)* that controls the microtubule growth during interphase [39]. Intuitively, genes including lead SNPs and at the same time presenting functional relevance with the trait under study are considered ideal functional candidates for the trait under study. Yet, it is important to bear in mind that, due to LD, the list with the plausible causative genes may eventually include tens or hundreds of genes. In line with this scenario, a total number of 34 modular genes implicated in developmental processes have been identified in strong LD genomic regions around markers *rs316794400, rs315329074* and *rs15608447* [40].

Another important aspect for successful detection of causative SNP when performing a XP-GWAS relates to MAF. Specifically, the lower the MAF of the causal SNP, the smaller the range of allele frequency in the genotyped SNPs which will result in LD between the two. Therefore, for low-MAF QTLs there are likely to be fewer genotyped SNPs which are in strong enough LD to detect the association. In line with this hypothesis, MacLeod et al. [41] demonstrated, via simulations, that QTLs with low MAFs were harder to detect than those with higher allele frequencies. This scenario may explain why *rs315329074* (*CACNB1* gene) with MAF=0.17 and effect size β*=2.9* g was consistently detected across almost all sampling cases, while *rs15425131* (*MYOM2* gene) with lower MAF=0.091 and highest effect size (β=3.9 g) could be detected only in EPS_50%. Even so, successful detection of low MAF variants is dependent not only on the selected fraction of the extreme tails but also on the detection methodology used. While a EPS_50% GWAS was required here to detect *rs15425131* and the *MYOM2* gene, this specific association could be detected when using F_{ST} genome scans even in low sized (10%) extreme samples [42].

1.6. Conclusions

In conclusion the use of EPS resulted in identification of 5 putative causal genetic variant residing in non-coding regulatory regions. Non coding variants constitute the majority of signals in GWAS [43]. Specific methods are needed to translate these results to elucidate the role of noncoding variants [44]. To this end, Claussnitzer et al. [44] have generated a

roadmap by utilizing combined public resources (epigenomic annotations, chromosome conformation, and regulatory motif conservation), targeted experiments for risk and non-risk haplotypes (enhancer screening, gene expression, and cellular profiling) and directed perturbations in primary cells and mouse models (regulator–target knockdown and overexpression and CRISPR–Cas9 genome editing). Finally, while the present study has delivered some practical guidance to perform cost-efficient GWAS, many issues still need to be addressed. These issues relate to the usefulness of alternative sampling strategies such as a two-stage design [12], asymmetrical sampling and the comparison between diverse methods such as signatures of selection in same or different sized samples.

1.7. References

[1] Steemers FJ, Chang W, Lee G, Barker DL, Shen R, Gunderson KL. Whole-genome genotyping with the single-base extension assay. Nat Methods. Nature Publishing Group; 2006;3: 31–33. doi:10.1038/nmeth842.

[2] Lebowitz RJ, Soller M, Beckmann JS. Trait-based analyses for the detection of linkage between marker loci and quantitative trait loci in crosses between inbred lines. Theor Appl Genet. Springer-Verlag; 1987;73: 556–562. doi:10.1007/BF00289194.

[3] Lander ES, Botstein D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics. Genetics Society of America; 1989;121: 185–99. Available: http://www.ncbi.nlm.nih.gov/pubmed/2563713.

[4] Darvasi A, Soller M. Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. Theor Appl Genet. Springer-Verlag; 1992;85–85: 353–359. doi:10.1007/BF00222881.

[5] Gallais A, Moreau L, Charcosset A. Detection of marker–QTL associations by studying change in marker frequencies with selection. Theor Appl Genet. Springer-Verlag; 2007;114: 669–681. doi:10.1007/s00122-006-0467-z.

[6] Sen S, Johannes F, Broman KW. Selective genotyping and phenotyping strategies in a complex trait context. Genetics. Genetics Society of America; 2009;181: 1613–26. doi:10.1534/genetics.108.094607.

[7] Barnett IJ, Lee S, Lin X. Detecting rare variant effects using extreme phenotype sampling in sequencing association studies. Genet Epidemiol. NIH Public Access; 2013;37: 142–51. doi:10.1002/gepi.21699.

[8] Rabier CE. On statistical inference for selective genotyping. J Stat Plan Inference. North-Holland; 2014;147: 24–52. doi:10.1016/J.JSPI.2013.11.010.

[9] Flint J, Valdar W, Shifman S, Mott R. Strategies for mapping and cloning quantitative trait genes in rodents. Nat Rev Genet. Nature Publishing Group; 2005;6: 271–286. doi:10.1038/nrg1576.

[10] Yang J, Jiang H, Yeh C-T, Yu J, Jeddeloh JA, Nettleton D, et al. Extreme-phenotype genome-wide association study (XP-GWAS): a method for identifying trait-associated variants by sequencing pools of individuals selected from a diversity panel. Plant J. John Wiley & Sons, Ltd (10.1111); 2015;84: 587–596. doi:10.1111/tpj.13029.

[11] Zhang Y, Qi G, Park J-H, Chatterjee N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. Nat Genet. Nature Publishing Group; 2018;50: 1318–1326. doi:10.1038/s41588-018-0193-x.

[12] Li D, Lewinger JP, Gauderman WJ, Murcray CE, Conti D. Using extreme phenotype sampling to identify the rare causal variants of quantitative traits in association studies. Genet Epidemiol. NIH Public Access; 2011;35: 790–9. doi:10.1002/gepi.20628.

[13] Wan L, Dong L, Xiao S, Han Z, Wang X, Wang Z. Genomewide association study for economic traits in the large yellow croaker with different numbers of extreme phenotypes. J Genet. 2018;97: 887–895. Available: http://www.ncbi.nlm.nih.gov/pubmed/30262700.

[14] Zhou Y-J, Wang Y, Chen L-L. Detecting the Common and Individual Effects of Rare Variants on Quantitative Traits by Using Extreme Phenotype Sampling. Genes (Basel). Multidisciplinary Digital Publishing Institute (MDPI); 2016;7. doi:10.3390/genes7010002.

[15] Bjørnland T, Bye A, Ryeng E, Wisløff U, Langaas M. Improving power of genetic association

studies by extreme phenotype sampling: a review and some new results. 2017; Available: http://arxiv.org/abs/1701.01286.
[16] Kranis A, Gheyas AA, Boschiero C, Turner F, Yu L, Smith S, et al. Development of a high density 600K SNP genotyping array for chicken. BMC Genomics. BioMed Central; 2013;14: 59. doi:10.1186/1471-2164-14-59.
[17] Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, et al. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet. Nature Publishing Group; 2012;44: 825–830. doi:10.1038/ng.2314.
[18] VanRaden PM. Efficient Methods to Compute Genomic Predictions. J Dairy Sci. 2008;91: 4414–4423. doi:10.3168/jds.2007-0980.
[19] Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of mixed-model association methods. Nat Genet. NIH Public Access; 2014;46: 100–6. doi:10.1038/ng.2876.
[20] Chen J, Chen Z. Extended Bayesian information criteria for model selection with large model spaces. Biometrika. Oxford University Press; 2008;95: 759–771. doi:10.1093/biomet/asn034.
[21] Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing [Internet]. Journal of the Royal Statistical Society. Series B (Methodological). WileyRoyal Statistical Society; 1995. pp. 289–300. doi:10.2307/2346101.
[22] Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, Willer CJ, et al. Genomic inflation factors under polygenic inheritance. Eur J Hum Genet. Nature Publishing Group; 2011;19: 807–12. doi:10.1038/ejhg.2011.39.
[23] Falconer DS, Mackay TFC. Introduction to Quantitative Genetics. 4th ed. Longmans Green, Harlow, Essex, UK.; 1996.
[24] Seabury C, Oldeschulte D, Saatchi M, Beever J, Decker J, Halley Y, et al. Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. BMC Genomics. BioMed Central; 2017;18: 386–386. doi:10.1186/S12864-017-3754-Y.
[25] Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S, Freimer NB, et al. Variance component model to account for sample structure in genome-wide association studies. Nat Genet. Nature Publishing Group; 2010;42: 348–354. doi:10.1038/ng.548.
[26] McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. BioMed Central; 2016;17: 122. doi:10.1186/s13059-016-0974-4.
[27] Beavis W. QTL analyses: power, precision, and accuracy.Molecular dissection of complex traits, 1. 1998. pp. 145–162.
[28] Zöllner S, Pritchard JK. Overcoming the Winner's Curse: Estimating Penetrance Parameters from Case-Control Data. Am J Hum Genet. Elsevier; 2007;80: 605. doi:10.1086/512821.
[29] Tak YG, Farnham PJ. Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. Epigenetics Chromatin. BioMed Central; 2015;8: 57. doi:10.1186/s13072-015-0050-4.
[30] Guey LT, Kravic J, Melander O, Burtt NP, Laramie JM, Lyssenko V, et al. Power in the phenotypic extremes: a simulation study of power in discovery and replication of rare variants. Genet Epidemiol. 2011;35: n/a-n/a. doi:10.1002/gepi.20572.
[31] Lumley T, Dupuis J, Rice KM, Barbalic M, Bis JC, Cupples LA, et al. TWO-PHASE SUBSAMPLING DESIGNS FOR GENOMIC RESEQUENCING STUDIES [Internet]. 2012. Available: https://pdfs.semanticscholar.org/c2cb/173513a8eb8a0adfb2cd59e90248333ce4c9.pdf.
[32] King EG, Long AD. The Beavis Effect in Next-Generation Mapping Panels in Drosophila melanogaster. G3 Genes, Genomes, Genet. G3: Genes, Genomes, Genetics; 2017;7: 1643–1652. doi:10.1534/G3.117.041426.
[33] Chen F, Liu Y, Sugiura Y, Allen PD, Gregg RG, Lin W. Neuromuscular synaptic patterning requires the function of skeletal muscle dihydropyridine receptors. Nat Neurosci. Nature Publishing Group; 2011;14: 570–577. doi:10.1038/nn.2792.
[34] Fukuzawa A, Lange S, Holt M, Vihola A, Carmignac V, Ferreiro A, et al. Interactions with titin and myomesin target obscurin and obscurin-like 1 to the M-band - implications for hereditary myopathies. J Cell Sci. 2008;121: 1841–1851. doi:10.1242/jcs.028019.
[35] Schoenauer R, Lange S, Hirschy A, Ehler E, Perriard J-C, Agarkova I. Myomesin 3, a Novel

Structural Component of the M-band in Striated Muscle. J Mol Biol. 2008;376: 338–351. doi:10.1016/j.jmb.2007.11.048.
[36] Beck L, Leroy C, Beck-Cormier S, Forand A, Salaün C, Paris N, et al. The Phosphate Transporter PiT1 (Slc20a1) Revealed As a New Essential Gene for Mouse Liver Development. Lewin A, editor. PLoS One. Public Library of Science; 2010;5: e9148. doi:10.1371/journal.pone.0009148.
[37] Yao H-S, Sun C, Li X-X, Wang Y, Jin K-Z, Zhang X-P, et al. Annexin A4-nuclear factor-κB feedback circuit regulates cell malignant behavior and tumor growth in gallbladder cancer. Sci Rep. Nature Publishing Group; 2016;6: 31056. doi:10.1038/srep31056.
[38] Nakashima K, Ishida A, Ijiri D, Ohtsuka A. Effect of dexamethasone on the expression of atrogin-1/MAFbx in chick skeletal muscle. Anim Sci J. John Wiley & Sons, Ltd (10.1111); 2016;87: 405–410. doi:10.1111/asj.12437.
[39] van der Vaart B, Manatschal C, Grigoriev I, Olieric V, Gouveia SM, Bjelić S, et al. SLAIN2 links microtubule plus end–tracking proteins and controls microtubule growth in interphase. J Cell Biol. 2011;193: 1083–1099. doi:10.1083/jcb.201012179.
[40] Tarsani E, Kranis A, Maniatis G, Avendano S, Hager-Theodorides AL, Kominakis A. Discovery and characterization of functional modules associated with body weight in broilers. Sci Rep. Nature Publishing Group; 2019;9: 9125. doi:10.1038/s41598-019-45520-5.
[41] MacLeod IM, Hayes BJ, Savin KW, Chamberlain AJ, McPartlan HC, Goddard ME. Power of a genome scan to detect and locate quantitative trait loci in cattle using dense single nucleotide polymorphisms. J Anim Breed Genet. 2010;127: 133–142. doi:10.1111/j.1439-0388.2009.00831.x.
[42] Tarsani E, Kominakis A, Theodorou G, Palamidi I. Exploiting extreme phenotypes to investigate haplotype structure and detect signatures of selection for body weight in broilers. THE XV th EUROPEAN POULTRY CONFERENCE DUBROVNIK, CROATIA Conference Information and Proceedings. 2018. p. 90. Available: https://pure.au.dk/portal/files/140225746/sbornik_docladov_epc_2018.pdf.
[43] 1000 Genomes Project Consortium T 1000 GP, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. Europe PMC Funders; 2012;491: 56–65. doi:10.1038/nature11632.
[44] Claussnitzer M, Dankel SN, Kim K-H, Quon G, Meuleman W, Haugen C, et al. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. N Engl J Med. NIH Public Access; 2015;373: 895. doi:10.1056/NEJMOA1502214.
[45] Anand L. chromoMap: An R package for Interactive Visualization and Annotation of Chromosomes. bioRxiv. Cold Spring Harbor Laboratory; 2019; 605600. doi:10.1101/605600.

Tables and Figures of Chapter 1

Table 1. Proportion of variance explained (PVE%), minor allele frequency (MAF) and allelic effects (β) per significant SNP detected in whole population sampling. The estimated phenotypic variance (σ_p^2) was as high as 171.225 g^2.

SNP ID	GGA	Position (bp)	p-value	-log10(p-value)	FDR p-value	PVE(%)	MAF	*β (g)*
rs13923872	1	114,049,481	1.8832E-06	5.7251	0.04058	0.3069	0.414	1.041
rs15425131	3	90,795,168	1.7989E-51	50.7450	3.88E-46	1.4411	0.091	3.861
rs313332188	3	99,991,484	4.8484E-13	12.3144	2.09E-08	0.6833	0.437	1.542
rs15608447	4	66,459,916	2.662E-14	13.5748	1.43E-09	0.9009	0.469	1.761
rs317014229	10	1,288,866	1.7144E-08	7.7659	0.00046	0.1351	0.407	0.692
rs316794400	22	5,149,585	9.6605E-20	19.0150	6.94E-15	0.6159	0.202	1.809
rs314844319	24	1,869,760	1.4352E-06	5.8431	0.03436	0.0595	0.066	0.907
rs312861757	25	3,326,746	2.4423E-09	8.6122	7.52E-05	0.2349	0.077	1.685
rs317777863	25	196,842	2.9874E-10	9.5247	1.07E-05	0.00896	0.364	0.182
rs315329074	27	6,920,352	2.4734E-22	21.6067	2.67E-17	1.4233	0.171	2.934

Table 2. Genome-wide significant SNPs across the sampling scenarios (RS: random sampling, EPS: extreme phenotype sampling, EPSB: extreme phenotypes binary case).

RS

Sample proportion (%)	SNP ID	GGA	Position (bp)[a]	p-value	-log10(p-value)	FDR p-value	PVE (%)	MAF
5	*rs314268898*	18	1,026,668	3.62E-11	10.4411	7.81E-06	12.4112	0.089
10	*rs316975706*	26	805,527	2.15E-15	14.6670	4.64E-10	7.5871	0.324
20	*rs314007348*	4	66,225,823	2.64E-08	7.5777	0.0019	2.1494	0.447
	rs318098582	11	18,407,493	7.93E-34	33.1007	1.71E-28	5.7627	0.133
	rs16192702	24	494,813	4.76E-12	11.3224	5.13E-07	1.5934	0.093
	rs317587988	28	4,553,250	6.48E-07	6.1880	0.03496	2.6190	0.465
30	*rs314007348*	4	66,225,823	3.44E-08	7.4628	0.00248	1.5643	0.438
	rs314066852	10	13,562,465	1.07E-06	5.9667	0.03879	0.1417	0.191
	rs312428343	22	2,016,328	1.06E-07	6.9720	0.00575	0.0212	0.294
	rs316794400	22	5,149,585	6.93E-07	6.1590	0.02989	1.2122	0.191
	rs317288536	25	2,173,372	6.15E-21	20.2109	1.33E-15	3.1489	0.089
	rs315329074	27	6,920,352	3.14E-15	14.5037	3.38E-10	2.8260	0.173
40	*rs315329074*	27	6,920,352	5.67E-26	25.2460	1.22E-20	3.3903	0.177

	rs317792664	28	4,336,570	1.94E-24	23.7121	2.09E-19	2.8250	0.079
50	*rs313097265*	2	93,140,828	7.79E-42	41.1081	1.68E-36	2.7875	0.101
	rs314007348	4	66,225,823	5.43E-13	12.2646	3.91E-08	1.4746	0.446
	rs312428343	22	2,016,328	5.88E-10	9.2302	3.17E-05	0.0368	0.302
	rs315329074	27	6,920,352	1.76E-30	29.7535	1.90E-25	2.4251	0.174
EPS								
Sample proportion (%)	**SNP ID**	**GGA**	**Position (bp)[a]**	**p-value**	**-log$_{10}$(p-value)**	**FDR p-value**	**PVE (%)**	**MAF**
5	*rs317414603*	20	6,729,013	1.47E-10	9.8301	3.19E-05	20.9532	0.384
	rs316714498	27	5,853,588	2.11E-10	9.6739	2.28E-05	21.0961	0.385
10	*rs312675887*	9	17,794,694	7.92E-14	13.1010	8.54E-09	9.2336	0.299
	rs317414603	20	6,729,013	5.17E-07	6.2864	0.03716	12.2069	0.385
	rs316714498	27	5,853,588	2.78E-23	22.555	6.01E-18	12.9160	0.385
20	*rs15272503*	1	54,426,960	5.98E-12	11.2233	4.30E-07	3.3308	0.410

	rs315936751	1	61,491,143	4.58E-07	6.3385	0.01412	1.7151	0.179
	rs14265664	2	138,095,717	2.59E-07	6.5851	0.00934	3.9939	0.078
	rs317466272	5	23,021,657	1.25E-07	6.9020	0.0054	1.8507	0.052
	rs315882280	6	8,228,013	1.63E-08	7.7858	0.00088	5.7124	0.444
	rs313850906	6	11,119,203	2.63E-27	26.5792	5.68E-22	3.5653	0.469
	rs315438523	8	8,011,741	1.49E-06	5.8267	0.03212	1.1993	0.324
	rs312675887	9	17,794,694	1.91E-16	15.7171	2.07E-11	5.4086	0.292
	rs318099392	19	7,931,078	0.0000011	5.9561	0.0265	3.4601	0.225
	rs314275684	23	4,557,952	2.52E-06	5.5976	0.04538	0.1370	0.112
	rs313580984	24	2,201,697	5.53E-07	6.2566	0.01492	1.6366	0.068
	rs315052836	25	3,861,162	2.59E-06	5.5859	0.04303	0.4813	0.051
	rs312334304	28	1,708,458	2.05E-06	5.6880	0.04019	0.6336	0.064
30	*rs15608447*	4	66,459,916	3.45E-07	6.4610	0.01491	1.1560	0.479
	rs318098582	11	18,407,493	5.19E-25	24.2844	1.12E-19	3.4025	0.167

	rs312732833	15	3,230,864	5.68E-10	9.2452	3.06E-05	1.0124	0.219
	rs316794400	22	5,149,585	2.11E-15	14.6739	1.52E-10	1.8574	0.162
	rs315329074	27	6,920,352	3.12E-22	21.5045	3.37E-17	4.5656	0.219
40	*rs15608447*	4	66,459,916	8.84E-08	7.0534	0.00381	1.0826	0.476
	rs318098582	11	18,407,493	1.15E-27	26.9357	2.50E-22	2.9356	0.154
	rs312732833	15	3,230,864	4.79E-10	9.3192	2.58E-05	0.6029	0.234
	rs316794400	22	5,149,585	2.12E-16	15.6727	1.53E-11	1.6535	0.171
	rs315329074	27	6,920,352	7.63E-26	25.1171	8.23E-21	3.9738	0.204
50	*rs14265664*	2	138,095,717	1.07E-08	7.9691	0.00046	1.3448	0.097
	rs15425131	3	90,795,168	1.94E-38	37.7102	4.20E-33	2.1237	0.121
	rs313332188	3	99,991,484	4.59E-08	7.3380	0.00165	0.9527	0.413
	rs15608447	4	66,459,916	6.91E-10	9.1601	3.73E-05	1.1722	0.471
	rs317466272	5	23,021,657	4.53E-07	6.3435	0.01086	0.6491	0.059
	rs312834930	7	11,436,451	1.88E-07	6.7243	0.00508	0.8102	0.061

	rs316794400	22	5,149,585	1.28E-14	13.8921	1.38E-09	1.2500	0.180
	rs313580984	24	2,201,697	1.26E-07	6.8975	0.0039	0.7371	0.078
	rs313194380	25	2,919,439	2.01E-06	5.6963	0.03943	0.7341	0.071
	rs317093585	25	2,890,348	6.43E-07	6.1911	0.01388	0.3532	0.100
	rs315329074	27	6,920,352	5.09E-10	9.2925	3.66E-05	3.1952	0.195
EPSB								
Sample proportion (%)	**SNP ID**	**GGA**	**Position (bp)[a]**	**p-value**	**-log$_{10}$(p-value)**	**FDR p-value**	**PVE (%)**	**MAF**
5	*rs317414603*	20	6,729,013	7.13E-11	10.1464	1.54E-05	21.6499	0.384
	rs316714498	27	5,853,588	6.28E-10	9.2020	6.77E-05	21.0301	0.385
10	*rs14265664*	2	138,095,717	4.08E-08	7.3887	0.00294	7.1316	0.070
	rs317668107	3	33,354,124	1.01E-16	15.9946	1.09E-11	9.0454	0.357
	rs317414603	20	6,729,013	9.02E-17	16.0446	1.95E-11	12.7644	0.385
20	*rs315961647*	6	11,017,574	1.04E-11	10.9814	2.25E-06	3.1560	0.220
	rs318061321	7	946,559	1.01E-06	5.9932	0.04379	1.5109	0.214

	rs312675887	9	17,794,694	5.72E-08	7.2423	0.00308	4.7857	0.292
	rs318098582	11	18,407,493	7.00E-10	9.1548	5.03E-05	4.4808	0.190
	rs316714498	27	5,853,588	3.36E-10	9.4724	3.63E-05	5.7656	0.366
30	*rs15608447*	4	66,459,916	2.51E-08	7.5997	0.00135	1.2989	0.479
	rs14512409	5	9,351,944	6.40E-17	16.1932	6.91E-12	2.1006	0.181
	rs312589151	14	1,126,506	3.64E-18	17.4382	7.86E-13	3.7173	0.463
	rs315411246	27	4,002,212	5.42E-13	12.2653	3.90E-08	2.5922	0.077
	rs315329074	27	6,920,352	3.19E-07	6.4959	0.01376	3.6633	0.219
40	*rs15608447*	4	66,459,916	1.29E-07	6.8887	0.00557	1.0654	0.476
	rs313093970	10	16,409,109	1.32E-09	8.8791	7.12E-05	0.6982	0.233
	rs318098582	11	18,407,493	2.45E-22	21.6102	5.29E-17	2.5404	0.154
	rs316794400	22	5,149,585	5.05E-17	16.2965	3.63E-12	1.7230	0.171
	rs315329074	27	6,920,352	5.42E-18	17.2652	5.85E-13	3.2505	0.204
50	*rs314723705*	1	109,156,565	9.26E-07	6.0330	0.03329	0.6008	0.497

	rs15608447	4	66,459,916	3.32E-09	8.4782	0.00018	1.0125	0.471
	rs318098582	11	18,407,493	1.03E-12	11.9857	1.11E-07	2.1187	0.145
	rs316794400	22	5,149,585	2.82E-12	11.5493	2.03E-07	1.2014	0.180
	rs316297839	25	1,567,147	3.99E-09	8.3982	0.00017	1.7609	0.106
	rs315329074	27	6,920,352	1.83E-15	14.7353	3.96E-10	2.5462	0.195

Table 3. SNP signals across the sampling scenarios in relation to position of the SNPs in whole population sampling (PS).

Random Sampling				
Sample proportion (%)	**SNP ID**	**GGA**	**Position (bp)**	**Distance from PS SNP (bp)**
20	*rs314007348*	4	66,225,823	234,093
30	*rs314007348*	4	66,225,823	234,093
	rs316794400	22	5,149,585	0
	rs315329074	27	6,920,352	0
40	*rs315329074*	27	6,920,352	0
50	*rs314007348*	4	66,225,823	234,093
	rs315329074	27	6,920,352	0
Extreme phenotype sampling				
Sample proportion (%)	**SNP ID**	**GGA**	**Position (bp)**	**Distance from PS SNP (bp)**
20	*rs313580984*	24	2,201,697	331,937
30	*rs15608447*	4	66,459,916	0
	rs316794400	22	5,149,585	0
	rs315329074	27	6,920,352	0
40	*rs15608447*	4	66,459,916	0
	rs316794400	22	5,149,585	0
	rs315329074	27	6,920,352	0
50	*rs15425131*	3	90,795,168	0
	rs313332188	3	99,991,484	0
	rs15608447	4	66,459,916	0
	rs316794400	22	5,149,585	0
	rs313580984	24	2,201,697	331,937
	rs313194380	25	2,919,439	407,307
	rs317093585	25	2,890,348	436,398
	rs315329074	27	6,920,352	0
Extreme phenotypes binary case				
Sample proportion (%)	**SNP ID**	**GGA**	**Position (bp)**	**Distance from PS SNP (bp)**
30	*rs15608447*	4	66,459,916	0

	rs315329074	27	6,920,352	0
40	*rs15608447*	4	66,459,916	0
	rs316794400	22	5,149,585	0
	rs315329074	27	6,920,352	0
50	*rs15608447*	4	66,459,916	0
	rs316794400	22	5,149,585	0
	rs315329074	27	6,920,352	0

Table 4. Putative causative SNPs in the most efficient sampling scenario.

SNP ID	**GGA**	**Position (bp)**	**p-value**	**-log$_{10}$(p-value)**	**FDR p-value**	**PVE (%)**	**MAF**
rs315329074	27	6,920,352	5.09E-10	9.2925	3.66E-05	3.1952	0.195
rs15425131	3	90,795,168	1.94E-38	37.7102	4.20E-33	2.1237	0.121
rs14265664	2	138,095,717	1.07E-08	7.9691	0.00046	1.3448	0.097
rs316794400	22	5,149,585	1.28E-14	13.8921	1.38E-09	1.2500	0.180
rs15608447	4	66,459,916	6.91E-10	9.1601	3.73E-05	1.1722	0.471

Table 5. Positional candidate genes and reported QTLs associated with putative causative SNPs in the most efficient sampling scenario.

SNP ID	GGA	Position (bp)	Allele	Consequence	Candidate Gene	QTL(s)
rs315329074	27	6,920,352	T	Downstream gene variant, Intron variant	*CACNB1*	Body weight hatch (135726, Animal QTLdb),
						Comb weight (127127, Animal QTLdb),
						Femur bone mineral content (130479, Animal QTLdb)
						Femur weight (130480, Animal QTLdb),
						Proventriculus weight (96672, Animal QTLdb),
						Wattles weight (127120, Animal QTLdb)
rs15425131	3	90,795,168	G	Synonymous variant	*MYOM2*	Comb weight (127114, Animal QTLdb)
rs14265664	2	138,095,717	A	Intergenic variant	*FBXO32*	Wattles weight (127117, Animal QTLdb)
					LOC112531900	
rs316794400	22	5,149,585	A	Upstream gene variant	*SLC20A1*	Breast muscle percentage (95429, Animal QTLdb),
				Downstream gene variant	*ANXA4*	Abdominal fat weight (96666, Animal QTLdb)
rs15608447	4	66,459,916	G	Intron variant	*SLAIN2*	-
				Downstream gene variant, intron variant,	*LOC107053243*	
				Non coding transcript variant,		
				Upstream gene variant	*LOC112532289*	

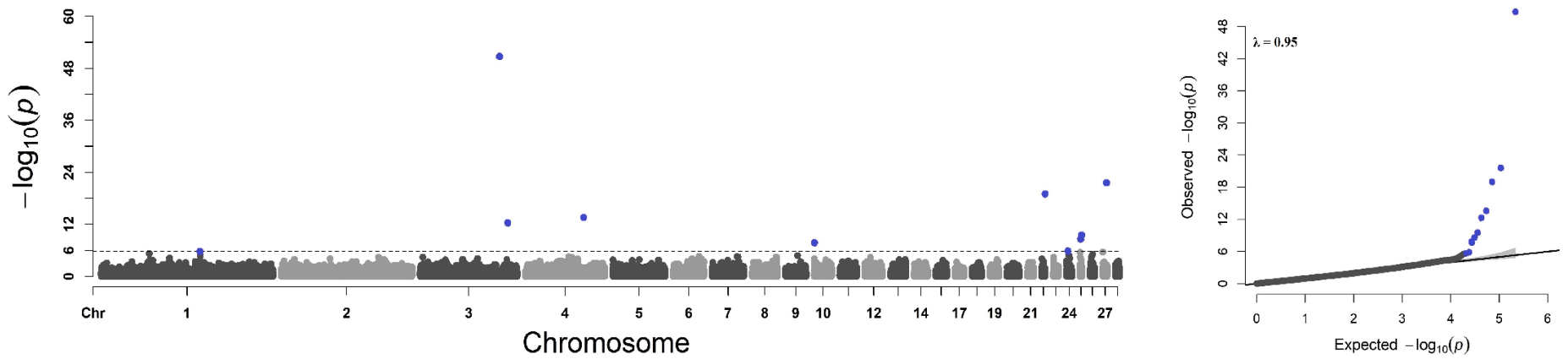


Figure 1. Manhattan plot (left) and quantile-quantile (Q-Q) plot (right) of SNP p-values in whole population. In Manhattan plot, y-axis presents the observed SNP $-\log_{10}$ (p-values) and the x-axis the SNP positions across the 27 autosomes. Horizontal line shows the genome-wide significant threshold. In Q-Q plot, y-axis and x-axis represent observed SNP $-\log_{10}$ (p-values) and expected $-\log_{10}$ (p-values), respectively. Estimation of λ is also shown on the top left in the Q-Q plot. Blue points represent the genome-wide significant SNPs for the trait. Both plots were constructed using the CMplot package (https://github.com/YinLiLin/R-CMplot) in R (http://www.r-project.org/).

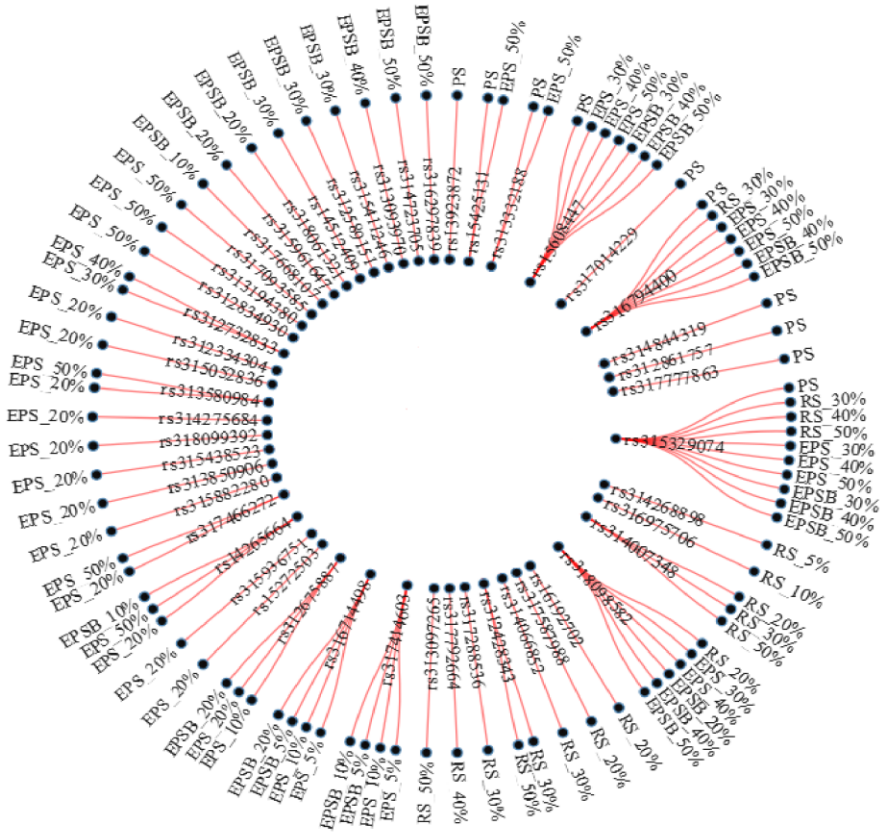


Figure 2. Radial network of the genome-wide significant SNPs detected across the various sampling scenarios (PS: whole population sampling, RS: random sampling, EPS: extreme phenotype sampling and EPSB: extreme phenotypes sampling binary case). Figure was constructed using the data.tree and networkD3 packages in R (http://www.r-project.org/).

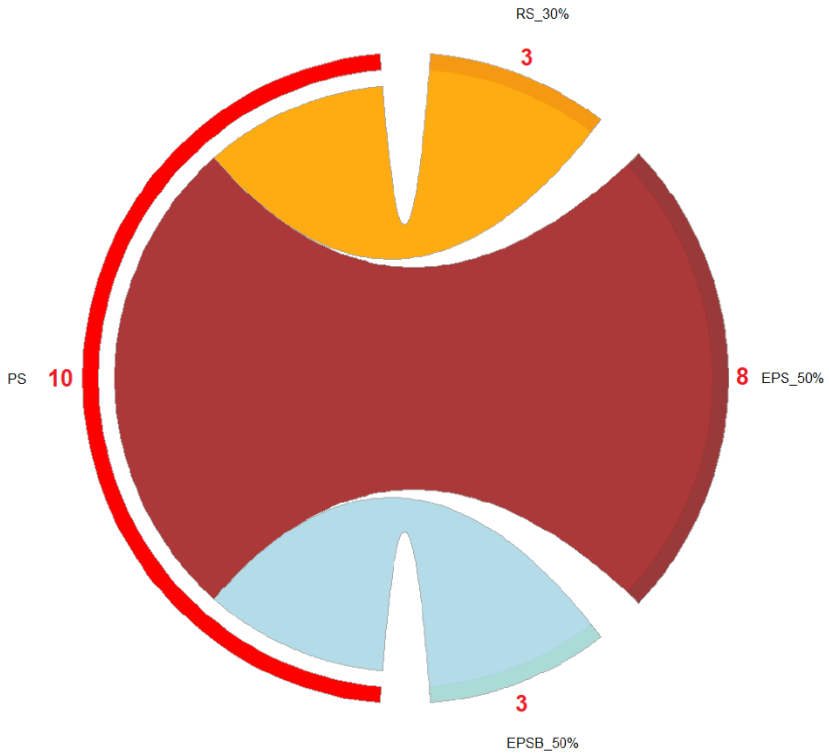


Figure 3. Chord diagram showing detection efficacy (DE) across sampling strategies (RS: random sampling, EPS: extreme phenotype sampling and EPSB: extreme phenotypes sampling binary case) in relation to whole population sampling (PS). Figure was constructed with the DescTools package in R (http://www.r-project.org/).

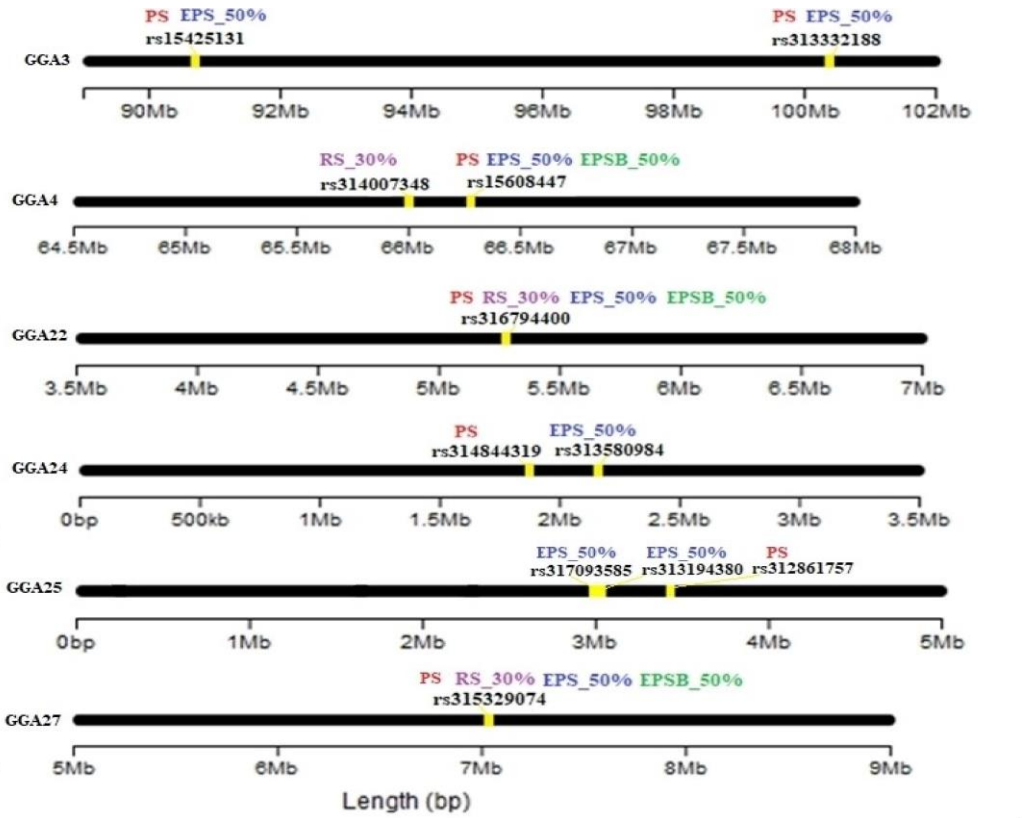


Figure 4. Positions of SNPs detected across sampling scenarios (RS: random sampling, EPS: extreme phenotype sampling and EPSB: extreme phenotypes sampling binary case) in relation to position of SNPs in whole population sampling (PS) on the same autosomes. SNP positions denoted by yellow color. Figure was constructed with the chromoMap [45] package in R (http://www.r-project.org/).

Chapter 2

Elucidating the functional role of 1,012 candidate genes revealed by a Genome Wide Association Study for body weight in broilers[1]

2.1. Abstract

Aim of the present study was first to identify genetic variants associated with body weight at 35 days of age (BW) in broilers, second to provide a list with positional candidate genes for the trait under study and third to prioritize candidate genes using various Bioinformatics methods. A genome-wide association study (GWAS) for the trait was performed using 6,598 broilers and dense genome wide SNP data (n=262,067). Application of an additive multi-locus mixed model resulted in 12 genome-wide significant SNPs, dispersed on 9 autosomes. A total number of 1,012 positional candidate genes (of which n=350 non-annotated and n=27 miRNA) were identified within 1Mb distances around the statistically significant SNPs. Functional enrichment analysis pointed out 49 candidate genes participating in system development and 17 more genes in skeletal system development. In addition, 28 candidates were members of S100 calcium binding proteins, HOXL subclass homeoboxes and type I Keratins gene families. A total number of 25 genes were members of functionally relevant to BW pathways such as MAPK6/MAPK4 signaling, signaling by EGFR, signaling by IGF1R, signaling by insulin receptor, TCF dependent signaling in response to WNT and NGF pathway. Gene prioritization analysis highlighted 248 prioritized genes with 10 top ranked genes (*SMAD4*, *CHRNB2*, *CDH1*, *NTRK1*, *RARA*, *STAT5B*, *SCARB1*, *NR1D1*, *SHC1* and *CYBB*). Topological network analysis revealed 22 highly connected genes of which 14 (*UBC*, *SMAD4*, *SHC1*, *NRAS*, *PSMD4*, *CDC6*, *PSMD7*, *RARA*, *PSMB4*, *CDH1*, *STAT5B*, *MED1*, *PSMD3* and *CDT1*) were also included in the prioritized genes. Current findings underlined the effectiveness of computational approaches in narrowing down the large list of positional candidate genes provided by GWAS. Nevertheless, detected or prioritized genes were method dependent.

2.2. Introduction

Body weight (BW) in broilers reflects the balance between the nutrient intake and expenditure, resulting in protein or fat deposition and skeletal growth. Apart from significant economic importance, the trait also presents considerable biological interest as it is a typical complex (polygenic) trait. To date, the ChickenQTLdb (https://www.animalgenome.org/cgi-bin/QTLdb/GG/index, accessed: 3rd September 2017) has over 7,812 QTL/SNP associations of which 3,582 are related to growth traits and 166 to BW. Several genome wide association studies (GWAS) have already been carried out for growth traits (e.g.[1,2]) in the species. The development of the chicken 600k SNP array [3] made it possible to efficiently screen for causal loci and genes with relevance to the trait. Despite the large number of GWAS findings, the genetic architecture of BW in chicken has still a limited understanding [4] since only a small number of positional candidate genes are confirmed as true functionally relevant to the trait (e.g. *HDAC2* [5] and *GNPDA2* [6] genes).

In almost all GWAS, genes that lie within or are in closest proximity to significant SNPs and have functional relevance to the trait under study are considered as the most plausible

[1] *Findings of the current Chapter have been published in the Proceedings of the World Congress on Genetics Applied to Livestock Production, vol. Species - Avian 1, p. 564, 2018.*

causative genes. When information on gene functions of positional candidate genes is limited, various computational approaches can be employed to predict the most functionally significant genes. Such approaches entail functional enrichment analysis (FEA) [7], pathway analysis (PA) [8], gene network analysis (GNA) [9] and gene prioritization analysis (GPA) [10]. While FEA and PA identifies terms (genes/pathways) based on their statistically over-representation within a given GO term or known pathway, GNA utilizes background knowledge interaction networks to predict gene interactions. During gene prediction, several interactions [11] (e.g. protein-protein interaction networks (PPINs), interactions by RNA co-expression, literature-curated interactions and interactions derived from high-throughput experiments are exploited, based on the 'guilt-by-association' (GBA [12]) principle. The GBA principle states that genes with related function(s) tend to be protein interaction partners or share features such as expression patterns. Based on GBA principle, a functional annotation-based approach known as GPA can also be employed. Candidate genes are prioritized based on their functional similarity to a list of already known genes [10] or keyword(s) [13] associated with a phenotype of interest. Several semantic annotations such as: Molecular Function, Biological Process, Cellular Component, Human Phenotype, Mouse Phenotype and Pathway etc. are used [10].

In the present study, first we conducted a GWAS for BW at 35 days of age to identify genomic variants i.e. SNPs associated with BW in broilers. Next, we searched for published QTLs/associations as well as positional candidate genes in 1Mb flanking regions around the significant SNPs. Finally, we employed various computational approaches such as FEA, PA, GNA and GPA to predict the most functionally significant genes for the trait under study. Current findings are expected to contribute to a better understanding of the genetic architecture underlying growth and development in the species.

2.3. Material and methods

2.3.1. Data

Genotypic and phenotypic records for 6,727 broilers (n=3,735 males and n=2,992 females) from a grand-grandparent (GGP) commercial line were made available by Aviagen Ltd. Phenotypic records for BW at 35 days of age ranged from 1,130 to 2,630 g with an average of 1840.2 g (SD=194 g). Animals were genotyped using the 600K Affymetrix® Axiom® high density genotyping array [3] resulting in a total number of 630,954 SNPs. Only n=547,784 SNPs located on the 28 autosomes (GGA1-28) were considered here. From the original number of animals, 72 females and 57 males were excluded because they had a call rate <0.99 and autosomal heterozygosity outside the 1.5 IQR (inter-quartile range) resulting in a number of n=6,598 samples. Furthermore, a number of 285,717 SNPs were excluded due to: call rate <0.99, MAF (minor allele frequency) <0.01 and linkage disequilibrium (LD) r^2 values greater than 0.99 within windows of 1 Mb inter-marker distance(s). A total of 6,598 samples and 262,067 SNPs were retained for GWAS. Quality control at the sample and marker level was performed using SNP & Variation Suite software (version 8.8.1).

2.3.2. Statistical analysis

An additive multi-locus mixed model (MLMM) [14] stepwise regression with forward inclusion and backward elimination was employed to identify the genome-wide significant markers associated with the trait. The following statistical model was applied to the data:

$$y = X\boldsymbol{\beta} + \boldsymbol{w\alpha} + Z\boldsymbol{u} + \boldsymbol{e}$$

where y is the n x 1 vector of phenotypic values of BW for n broilers, X is the n x 55 matrix of fixed effects: sex (2 classes), hatch (36 classes) and mating group (17 classes), $\boldsymbol{\beta}$ is the 55 x 1 vector of corresponding coefficients of fixed effects, $\boldsymbol{w}$ is the vector with elements of 0, 1,

and 2 for the homozygote of the minor allele, heterozygote, and homozygote of the major allele, $\boldsymbol{\alpha}$ is the vector of the fixed effect for the minor allele of the candidate SNP to be tested for association, Z is the incidence matrix relating observations to the polygenic random effects, $\boldsymbol{u}$ is the vector of polygenic random effects and $\boldsymbol{e}$ is the vector of random residuals. The random effects were assumed to be normally distributed with zero means and the following covariance structure:

$$Var\begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} G\sigma_u^2 & 0 \\ 0 & I\sigma_e^2 \end{bmatrix}$$

where σ_u^2 and $\mathbf{I}\sigma_{\varepsilon}^{\mathbf{2}}$ σ_e^2 are the polygenic and error variance components, I is the nxn identity matrix, and G is the nxn genomic relationship matrix (GRM, [15]) with elements of pairwise relationship coefficient using the 262,067 SNPs. The genomic relationship coefficient between two individuals j and k, was estimated as follows:

$$\frac{1}{262{,}067}\sum_{i=1}^{262{,}067}\frac{(x_{ij}-2p_i)(x_{ik}-2p_i)}{2p_i(1-2p_i)}$$

where x_{ij} and x_{ik} represent the number (0, 1, or 2) of the minor allele for the i_{th} SNP of the j_{th} and k_{th} individuals, and p_i is the frequency of the minor allele [15].
Statistically significant markers were selected at the optimal step of the MLMM stepwise regression according to extended Bayesian Information Criterion (eBIC [16]). P-values of these SNPs were then corrected for multiple comparisons using the false-discovery rate (FDR [17]) and 0.05 significance threshold. The genomic inflation factor (λ) was also calculated. λ values less than or close to 1 denote no potential systematic bias due to population structure or to the analytical approach [18]. All analyses were performed using the SNP & Variation Suite software (version 8.8.1).

2.3.3. Identification of QTL and positional candidate genes

Since the current population displays considerable LD levels for markers distanced up to 1Mb, we searched for growth related QTL/associations as well as positional candidate genes within 1Mb distances around the significant SNPs using the ChickenQTLdb [19] and the NCBI database (http://www.ncbi.nlm.nih.gov/snp/?term=gallus+gallus and http://www.ncbi.nlm.nih.gov/gene), respectively. The positions of QTL were remapped from *Gallus gallus 4* to *Gallus _gallus-5.0* assembly using the Genome remapping service from NCBI database (https://www.ncbi.nlm.nih.gov/genome/tools/remap).

2.3.4. Functional enrichment and pathway analysis

Functional enrichment analysis (FEA) of candidate genes was carried out using the PANTHER (Protein ANalysis THrough Evolutionary Relationships http://pantherdb.org/) data base [20] and the following settings: reference gene list in *Homo sapiens*, PANTHER Gene Ontology (GO) - Biological Process (BP) and the Bonferroni multiple testing correction. In total, 575 out of the 1,012 candidate genes were identified by PANTHER and used during FEA. GO BPs with p-values lower than 0.05 were considered as significantly enriched. Next, the ToppFun portal [10] was employed to detect the significantly enriched gene families (i.e. groups of homologous genes with common evolutionary origin and biological functions) using a threshold FDR p-value of 0.05. Pathway analysis (PA) followed, using the Reactome pathway database and the ReactomeFIViz application [21] of Cytoscape

(http://www.cytoscape.org/). Also here a threshold FDR p-value of 0.05 was used to identify significant pathways.

2.3.5. Gene prioritization

Two more computational approaches i.e. GPA and GNA were applied to assess the functional relevance or connectivity of the positional candidate genes and prioritize these accordingly. First, *in silico* GPA was performed, based on the functional similarity of candidate genes to a list of training genes (n=763) with annotated functions. These genes were extracted from the NCBI data base using the following search terms: body weight, body size and BMI in human and mouse. GPA was performed with the ToppGene portal [10]. The portal performs functional annotation-based candidate gene prioritization using fuzzy-based similarity measures to compute the similarity between any two genes based on semantic annotations. In the present study, the following semantic annotations were used: Molecular Function, Biological Process, Cellular Component, Human Phenotype, Mouse Phenotype and Pathway. For each test gene, a p-value for each annotation was derived by random sampling of 5,000 genes from the whole genome and these partial p-values were combined into an overall score using statistical meta-analysis. Prioritized genes were ranked in diminishing order of the combined (overall) p-value using a threshold value of 0.05 for significance.

Finally, the NetworkAnalyst (http://www.networkanalyst.ca/, version 3.0) [22] portal was employed to perform GNA. The aim of this analysis was to detect hub nodes i.e. genes with high connectivity (or node degree as it is known in graph and network analysis) in the PPIN. Here, default networks are created by searching for direct interaction partners in the molecular interaction knowledgebase which are generally known as the first-order interaction networks. In our case, the use of n=635 query genes ('seeds') resulted in a huge network rendering its topological analysis impossible. For this reason, we extracted, from the first-order network, a minimal sub-network with maximal connections between genes, using the Prize-collecting Steiner Forest (PCSF) algorithm. The expanded network was constructed using 635 genes (LOC and MIR genes were excluded) and the IMEx Interactome data base that exploits literature-curated comprehensive data from InnateDB.

2.4. Results

2.4.1. Significant SNPs and positional candidate genes

Genomic inflation factor λ was as high as 0.93 indicating no systematic bias due to population structure or analytical approach. Figure 1 presents the global view of SNP p-values (at the $-\log_{10}$ scale) across the 28 autosomes (GGA1-28). In total, 12 SNPs dispersed across 9 autosomes (GGA1, GGA4, GGA10, GGA11, GGA15, GGA22, GGA25, GGA26 and GGA27) reached genome-wide significance (FDR p-value<0.05, Table 1) while 1,012 positional candidate genes (see Supplementary Table S1) were identified within the searched genomic regions. Of the candidate genes, n=350 were non-annotated (LOC genes) and n=27 involved microRNA (miRNA) genes resulting in a total number n= 635 annotated genes. The maximum number of candidate genes (n=198) was observed for *rs312758346 (*GGA25) and the minimum (n=25) for *rs316794400* (GGA22). Seven significant markers were located within the following genes: *SLAIN2, ZC3H18, TMEM132D, F-KER, FCRL4, LEMD2* and *CACNB1* (Supplementary Table S1).

2.4.2. Detection of QTL/associations

A total number of n=197 growth related QTL/associations are reported within the searched genomic regions around the significant markers (Supplementary Table S2). These QTL/associations are distributed across eight chromosomes (GGA1, GGA4, GGA10, GGA11, GGA15, GGA22, GGA26 and GGA27) and pertain to growth traits such as carcass

weight, abdominal fat percentage, breast muscle percentage etc. The maximum number (n=65) of QTL/associations is reported around marker *rs315329074* (GGA27) and the minimum number (n=1) for *rs316794400* (GGA22*).* No QTLs or SNP associations are reported around *rs317288536* and *rs312758346.*

2.4.3. Functional enrichment and pathway analysis

FEA revealed 49 candidate genes as participating in the BP of system development (GO:0048731, Bonferroni p-value=0.0264) and 17 genes participating in the skeletal system development (GO:0001501, Bonferroni p-value=0.00398) (Table 2). Furthermore, 88 out of 1,012 genes were members of 14 gene families. Amongst them, the three most significant gene families were the following: S100 calcium binding proteins, HOXL subclass homeoboxes and type I Keratins (Table 3) with a total number of 28 genes. A total number of 25 unique genes were identified across functionally relevant to BW pathways such as MAPK6/MAPK4 signaling, signaling by EGFR, signaling by Type 1 *Insulin-like Growth Factor 1 Receptor* (*IGF1R*), signaling by insulin receptor, TCF dependent signaling in response to WNT and nerve growth factor (NGF) pathway (Table 4). The highest number of genes (n=22) was observed for NGF pathway. Note that a lot of genes were enriched in many pathways, simultaneously. For instance, genes *UBC, PSME3, PSMD7, PSMD4, PSMB4, PSMB3* and *PSMD3* were enriched in six pathways (Table 4).

2.4.4. Gene prioritization and network analysis

A total number of n=559 positional candidate genes were submitted to GPA as the rest genes were non-annotated or could not be mapped to any human annotated gene. Results of GPA are on Table S3. In total, 248 genes were prioritized (p-value <0.05). Among them, the first 10 top ranked genes were: *SMAD4, CHRNB2, CDH1, NTRK1, RARA, STAT5B, SCARB1, NR1D1, SHC1* and *CYBB*. The sub-network constructed during GNA consisted of 447 nodes (genes), 1,022 edges (connections) and 208 seeds (proteins). A graphical depiction of this sub-network is shown in Figure 2. Detailed information of node degrees is provided in Supplementary Table S4. The average node degree of this network was 4.6 (SD=5.7, min=1, max=68). In total, there were 22 hub genes detected with node degree ≥15 (range: 15 to 68). A total number of 31 common genes were detected between prioritized genes and genes with node degree higher than average (>5). The ranking (Spearman) correlation of these genes was as high as 0.53 implying a moderate rank similarity between the two analyses. Of the 22 hub genes, 14 genes (*UBC, SMAD4, SHC1, NRAS, PSMD4, CDC6, PSMD7, RARA, PSMB4, CDH1, STAT5B, MED1, PSMD3* and *CDT1)* were also prioritized during GPA implying that highly connected genes tend to be prioritized as well.

2.5. Discussion

Present results confirm previous findings suggesting that GGA1 and GGA4 [23] as well as GGA10, GGA15, GGA22, GGA26 [2] and GGA27 [24] harbor QTLs related to BW. Current findings also confirm the importance of Wnt-signaling, MAPK and insulin signaling pathways for growth traits in the species [23] as well as genes (*STAT5B* [25], *TXK* [26], *GABRG1* [26] and *SGCB* [27]) which have previously been reported as significant for BW in chickens.

Significant markers that fall within genes pointed out the following seven genes: *SLAIN2, ZC3H18, TMEM132D, F-KER, FCRL4, LEMD2 and CACNB1*. Of these, *SLAIN2, LEMD2, F-KER* and *CACNB1* have documented involvement in growth or body structure. Specifically, *SLAIN2* (*SLAIN motif family member 2*) gene is necessary for the normal structure of microtubule cytoskeleton as it controls the microtubule growth during interphase [28]. *F-KER (feather keratin I)* is a feather keratin affecting epidermal structure [29] while *LEMD2 (LEM domain containing 2)* plays an important role in mouse embryonic development by regulating various signaling pathways such as MAPK (mitogen-activated protein kinase) and AKT (also

known as Protein Kinase B) [30]. Finally, the murine *Cacnb1 (calcium voltage-gated channel auxiliary subunit beta 1)* gene is known to affect skeletal muscle development [31].
Functional enrichment analysis and gene family analysis pointed out genes (*MEOX1, HOXB3, HOXB4, HOXB5, HOXB9 and HOXB13*) that were also found to participate in BP of skeletal system development. These genes belong to the super-family of homeobox genes that play a fundamental role in embryonic development, cell proliferation and metabolic processes [32,33,34]. Additional genes in the same BP were *PHOSPHO1 (phosphoethanolamine/phosphocholine phosphatase)* that contributes to bone mineralization during embryonic bone development in chickens [35], *MFGE8 (milk fat globule-EGF factor 8 protein)* that promotes obesity in mice [36] and *SCUBE3* (*signal peptide, CUB domain and EGF like domain containing 3)* that is implicated in murine embryonic development [37].
From the rest significant gene families, S100 proteins have been reported as regulators in several functions such as Ca^{2+} homeostasis, energy metabolism, proliferation and differentiation [38] and type I keratins are filament-forming proteins of epithelial cells that are necessary for the normal structure and function of the tissues [39]. In chickens, *KRT14* and *KRT15* are reported to participate in keratinocytes proliferation [40] and pigmentation of muscle tissues [41] , respectively.
Results of pathway analysis were also relevant to the trait under study. Specifically, the nerve growth factor (NGF) belongs to neurotrophins that play an important role in regulation of growth and survival of nerve cells [42]. The epidermal growth factor receptor (EGFR) signaling pathway regulate various functions such as growth, survival, proliferation, and differentiation in mammalian cells [43]. The Wnt-signaling is required in several embryonic developmental processes e.g. skeletal morphogenesis [44]. MAPK6 and MAPK4 (also known as ERK3 and ERK4) are reported to contribute to cell differentiation and cell cycle regulation [45] and IGF1R signaling regulates skin development and differentiation [46] as well as glucose, lipid, and energy homeostasis [47].
Pathway analysis also highlighted the importance of *UBC, PSME3, PSMD7, PSMD4, PSMB4, PSMB3 and PSMD3* as growth-related genes. Specifically, *UBC* (*ubiquitin C*) is significant for liver development in mice [48], the absence of PSME3 (*proteasome activator subunit 3*) causes retardation of cell proliferation and body growth [49], *PSMD7* (*proteasome 26S subunit, non-ATPase 7*) regulates cell proliferation, cell cycle and cell apoptosis [50] while *PSMD4* (*proteasome 26S subunit, non-ATPase 4*, also known as *Rpn10*) is essential for embryonic development [51]. *PSMB4* (*proteasome subunit beta 4*) regulates cell growth [52] while *PSMB3 (proteasome subunit beta 3)* is involved in an ATP/ubiquitin-dependent process [53] which is responsible for muscle atrophy in rats [54]. Finally, *PSMD3 (proteasome 26S subunit, non-ATPase 3)* inhibits cell proliferation and induces cellular apoptosis [55].
The following 14 genes *UBC, SMAD4, SHC1, NRAS, PSMD4, CDC6, PSMD7, RARA, PSMB4, CDH1, STAT5B, MED1, PSMD3 and CDT1* were highlighted by both GPA and GNA. Aside from *UBC, PSMD4, PSMD7, PSMB4, STAT5B and PSMD3* that were discussed earlier, in this group fall additional highly promising causal genes. Specifically, *SMAD4 (SMAD family member 4)* is a central mediator of the transforming growth factor *β* signaling pathway which affects among others the cell growth [56] while *SHC1 (SHC adaptor protein 1),* mediates the IGF-1 pathway and contributes to the activation of Ras/MAPK pathway leading to cell proliferation [57]. *NRAS (neuroblastoma RAS viral oncogene homolog*) controls cell growth, differentiation, and survival by facilitating signal transduction [58] while *CDC6 (cell division cycle 6*) is an essential DNA replication factor and its loss-of-function leads to aberrant cell proliferation [59]. *RARA (retinoic acid receptor alpha)* affects the hippocampal development [60] while *CDH1 (cadherin 1*) mediates early embryonic development and cell differentiation [61]. Finally, *MED1 (mediator complex subunit 1)* has a key role in mammary epithelial cell growth [62] and *CDT1 (chromatin licensing and DNA replication factor 1*) deregulation affects cell proliferation [63].

Another interesting finding of the present study was the inclusion of miRNAs in the list of candidate genes for the trait under study. miRNAs act as post-transcriptional regulators in gene expression and play an essential role in a wide variety of biological processes including cell proliferation, differentiation, metabolism and growth [64]. Of the detected microRNAs, miR3529 is involved in follicular development-related pathways in chickens [65], miR140 promotes myoblast proliferation in chickens [66] while miR10A is expressed in chicken breast muscles [67] and is implicated in muscle development and myogenesis regulation in chickens [68].

In general, results of the computational approaches employed in the present study have pointed out several known or novel most promising causal genes for the trait under study. Nevertheless, the presence of different functional candidate genes across analyses indicates that gene candidacy prediction was method specific. This is a rather expected finding as each analysis uses different principle foundation(s) to infer functional relevance of the candidate genes. In over-representation approaches (functional enrichment analysis, gene family analysis and pathway analysis), identification of genes or pathways is based on statistical testing of over-represented for terms within given functions or known pathways. Here, results may be test dependent and biased toward well-known pathways or multi-functional genes because of many over-represented pathways, leading to false positives [69].

On the other hand, functional annotation-based (GPA) or network-based gene (GNA) prioritization displays important advantages. In the first (GPA), a plethora of data sources are used to infer functional similarity of candidate genes to known genes or phenotypes, including GO, pathway annotations, published and gene expression data [10]. An obvious limitation here is the poor functional prediction for genes with limited or unknown annotated functions [70]. In the second (GNA), high node degree of hubs is widely exploited, but this salient feature may also lead to false positives as highly connected genes tend to have many annotated functions [11]. As hub genes tend to have many semantic annotations, they are also expected to be highly prioritized during GPA.

In general, the performance of prioritization methods is dependent on the mined databases and the quality of interaction data that may suffer from incompleteness and unreliability with missing interactions [10]. A final critical point with regard to all methods applied here relates to data mined to infer functional relevance of candidate genes. In all tools, only information and/or PPINs referring to human or murine proteins are used [10] and this information is used to infer gene functions in other species such as that studied here.

In conclusion, gene enrichment and/or prioritization methods are useful to narrow down the large list of positional candidate genes provided by GWAS findings. Nevertheless, prioritized genes appear to be method dependent and each method is not flawless. Identification of top prioritized or common genes among methods offer a good strategy to identify highly promising candidate genes in an attempt to reduce time and costs of experimental validation of functional analyses. Finally, present findings are supportive of the hypothesis that the genetic architecture of the trait approximates the infinitesimal model.

2.6. References

[1] Xie L, Luo C, Zhang C, Zhang R, Tang J, Nie Q, et al. Genome-Wide Association Study Identified a Narrow Chromosome 1 Region Associated with Chicken Growth Traits. Liu Z, editor. PLoS One. Public Library of Science; 2012;7: e30910. doi:10.1371/journal.pone.0030910

[2] Van Goor A, Bolek KJ, Ashwell CM, Persia ME, Rothschild MF, Schmidt CJ, et al. Identification of quantitative trait loci for body temperature, body weight, breast yield, and digestibility in an advanced intercross line of chickens under heat stress. Genet Sel Evol. BioMed Central; 2015;47: 96. doi:10.1186/s12711-015-0176-7

[3] Kranis A, Gheyas AA, Boschiero C, Turner F, Yu L, Smith S, et al. Development of a high density 600K SNP genotyping array for chicken. BMC Genomics. BioMed Central; 2013;14: 59. doi:10.1186/1471-2164-14-59

[4] Pettersson ME, Carlborg Ö. Dissecting the genetic architecture of complex traits and its impact on genetic improvement programs: lessons learnt from the Virginia chicken lines. Rev Bras Zootec. Sociedade Brasileira de Zootecnia; 2010;39: 256–260. doi:10.1590/S1516-35982010001300028
[5] Shahjahan M, Liu R, Zhao G, Wang F, Zheng M, Zhang J, et al. Identification of Histone Deacetylase 2 as a Functional Gene for Skeletal Muscle Development in Chickens. Asian-Australasian J Anim Sci. 2016;29: 479–486. doi:10.5713/ajas.15.0252
[6] Ouyang H, Zhang H, Li W, Liang S, Jebessa E, Abdalla BA, et al. Identification, expression and variation of the *GNPDA2* gene, and its association with body weight and fatness traits in chicken. PeerJ. PeerJ Inc.; 2016;4: e2129. doi:10.7717/peerj.2129
[7] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. National Academy of Sciences; 2005;102: 15545–50. doi:10.1073/pnas.0506580102
[8] García-Campos MA, Espinal-Enríquez J, Hernández-Lemus E. Pathway Analysis: State of the Art. Front Physiol. Frontiers; 2015;6: 383. doi:10.3389/fphys.2015.00383
[9] Chen J, Aronow BJ, Jegga AG. Disease candidate gene identification and prioritization using protein interaction networks. BMC Bioinformatics. 2009;10: 73. doi:10.1186/1471-2105-10-73
[10] Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res. 2009;37: W305–W311. doi:10.1093/nar/gkp427
[11] Gillis J, Pavlidis P. The Impact of Multifunctional Genes on "Guilt by Association" Analysis. Bader J, editor. PLoS One. 2011;6: e17258. doi:10.1371/journal.pone.0017258
[12] Walker MG, Volkmuth W, Sprinzak E, Hodgson D, Klingler T. Prediction of gene function by genome-scale expression analysis: prostate cancer-associated genes. Genome Res. 1999;9: 1198–203. Available: http://www.ncbi.nlm.nih.gov/pubmed/10613842
[13] Stelzer G, Plaschkes I, Oz-Levi D, Alkelai A, Olender T, Zimmerman S, et al. VarElect: the phenotype-based variation prioritizer of the GeneCards Suite. BMC Genomics. BioMed Central; 2016;17: 444. doi:10.1186/s12864-016-2722-2
[14] Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, et al. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet. Nature Publishing Group; 2012;44: 825–830. doi:10.1038/ng.2314
[15] VanRaden PM. Efficient Methods to Compute Genomic Predictions. J Dairy Sci. 2008;91: 4414–4423. doi:10.3168/jds.2007-0980
[16] Chen J, Chen Z. Extended Bayesian information criteria for model selection with large model spaces. Biometrika. Oxford University Press; 2008;95: 759–771. doi:10.1093/biomet/asn034
[17] Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing [Internet]. Journal of the Royal Statistical Society. Series B (Methodological). WileyRoyal Statistical Society; 1995. pp. 289–300. doi:10.2307/2346101
[18] Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, Willer CJ, et al. Genomic inflation factors under polygenic inheritance. Eur J Hum Genet. Nature Publishing Group; 2011;19: 807–12. doi:10.1038/ejhg.2011.39
[19] Chicken QTL database [Internet]. [cited 21 Apr 2019]. Available: https://www.animalgenome.org/cgi-bin/QTLdb/GG/download?tmpname=mapDwnLd&file=cM
[20] Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, et al. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. 2017;45: D183–D189. doi:10.1093/nar/gkw1138
[21] Wu G, Dawson E, Duong A, Haw R, Stein L. ReactomeFIViz: A cytoscape app for pathway and network-based data analysis. F1000Research. Faculty of 1000 Ltd; 2014;3. doi:10.12688/f1000research.4431.2
[22] Xia J, Benner MJ, Hancock REW. NetworkAnalyst--integrative approaches for protein-protein interaction network analysis and visual exploration. Nucleic Acids Res. 2014;42: W167-74. doi:10.1093/nar/gku443
[23] Xu Z, Nie Q, Zhang X. Overview of Genomic Insights into Chicken Growth Traits Based on Genome- Wide Association Study and microRNA Regulation. Curr Genomics. 2013;14: 137–146. doi:10.2174/1389202911314020006
[24] Lien C-Y, Tixier-Boichard M, Wu S-W, Wang W-F, Ng CS, Chen C-F. Detection of QTL for traits related to adaptation to sub-optimal climatic conditions in chickens. Genet Sel Evol. BioMed Central; 2017;49: 39. doi:10.1186/s12711-017-0314-5

[25] Zhao XH, Wang JY, Zhang GX, Wei Y, Gu YP, Yu YB. Single nucleotide polymorphism in the STAT5b gene is associated with body weight and reproductive traits of the Jinghai Yellow chicken. Mol Biol Rep. Springer Netherlands; 2012;39: 4177–4183. doi:10.1007/s11033-011-1202-7
[26] Gu X, Feng C, Ma L, Song C, Wang Y, Da Y, et al. Genome-Wide Association Study of Body Weight in Chicken F2 Resource Population. Mailund T, editor. PLoS One. Public Library of Science; 2011;6: e21872. doi:10.1371/journal.pone.0021872
[27] Zhang GX, Fan QC, Zhang T, Wang JY, Wang WH, Xue Q, et al. Genome-wide association study of growth traits in the Jinghai Yellow chicken. Genet Mol Res. 2015;14: 15331–15338. doi:10.4238/2015.November.30.10
[28] van der Vaart B, Manatschal C, Grigoriev I, Olieric V, Gouveia SM, Bjelić S, et al. SLAIN2 links microtubule plus end–tracking proteins and controls microtubule growth in interphase. J Cell Biol. 2011;193: 1083–1099. doi:10.1083/jcb.201012179
[29] Greenwold MJ, Sawyer RH. Genomic organization and molecular phylogenies of the beta (beta) keratin multigene family in the chicken (Gallus gallus) and zebra finch (Taeniopygia guttata): implications for feather evolution. BMC Evol Biol. BioMed Central; 2010;10: 148. doi:10.1186/1471-2148-10-148
[30] Tapia O, Fong LG, Huber MD, Young SG, Gerace L. Nuclear Envelope Protein Lem2 is Required for Mouse Development and Regulates MAP and AKT Kinases. Palazzo AF, editor. PLoS One. 2015;10: e0116196. doi:10.1371/journal.pone.0116196
[31] Chen F, Liu Y, Sugiura Y, Allen PD, Gregg RG, Lin W. Neuromuscular synaptic patterning requires the function of skeletal muscle dihydropyridine receptors. Nat Neurosci. Nature Publishing Group; 2011;14: 570–577. doi:10.1038/nn.2792
[32] Reijntjes S, Stricker S, Mankoo BS. A comparative analysis of Meox1 and Meox2 in the developing somites and limbs of the chick embryo. Int J Dev Biol. 2007;51: 753–759. doi:10.1387/ijdb.072332sr
[33] Procino A, Cillo C. The HOX genes network in metabolic diseases. Cell Biol Int. 2013;37: n/a-n/a. doi:10.1002/cbin.10145
[34] Gouveia A, Marcelino HM, Gonçalves L, Palmeirim I, Andrade RP. Patterning in time and space: HoxB cluster gene expression in the developing chick embryo. Cell Cycle. Taylor & Francis; 2015;14: 135. doi:10.4161/15384101.2014.972868
[35] MacRae VE, Davey MG, McTeir L, Narisawa S, Yadav MC, Millan JL, et al. Inhibition of PHOSPHO1 activity results in impaired skeletal mineralization during limb development of the chick. Bone. 2010;46: 1146–1155. doi:10.1016/j.bone.2009.12.018
[36] Khalifeh-Soltani A, McKleroy W, Sakuma S, Cheung YY, Tharp K, Qiu Y, et al. Mfge8 promotes obesity by mediating the uptake of dietary fats and serum fatty acids. Nat Med. 2014;20: 175–183. doi:10.1038/nm.3450
[37] Haworth K, Smith F, Zoupa M, Seppala M, Sharpe PT, Cobourne MT. Expression of the Scube3 epidermal growth factor-related gene during early embryonic development in the mouse. Gene Expr Patterns. 2007;7: 630–634. doi:10.1016/j.modgep.2006.12.004
[38] Donato R, Cannon BR, Sorci G, Riuzzi F, Hsu K, Weber DJ, et al. Functions of S100 proteins. Curr Mol Med. NIH Public Access; 2013;13: 24–57. Available: http://www.ncbi.nlm.nih.gov/pubmed/22834835
[39] Schweizer J, Bowden PE, Coulombe PA, Langbein L, Lane EB, Magin TM, et al. New consensus nomenclature for mammalian keratins. J Cell Biol. 2006;174: 169–174. doi:10.1083/jcb.200603161
[40] Couteaudier M, Trapp-Fragnet L, Auger N, Courvoisier K, Pain B, Denesvre C, et al. Derivation of keratinocytes from chicken embryonic stem cells: Establishment and characterization of differentiated proliferative cell populations. Stem Cell Res. Elsevier; 2015;14: 224–237. doi:10.1016/J.SCR.2015.01.002
[41] Tarique T, Yang S, Mohsina Z, Qiu J, Yan Z, Chen G, et al. Identification of genes involved in regulatory mechanism of pigments in broiler chickens. Genet Mol Res. 2014;13: 7201–7216. doi:10.4238/2014.September.5.6
[42] Wang H, Wang R, Thrimawithana T, Little PJ, Xu J, Feng Z-P, et al. The nerve growth factor signaling and its potential as therapeutic target for glaucoma. Biomed Res Int. 2014;2014: 759473. doi:10.1155/2014/759473
[43] Oda K, Matsuoka Y, Funahashi A, Kitano H. A comprehensive pathway map of epidermal growth factor receptor signaling. Mol Syst Biol. EMBO; 2005;1. doi:10.1038/msb4100014

[44] Yang Y. Wnt signaling in development and disease. Cell and Bioscience. BioMed Central; 2012. p. 14. doi:10.1186/2045-3701-2-14
[45] Long W, Foulds CE, Qin J, Liu J, Ding C, Lonard DM, et al. ERK3 signals through SRC-3 coactivator to promote human lung cancer cell invasion. J Clin Invest. 2012;122: 1869–1880. doi:10.1172/JCI61492
[46] Sadagurski M, Yakar S, Weingarten G, Holzenberger M, Rhodes CJ, Breitkreutz D, et al. Insulin-Like Growth Factor 1 Receptor Signaling Regulates Skin Development and Inhibits Skin Keratinocyte Differentiation. Mol Cell Biol. American Society for Microbiology; 2006;26: 2675–2687. doi:10.1128/mcb.26.7.2675-2687.2006
[47] Boucher J, Kleinridders A, Ronald Kahn C. Insulin receptor signaling in normal and insulin-resistant states. Cold Spring Harb Perspect Biol. 2014;6. doi:10.1101/cshperspect.a009191
[48] Hallengren J, Chen PC, Wilson SM. Neuronal Ubiquitin Homeostasis. Cell Biochem Biophys. NIH Public Access; 2013;67: 67–73. doi:10.1007/s12013-013-9634-4
[49] Murata S, Kawahara H, Tohma S, Yamamoto K, Kasahara M, Nabeshima Y, et al. Growth retardation in mice lacking the proteasome activator PA28gamma. J Biol Chem. American Society for Biochemistry and Molecular Biology; 1999;274: 38211–5. doi:10.1074/JBC.274.53.38211
[50] Shi K, Zhang JZ, Zhao RL, Yang L, Guo D. PSMD7 downregulation induces apoptosis and suppresses tumorigenesis of esophageal squamous cell carcinoma via the mTOR/p70S6K pathway. FEBS Open Bio. Wiley Blackwell; 2018;8: 533–543. doi:10.1002/2211-5463.12394
[51] Hamazaki J, Sasaki K, Kawahara H, Hisanaga S -i., Tanaka K, Murata S. Rpn10-Mediated Degradation of Ubiquitinated Proteins Is Essential for Mouse Development. Mol Cell Biol. American Society for Microbiology; 2007;27: 6629–6638. doi:10.1128/mcb.00509-07
[52] Wang H, He Z, Xia L, Zhang W, Xu L, Yue X, et al. PSMB4 overexpression enhances the cell growth and viability of breast cancer cells leading to a poor prognosis. Oncol Rep. Spandidos Publications; 2018;40: 2343–2352. doi:10.3892/or.2018.6588
[53] Choong LY, Lim S, Chong PK, Wong CY, Shah N, Lim YP. Proteome-Wide Profiling of the MCF10AT Breast Cancer Progression Model. Aziz SA, editor. PLoS One. Public Library of Science; 2010;5: e11030. doi:10.1371/journal.pone.0011030
[54] Temparis S, Asensi M, Estrela JM, Ferrara M, Larbaud D, Attaix D. Increased ATP-Ubiquitin-dependent Proteolysis in Skeletal Muscles of Tumor-bearing Rats. Cancer Res. 1994;54: 5568–5573.
[55] Fararjeh AS, Chen LC, Ho YS, Cheng TC, Liu YR, Chang HL, et al. Proteasome 26s subunit, non-atpase 3 (Psmd3) regulates breast cancer by stabilizing her2 from degradation. Cancers (Basel). MDPI AG; 2019;11. doi:10.3390/cancers11040527
[56] Zhao M, Mishra L, Deng C-X. The role of TGF-β/SMAD4 signaling in cancer. Int J Biol Sci. 2018;14: 111–123. doi:10.7150/ijbs.23230
[57] Wagner K, Hemminki K, Grzybowska E, Klaes R, Butkiewicz D, Pamula J, et al. The insulin-like growth factor-1 pathway mediator genes: SHC1 Met300Val shows a protective effect in breast cancer. Carcinogenesis. 2004;25: 2473–2478. doi:10.1093/carcin/bgh263
[58] Eisfeld A-K, Schwind S, Hoag KW, Walker CJ, Liyanarachchi S, Patel R, et al. NRAS isoforms differentially affect downstream pathways, cell growth, and cell transformation. doi:10.1073/pnas.1401727111
[59] Borlado LR, Méndez J. CDC6: from DNA replication to cell cycle checkpoints and oncogenesis. Carcinogenesis. 2008;29: 237–43. doi:10.1093/carcin/bgm268
[60] Huang H, Wei H, Zhang X, Chen K, Li Y, Qu P, et al. Changes in the expression and subcellular localization of RARα in the rat hippocampus during postnatal development. Brain Res. 2008;1227: 26–33. doi:10.1016/j.brainres.2008.06.073
[61] Stemmler MP, Bedzhov I. A Cdh1HA knock-in allele rescues the Cdh1-/- phenotype but shows essential Cdh1 function during placentation. Dev Dyn. John Wiley & Sons, Ltd; 2010;239: 2330–2344. doi:10.1002/dvdy.22375
[62] Hasegawa N, Sumitomo A, Fujita A, Aritome N, Mizuta S, Matsui K, et al. Mediator Subunits MED1 and MED24 Cooperatively Contribute to Pubertal Mammary Gland Development and Growth of Breast Carcinoma Cells. Mol Cell Biol. 2012;32: 1483–1495. doi:10.1128/MCB.05245-11
[63] Pozo P, Cook J. Regulation and Function of Cdt1; A Key Factor in Cell Proliferation and Genome Stability. Genes (Basel). MDPI AG; 2016;8: 2. doi:10.3390/genes8010002
[64] Zhang M, Li D-H, Li F, Sun J-W, Jiang R-R, Li Z-J, et al. Integrated Analysis of MiRNA and Genes Associated with Meat Quality Reveals that Gga-MiR-140-5p Affects Intramuscular Fat

Deposition in Chickens. Cell Physiol Biochem. S. Karger AG; 2018;46: 2421–2433. doi:10.1159/000489649
[65] Wang W, Wu K, Jia M, Sun S, Kang L, Zhang Q, et al. Dynamic Changes in the Global MicroRNAome and Transcriptome Identify Key Nodes Associated With Ovarian Development in Chickens. Front Genet. Frontiers Media SA; 2018;9: 491–491. doi:10.3389/fgene.2018.00491
[66] Luo W, Li E, Nie Q, Zhang X. Myomaker, regulated by MYOD, MYOG and miR-140-3P, promotes chicken myoblast fusion. Int J Mol Sci. MDPI AG; 2015;16: 26186–26201. doi:10.3390/ijms161125946
[67] Ouyang H, He X, Li G, Xu H, Jia X, Nie Q, et al. Deep sequencing analysis of miRNA expression in breast muscle of fast-growing and slow-growing broilers. Int J Mol Sci. MDPI AG; 2015;16: 16242–16262. doi:10.3390/ijms160716242
[68] Khatri B, Seo D, Shouse S, Pan JH, Hudson NJ, Kim JK, et al. MicroRNA profiling associated with muscle growth in modern broilers compared to an unselected chicken breed. BMC Genomics. BioMed Central Ltd.; 2018;19: 683. doi:10.1186/s12864-018-5061-7
[69] Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37: 1–13. doi:10.1093/nar/gkn923
[70] Chen J, Aronow BJ, Jegga AG. Disease candidate gene identification and prioritization using protein interaction networks. BMC Bioinformatics. BioMed Central; 2009;10: 73. doi:10.1186/1471-2105-10-73

Tables and Figures of Chapter 2

Table 1. Genome-wide significant SNPs for BW. 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
rs13923872	1	112741685	0.0112	45
rs15608447	4	66885210	4.25E-09	38
rs312691174	4	29074989	0.00037	26
rs318199727	10	13536548	0.04111	44
rs318098582	11	18651449	0.00012	87
rs317945754	15	3557083	0.04594	29
rs316794400	22	4594855	6.07E-07	25
rs317288536	25	976833	8.05E-09	113
rs312758346	25	2412866	1.59E-05	198
rs317627533	26	4597439	2.12E-05	105
rs314452928	27	104022	0.0105	110
rs315329074	27	4528275	8.05E-16	192

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

Table 2. Significantly enriched GO biological processes.

GO_ID	GO BP term	p-value	Number of genes	Genes involved
GO:0001501	skeletal system development	0.00398	17	*HOXB4, BGLAP, MFGE8, HOXB9, ACAN, HOXB3, HAPLN3, ADAMTS4, MEOX1, PHOSPHO1, CNTNAP1, SCUBE3, PRICKLE4, HOXB13, BCAN, HOXB5, HAPLN2*
GO:0048731	system development	0.0264	49	*TPM3, CASK ,RIT1, HOXB4, BGLAP, PLEKHH3, FCER1G, PCDH18, CACNB1, NTRK3, KIRREL, MFGE8, DBNDD1, ZFHX3, ERBB2, HOXB9, CDH1, ACAN, HAP1, HOXB3, HAPLN3, ADAMTS4, ADGRD1, MEOX1, PHOSPHO1, CNTNAP1, CDK12, SCUBE3, PRICKLE4, HOXB13, FRYL, OLFML3, BCAN, PYGO2, NEUROD2, NGF, HOXB5, ZFPM1, ULK1, FMOD, PACSIN1, STAT3, HAPLN2, CSF3, NTRK1, IL6R, CDH3, PRELP, MEF2D*

Table 3. Gene families and related candidate genes.

Gene family	FDR p-value	Number of genes	Genes involved
S100 calcium binding proteins |EF-hand domain containing	4.45E-08	9	*S100A14, S100A1, S100A4, S100A6, S100A9, S100A10, S100A11, S100A13, S100A16*

HOXL subclass homeoboxes	1.13E-06	11	*MEOX1, HOXB13,HOXB1,HOXB2, HOXB3, HOXB4,HOXB5,HOXB6, HOXB7, HOXB8,HOXB9*
Keratins, type I	6.52E-06	8	*KRT20, KRT23,KRT10,KRT12, KRT14,KRT15, KRT17, KRT19*

Table 4. Candidate genes involved in significant pathways.

Pathway	**FDR**	**Number of genes**	**Genes involved**
Signaling by NGF	2.37E-06	22	*NTRK1, SHC1, PHB, NGF, NRAS, LAMTOR2, NCSTN, PIP5K1A, UBC, PIP4K2B,THEM4, PH1A, 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*

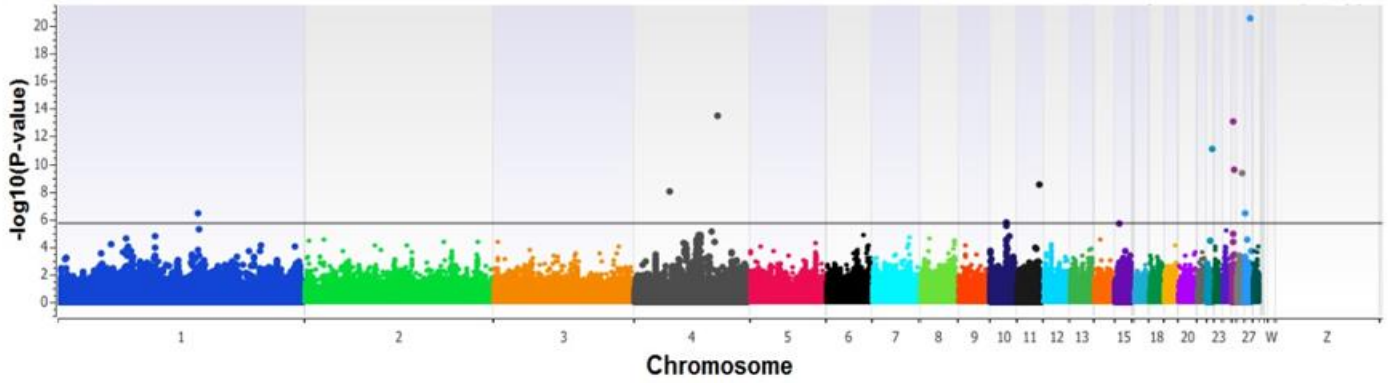


Figure 1. Manhattan plot displaying the $-\log_{10}$ (observed p-values) of the genome-wide SNPs (y-axis) across the 28 autosomes (x-axis). The horizontal line denotes the genome-wide significant threshold.

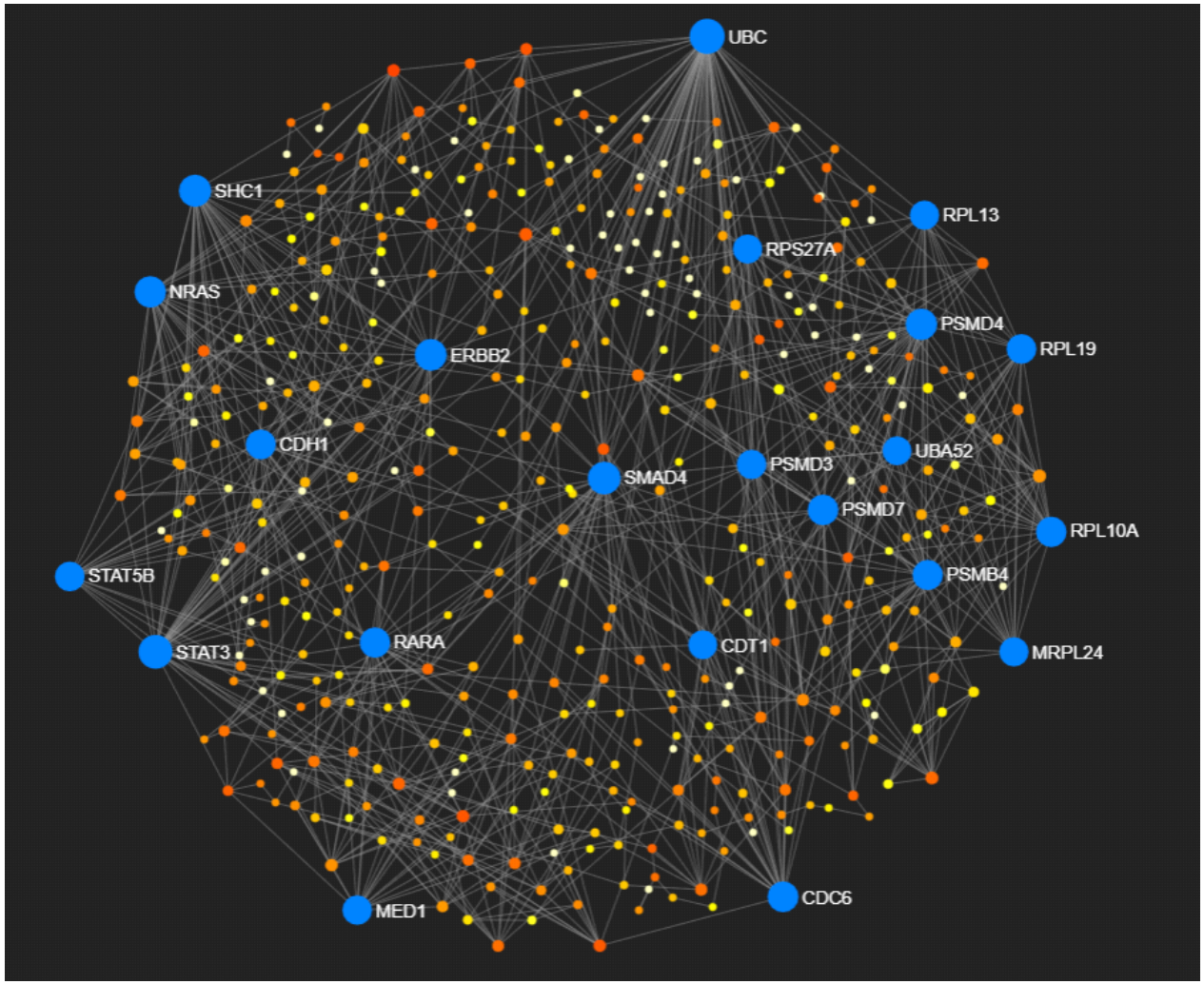


Figure 2. Depiction of a minimum gene network comprised of 447 nodes, 1,022 edges and 208 seed proteins. Genes with blue color (*UBC, STAT3, SMAD4, SHC1, ERBB2, NRAS, PSMD4, CDC6, PSMD7, RARA, RPL10A, PSMB4, CDH1, RPL19, STAT5B, MED1, PSMD3, RPS27A, RPL13, MRPL24, UBA52 and CDT1*) represent hub genes (node degree>15). Orange, yellow and white colors represent genes with node degree <15.

Chapter 3

Discovery and characterization of functional modules associated with body weight in broilers[1]

3.1. Abstract

Aim of the present study was to investigate whether body weight (BW) in broilers is associated with functional modular genes. To this end, first a GWAS for BW was conducted using 6,598 broilers and the high density SNP array. The next step was to search for positional candidate genes and QTLs within strong LD genomic regions around the significant SNPs. Using all positional candidate genes, a network was then constructed and community structure analysis was performed. Finally, functional enrichment analysis was applied to infer the functional relevance of modular genes. A total number of 645 positional candidate genes were identified in strong LD genomic regions around 11 genome-wide significant markers. 428 of the positional candidate genes were located within growth related QTLs. Community structure analysis detected 5 modules while functional enrichment analysis showed that 52 modular genes participated in developmental processes such as skeletal system development. An additional number of 14 modular genes (*GABRG1, NGF, APOBEC2, STAT5B, STAT3, SMAD4, MED1, CACNB1, SLAIN2, LEMD2, ZC3H18, TMEM132D, FRYL* and *SGCB*) were also identified as related to body weight. Taken together, current results suggested a total number of 66 genes as most plausible functional candidates for the trait examined.

3.2. Introduction

Body weight (BW) is an economically important trait for the broiler industry. This trait also presents considerable biological interest as it is a typical complex (polygenic) trait. To date, the ChickenQTLdb [1] contains over 7,812 QTL/SNP associations of which 3,582 are related to growth. Several genome wide association studies (GWAS) have already been performed for growth traits (e.g.[2,3]) in the species. The development of the chicken 600k SNP array [4] facilitates efficient screening for causal loci and genes with relevance to target traits due to the uniform coverage across chromosomes and the inclusion of markers within coding regions. Despite the large number of findings by GWAS, understanding of the genetic architecture of BW in chicken remains limited [5], since only a small number of positional candidate genes are confirmed as truly functionally relevant to the trait (e.g. *HDAC2* and *GNPDA2*[6,7]). The use of various Bioinformatics tools such as gene enrichment analysis [8], pathway analysis [9] and gene network analysis [10] can tackle this problem and aid in identifying the most promising functional candidate genes for the trait under study. Moreover, applications such as GeneMANIA [11] that is based on the guilt-by-association (GBA) principle [12] may also facilitate the identification of true causative genetic variants. The GBA principle states that gene products, which are protein interaction partners, tend to be functionally related [13]. Furthermore, genes in protein–protein interaction networks (PPINs) are organized into densely linked clusters i.e. communities or modules [14]. Modules present a structurally independent gene sub-network with more interior connections and consist of proteins which have the same or similar biological function(s) [15]. Modules could be further distinguished in protein complexes and in dynamic functional modules. Protein complexes are formed by several proteins which interact at the same place and time while dynamic functional modules are composed of few proteins participating in a specific cellular function not necessarily at the same place and time [16]. Moreover, functional modules consist of one or multiple protein

1 *published in Sci. Rep. 9, 9125 (2019). https://doi.org/10.1038/s41598-019-45520-5.*

complexes participating in a common biological process [17]. Since modules do not emerge by chance, they can reveal interactions with biological importance within large PPINs ([16,18]). The module-based approach has already been used to cluster genes into functional groups and to predict protein functions [19]. Investigation of functional modules has mainly been focused on human diseases such as obesity [20], breast cancer ([21,22]), coronary artery disease [23] and asthma [24]. Apart from human, functional modules have been identified in other species as well, such as in *Mus musculus* for discrete and rhythmic forelimb movements in motor cortex [25] and in *Gallus gallus* for muscle development and intramuscular fat accumulation at different post-hatching ages [26].

Driven from findings in other species and traits, aim of the present study was first to investigate whether body weight in broilers is associated with functional modules and second to propose novel candidate genes for the trait in question.

3.3. Methods

3.3.1. Ethics Statement

All animals included in this study were not subjected to any invasive procedures.

3.3.2. Data and quality control

In total, n= 6,727 broilers (n=3,735 males and n=2,992 females) from a grand-grandparent (GGP) commercial line with records on BW at 35 days of age were made available by Aviagen Ltd. Phenotypic records for BW ranged from 1,130 to 2,630 g with an average of 1840.2 g (SD=194 g). Animals were genotyped using the 600k Affymetrix® Axiom® high density genotyping array [4] resulting in a total number of 578,815 SNPs. Only autosomal SNPs (n=547,705) were considered. Quality control was performed first at a sample and second at a marker level. At a sample level, 72 females and 57 males were excluded due to call rate <0.99 and autosomal heterozygosity outside the 1.5 IQR (inter-quartile range) resulting in a number of n=6,598 samples. At a marker level, a number of 285,717 SNPs were excluded due to: call rate <0.99, MAF (minor allele frequency) <0.01 and linkage disequilibrium (LD) r^2 values greater than 0.99 within windows of 1 Mb inter-marker distance(s). A total of 6,598 samples and 262,067 SNPs were retained for GWAS. Quality control was performed using the SNP & Variation Suite software (version 8.8.1) of Golden Helix (http://www.goldenhelix.com).

3.3.3. Statistical analysis

A multi-locus mixed-model (MLMM) stepwise regression with forward inclusion and backward elimination [27] of SNPs was employed to identify genome-wide significant markers associated with the trait. The following statistical model was applied to the data:

$$y = X\boldsymbol{\beta} + \boldsymbol{w\alpha} + Z\boldsymbol{u} + \boldsymbol{e}$$

where y is the n x 1 vector of phenotypic values of BW for n broilers, X is the n x 55 matrix of fixed effects: sex (2 classes), hatch (36 classes) and mating group (17 classes), $\boldsymbol{\beta}$ is the 55 x 1 vector of corresponding coefficients of fixed effects, $\boldsymbol{w}$ is the vector with elements of 0, 1, and 2 for the homozygote of the minor allele, heterozygote, and homozygote of the major allele, $\boldsymbol{\alpha}$ is the vector of the fixed effect for the minor allele of the candidate SNP to be tested for association, Z is the incidence matrix relating observations to the polygenic random effects, $\boldsymbol{u}$ is the vector of polygenic random effects and $\boldsymbol{e}$ is the vector of random residuals.
The random effects were assumed to be normally distributed with zero means and the following covariance structure:

$$Var\begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} G\sigma_u^2 & 0 \\ 0 & I\sigma_e^2 \end{bmatrix}$$

where σ_u^2 and σ_e^2 are the polygenic and error variance components, I is the nxn identity matrix, and G is the nxn genomic relationship matrix (GRM [28]) with elements of pairwise relationship coefficient using the 262,067 SNPs. The genomic relationship coefficient between two individuals j and k, was estimated as follows:

$$\frac{1}{262{,}067}\sum_{i=1}^{262{,}067}\frac{(x_{ij}-2p_i)(x_{ik}-2p_i)}{2p_i(1-2p_i)}$$

where x_{ij} and x_{ik} represent the number (0, 1, or 2) of the minor allele for the i_{th} SNP of the j_{th} and k_{th} individuals, and p_i is the frequency of the minor allele [28].
Statistically significant markers were selected at the optimal step of the MLMM stepwise regression according to extended Bayesian Information Criterion (eBIC [29]). P-values of these SNPs were then corrected for multiple comparisons using the false-discovery rate (FDR [30]) correction method. Here, a cut-off FDR p-value less than 0.05 [31] was considered as significant. The FDR p-value of 0.05 states that, among all observed results, 5% would be false positives.

A Quantile-quantile (Q-Q) plot was also used to analyze the extent to which the observed distribution of the test statistic followed the expected (null) distribution. This plot along with the estimation of the genomic inflation factor (λ) was done to assess potential systematic bias due to population structure or to the analytical approach [32]. This analysis was performed using the SNP & Variation Suite (version 8.8.1) software (Golden Helix: http://www.goldenhelix.com).

3.3.4. Detection of candidate genomic regions with strong LD

We first estimated LD levels around each lead i.e. significant SNP. We then searched for genomic regions with strong LD around the lead SNPs defined as the maximum distance between the lead and the last SNP with D'≥0.8 [33]. Note that, the D', instead of the r^2 LD measurement, was preferably used here as the first one is reported to be independent [34] or less dependent [35] on MAF. All LD calculations were performed using the SNP & Variation Suite (version 8.8.1) software (Golden Helix: http://www.goldenhelix.com).

3.3.5. Identification of reported QTL and positional candidate genes

Next, we searched for growth/fatness related QTL in the ChickenQTLdb[1] and positional candidate genes in the NCBI database ([36,37]), within the strong LD genomic regions. Positions of QTL were remapped from *Gallus gallus 4* to *Gallus _gallus-5.0* assembly using the Genome Remapping Service from NCBI database [38].

3.3.6. Detection of community structure and functional module characterization

A gene network using all positional candidate genes was first constructed integrating the available *Homo sapiens* genes database (updated 17/3/2017) via the GeneMANIA V.3.4.1 plug-in [11] in Cytoscape V3.6.0 (http://cytoscape.org/ [39]). The gene network was built according to 7 types of interaction terms i.e. co-expression, co-localization, genetic interaction, pathway, physical interaction, predicted and shared protein domains. The automatic weighting method for network construction was also used while the number of related genes was set to zero.

Detection of community structure i.e. the appearance of densely interconnected nodes (modules) was then performed using the Girvan and Newman's clustering algorithm [40] via the GLay [41] of clusterMaker [42] plugin in Cytoscape [39]. This algorithm identifies modules within networks by repetitively removing edges with the highest "betweeness" i.e. edges between modules with higher values of betweeness rather than edges within modules. The strength of the network division into modules was also quantified using the modularity measure [40]. Typically, modularity values ranging from 0.3 to 0.7 are indicative of strong community structure [40].

Modular genes were then subjected to GO Biological Process (BP) term enrichment analysis using the DAVID functional annotation tool (https://david.ncifcrf.gov/, version 6.8) [43]. The *Homo sapiens* species was also selected for the input gene list and as whole genome background for enrichment analysis. The following settings were used during this analysis: an EASE score (a modified Fisher exact p-value [44]) cut-off=0.05 and a minimum number of genes per GO BP term=2. GO biological processes with p-values lower than 0.05 were considered as significantly enriched. The QuickGO [45] web-based tool was subsequently used to examine each resulting significantly enriched GO BP through browsing the hierarchical structure in the GO annotation database. GO BPs associated with developmental process or growth parent term(s) were considered as functionally relevant to the trait under study.

3.4. Data Availability

The data that support the findings of this study are available from Aviagen Ltd. but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Aviagen Ltd.

3.5. Results

3.5.1. Significant SNPs and positional candidate genes

Figure 1 shows the Q-Q plot of the expected and the observed p values ($-\log_{10}$ p values) of all SNPs. The genomic inflation factor (λ) was also estimated as high as 0.93. According to Kang et al. [46], λ values that lie outside of the conservative 95% confidence interval (0.992 to 1.008) denote dependency of SNPs. However, as the Q-Q plot clearly shows, there is no evidence of any systematic bias due to population structure or analytical approach in our case. As Yang et al. [32] emphasize in their paper, it is reasonable to expect deviation(s) of λ from 1 for purely polygenic traits such as that examined here in the absence of any systematic bias. The Q-Q plot also shows that some SNPs depart from the expected probability and thus might be associated with the trait. These SNPs are also displayed in Figure 1 in a form of a Manhattan plot.

Specifically, there were 12 SNPs detected, across nine autosomes (1, 4, 10, 11, 15, 22, 25, 26 and 27) reaching genome-wide significance (FDR p-value<0.05). A detailed description of the significant (lead) SNPs is provided in Table 1. Table 2 displays the extent of genomic regions displaying strong LD (D'>0.8) around the lead markers that were searched for positional candidate genes. Note that marker *rs312758346* (GGA25) was omitted here as LD levels around this marker were below the threshold LD value (D'<0.8). In total, 645 positional candidate genes were identified within the searched genomic regions (Supplementary Table S1). From the candidate genes, n=15 were microRNAs with 13 of them (*MIR6672, MIR1720, MIR7-2, MIR3529, MIR1571, MIR1560, MIR1785, MIR6662, MIR7454, MIR10A, MIR6663, MIR1735 and MIR6547*) published in the miRBase database (http://www.mirbase.org/) for *Gallus gallus*. Moreover, 190 candidate genes were unannotated (LOC) resulting in a total

number of 455 annotated positional candidate genes. The maximum number of candidate genes (n=192) was identified in a region spanning 998.5 kb (average D'=0.98) around marker *rs315329074 on* GGA27. At the other extreme, the smallest number of candidate genes was identified for *rs316794400* within a narrow region spanning 26.6 kb (average D'=0.96) on GGA22. Six out of the 11 lead markers were located within annotated genes i.e. *SLAIN2* (GGA4), *ZC3H18* (GGA11), *TMEM132D* (GGA15), *F-KER* (GGA25), *LEMD2* (GGA26) and *CACNB1* (GGA27).

3.5.2. Reported QTL/associations

Table 2 shows the number of published QTL/associations reported within the searched genomic regions. A total of 186 QTL/associations related to growth traits or carcass traits (carcass weight, abdominal fat percentage, breast muscle percentage and average daily gain) were identified within the searched regions. QTL/associations were distributed across eight chromosomes (1, 4, 10, 11, 15, 22, 26 and 27) and a detailed description of the reported QTL can be found in Supplementary Table S2. Note that the searched region around *rs317288536 (*GGA25) is not reported to harbor any QTL/association (Table 2). Furthermore, the only QTL reported on GGA22 as well as two additional QTL on GGA26 and GGA27 could not be remapped in *Gallus gallus*-5.0 by the Genome remapping service tool from NCBI database. Nevertheless, based on the *Gallus gallus-4* genome assembly, the searched regions around *rs316794400, rs317627533* and *rs314452928* overlapped with three QTL (IDs: 95429, 30883 and 55944). The maximum number (n=65) of QTL/associations was reported around *rs315329074* (GGA27) and the minimum number (n=1) around *rs316794400 (*GGA22*).* Nine out of the 12 lead SNPs on autosomes 1, 4, 10, 11, 15, 26 and 27 lie within 96 out of the 186 growth-related QTL (Supplementary Table S2). In addition, nearly all reported QTL on the searched regions on GGA4 (n=49/50) and GGA11 (n=9/9) contain at least one of the lead markers (*rs312691174*, *rs15608447* and *rs318098582* respectively).

We further sought to examine the locations of the positional candidate genes in the relation to the positions of the reported QTL. These results are illustrated in Fig. 2 in forms of circular maps for seven autosomes (1, 4, 10, 11, 15, 26 and 27). On GGA1, all 33 candidate genes (around *rs13923872*) are lying in a genomic region spanning from 2421 to 196203 kb where 17 relevant QTL have been reported. On GGA4, all 16 candidate genes around *rs312691174* are located within a region spanning from 4965 to 91268 kb, where all 14 published QTL reside. On the same autosome, all genes (n=36) around the second significant marker (*rs15608447*) are located in a region spanning from 4965 to 91268 kb where 35 QTL relevant to body weight, liver weight, carcass weight, total white fat weight have been reported. On GGA10, the region spanning from 693 to 20423 kb around the 'lead' marker (*rs318199727)* harbours all the 33 candidate genes and overlaps with 6 growth related QTL. On GGA11, in a region spanning from 953 to 20209 kb around *rs318098582* there have been 7 reported QTL and 27 candidate genes identified. On GGA15, in a region spanning from 1932 to 10689 kb around *rs317945754),* 6 QTL related to growth traits (visceral fat weight, abdominal fat weight and breast muscle weight) are reported and 20 candidate genes were identified. Moreover, on GGA26 *(rs317627533),* 64 out of the 93 candidate genes lie in a narrow region (1264 to 4918 kb) where QTL associated with growth traits such as body weight and shank weight, are reported. On GGA27 in a regions spanning from 55 to 4520 kb around *rs314452928,* 2 related QTL were identified including 7 out of the 12 candidate genes. All 192 genes around the second marker (*rs315329074*) on GGA27 were located within one published QTL spanning 3788 to 5630 kb that has been associated with thigh percentage. In total, 428 out of the 462 positional candidate genes (genes on GGA22 and GGA25 were not included here) were located within regions with reported QTL/associations.

3.5.3. Detection of community structure

A network including 402 genes (nodes) and 5294 interactions (edges) was generated. Note that for *APOA1BP* and *LOH11CR2A* genes the homologous human gene descriptions (*NAXE* and *VWA5A)* were used, respectively. Community structure analysis detected 5 modules, formed by 401 genes (see Supplementary Table S3). One more module was also detected but this was consisted by only one gene (*NIPAL1*). Thus it cannot be considered as a typical module. Note that this gene network had a strong community structure as indicated by the high (0.59) estimated modularity value [40]. Distribution of the 401 genes across the 5 modules is displayed in Figure 3. Module_2 consisted of 187 genes, module_3 of 22 genes, module_4 of 18 genes, module_5 of 152 genes and module_6 of 22 genes.

3.5.4. Functional enrichment analysis per module

Four (module_ID: 2-5) out of the five modules exhibited enriched GO BPs while only three modules were associated with developmental processes according to QuickGO. Specifically, in module_2, a total number of 21 enriched GO BPs (Supplementary Table S4) and 78 participating genes were identified. According to QuickGO (see Figure 3 and 4), 9 out of the 21 GO BPs were related to development with 42 member genes (Supplementary Table S4). In the same module, 8 genes belonging to the homeobox B family genes along with *MDFI* were found to be enriched in embryonic skeletal system morphogenesis (GO:0048704). In module_3, none of the enriched GO BP terms were related to development (Supplementary Table S4). In module_4, three significantly enriched BPs (Supplementary Table S4) were identified in 7 member genes. Here, the only GO BP term that was associated with development through QuickGO was cell differentiation (GO:0030154) with 4 member genes (*PPARD, ELF2, ETV3L and ETV4,* Figure 4). A total number of 29 GO BPs were found as significantly enriched in module_5 (Supplementary Table S4). Here, two development related processes i.e. multicellular organism growth (GO:0035264) and bone development (GO:0060348) were identified by QuickGO (Figure 4) with 6 involved genes (*KAT2A, SP2, ANKRD11, RARA, BGLAP and AKAP13*).

3.5.5. Functional candidate genes

An exhaustive list, including 66 modular genes, of the most plausible candidate genes for BW is provided in Table 3. From these genes, 52 were participating in enriched developmental processes, 7 were growth related genes that were not enriched to any developmental process and 7 were growth related genes identified in previous studies. These 66 modular genes were distributed across 7 chromosomes (GGA4, GGA10, GGA11, GGA15, GGA25, GGA26 and GGA27) with 47 of them detected in module_2. The *KRT (keratins)* family and B cluster of *HOX (homeobox)* family genes were also included here.

3.6. Discussion

Results of the present study have shown that a typical quantitative trait such as that examined here is associated with modular genes exhibiting functional relevance to developmental processes. This means that application of functional enrichment analysis on modular genes can facilitate the identification of true causative genes for the trait under study. Following this approach, a total number of 52 functional candidate genes could be identified in the present study. Example genes that fall in this category were the following: *BTG2, ZAR1, MEOX1, KRT14, KRT15, TXK, CSF3, ACAN, HOXB, MDFI, NES, IGFBP4, PRELP, PPARD, ELF2, KAT2A, RARA and BGLAP*. Specifically, *BTG2, ZAR1, MEOX1, KRT14* and *KRT15* have been reported to participate in cerebellar development [47], development of follicular oocytes [48], somite differentiation [49], keratinocytes proliferation [50] and pigmentation of muscle tissues [51], in chickens, respectively. *TXK (TXK tyrosine kinase)* has been reported as BW related gene [52] while *CSF3 (colony stimulating factor 3)* has been described as a

myelomonocytic growth factor in chickens [53]. *ACAN (aggrecan)* is essential for cartilage formation during development in chicken and mouse mutants [54] and the *HOX B* cluster genes are expressed in chick embryonic development [55]. The *MDFI* (MyoD family inhibitor) tumor suppressor gene is known to have a negative effect on myogenic regulatory factors [56] while *NES (nestin)* is known as a neural progenitor cell marker during central nervous system development and a marker protein for neovascularization [57]. Furthermore, *IGFBP4 (insulin like growth factor binding protein 4)* is required for the adipose tissue development [58] while *PRELP (proline and arginine rich end leucine rich repeat protein)* is highly expressed in cartilage, basement membranes, and participates in bone development [59]. *PPARD (peroxisome proliferator activated receptor delta)* is a critical gene for normal adipose development and lipid homeostasis [60] while *ELF2 (E74 like ETS transcription factor 2)* plays a key role in the development of lymphocytes [61]. *KAT2A (lysine acetyltransferase 2A)* is necessary for growth and differentiation of craniofacial cartilage and bone in zebrafish and mice [62], *RARA (retinoic acid receptor alpha)* affects the hippocampal development [63] and finally *BGLAP* is produced by osteoblasts shaping new bones in chickens [64].

However, the search for modular genes that are exclusively enriched in functionally relevant terms has not proved to be efficient in identifying all true functional candidate genes. This finding may be fairly supported by the fact that 7 more genes (*GABRG1, NGF, APOBEC2, STAT5B, STAT3, SMAD4 and MED1*) that despite having well documented relevance to development were found to be enriched in other but developmental GO BP terms. Specifically, *GABRG1(gamma-aminobutyric acid type A receptor gamma1 subunit)* is reported as a BW related gene [52] and *NGF (nerve growth factor)* is a regulator of the somite survival and axial rotation during early chicken embryo development [65]. *APOBEC2 (apolipoprotein B mRNA editing enzyme catalytic subunit 2)* is known as a critical regulator and maintainer of muscle development in mammals and might affect muscle development in chickens [66]. In chickens, *STAT5B (signal transducer and activator of transcription 5B)* is associated with growth [67]. *STAT3 (signal transducer and activator of transcription 3)* plays a central role in development [68], *SMAD4 (SMAD family member 4)* is a central mediator of the transforming growth factor β signaling pathway which affects among others the cell growth [69] and finally *MED1 (mediator complex subunit 1)* has a key role in mammary epithelial cell growth [70].
The list with the most plausible candidate genes for the trait was, however, not exhausted in the previous two categories since 7 more genes (*CACNB1, SLAIN2, LEMD2, ZC3H18, TMEM132D, FRYL and SGCB)* with well documented implication to BW, were completely omitted in any enrichment analysis. Most interestingly, five of the above genes (*CACNB1, SLAIN2, LEMD2, ZC3H18* and *TMEM132D)* contained lead SNPs. *CACNB1 (calcium voltage-gated channel auxiliary subunit beta 1)* has been reported to affect skeletal muscle development [71] in mice. *SLAIN2 (SLAIN motif family member 2)* is necessary for the normal structure of microtubule cytoskeleton as it controls the microtubule growth during interphase [72]. *LEMD2 (LEM domain containing 2*) participates in nuclear structure organization [73] and plays an important role in mouse embryonic development by regulating various signaling pathways such as MAPK (mitogen-activated protein kinase) and AKT (also known as Protein Kinase B) [74]. *ZC3H18 (zinc finger CCCH-type containing 18)* participates in RNA degradation [75] and affects mRNA metabolism [76]. Finally, *TMEM132D (transmembrane protein 132D*) may function as a tumor suppressor gene [77]. Finally, both *FRYL (FRY like transcription coactivator)* and *SGCB (sarcoglycan beta)* have been associated with growth ([78,79]) in chickens.
The two aforementioned gene lists underline the potential limitations of a cluster based method such as that used here to assess the biological properties of the candidate gene sets. Specifically, these limitations relate to i) grouping of similar terms into a cluster and

evaluating the enrichment of functional clusters instead of each individual term within the clusters and ii) the evaluation of the identified term clusters separately, while not taken into consideration the relationships between clusters [80].

Apart from functional enrichment analysis, other analyses such as pathway analysis, gene network analysis and GBA gene prioritization analysis could also assist in identifying true causative genetic variants for the trait under study. For instance, in a previous study [81], the use of GBA gene prioritization analysis on 1,012 positional candidate genes revealed 248 functional candidate genes for the same trait. However, fixed genomic regions (of 1Mb) around the lead genomic markers were used in that study. A final interesting result of the present study was the discovery of 15 microRNAs within the 645 candidate genes for the trait under investigation. One of these, i.e. *MIR10A* has been reported as significant for feed intake in broilers [82]. *MIR10A* together with *MIR10B* have been reported to inhibit the development of human, mouse and rat granulosa cells during folliculogenesis [83]. Finally, *MIR7-2* has been reported as genomic locus for peroxisome proliferator activated receptor regulation [84] and may have a functional role in hepatic lipid homeostasis. MicroRNAs have emerged as important regulators of gene expression post-transcriptionally and in *Gallus gallus* are known to play crucial roles in various biological processes such as the accumulation of abdominal fat [85] and the lipid metabolism [86].
In conclusion, the present GWAS revealed a large number of genomic regions and genes implicated in the genetic architecture of a complex trait such as the BW that fully complies with the Fisher's infinitesimal model of inheritance. Exploitation of both community structure and functional enrichment analyses highlighted 3 modules as related to development. Current findings also indicated 52 modular genes participating in developmental processes and 14 more modular genes related to BW. Finally, the present study proposed 66 functional candidate genes for BW, some of which are novel and some identified candidates in previous studies.

3.7. References

[1] Chicken QTL Database [Internet]. [cited 3 Sep 2017]. Available: https://www.animalgenome.org/cgi-bin/QTLdb/GG/download?tmpname=mapDwnLd&file=cM

[2] Xie L, Luo C, Zhang C, Zhang R, Tang J, Nie Q, et al. Genome-Wide Association Study Identified a Narrow Chromosome 1 Region Associated with Chicken Growth Traits. Liu Z, editor. PLoS One. Public Library of Science; 2012;7: e30910. doi:10.1371/journal.pone.0030910

[3] Van Goor A, Bolek KJ, Ashwell CM, Persia ME, Rothschild MF, Schmidt CJ, et al. Identification of quantitative trait loci for body temperature, body weight, breast yield, and digestibility in an advanced intercross line of chickens under heat stress. Genet Sel Evol. BioMed Central; 2015;47: 96. doi:10.1186/s12711-015-0176-7

[4] Kranis A, Gheyas AA, Boschiero C, Turner F, Yu L, Smith S, et al. Development of a high density 600K SNP genotyping array for chicken. BMC Genomics. BioMed Central; 2013;14: 59. doi:10.1186/1471-2164-14-59

[5] Pettersson ME, Carlborg Ö. Dissecting the genetic architecture of complex traits and its impact on genetic improvement programs: lessons learnt from the Virginia chicken lines. Rev Bras Zootec. Sociedade Brasileira de Zootecnia; 2010;39: 256–260. doi:10.1590/S1516-35982010001300028

[6] Shahjahan M, Liu R, Zhao G, Wang F, Zheng M, Zhang J, et al. Identification of Histone Deacetylase 2 as a Functional Gene for Skeletal Muscle Development in Chickens. Asian-Australasian J Anim Sci. 2016;29: 479–486. doi:10.5713/ajas.15.0252

[7] Ouyang H, Zhang H, Li W, Liang S, Jebessa E, Abdalla BA, et al. Identification, expression and variation of the *GNPDA2* gene, and its association with body weight and fatness traits in chicken. PeerJ. 2016;4: e2129. doi:10.7717/peerj.2129

[8] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set

enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. National Academy of Sciences; 2005;102: 15545–50. doi:10.1073/pnas.0506580102
[9] García-Campos MA, Espinal-Enríquez J, Hernández-Lemus E. Pathway Analysis: State of the Art. Front Physiol. Frontiers; 2015;6: 383. doi:10.3389/fphys.2015.00383
[10] Chen J, Aronow BJ, Jegga AG. Disease candidate gene identification and prioritization using protein interaction networks. BMC Bioinformatics. BioMed Central; 2009;10: 73. doi:10.1186/1471-2105-10-73
[11] Montojo J, Zuberi K, Rodriguez H, Kazi F, Wright G, Donaldson SL, et al. GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. Bioinformatics. 2010;26: 2927–2928. doi:10.1093/bioinformatics/btq562
[12] Walker MG, Volkmuth W, Sprinzak E, Hodgson D, Klingler T. Prediction of gene function by genome-scale expression analysis: prostate cancer-associated genes. Genome Res. 1999;9: 1198–203. Available: http://www.ncbi.nlm.nih.gov/pubmed/10613842
[13] Oliver S. Guilt-by-association goes global. Nature. 2000;403: 601–602. doi:10.1038/35001165
[14] Liu G, Wang H, Chu H, Yu J, Zhou X. Functional diversity of topological modules in human protein-protein interaction networks. Sci Rep. Nature Publishing Group; 2017;7: 16199. doi:10.1038/s41598-017-16270-z
[15] Terentiev AA, Moldogazieva NT, Shaitan K V. Dynamic proteomics in modeling of the living cell. Protein-protein interactions. Biochemistry (Mosc). 2009;74: 1586–607. Available: http://www.ncbi.nlm.nih.gov/pubmed/20210711
[16] Spirin V, Mirny LA. Protein complexes and functional modules in molecular networks. Proc Natl Acad Sci U S A. National Academy of Sciences; 2003;100: 12123–8. doi:10.1073/pnas.2032324100
[17] Lu H, Shi B, Wu G, Zhang Y, Zhu X, Zhang Z, et al. Integrated analysis of multiple data sources reveals modular structure of biological networks. Biochem Biophys Res Commun. Academic Press; 2006;345: 302–309. doi:10.1016/J.BBRC.2006.04.088
[18] Zhang S, Ning X, Zhang X-S. Identification of functional modules in a PPI network by clique percolation clustering. Comput Biol Chem. Elsevier; 2006;30: 445–451. doi:10.1016/J.COMPBIOLCHEM.2006.10.001
[19] Sharan R, Ulitsky I, Shamir R. Network-based prediction of protein function. Mol Syst Biol. 2007;3: 88. doi:10.1038/msb4100129
[20] Jagannadham J, Jaiswal HK, Agrawal S, Rawal K. Comprehensive Map of Molecules Implicated in Obesity. Castiglione F, editor. PLoS One. Public Library of Science; 2016;11: e0146759. doi:10.1371/journal.pone.0146759
[21] Sharma P, Bhattacharyya DK, Kalita J. Disease biomarker identification from gene network modules for metastasized breast cancer. Sci Rep. Nature Publishing Group; 2017;7: 1072. doi:10.1038/s41598-017-00996-x
[22] Hallett RM, Cockburn JG, Li B, Dvorkin-Gheva A, Hassell JA, Bane A. Identification and evaluation of network modules for the prognosis of basal-like breast cancer. Oncotarget. 2015;6: 17713–24. doi:10.18632/oncotarget.4034
[23] Lempiäinen H, Brænne I, Michoel T, Tragante V, Vilne B, Webb TR, et al. Network analysis of coronary artery disease risk genes elucidates disease mechanisms and druggable targets. Sci Rep. Nature Publishing Group; 2018;8: 3434. doi:10.1038/s41598-018-20721-6
[24] Sharma A, Menche J, Huang CC, Ort T, Zhou X, Kitsak M, et al. A disease module in the interactome explains disease heterogeneity, drug response and captures novel pathways and genes in asthma. Hum Mol Genet. 2015;24: 3005–3020. doi:10.1093/hmg/ddv001
[25] Hira R, Terada S-I, Kondo M, Matsuzaki M. Distinct Functional Modules for Discrete and Rhythmic Forelimb Movements in the Mouse Motor Cortex. J Neurosci. 2015;35: 13311–13322. doi:10.1523/JNEUROSCI.2731-15.2015
[26] Liu J, Fu R, Liu R, Zhao G, Zheng M, Cui H, et al. Protein Profiles for Muscle Development and Intramuscular Fat Accumulation at Different Post-Hatching Ages in Chickens. PENA i SUBIRÀ RN, editor. PLoS One. 2016;11: e0159722. doi:10.1371/journal.pone.0159722
[27] Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, et al. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet. Nature Publishing Group; 2012;44: 825–830. doi:10.1038/ng.2314

[28] VanRaden PM. Efficient Methods to Compute Genomic Predictions. J Dairy Sci. 2008;91: 4414–4423. doi:10.3168/jds.2007-0980
[29] Chen J, Chen Z. Extended Bayesian information criteria for model selection with large model spaces. Biometrika. Oxford University Press; 2008;95: 759–771. doi:10.1093/biomet/asn034
[30] Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing [Internet]. Journal of the Royal Statistical Society. Series B (Methodological). WileyRoyal Statistical Society; 1995. pp. 289–300. doi:10.2307/2346101
[31] Qu H-Q, Tien M, Polychronakos C. Statistical significance in genetic association studies. Clin Invest Med. NIH Public Access; 2010;33: E266-70. Available: http://www.ncbi.nlm.nih.gov/pubmed/20926032
[32] Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, Willer CJ, et al. Genomic inflation factors under polygenic inheritance. Eur J Hum Genet. Nature Publishing Group; 2011;19: 807–12. doi:10.1038/ejhg.2011.39
[33] Yi G, Shen M, Yuan J, Sun C, Duan Z, Qu L, et al. Genome-wide association study dissects genetic architecture underlying longitudinal egg weights in chickens. BMC Genomics. BioMed Central; 2015;16: 746. doi:10.1186/s12864-015-1945-y
[34] Hedrick PW. Gametic disequilibrium measures: proceed with caution. Genetics. 1987;117: 331–41. Available: http://www.ncbi.nlm.nih.gov/pubmed/3666445
[35] McRae AF, McEwan JC, Dodds KG, Wilson T, Crawford AM, Slate J. Linkage disequilibrium in domestic sheep. Genetics. 2002;160: 1113–22. Available: http://www.ncbi.nlm.nih.gov/pubmed/11901127
[36] Gallus gallus Annotation Release 103 [Internet]. [cited 3 Sep 2017]. Available: https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Gallus_gallus/103/
[37] Gallus_gallus-5.0 in NCBI assembly data base [Internet]. [cited 3 Sep 2017]. Available: https://www.ncbi.nlm.nih.gov/assembly/GCF_000002315.4/
[38] Coordinate remapping service: NCBI [Internet]. [cited 3 Sep 2017]. Available: https://www.ncbi.nlm.nih.gov/genome/tools/remap
[39] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. 2003;13: 2498–2504. doi:10.1101/gr.1239303
[40] Newman MEJ, Girvan M. Finding and evaluating community structure in networks. Phys Rev E. 2004;69: 026113. doi:10.1103/PhysRevE.69.026113
[41] Su G, Kuchinsky A, Morris JH, States DJ, Meng F. GLay: community structure analysis of biological networks. Bioinformatics. Oxford University Press; 2010;26: 3135–3137. doi:10.1093/bioinformatics/btq596
[42] Morris JH, Apeltsin L, Newman AM, Baumbach J, Wittkop T, Su G, et al. clusterMaker: a multi-algorithm clustering plugin for Cytoscape. BMC Bioinformatics. BioMed Central; 2011;12: 436. doi:10.1186/1471-2105-12-436
[43] Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane H, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. BioMed Central; 2003;4: R60. doi:10.1186/gb-2003-4-9-r60
[44] Hosack DA, Dennis G, Sherman BT, Lane HC, Lempicki RA, Lempicki RA. Identifying biological themes within lists of genes with EASE. Genome Biol. BioMed Central; 2003;4: R70. doi:10.1186/gb-2003-4-10-r70
[45] Huntley RP, Binns D, Dimmer E, Barrell D, O'Donovan C, Apweiler R. QuickGO: a user tutorial for the web-based Gene Ontology browser. Database (Oxford). Oxford University Press; 2009;2009: bap010. doi:10.1093/database/bap010
[46] Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S, Freimer NB, et al. Variance component model to account for sample structure in genome-wide association studies. Nat Genet. Nature Publishing Group; 2010;42: 348–354. doi:10.1038/ng.548
[47] Kamaid A, Giráldez F. Btg1 and Btg2 gene expression during early chick development. Dev Dyn. John Wiley & Sons, Ltd; 2008;237: 2158–2169. doi:10.1002/dvdy.21616
[48] Sun C, Lu J, Yi G, Yuan J, Duan Z, Qu L, et al. Promising Loci and Genes for Yolk and Ovary Weight in Chickens Revealed by a Genome-Wide Association Study. Veitia RA, editor. PLoS One. Public Library of Science; 2015;10: e0137145. doi:10.1371/journal.pone.0137145
[49] Reijntjes S, Stricker S, Mankoo BS. A comparative analysis of Meox1 and Meox2 in the developing somites and limbs of the chick embryo. Int J Dev Biol. 2007;51: 753–759.

doi:10.1387/ijdb.072332sr
[50] Couteaudier M, Trapp-Fragnet L, Auger N, Courvoisier K, Pain B, Denesvre C, et al. Derivation of keratinocytes from chicken embryonic stem cells: Establishment and characterization of differentiated proliferative cell populations. Stem Cell Res. Elsevier; 2015;14: 224–237. doi:10.1016/J.SCR.2015.01.002
[51] Tarique T, Yang S, Mohsina Z, Qiu J, Yan Z, Chen G, et al. Identification of genes involved in regulatory mechanism of pigments in broiler chickens. Genet Mol Res. 2014;13: 7201–7216. doi:10.4238/2014.September.5.6
[52] Gu X, Feng C, Ma L, Song C, Wang Y, Da Y, et al. Genome-Wide Association Study of Body Weight in Chicken F2 Resource Population. Mailund T, editor. PLoS One. Public Library of Science; 2011;6: e21872. doi:10.1371/journal.pone.0021872
[53] Gibson MS, Kaiser P, Fife M. Identification of Chicken Granulocyte Colony-Stimulating Factor (G-CSF/CSF3): The Previously Described Myelomonocytic Growth Factor Is Actually CSF3. J Interf Cytokine Res. Mary Ann Liebert, Inc. publishers140 Huguenot Street, 3rd FloorNew Rochelle, NY 10801-5215USA; 2009;29: 339–344. doi:10.1089/jir.2008.0103
[54] Bou-Gharios G, Liu K, Li I, De Val S. Three new functionally conserved cis-regulatory elements in the Acan gene. Osteoarthr Cartil. Elsevier; 2014;22: S143–S144. doi:10.1016/j.joca.2014.02.267
[55] Gouveia A, Marcelino HM, Gonçalves L, Palmeirim I, Andrade RP. Patterning in time and space: HoxB cluster gene expression in the developing chick embryo. Cell Cycle. Taylor & Francis; 2015;14: 135. doi:10.4161/15384101.2014.972868
[56] Li J, Chen C, Bi X, Zhou C, Huang T, Ni C, et al. DNA methylation of CMTM3 , SSTR2 , and MDFI genes in colorectal cancer. Gene. 2017;630: 1–7. doi:10.1016/j.gene.2017.07.082
[57] Suzuki S, Namiki J, Shibata S, Mastuzaki Y, Okano H. The Neural Stem/Progenitor Cell Marker Nestin Is Expressed in Proliferative Endothelial Cells, but Not in Mature Vasculature. J Histochem Cytochem. 2010;58: 721–730. doi:10.1369/jhc.2010.955609
[58] Maridas DE, DeMambro VE, Le PT, Mohan S, Rosen CJ. IGFBP4 Is Required for Adipogenesis and Influences the Distribution of Adipose Depots. Endocrinology. The Endocrine Society; 2017;158: 3488–3500. doi:10.1210/en.2017-00248
[59] Rucci N, Rufo A, Alamanou M, Capulli M, Del Fattore A, Ahrman E, et al. The glycosaminoglycan-binding domain of PRELP acts as a cell type-specific NF-kappaB inhibitor that impairs osteoclastogenesis. J Cell Biol. The Rockefeller University Press; 2009;187: 669–83. doi:10.1083/jcb.200906014
[60] Shiue Y-L, Chen L-R, Tsai C-J, Yeh C-Y, Huang C-T. Emerging roles of peroxisome proliferator-activated receptors in the pituitary gland in female reproduction. Biomarkers Genomic Med. Elsevier; 2013;5: 1–11. doi:10.1016/J.GMBHS.2013.04.008
[61] Guan FHX, Bailey CG, Metierre C, O'Young P, Gao D, Khoo TL, et al. The antiproliferative ELF2 isoform, ELF2B, induces apoptosis in vitro and perturbs early lymphocytic development in vivo. J Hematol Oncol. BioMed Central; 2017;10: 75. doi:10.1186/s13045-017-0446-7
[62] Sen R, Pezoa S, Carpio Shull L, Hernandez-Lagunas L, Niswander L, Artinger K. Kat2a and Kat2b Acetyltransferase Activity Regulates Craniofacial Cartilage and Bone Differentiation in Zebrafish and Mice. J Dev Biol. 2018;6: 27. doi:10.3390/jdb6040027
[63] Huang H, Wei H, Zhang X, Chen K, Li Y, Qu P, et al. Changes in the expression and subcellular localization of RARα in the rat hippocampus during postnatal development. Brain Res. 2008;1227: 26–33. doi:10.1016/j.brainres.2008.06.073
[64] Jiang S, Cheng HW, Hester PY, Hou J-F. Development of an enzyme-linked immunosorbent assay for detection of chicken osteocalcin and its use in evaluation of perch effects on bone remodeling in caged White Leghorns. Poult Sci. Oxford University Press; 2013;92: 1951–1961. doi:10.3382/ps.2012-02909
[65] Manca A, Capsoni S, Di Luzio A, Vignone D, Malerba F, Paoletti F, et al. Nerve growth factor regulates axial rotation during early stages of chick embryo development. Proc Natl Acad Sci. 2012;109: 2009–2014. doi:10.1073/pnas.1121138109
[66] Li J, Zhao X-L, Gilbert ER, Li D-Y, Liu Y-P, Wang Y, et al. APOBEC2 mRNA and protein is predominantly expressed in skeletal and cardiac muscles of chickens. Gene. 2014;539: 263–269. doi:10.1016/j.gene.2014.01.003
[67] Zhao XH, Wang JY, Zhang GX, Wei Y, Gu YP, Yu YB. Single nucleotide polymorphism in the STAT5b gene is associated with body weight and reproductive traits of the Jinghai Yellow

chicken. Mol Biol Rep. Springer Netherlands; 2012;39: 4177–4183. doi:10.1007/s11033-011-1202-7
[68] Johnston PA, Grandis JR. STAT3 SIGNALING: Anticancer Strategies and Challenges. Mol Interv. 2011;11: 18–26. doi:10.1124/mi.11.1.4
[69] Zhao M, Mishra L, Deng C-X. The role of TGF-β/SMAD4 signaling in cancer. Int J Biol Sci. 2018;14: 111–123. doi:10.7150/ijbs.23230
[70] Hasegawa N, Sumitomo A, Fujita A, Aritome N, Mizuta S, Matsui K, et al. Mediator Subunits MED1 and MED24 Cooperatively Contribute to Pubertal Mammary Gland Development and Growth of Breast Carcinoma Cells. Mol Cell Biol. 2012;32: 1483–1495. doi:10.1128/MCB.05245-11
[71] Chen F, Liu Y, Sugiura Y, Allen PD, Gregg RG, Lin W. Neuromuscular synaptic patterning requires the function of skeletal muscle dihydropyridine receptors. Nat Neurosci. Nature Publishing Group; 2011;14: 570–577. doi:10.1038/nn.2792
[72] van der Vaart B, Manatschal C, Grigoriev I, Olieric V, Gouveia SM, Bjelić S, et al. SLAIN2 links microtubule plus end–tracking proteins and controls microtubule growth in interphase. J Cell Biol. 2011;193: 1083–1099. doi:10.1083/jcb.201012179
[73] Brachner A, Reipert S, Foisner R, Gotzmann J. LEM2 is a novel MAN1-related inner nuclear membrane protein associated with A-type lamins. J Cell Sci. The Company of Biologists Ltd; 2005;118: 5797–810. doi:10.1242/jcs.02701
[74] Tapia O, Fong LG, Huber MD, Young SG, Gerace L. Nuclear Envelope Protein Lem2 is Required for Mouse Development and Regulates MAP and AKT Kinases. Palazzo AF, editor. PLoS One. 2015;10: e0116196. doi:10.1371/journal.pone.0116196
[75] Giacometti S, Benbahouche NEH, Domanski M, Robert M-C, Meola N, Lubas M, et al. Mutually Exclusive CBC-Containing Complexes Contribute to RNA Fate. Cell Rep. Elsevier; 2017;18: 2635–2650. doi:10.1016/j.celrep.2017.02.046
[76] Winczura K, Schmid M, Iasillo C, Molloy KR, Harder LM, Andersen JS, et al. Characterizing ZC3H18, a Multi-domain Protein at the Interface of RNA Production and Destruction Decisions. Cell Rep. 2018;22: 44–58. doi:10.1016/j.celrep.2017.12.037
[77] Iwakawa R, Kohno T, Totoki Y, Shibata T, Tsuchihara K, Mimaki S, et al. Expression and clinical significance of genes frequently mutated in small cell lung cancers defined by whole exome/RNA sequencing. Carcinogenesis. 2015;36: 616–621. doi:10.1093/carcin/bgv026
[78] Zhang GX, Fan QC, Zhang T, Wang JY, Wang WH, Xue Q, et al. Genome-wide association study of growth traits in the Jinghai Yellow chicken. Genet Mol Res. 2015;14: 15331–15338. doi:10.4238/2015.November.30.10
[79] Pampouille E, Berri C, Boitard S, Hennequet-Antier C, Beauclercq SA, Godet E, et al. Mapping QTL for white striping in relation to breast muscle yield and meat quality traits in broiler chickens. BMC Genomics. 2018;19: 202. doi:10.1186/s12864-018-4598-9
[80] Sun D, Liu Y, Zhang X-S, Wu L-Y. CEA: Combination-based gene set functional enrichment analysis. Sci Rep. Nature Publishing Group; 2018;8: 13085. doi:10.1038/s41598-018-31396-4
[81] Tarsani E, Kranis A, Maniatis G, Kominakis A. Investigating the functional role of 1,012 candidate genes identified by a Genome Wide Association Study for body weight in broilers. Proc World Congr Genet Appl to Livest Prod. World Congress on Genetics Applied to Livestock Production; 2018;Species-Avian 1: 564. Available: http://www.wcgalp.org/proceedings/2018/investigating-functional-role-1012-candidate-genes-identified-genome-wide
[82] Yuan J, Wang K, Yi G, Ma M, Dou T, Sun C, et al. Genome-wide association studies for feed intake and efficiency in two laying periods of chickens. Genet Sel Evol. BioMed Central; 2015;47: 82. doi:10.1186/s12711-015-0161-1
[83] Jiajie T, Yanzhou Y, Hoi-Hung AC, Zi-Jiang C, Wai-Yee C. Conserved miR-10 family represses proliferation and induces apoptosis in ovarian granulosa cells. Sci Rep. Nature Publishing Group; 2017;7: 41304. doi:10.1038/srep41304
[84] Singaravelu R, Quan C, Powdrill MH, Shaw TA, Srinivasan P, Lyn RK, et al. MicroRNA-7 mediates cross-talk between metabolic signaling pathways in the liver. Sci Rep. Nature Publishing Group; 2018;8: 361. doi:10.1038/s41598-017-18529-x
[85] Huang HY, Liu RR, Zhao GP, Li QH, Zheng MQ, Zhang JJ, et al. Integrated analysis of microRNA and mRNA expression profiles in abdominal adipose tissues in chickens. Sci Rep. Nature Publishing Group; 2015;5: 16132. doi:10.1038/srep16132
[86] Li H, Ma Z, Jia L, Li Y, Xu C, Wang T, et al. Systematic analysis of the regulatory functions of microRNAs in chicken hepatic lipid metabolism. Sci Rep. Nature Publishing Group; 2016;6: 31766.

doi:10.1038/srep31766
[87] Conant GC, Wolfe KH. GenomeVx: simple web-based creation of editable circular chromosome maps. Bioinformatics. 2008;24: 861–862. doi:10.1093/bioinformatics/btm598

Tables and Figures of Chapter 3

Table 1. Genome-wide significant SNPs (FDR p-value<0.05) for BW.

SNP ID	**GGA**	**Position (bp)[1]**	**-log$_{10}$(p-value)**	**FDR p-value**
rs13923872	1	112,741,685	6.415	0.0112
rs312691174	4	29,074,989	7.948	0.00037
rs15608447	4	66,885,210	13.489	4.25E-09
rs318199727	10	13,536,548	5.763	0.04111
rs318098582	11	18,651,449	8.513	0.00012
rs317945754	15	3,557,083	5.677	0.04594
rs316794400	22	4,594,855	11.033	6.07E-07
rs317288536	25	976,833	13.035	8.05E-09
rs312758346	25	2,412,866	9.517	1.59E-05
rs317627533	26	4,597,439	9.313	2.12E-05
rs314452928	27	104,022	6.398	0.0105
rs315329074	27	4,528,275	20.513	8.05E-16

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

Table 2. Number of positional candidate genes and QTL/associations within the searched genomic regions (± maximum distance of the farest SNP being in strong LD (D'>0.8) with the lead SNP; D': average D' values within the searched genomic region).

SNP ID	GGA	Position (bp)[1]	Searched genomic range around 'lead' SNP (± bp)	D'	Number of positional candidate genes	Number of QTL/ associations
rs13923872	1	112,741,685	613,054	0.91	33	20
rs312691174	4	29,074,989	650,472	1	16	14
rs15608447	4	66,885,210	718,407	0.88	36	36
rs318199727	10	13,536,548	737,906	0.83	33	11
rs318098582	11	18,651,449	300,257	0.81	27	9
rs317945754	15	3,557,083	935,183	0.99	20	21
rs316794400	22	4,594,855	26,589	0.96	7	1
rs317288536	25	976,833	1,004,513	0.83	176	-
rs317627533	26	4,597,439	773,988	0.9	93	6
rs314452928	27	104,022	140,067	0.94	12	3
rs315329074	27	4,528,275	998,553	0.98	192	65

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

Table 3. List of 66 most plausible candidate genes for BW according to the following criteria: modular genes participating in enriched developmental processes, growth related modular genes not significantly enriched to any developmental process and growth related modular genes reported in previous studies.

Criterion	Gene	Description	Module_ID	GGA
modular genes participating in enriched developmental processes	*BTG2*	*BTG anti-proliferation factor 2*	module_2	26
	ZAR1	*zygote arrest 1*	module_2	4
	MEOX1	*mesenchyme homeobox 1*	module_2	27
	KRT14	*keratin 14*	module_2	27
	KRT15	*keratin 15*	module_2	27
	TXK	*TXK tyrosine kinase*	module_2	4
	CSF3	*colony stimulating factor 3*	module_2	27
	ACAN	*aggrecan*	module_2	10
	HOXB1	*homeobox B1*	module_2	27
	HOXB2	*homeobox B2*	module_2	27
	HOXB3	*homeobox B3*	module_2	27
	HOXB4	*homeobox B4*	module_2	27
	HOXB5	*homeobox B5*	module_2	27
	HOXB6	*homeobox B6*	module_2	27
	HOXB7	*homeobox B7*	module_2	27
	HOXB8	*homeobox B8*	module_2	27
	HOXB9	*homeobox B9*	module_2	27
	HOXB13	*homeobox B13*	module_2	27
	MDFI	*MyoD family inhibitor*	module_2	26
	NES	*nestin*	module_2	25
	TBX21	*T-box 21*	module_2	27
	IGFBP4	*insulin like growth factor binding protein 4*	module_2	27
	PRELP	*proline and arginine rich end leucine rich repeat protein*	module_2	26
	HAPLN2	*hyaluronan and proteoglycan link protein 2*	module_2	25

	HAPLN3	*hyaluronan and proteoglycan link protein 3*	module_2	10
	GABRA4	*gamma-aminobutyric acid type A receptor alpha4 subunit*	module_2	4
	BCAN	*brevican*	module_2	25
	NHLH1	*nescient helix-loop-helix 1*	module_2	25
	ZBTB7B	*zinc finger and BTB domain containing 7B*	module_2	25
	FZD10	*frizzled class receptor 10*	module_2	15
	TCP11	*t-complex 11*	module_2	26
	PIWIL1	*piwi like RNA-mediated gene silencing 1*	module_2	15
	SPDEF	*SAM pointed domain containing ETS transcription factor*	module_2	26
	ZFPM1	*zinc finger protein, FOG family member 1*	module_2	11
	CBFA2T3	*CBFA2/RUNX1 translocation partner 3*	module_2	11
	KRT17	*keratin 17*	module_2	27
	CRABP2	*cellular retinoic acid binding protein 2*	module_2	25
	SH2D2A	*SH2 domain containing 2A*	module_2	25
	NR1D1	*nuclear receptor subfamily 1 group D member 1*	module_2	27
	STX2	*syntaxin 2*	module_2	15
	TEC	*tec protein tyrosine kinase*	module_2	4
	ETV3	*ETS variant 3*	module_2	25
	PPARD	*peroxisome proliferator activated receptor delta*	module_4	26
	ELF2	*E74 like ETS transcription factor 2*	module_4	4
	ETV3L	*ETS variant 3 like*	module_4	25
	ETV4	*ETS variant 4*	module_4	27
	KAT2A	*lysine acetyltransferase 2A*	module_5	27
	RARA	*retinoic acid receptor alpha*	module_5	27
	BGLAP	*bone gamma-carboxyglutamate protein*	module_5	25
	SP2	*Sp2 transcription factor*	module_5	27

	ANKRD11	*ankyrin repeat domain 11*	module_5	11
	AKAP13	*A-kinase anchoring protein 13*	module_5	10
growth related modular genes not significantly enriched to any developmental process	*GABRG1*	*gamma-aminobutyric acid type A receptor gamma1 subunit*	module_2	4
	NGF	*nerve growth factor*	module_2	26
	APOBEC2	*apolipoprotein B mRNA editing enzyme catalytic subunit 2*	module_5	26
	STAT5B	*signal transducer and activator of transcription 5B*	module_3	27
	STAT3	*signal transducer and activator of transcription 3*	module_5	27
	SMAD4	*SMAD family member 4*	module_5	25
	MED1	*mediator complex subunit 1*	module_5	27
growth related modular genes reported in previous studies	*CACNB1*	*calcium voltage-gated channel auxiliary subunit beta 1*	module_2	27
	SLAIN2	*SLAIN motif family member 2*	module_5	4
	LEMD2	*LEM domain containing 2*	module_5	26
	ZC3H18	*zinc finger CCCH-type containing 18*	module_5	11
	TMEM132D	*transmembrane protein 132D*	module_2	15
	FRYL	*FRY like transcription coactivator*	module_4	4
	SGCB	*sarcoglycan beta*	module_2	4

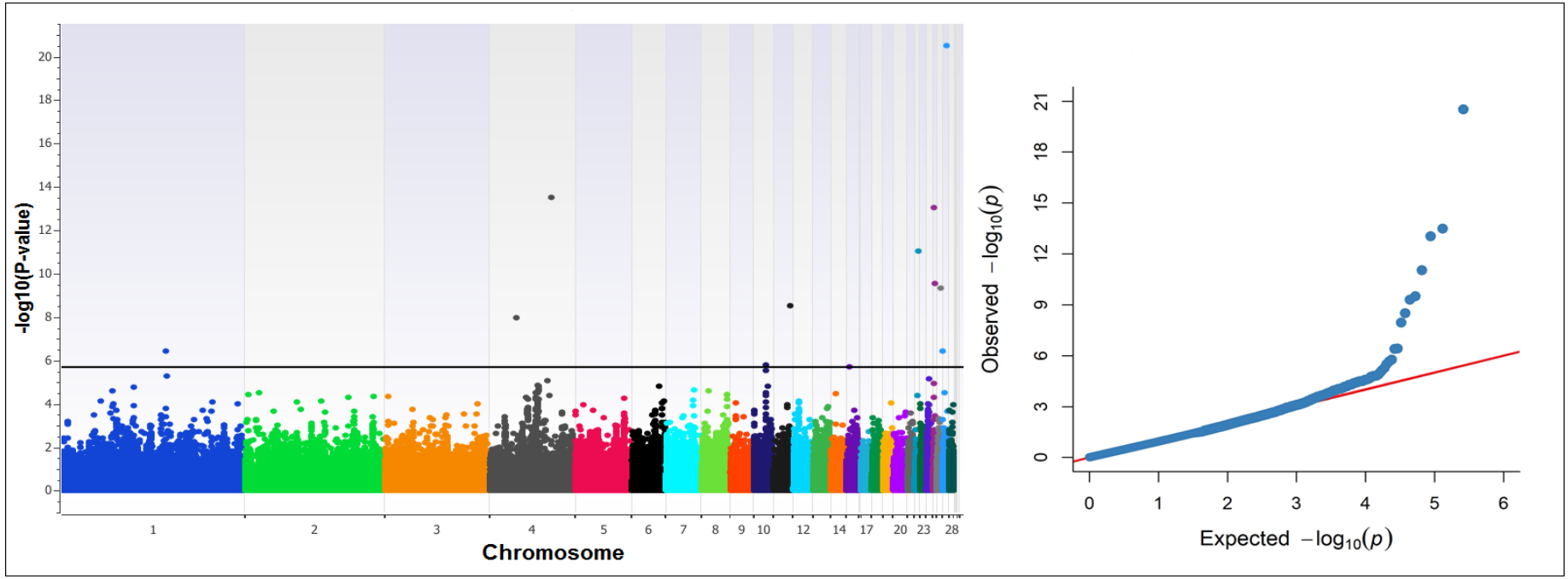
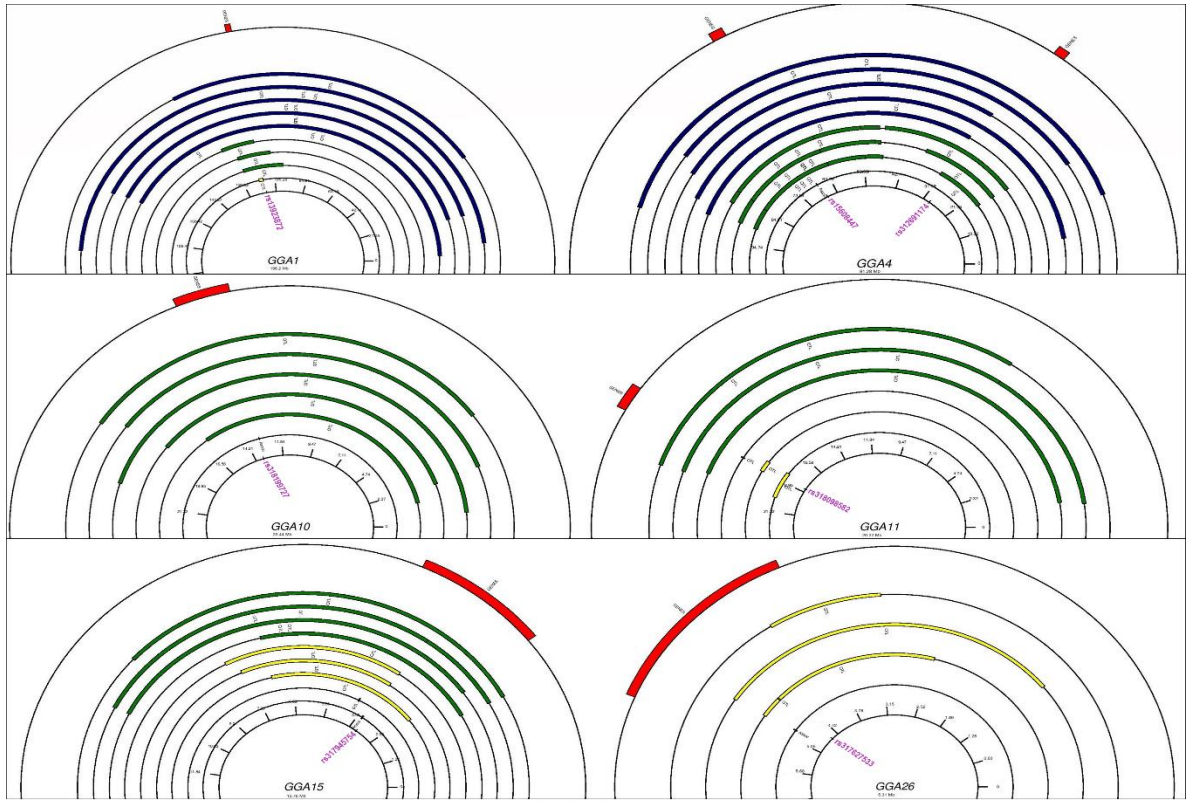


Figure 1. Manhattan plot (left) and quantile-quantile plot (right) for BW. Manhattan plot shows the $-\log_{10}$ (observed p-values) of the genome-wide SNPs (y-axis) across the 28 autosomes (x-axis), and the horizontal line denotes the genome-wide significant threshold. With regard to the Q-Q plot, the y-axis represents the observed $-\log_{10}$ (p-values) and the x-axis shows the expected $-\log_{10}$ (p-values). Manhattan plot was constructed with SNP & Variation Suite (version 8.8.1) software (Golden Helix: http://www.goldenhelix.com) while Q-Q plot with the CMplot package (https://github.com/YinLiLin/R-CMplot) in R (http://www.r-project.org/).


GGA1
GGA4
GGA10
GGA11
GGA15
GGA26

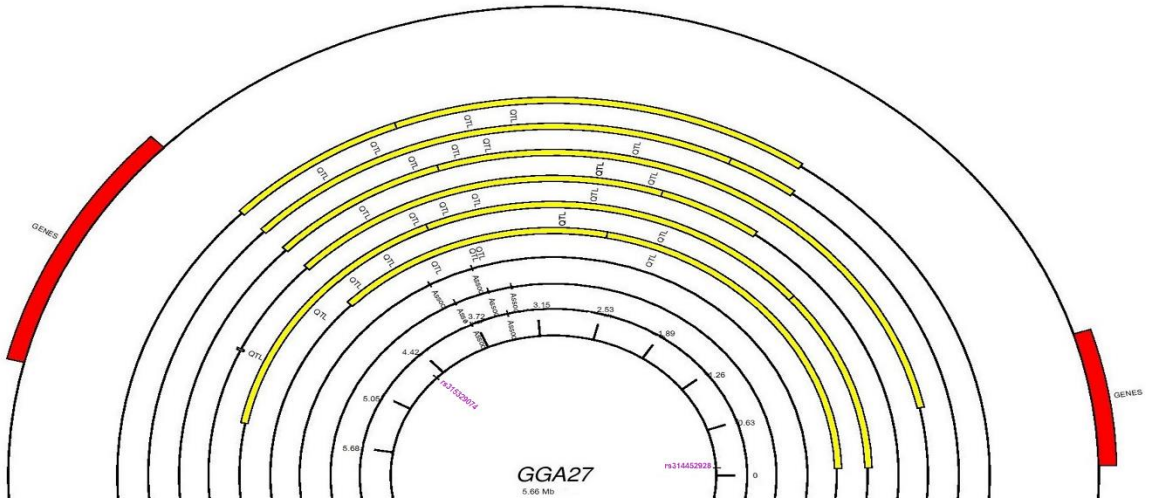

GENES
GENES
QTL
GGA27
5.66 Mb

Figure 2. Circular chromosome maps for seven autosomes presenting combined data of reported QTL (n=183) and positional candidate genes (n=462). Blue color represents the extent of large sized QTL (50 - 196.2 Mb), green color the medium sized QTL (5-50 Mb) and the yellow color is indicative of the small QTL (0-5 Mb). Red color indicates the starting and ending positions of positional candidate genes. The position(s) of the significant SNPs (labeled in purple color) is also given. The figure was constructed using GenomeVx [87].

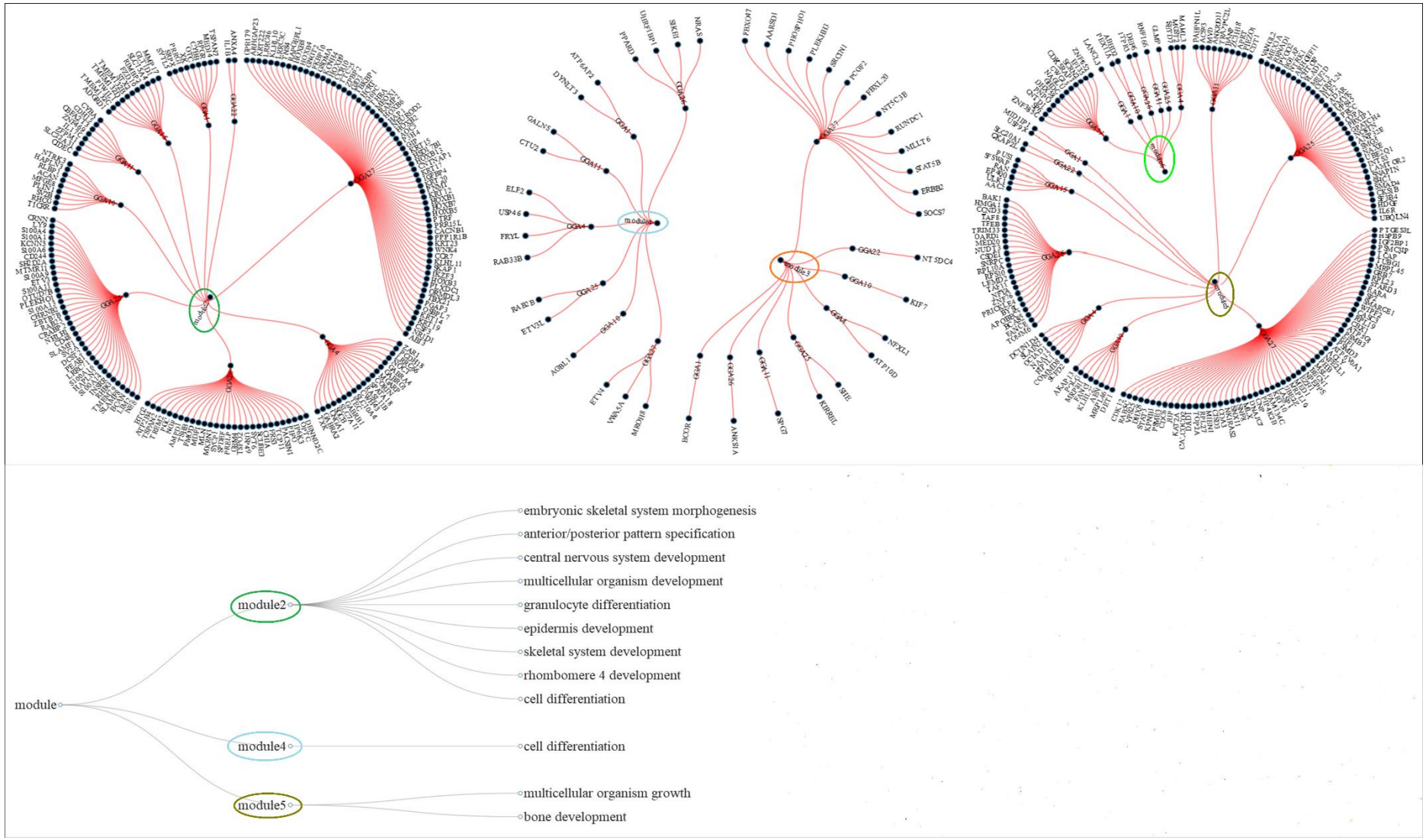


Figure 3. Network modules along with the significantly enriched developmental processes per module. The five modules are presented in the three radial networks (on the top) as circles/ellipses with different color together with their member genes and the corresponding chromosomes. The diagonal network at the bottom

provides the significantly enriched developmental processes per module. Figure was constructed using the data.tree and networkD3 packages in R (http://www.r-project.org/)·

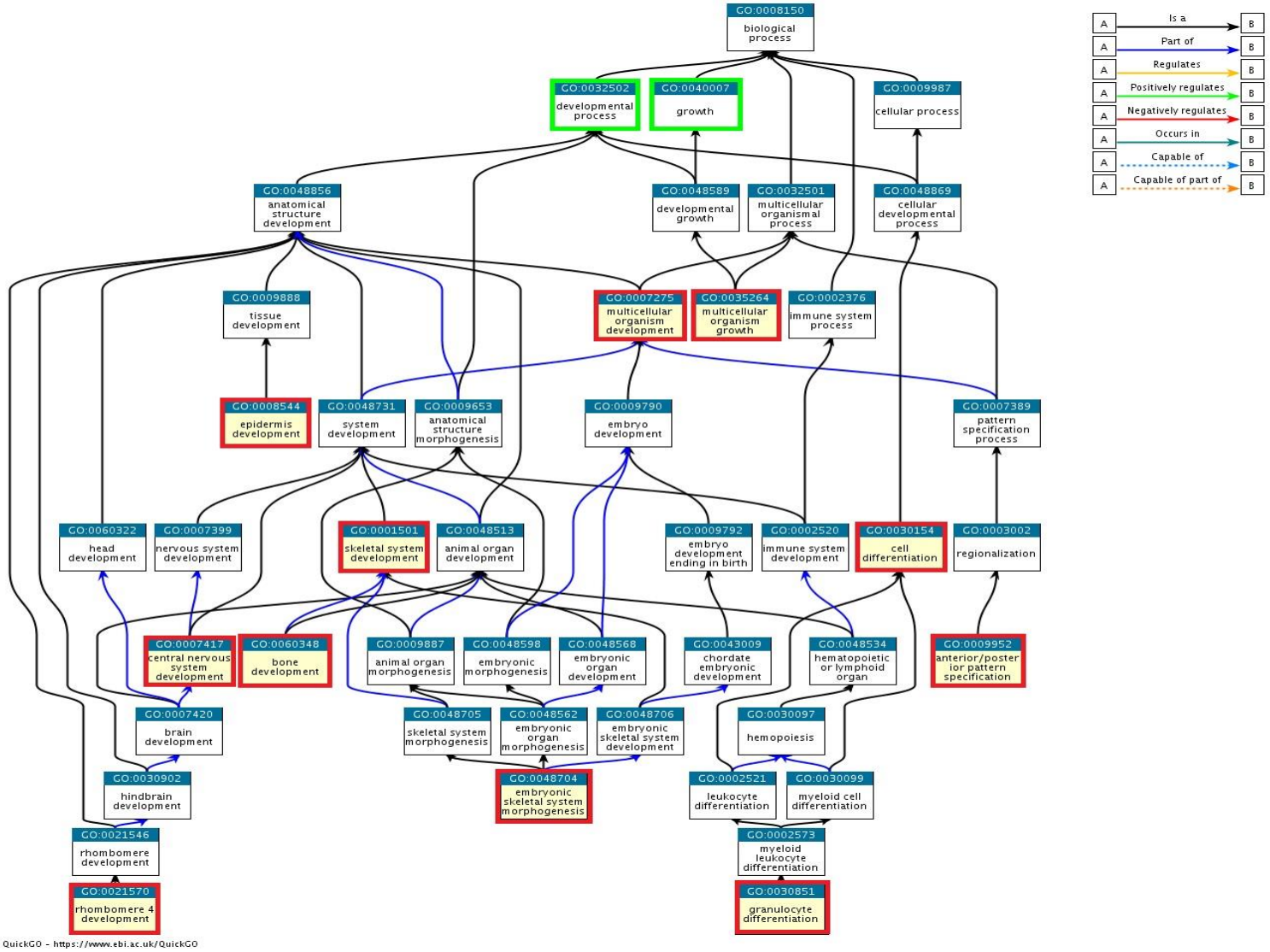


Figure 4. GO hierarchical structure for the eleven significantly enriched BPs (denoted with red color) associated with developmental process/growth term (denoted with green color). This GO tree was created and extracted by QuickGO [45].

Chapter 4

Deciphering the mode of action and position of genetic variants impacting on egg number in broiler breeders[1]

4.1. Abstract

Aim of the present study was first to identify genetic variants associated with egg number (EN) in female broilers, second to describe the mode of their gene action (additive and/or dominant) and third to provide a list with implicated candidate genes for the trait. A number of 2,586 female broilers genotyped with the high density (~600k) SNP array and with records on EN (mean=132.4 eggs, SD=29.8 eggs) were used. Data were analyzed with application of additive and dominant multi-locus mixed models. A number of 7 additive, 4 dominant and 6 additive plus dominant marker-trait significant associations were detected. A total number of 57 positional candidate genes were detected within 50 kb downstream and upstream flanking regions of the 17 significant markers. Functional enrichment analysis pinpointed two genes (*BHLHE40* and *CRTC1*) to be involved in the 'entrainment of circadian clock by photoperiod' biological process. Gene prioritization analysis of the positional candidate genes identified 10 top ranked genes (*GDF15, BHLHE40, JUND, GDF3, COMP, ITPR1, ELF3, ELL, CRLF1* and *IFI30).* Seven prioritized genes (*GDF15, BHLHE40, JUND, GDF3, COMP, ELF3, CRTC1)* have documented functional relevance to reproduction, while two more prioritized genes (*ITPR1* and *ELL*) are reported to be related to egg quality in chickens. Present results have shown that detailed exploration of phenotype-marker associations can disclose the mode of action of genetic variants and help in identifying causative genes associated with reproductive traits in the species.

4.2. Introduction

The breeding objective used for selection in broilers is balanced between reproduction, welfare and production traits [1]. Modern broiler breeding programs strive to optimize the overall reproductive efficiency, which is defined as the number of viable chicks per breeder hen and is determined by the egg production in combination with fertility and hatchability. Among the different metrics to describe egg production, egg number (EN), defined as the number of eggs laid over the duration of the laying period (from 28 to 54 weeks), is one of the most commonly used ones for selection purposes in commercial broilers [2,3].

As a typical reproductive trait, EN presents low to medium additive heritability estimates. In broiler hens, pedigree-based additive heritability for the trait has been estimated as high as 0.32, while respective estimates are in the range from 0.13 to 0.36 when using genomic relationship matrices [3,4]. The contribution of dominance may also be of importance for the trait, as estimates of the genomic dominant heritability has been found as high as 0.06 [3].

High-density SNP (single nucleotide polymorphism) genotyping arrays have greatly facilitated the detection of candidate causal variants in genome-wide association studies (GWAS) for various traits related to egg production and egg quality. Most GWAS have, so far, focused on the detection of additive SNPs for egg production [5-7] and egg quality traits [5,7-10]. It is noted that these studies have been focusing on EN in layer chickens and not broiler breeders. Moreover, to our knowledge, there is only one study [7] that aimed at identifying dominant SNPs for egg production and quality traits in chickens.

---

[1] Results of this Chapter have been presented in the XI[th] European symposium on Poultry Genetics. This Chapter has been accepted for publication in BMC Genomics.

Driven from the scarcity of published reports for broiler breeders, we elaborated the present study with the primary aim to detect genetic variants impacting on EN. Next, we sought to describe the mode of gene action of the significant genetic variants and finally attempted to provide a list with most likely candidate genes for the trait under investigation. Current findings are expected to contribute to a better understanding of the genetic mechanism(s) underlying the EN phenotype in the species.

4.3. Material and methods

4.3.1. Data

Genotypic and phenotypic records for 2,992 female broiler breeders from a purebred line were made available by Aviagen Ltd. Phenotypic records for EN were collected from 28 to 50 weeks of age and ranged from 26 to 196 eggs per female broiler with an average of 132.4 (SD=29.8). Animals were genotyped with the 600k Affymetrix® Axiom® high density genotyping array [11] resulting in a total number of 544,927 SNPs dispersed in 29 autosomes (GGA1-28 and GGA33). Quality control (QC) was performed first at a sample and second at a marker level. At a sample level, 406 animals were excluded due to call rate <0.99 and autosomal heterozygosity outside the 1.5 IQR (inter-quartile range: 0.013). At a marker level, a number of 305,660 SNPs were excluded due to: call rate <0.95, MAF (minor allele frequency) <0.05 and linkage disequilibrium (LD) r^2 values greater than 0.99 within windows of 1 Mb inter-marker distance(s). A total of 2,586 samples and 239,267 SNPs across 28 autosomes (GGA1-28) were retained for further analyses. QC was performed using the SNP & Variation Suite software (version 8.8.3).

4.3.2. Marker-trait association analysis

Multi-locus mixed-model (MLMM) stepwise regression with forward inclusion and backward elimination [12] of SNPs was employed to identify markers associated with the trait, assuming first an additive and second a dominant gene action for the SNP effects.

Specifically, the following statistical model was used for EN data:

$$y = X\boldsymbol{\beta} + \boldsymbol{w}\boldsymbol{a} + Z\boldsymbol{u} + \boldsymbol{e}$$

where y is the n x 1 vector of phenotypic values of EN for n female broilers, X is the n x 53 matrix of fixed effects: hatch (36 classes) and mating group (17 classes), $\boldsymbol{\beta}$ is the 53 x 1 vector of corresponding coefficients of fixed effects, $\boldsymbol{a}$ is the fixed effect for the minor allele of the candidate SNP to be tested for association, $\boldsymbol{w}$ is the incidence vector relating observations to SNP effects with elements coded as 0 for the major homozygous genotype, 1 for the heterozygote genotype and 2 for the minor homozygous genotype (additive genetic model) and 0 for the major homozygous genotype and 1 for the heterozygous and minor homozygous genotypes (dominant genetic model). Z is the incidence matrix relating observations to the polygenic random effects, $\boldsymbol{u}$ is the vector of polygenic random effects and $\boldsymbol{e}$ is the vector of random residuals.

The random effects were assumed to be normally distributed with zero means and the following covariance structure:

$$Var\begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} G\sigma_u^2 & 0 \\ 0 & I\sigma_e^2 \end{bmatrix}$$

where σ_u^2 and σ_e^2 are the polygenic and error variance components, I is the nxn identity

matrix, and G is the n x n genomic relationship matrix (GRM [13]) with elements of pairwise relationship coefficient using the 239,267 SNPs. The genomic relationship coefficient between two individuals j and k, was estimated as follows:

$$\frac{1}{239{,}267}\sum_{i=1}^{239{,}267}\frac{(x_{ij}-2p_i)(x_{ik}-2p_i)}{2p_i(1-2p_i)}$$

where x_{ij} and x_{ik} represent the number (0, 1, 2 in the additive model and 0, 1, 1 in the dominant model) of the minor allele of the i_{th} SNP of the j_{th} and k_{th} individuals, and p_i is the frequency of the minor allele [13].

Statistically significant SNPs per genetic model were selected at the optimal step of the MLMM stepwise regression according to extended Bayesian Information Criterion (eBIC[14]). SNP p-values were then corrected for multiple comparisons using the Bonferroni correction method. A SNP was considered as significant at the genome-wide level when its p-value was lower than the threshold value 2.09E-07 (0.05/239,267) while a chromosome-wide significant SNP had a p-value lower than 0.05/N, where N is the number of markers on a given chromosome. All analyses were performed using the SNP & Variation Suite software (version 8.8.3). SNP positions were based on *GRCg6a* assembly [15,16].

4.3.3. Quantile-quantile plots and estimation of the genomic inflation factor

Quantile-quantile (Q-Q) plots were used to analyze the extent to which the observed distribution of the test statistic followed the expected (null) distribution. These plots along with the estimation of the genomic inflation factor (λ) were used to assess potential systematic bias due to population structure or to the analytical approach [17]. Estimation of λ was performed using the SNP & Variation Suite (version 8.8.3).

4.3.4. Estimation of genomic heritability and proportion of variance explained

Estimation of the genomic heritability was implemented via the estimated GRM of 2,586 animals derived from 239,267 SNPs. The proportion of variance explained by a SNP *k* (PVE_k) was also calculated as follows:

$$PVE_k = \frac{mrss_{h0} - mrss_k}{mrss_{h0}}$$

where $mrss_{h0}$ is the Mahalonobis root sum of squares (mrss) for the null hypothesis and $mrss_k$ is the same for marker *k*. All above estimations were performed using the SNP & Variation Suite software (version 8.8.3).

4.3.5. Identification of significant SNPs under multicollinearity conditions

When multiple markers were present in a specific genomic region, a variable selection method i.e. the Least Absolute Shrinkage and Selection Operator (LASSO) [18] as implemented in procedure GLMSELECT in SAS 9.3 (2012) was applied to identify the most significant markers in the area.

4.3.6. Estimation of the degree of dominance

Significant SNPs associated with dominant or dominant and additive gene action(s) were

further analysed toward estimation of additive allelic effects, dominance deviation and degree of dominance. This analysis was based on estimates of genotype least squares (LS) means by application of a mixed model to the EN data fitting hatch, mating group and the marker as fixed effects and the animal as a random effect. The Satterthwaite method was used for estimating the degrees of freedom and the Tukey-Kramer method for adjusting the p-values because of multiple comparisons between genotype means. Results of the mixed model analysis are presented as LS means (μ) with standard errors (SE). Additive allelic effect (a) was defined as half the difference between LS means of the two homozygous genotypes, using the minor homozygous genotypes as reference. Dominance deviation (d) was the heterozygous genotype LS mean minus the average of the two homozygous genotype LS means. Finally, degree of dominance was determined as $|d/a|$, where additive= 0-0.20, partial dominance = 0.21-0.80, complete dominance = 0.81-1.20 and overdominance>1.20 [19,20]. This analysis was performed by the MIXED procedure in SAS 9.3 (2012).

4.3.7. Detection, functional characterization and prioritization of positional candidate genes

We searched within 50 kb downstream and upstream flanking regions of each significant marker for positional candidate genes using the NCBI database [21] and the *GRCg6a* assembly [15,16]. Subsequently, the total number of positional candidate genes was subjected to the following analyses: Gene Ontology (GO) Biological Process (BP) enrichment analysis and gene prioritization analysis (PA).

GO enrichment analysis for BP was performed using the DAVID functional annotation tool (https://david.ncifcrf.gov/, version 6.8) [22]. Specifically, we selected the *Gallus gallus* species for the input gene list and as whole genome background for enrichment analysis. The following settings were used in this analysis: EASE score (modified Fisher's exact p-value [23]) cutoff=0.05 and minimum number of genes per GO BP term=2. GO biological processes with p-values lower than 0.05 were considered as significantly enriched.

Next, gene prioritization analysis (PA) of the positional candidate genes was performed using the ToppGene portal (https://toppgene.cchmc.org/prioritization.jsp [24]). PA was based on the functional similarity of the positional candidate genes (test genes) to a training gene list including a total number of 31 genes (Supplementary Table 1). The latter genes were retrieved from the NCBI database [21] using the search terms 'reproduction' and 'egg production' in *Gallus gallus*. The portal performs functional annotation-based candidate gene prioritization using fuzzy-based similarity measures to compute the similarity between any two genes based on semantic annotations. In our study two semantic annotations: 'GO: Biological Process' and 'Coexpression' were used. A p-value for each annotation of a test gene was derived by random sampling of 5,000 genes from the whole genome and these partial p values were combined into an overall p value using the probability density function. For gene prioritization, there were 30 training genes *(ZNF764L* was omitted) and 43 test genes (positional candidate genes). Not all of the 57 positional candidate genes were included in the analysis because the human homologs could not be found for all of them, especially for LOC genes (n=14). Genes with an overall p-value lower than 0.05 were considered as highly prioritized.

4.4. Results

4.4.1. Significant SNPs and PVE

Additive and dominant genomic heritability estimates were identical and equal to 0.167 (SE=0.03) for the trait. The Q-Q plots (see Supplementary Figure 1) of the expected and observed SNP p-values along with the estimations of the genomic inflation factors (λ=0.95 and 0.97 for the respective additive and dominant genetic model) were indicate of no

systematic bias due to population structure or analytical approach. Profiles of the SNP p-values (expressed as $-\log_{10}$) for the additive and dominant genetic model are presented in form of circular Manhattan plots in Figure 1. No SNP was found to reach genome-wide significance (p<2.09E-07) using the Bonferroni correction method. Nevertheless, using the same correction method, a total number of 17 SNPs reached chromosome-wide significance across four autosomes (12, 22, 26 and 28) (Table 1). Specifically, one marker (*rs313298834)* was detected on GGA12 (threshold p=0.05/7,475=6.68896E-06), one (*rs314011910)* on GGA22 (threshold p=0.05/1,870=2.6738E-05), one (*rs313045367)* on GGA26 (threshold p=0.05/3,013=1.65948E-05) and 14 on GGA28 (threshold p=0.05/2,268=2.20459E-05). Of the 17 SNPs, 7 were associated with additive, 4 with dominant and 6 markers with both gene actions (Table 1). Of the additive SNPs, one marker (*rs313045367*) resided on GGA26 while 6 were located on GGA28. One dominant SNP (*rs314011910)* was detected on GGA22 while 3 dominant SNPs (*rs15250929, rs314052602* and *rs318126353*) were located on GGA28. Of markers displaying both gene actions, one marker (*rs313298834*) resided on GGA12 and 5 were located on GGA28 (Table 1). Note that the 14 significant SNPs residing on GGA28 were co-localized in a region spanning 240,432 bp (3,818,934-4,059,366 bp) with high LD (r^2) levels. A detailed view of these SNPs along with LD (r^2) levels between markers is depicted in Figure 2. As the LD heatmap shows, there are two haplotype blocks (r^2>0.70) formed by marker pairs *rs15250929-rs16212041* and *rs314418757-rs318126353* (Figure 2). PVE by significant markers ranged from 0.70% (*rs10724922, rs317783777*) to 0.85% (*rs314418757*) for the additive markers and from 0.69% (*rs314011910*) to 0.84% (*rs16212031, rs16212040, rs16212041*) for the dominant markers (Table 1). All together, the significant additive and dominant SNPs explained a considerable part i.e. 60% and 47% of the additive and genomic heritability, respectively. Nevertheless, as many of the significant markers were localized in nearby locations on GGA28, PVE by markers are biased upwards.

4.4.2. Estimation of the degree of dominance

Application of the LASSO method on the 14 co-localized SNPs on GGA28 resulted in selection of two markers i.e. *rs16212040* and *rs318126353* each one residing per different LD block (Figure 2). Of these, *rs16212040* was associated with both gene actions while *rs318126353* was associated only with dominant gene action. Two more SNPs i.e. *rs313298834 (*GGA12) and *rs314011910 (*GGA22) were detected as additive/dominant or dominant markers, respectively. Estimates of *a*, *d* and |*d/a*| for the four SNPs (*rs16212040, rs318126353, rs313298834* and *rs314011910*) are shown in Table 2. In line with a purely dominant model where genotypic values are solely determined by the presence or absence of the dominant allele, genotypic means of the minor homozygous and minor heterozygous were found to significantly differ from the major homozygous genotypic means (Table 2). Degree of dominance for the four SNPs ranged from 0.42-0.76 (partial dominance, markers: *rs16212040, rs313298834* and *rs318126353)* to 1.1 (complete dominance, marker: *rs314011910)*. Notably, no marker was associated with overdominance. We furthermore sought to quantify the joint effect of the combined genotype of the two markers (*rs16212040* and *rs318126353*) retained by LASSO on GGA28 by estimating LS means for the combined genotypes (Table 3). This exercise delivered interesting results as highest EN values were attained for AABB (μ=138.8, n=9) and ABAA (μ=138.9, n=71) that could not be attributed to additive effects of individual markers. Specifically, the highest EN estimate for AABB is suggestive of additive-by-additive (AABB) interaction (epistasis) while that of ABAA of additive-by-dominance (ABAA) epistasis. However, due to limited number of observations, especially for the AABB combined genotype (n=9), current results should be treated with caution.

4.4.3. Positional candidate genes

A total number of 57 positional candidate genes (i.e. 43 annotated and 14 LOC genes) were identified within the searched genomic regions (Supplementary Table 2). The maximum number of genes (n=16) were detected around dominant *rs318126353* (GGA28) while the minimum number of genes (n=6) were identified around 5 SNPs (*rs317783777, rs314011910, rs16212040, rs16212041* and *rs314418757*). Four additive SNPs (*rs313045367, rs10724922, rs317783777* and *rs315316434*) were located within genes *ARL8A* (GGA26), *UPF1* (GGA28*), CRTC1* (GGA28) and *TMEM59L* (GGA28), while 2 more markers (*rs313312915* and *rs14307369*) resided in gene *ELL* (GGA28). Three dominant SNPs (*rs15250929, rs314052602* and *rs318126353*) lied within genes *DDX49, KXD1 PGPEP1 (*GGA28). Of additive/dominant SNPs, 3 co-localized markers (*rs314228493, rs16212040* and *rs16212041*) were detected within *COMP* (GGA28) and one more (*rs314418757*) within *CRTC1* (GGA28). As 14 significant markers resided in nearby locations on GGA28, 26 out of the 36 positional candidate genes were associated with more than one marker (Figure 3). The maximum number of SNPs (n=10) were associated with gene *CRTC1*.

4.4.4. Functional enrichment analysis

A total number of 50 out of the 57 positional candidate genes were recognized by the DAVID tool and used for functional enrichment analysis. The latter analysis revealed the 'entrainment of circadian clock by photoperiod' (GO:0043153) as the only significantly (p=0.028) enriched BP with two participating genes (*CRTC1* and *BHLHE40)* (results not shown).

4.4.5. Prioritized genes

Results of PA are displayed on Table 4. A total number of 10 out of the 43 positional candidate genes were prioritized (overall p-value<0.05) according to the semantic annotations imposed. The majority (n=7) of the prioritized genes resided on GGA28, followed by two genes (*BHLHE40* and *ITPR1*) on GGA12 and one (*ELF3*) on GGA26. On GGA28, the first ranked gene was *GDF15*, followed by *JUND, GDF3, COMP, ELL, CRLF1* and *IFI30*. Notably, two highly ranked genes i.e. *GDF15* (1^{st}) and *GDF3* (4^{th}) belong to the transforming growth factor beta (TGF-β) superfamily. The two genes (*BHLHE40* and *CRTC1)* that participated in GO:0043153 'entrainment of circadian clock by photoperiod' were also prioritized and ranked 2^{nd} and 13^{th}, respectively.

4.5. Discussion

4.5.1. Mode of gene action

This is the first GWAS enlisting a significant number of animals (n~2600) and reporting on genetic variants implicated in the genetic control of EN in broiler breeders. Present results have demonstrated the need to thoroughly exploring the applicability of all possible genetic models when conducting a GWAS. This is particularly important when analyzing quantitative traits such as EN where not only additive but also non-additive e.g. dominant gene action of the causative loci may be fairly anticipated [3]. In line with this expectation, 4 of the 17 significant variants were dominant while 6 more were additive and dominant associations. The latter seems to be a controversial finding, but it can be fairly explained by examining the genotypic means across the examined variants of Table 2. A 'complete dominant' genetic model is when $|d|=|\alpha|$ meaning equal genotypic values for the minor homozygous (μ_{AA}) and the minor heterozygous genotypes (μ_{AB}) that both differ from the major homozygous genotypic mean (μ_{BB}). This was exactly the case for marker *rs314011910* that was detected only as dominant variant. But what happens in the case of partial dominance ($0<|d|<|a|$)? In such cases (see markers *rs313298834* and *rs16212040* in Table 2) all genotypic means differ

($\mu_{AA} \neq \mu_{AB}$, $\mu_{AA} \neq \mu_{BB}$ and $\mu_{AB} \neq \mu_{BB}$) meaning that apart from the dominant model, a linear relationship between the genotypic mean values and the number of copies of the minor allele i.e. the additive genetic model might also be applicable. For an excellent interpretation of how least squares regression performs in GWAS in additive and dominant models we refer to Huang and Mackay [25]. So far, we have discussed the applicability of the additive and dominant model, but we have neglected the case of overdominance ($|d|>|a|$). In the latter case, $\mu_{AB} > \mu_{AA}$ and $\mu_{AB} > \mu_{BB}$ implying the need of using a different model parameterization by coding the heterozygous genotypes as 1 and the two homozygous genotypes as 0. Due to model parameterization difficulties we could not explore the validity of an over-dominant genetic model here and this may be the reason why no marker has been associated with over-dominance in the current study.

While estimates of genetic effects (additive and/or dominant) are expected unbiased for a few, independent variants, this may not be case for multiple, highly correlated variants residing on the same haplotype block(s) because the effect(s) may be 'shared' by many markers. Under such conditions, it is vital to have a parsimonious model involving limited number of regressors (SNPs). To this end, application of the LASSO technique has proved particularly helpful as it has selected only two markers, each one residing in the two LD blocks on GGA28. Then, the next step was to explore whether the two variants interact and, if yes, to portray the exact type of interaction. This exploration has delivered interesting results since non-additive genetic interaction(s) between the two variants could also be detected. Although these findings are based on limited number of observations, they are indicative of potential importance of epistasis in the inheritance of the trait.

4.5.2. Functional candidate genes

Another intriguing problem that needed to be addressed in the present study was as how to narrow down the list with the 43 positional candidate genes. This post-GWAS step presents an important problem, because the experimental validation of the true causal genetic variants requires considerable costs, effort and time. To address this issue, we performed *in silico* prioritization analysis (PA) using explicitly two semantic annotations: GO: BP and co-expression. This approach was based on the assumption that co-expressed genes tend to be involved in the same biological process and that expression of functionally related genes should vary concordantly across the various tissues. Although gene co-expression networks typically do not provide information about causality, they can serve as first proof of their involvement in a particular biological process [26] and can be effectively used for the identification of regulatory genes underlying phenotypes [27]. Following this approach, 10 highly prioritized genes (*GDF15, BHLHE40, JUND, GDF3, COMP, ITPR1, ELF3, ELL, CRLF1* and *IFI30)* with interesting biological properties were highlighted. Genes *GDF15 (growth differentiation factor 15*, placed 1st) and *GDF3 (growth differentiation factor 3, placed 4th*) serve as good examples here since they both belong to the TGF-*β* superfamily genes. In rodents and humans, many factors belonging to the TGF-β superfamily are expressed by ovarian somatic cells and oocytes in a developmental manner and function as intraovarian regulators of folliculogenesis [28]. In humans, *GDF15* is involved in placentation [29], while *GDF3* might affect folliculogenesis by inhibiting the bone morphogenetic protein cytokines [30]. In chickens, *GDF3 (*also known as cVg1) is expressed at the early blastoderm stages of embryonic development [31] while another TGF-β member i.e. *GDF9* is expressed in the ovary and functions on hen granulosa cell proliferation as in mammals [32]. Expression of *BHLHE40 (basic helix-loop-helix family member e40)* in the mouse ovary leads to a circadian gating of cellular processes in the ovary as well as in the hypothalamus during ovulation [33]. *JUND (JunD proto-oncogene, AP-1 transcription factor subunit*) is important for maturation of human ovarian cells [34]. *COMP (cartilage oligomeric matrix protein)* is

involved in ovarian follicle development in mice [35] while mutations of *COMP* gene affect chondrogenesis in chickens [36]. *ITPR1 (inositol 1,4,5-trisphosphate receptor type 1)* is involved in the Ca^{2+} transport for supplying eggshell mineral precursors in chicken uterus [37,38] while *ELF3 (E74 like ETS transcription factor 3)* has been related to the development of chicken oviducts [39] and *ELL (elongation factor for RNA polymerase II)* has been associated with yolk weight [40] in chickens. Notably, the final two nominated candidates i.e. *CRLF1 (cytokine receptor like factor 1)* and *IFI30 (IFI30, lysosomal thiol reductase)* had no documented involvement in reproduction. Such a finding underscores the limitations of *in silico* PA. In almost every guilt-by-association (GBA)-based prioritization tool, functional annotations of genes refer mainly to human and mouse PPINs (protein-protein interaction networks) [41] neglecting relevant information on livestock species [42] such as that examined here. One more limitation of GBA-based networks relates to their degraded predictive performance for genes with unknown or multiple functions [41].

Of particular interest in this study were genes *BHLHE40* and *CRTC1(CREB regulated transcription coactivator 1).* Both genes were enriched in the BP of 'entrainment of circadian clock by photoperiod' raising the intriguing question as what might be the exact mechanism of their implication in egg production. To answer the question, first we have to provide a short description of the molecular mechanism underlying circadian rhythms (CR). CR are regulated by a pacemaker located in the suprachiasmatic nucleus of the hypothalamus that is entrained to the external light–dark cycle via light input from the retina conveyed via the retinohypothalamic tract [43]. In hens, as in many avian species, exposure to photoperiods of longer than 11.5 hrs/day results in the rapid induction of the hypothalamo-hypophysial-gonad axis, causing development and growth of testes and ovarian follicles [44]. At the intracellular level, four clock-gene families have been found to be involved in a transcription–translation feedback loop that generates the CR. Gene products of *Clock* and *Bmal1* act as positive components, whereas those of the *Per* and *Cry* genes act as negative ones [45]. With regard to our candidate genes, *BHLHE40* (also known as *BHLHB2*) acts as a suppressor of *Clock* and *Bmal1* genes [46] while an entrainment stimulus causes *CRTC1* to induce expression of *Per1* and *Sik1* [47]. As the molecular bases for circadian clocks are highly conserved it is likely that the avian molecular mechanisms are similar to those expressed in mammals, including humans [44].

In total, 7 (*GDF15, BHLHE40, JUND, GDF3, COMP, ELF3* and *CRTC1*) of the prioritized genes were associated with reproductive traits while 2 (*ITPR1* and *ELL*) were related to egg quality traits. From the above, only 3 genes i.e. *COMP, ELL* and *CRTC1* included significant SNPs. We finally, compared our candidate genes list (Supplementary Table 2) to a compiled gene list including 271 genes (Supplementary Table 3) identified in previous GWAS for chicken egg and reproductive traits. This comparison highlighted two common genes i.e. *ELL* and *ARL8A.* Note that the first is among the prioritized candidates (ranked 8^{th}) while the second *ARL8A (ADP ribosylation factor like GTPase 8A)* has been associated with eggshell thickness and eggshell formation [5] in chickens.

4.6. Conclusions

Current results have shown that apart from the additive also the dominant genetic model was of importance for EN in broilers. These results underline the need to thoroughly exploring the applicability of all possible genetic models when performing GWAS for a trait such as that examined here. Detailed follow-up studies are warranted to verify whether the identified genomic markers and the associated candidate genes present true causal genetic entities impacting on the trait. Such studies would entail targeted re-sequencing and molecular characterization of the candidate variants to facilitate the identification of true causal variants.

4.7. References

[1] Hiemstra SJ, Napel J Ten. Study of the impact of genetic selection on the welfare of chickens bred and kept for meat production. Final report of a project commissioned by the European Commission (DG SANCO/2011/12254); 2013.

[2] Abdollahi-Arpanahi R, Pakdel A, Nejati-Javaremi A, Moradi Shahrbabak M, Morota G, Valente BD, et al. Dissection of additive genetic variability for quantitative traits in chickens using SNP markers. J Anim Breed Genet. Wiley/Blackwell (10.1111); 2014;131: 183–193. doi:10.1111/jbg.12079

[3] Abdollahi-Arpanahi R, Morota G, Valente BD, Kranis A, Rosa GJM, Gianola D. Differential contribution of genomic regions to marked genetic variation and prediction of quantitative traits in broiler chickens. Genet Sel Evol. BioMed Central; 2016;48: 10. doi:10.1186/s12711-016-0187-z

[4] Momen M, Mehrgardi AA, Sheikhy A, Esmailizadeh A, Fozi MA, Kranis A, et al. A predictive assessment of genetic correlations between traits in chickens using markers. Genet Sel Evol. 2017;49: 16. doi:10.1186/s12711-017-0290-9

[5] Liao R, Zhang X, Chen Q, Wang Z, Wang Q, Yang C, et al. Genome-wide association study reveals novel variants for growth and egg traits in Dongxiang blue-shelled and White Leghorn chickens. Anim Genet. 2016;47: 588–596. doi:10.1111/age.12456

[6] Yuan J, Sun C, Dou T, Yi G, Qu L, Qu L, et al. Identification of Promising Mutants Associated with Egg Production Traits Revealed by Genome-Wide Association Study. Xu P, editor. PLoS One. Public Library of Science; 2015;10: e0140615. doi:10.1371/journal.pone.0140615

[7] Liu W, Li D, Liu J, Chen S, Qu L, Zheng J, et al. A Genome-Wide SNP Scan Reveals Novel Loci for Egg Production and Quality Traits in White Leghorn and Brown-Egg Dwarf Layers. Christoffels A, editor. PLoS One. Public Library of Science; 2011;6: e28600. doi:10.1371/journal.pone.0028600

[8] Liu Z, Sun C, Yan Y, Li G, Wu G, Liu A, et al. Genome-Wide Association Analysis of Age-Dependent Egg Weights in Chickens. Front Genet. Frontiers; 2018;9: 128. doi:10.3389/fgene.2018.00128

[9] Sun C, Qu L, Yi G, Yuan J, Duan Z, Shen M, et al. Genome-wide association study revealed a promising region and candidate genes for eggshell quality in an F2 resource population. BMC Genomics. BioMed Central; 2015;16: 565. doi:10.1186/s12864-015-1795-7

[10] Yi G, Shen M, Yuan J, Sun C, Duan Z, Qu L, et al. Genome-wide association study dissects genetic architecture underlying longitudinal egg weights in chickens. BMC Genomics. BioMed Central; 2015;16: 746. doi:10.1186/s12864-015-1945-y

[11] Kranis A, Gheyas AA, Boschiero C, Turner F, Yu L, Smith S, et al. Development of a high density 600K SNP genotyping array for chicken. BMC Genomics. BioMed Central; 2013;14: 59. doi:10.1186/1471-2164-14-59

[12] Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, et al. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet. Nature Publishing Group; 2012;44: 825–830. doi:10.1038/ng.2314

[13] VanRaden PM. Efficient Methods to Compute Genomic Predictions. J Dairy Sci. 2008;91: 4414–4423. doi:10.3168/jds.2007-0980

[14] Chen J, Chen Z. Extended Bayesian information criteria for model selection with large model spaces. Biometrika. Oxford University Press; 2008;95: 759–771. doi:10.1093/biomet/asn034

[15] GRCg6a [Internet]. [cited 21 Apr 2019]. Available: https://www.ncbi.nlm.nih.gov/assembly/GCF_000002315.6

[16] NCBI Gallus gallus Annotation Release 104 [Internet]. [cited 21 Apr 2019]. Available: https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Gallus_gallus/104/

[17] Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, Willer CJ, et al. Genomic inflation factors under polygenic inheritance. Eur J Hum Genet. Nature Publishing Group; 2011;19: 807–12. doi:10.1038/ejhg.2011.39

[18] Zou H. The Adaptive Lasso and Its Oracle Properties. 2006; doi:10.1198/016214506000000735

[19] Babu R, Nair SK, Kumar A, Rao HS, Verma P, Gahalain A, et al. Mapping QTLs for popping ability in a popcorn × flint corn cross. Theor Appl Genet. Springer-Verlag; 2006;112: 1392–1399. doi:10.1007/s00122-006-0242-1

[20] Mori A, Tsuda Y, Takagi M, Higa Y, Severson DW. Multiple QTL Determine Dorsal Abdominal Scale Patterns in the Mosquito Aedes aegypti. J Hered. Oxford University Press; 2016;107: 438–44. doi:10.1093/jhered/esw027

[21] NCBI [Internet]. [cited 21 Apr 2019]. Available: https://www.ncbi.nlm.nih.gov/gene/

[22] Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane H, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. BioMed Central; 2003;4: R60. doi:10.1186/gb-2003-4-9-r60

[23] Hosack DA, Dennis G, Sherman BT, Lane HC, Lempicki RA, Lempicki RA. Identifying biological themes within lists of genes with EASE. Genome Biol. BioMed Central; 2003;4: R70. doi:10.1186/gb-2003-4-10-r70

[24] Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res. Oxford University Press; 2009;37: W305-11. doi:10.1093/nar/gkp427

[25] Huang W, Mackay TFC. The Genetic Architecture of Quantitative Traits Cannot Be Inferred from Variance Component Analysis. Zhu X, editor. PLOS Genet. Public Library of Science; 2016;12: e1006421. doi:10.1371/journal.pgen.1006421

[26] Liseron-Monfils C, Olson A, Ware D. NECorr, a Tool to Rank Gene Importance in Biological Processes using Molecular Networks and Transcriptome Data. bioRxiv. Cold Spring Harbor Laboratory; 2018; 326868. doi:10.1101/326868

[27] van Dam S, Võsa U, van der Graaf A, Franke L, de Magalhães JP. Gene co-expression analysis for functional classification and gene–disease predictions. Brief Bioinform. Narnia; 2017;19: bbw139. doi:10.1093/bib/bbw139

[28] Knight PG, Glister C. TGF-β superfamily members and ovarian follicle development. Reproduction. 2006;132: 191–206. doi:10.1530/rep.1.01074

[29] Fejzo MS, Sazonova O V, Sathirapongsasuti JF, Hallgrímsdóttir IB, Vacic V, MacGibbon KW, et al. Placenta and appetite genes GDF15 and IGFBP7 are associated with hyperemesis gravidarum. Nat Commun. Nature Publishing Group; 2018;9: 1178. doi:10.1038/s41467-018-03258-0

[30] Shi J, Yoshino O, Osuga Y, Akiyama I, Harada M, Koga K, et al. Growth differentiation factor 3 is induced by bone morphogenetic protein 6 (BMP-6) and BMP-7 and increases luteinizing hormone receptor messenger RNA expression in human granulosa cells. Fertil Steril. Elsevier; 2012;97: 979–983. doi:10.1016/J.FERTNSTERT.2012.01.100

[31] Seleiro EAP, Connolly DJ, Cooke J. Early developmental expression and experimental axis determination by the chicken Vg1 gene. Curr Biol. Cell Press; 1996;6: 1476–1486. doi:10.1016/S0960-9822(96)00752-X

[32] Johnson PA, Dickens MJ, Kent TR, Giles JR. Expression and Function of Growth Differentiation Factor-9 in an Oviparous Species, Gallus domesticus1. Biol Reprod. 2005;72: 1095–1100. doi:10.1095/biolreprod.104.036822

[33] Boden MJ, Varcoe TJ, Voultsios A, Kennaway DJ. Reproductive biology of female Bmal1 null mice. REPRODUCTION. 2010;139: 1077–1090. doi:10.1530/REP-09-0523

[34] Günthert AR, Gründker C, Hollmann K, Emons G. Luteinizing hormone-releasing hormone induces JunD–DNA binding and extends cell cycle in human ovarian cancer cells. Biochem Biophys Res Commun. 2002;294: 11–15. doi:10.1016/S0006-291X(02)00427-8

[35] Galdones E, Penalver Bernabe B, Skory RM, Mackovic D, Broadbelt LJ, Shea LD, et al. Ovarian Expression of Cartilage Oligomeric Matrix Protein as a Potential Biomarker of Antral Follicle Development in the Mouse. Biol Reprod. Narnia; 2011;85: 647–647. doi:10.1093/biolreprod/85.s1.647

[36] Roman-Blas J, Dion AS, Seghatoleslami MR, Giunta K, Oca P, Jimenez SA, et al. MED and PSACH COMP mutations affect chondrogenesis in chicken limb bud micromass cultures. J Cell Physiol. 2010;224: 817–826. doi:10.1002/jcp.22185

[37] Duan Z, Chen S, Sun C, Shi F, Wu G, Liu A, et al. Polymorphisms in Ion Transport Genes Are Associated with Eggshell Mechanical Property. Yue GH, editor. PLoS One. Public Library of Science; 2015;10: e0130160. doi:10.1371/journal.pone.0130160

[38] Jonchère V, Brionne A, Gautron J, Nys Y. Identification of uterine ion transporters for mineralisation precursors of the avian eggshell. BMC Physiol. BioMed Central; 2012;12: 10. doi:10.1186/1472-6793-12-10

[39] Song G, Seo HW, Choi JW, Rengaraj D, Kim TM, Lee BR, et al. Discovery of Candidate Genes and Pathways Regulating Oviduct Development in Chickens1. Biol Reprod. 2011;85: 306–314. doi:10.1095/biolreprod.110.089227

[40] Sun C, Lu J, Yi G, Yuan J, Duan Z, Qu L, et al. Promising Loci and Genes for Yolk and Ovary Weight in Chickens Revealed by a Genome-Wide Association Study. Veitia RA, editor. PLoS One. Public Library of Science; 2015;10: e0137145. doi:10.1371/journal.pone.0137145

[41] Gillis J, Pavlidis P. “Guilt by Association” Is the Exception Rather Than the Rule in Gene Networks. Rzhetsky A, editor. PLoS Comput Biol. 2012;8: e1002444. doi:10.1371/journal.pcbi.1002444

[42] Kominakis A, Hager-Theodorides AL, Zoidis E, Saridaki A, Antonakos G, Tsiamis G. Combined GWAS and ’guilt by association’-based prioritization analysis identifies functional candidate genes for body size in sheep. Genet Sel Evol. BioMed Central; 2017;49: 41. doi:10.1186/s12711-017-0316-3

[43] Hughes S, Jagannath A, Hankins MW, Foster RG, Peirson SN. Photic Regulation of Clock Systems. Methods in enzymology. 2015. pp. 125–143. doi:10.1016/bs.mie.2014.10.018

[44] Cassone VM. Avian circadian organization: A chorus of clocks. Front Neuroendocrinol. Academic Press; 2014;35: 76–88. doi:10.1016/J.YFRNE.2013.10.002

[45] Honma S, Kawamoto T, Takagi Y, Fujimoto K, Sato F, Noshiro M, et al. Dec1 and Dec2 are regulators of the mammalian molecular clock. Nature. 2002;419: 841–844. doi:10.1038/nature01123

[46] Nakashima A, Kawamoto T, Honda KK, Ueshima T, Noshiro M, Iwata T, et al. DEC1 Modulates the Circadian Phase of Clock Gene Expression. Mol Cell Biol. American Society for Microbiology Journals; 2008;28: 4080–4092. doi:10.1128/MCB.02168-07

[47] Jagannath A, Butler R, Godinho SIH, Couch Y, Brown LA, Vasudevan SR, et al. The CRTC1-SIK1 pathway regulates entrainment of the circadian clock. Cell. Elsevier; 2013;154: 1100–1111. doi:10.1016/j.cell.2013.08.004

[48] Shin J-H, Blay S, Graham J, McNeney B. LDheatmap : An *R* Function for Graphical Display of Pairwise Linkage Disequilibria Between Single Nucleotide Polymorphisms. J Stat Softw. 2006;16: 1–9. doi:10.18637/jss.v016.c03

[49] Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv. 2014;:005165. doi:10.1101/005165.

Tables and Figures of Chapter 4

Table 1. Chromosome-wide significant SNPs identified by additive (add), dominant (dom) or both additive and dominant (add/dom) genetic models. (MAF: Minor Allele Frequency).

SNP ID	GGA[a]	Position (bp)[b]	p-value (add/dom)	-log$_{10}$(p-value) (add/dom)	Minor allele	MAF	PVE[c](%)		Genetic model
							add	dom	
rs313298834	12	18,995,645	5.03832E-06/3.37289E-06	5.298/5.472	B	0.34	0.8	0.83	add/dom
rs314011910	22	1,711,605	2.30566E-05	4.637	A	0.14	-	0.69	dom
rs313045367	26	362,590	8.41209E-06	5.075	B	0.15	0.77	-	add
rs15250929	28	3,818,934	1.93573E-05	4.713	B	0.2	-	0.7	dom
rs10724922	28	3,855,714	2.0557E-05	4.687	A	0.21	0.7	-	add
rs15251036	28	3,875,127	1.74651E-05	4.758	B	0.21	0.71	-	add
rs16212031	28	3,885,458	5.56121E-06/3.19578E-06	5.255/5.495	A	0.2	0.80	0.84	add/dom
rs314228493	28	3,888,943	4.69378E-06/4.46165E-06	5.328/5.351	B	0.2	0.81	0.81	add/dom
rs16212040	28	3,892,786	3.35519E-06/3.06323E-06	5.474/5.514	B	0.2	0.83	0.84	add/dom
rs16212041	28	3,892,872	3.24649E-06/3.06323E-06	5.489/5.514	A	0.2	0.83	0.84	add/dom
rs317783777	28	3,919,505	2.02328E-05	4.694	A	0.14	0.70	-	add
rs314418757	28	3,921,905	2.70727E-06/1.62505E-05	5.567/4.789	A	0.21	0.85	0.72	add/dom
rs315316434	28	3,971,928	1.71745E-05	4.765	A	0.22	0.71	-	add
rs314052602	28	3,990,564	1.30589E-05	4.884	B	0.21	-	0.73	dom

rs313312915	28	3,999,772	8.56382E-06	5.067	B	0.22	0.76	-	add
rs14307369	28	4,003,865	1.37243E-05	4.863	A	0.21	0.73	-	add
rs318126353	28	4,059,366	5.35831E-06	5.271	B	0.23	-	0.80	dom

[a]Chromosome for *Gallus gallus*

[b]Positions were based on *GRCg6a* assembly

[c] Proportion of variance explained

Table 2. Estimation of genotypic means (μ ± SE) for EN, additive allelic effects (α), dominance deviation (d) and degree of dominance (|d/α|) for the significant additive/dominant markers.

Marker	Genotype (coded as)	Sample size	μ ± SE	α± SE	d± SE	|d/α|
rs313298834 (add/dom)	AA (0)	1595	$129.9^{b} \pm 1.0$	$3.6^{*} \pm 0.8$	$2.6^{NS} \pm 2.0$	|2.6/3.6|=0.72
	AB (1)	245	$136.0^{a} \pm 2.0$			
	BB (2)	746	$137.1^{a} \pm 1.4$			
rs314011910 (dom)	BB (0)	2167	$133.7^{b} \pm 1.0$	$-3.5^{*} \pm 1.0$	$-3.8^{NS} \pm 3.2$	|3.8/3.5|=1.1
	AB (1)	91	$126.4^{a} \pm 3.2$			
	AA (2)	328	$126.7^{a} \pm 2.1$			
rs16212040 (add/dom)	AA (0)	1695	$135.2^{b} \pm 1.0$	$-5.0^{*} \pm 1.3$	$-2.1^{NS} \pm 1.7$	|2.1/5.0|=0.42
	AB (1)	758	$128.1^{a} \pm 1.3$			
	BB (2)	133	$125.2^{a} \pm 2.6$			
rs318126353 (dom)	AA (0)	1583	$135.5^{b} \pm 1.0$	$-4.1^{*} \pm 1.2$	$-3.1^{*} \pm 1.6$	|3.1/4.1|=0.76
	AB (1)	838	$128.3^{a} \pm 1.2$			
	BB (2)	165	$127.3^{a} \pm 2.4$			

^{a,b} means with different superscripts are statistically different ($p<0.05$)
*statistically significant with $p<0.05$
^{NS} non statistically significant

Table 3. Least squares mean (μ ± SE) for EN for combined genotype of markers *rs16212040* and *rs318126353* on GGA28. N is the sample size.

Combined genotype	**N**	**μ ± SE**
AA/AA	1512	135.3 ± 1.0
AA/AB	174	134.3 + 2.3
AA/BB	9	138.8 ± 9.6
AB/AA	71	138.9 ± 3.5
AB/AB	639	126.7 ± 1.3
AB/BB	48	129.9 ± 4.2
BB/AB	25	125.7 ± 5.8
BB/BB	108	125.0 ± 2.9

Table 4. List of prioritized genes.

Rank	*Gene ID*	*Description*	GGA	Overall p-value
1	*GDF15*	*growth differentiation factor 15*	28	0.019
2	*BHLHE40*	*basic helix-loop-helix family member e40*	12	0.027
3	*JUND*	*JunD proto-oncogene, AP-1 transcription factor subunit*	28	0.029
4	*GDF3*	*growth differentiation factor 3*	28	0.030
5	*COMP*	*cartilage oligomeric matrix protein*	28	0.037
6	*ITPR1*	*inositol 1,4,5-trisphosphate receptor type 1*	12	0.039
7	*ELF3*	*E74 like ETS transcription factor 3*	26	0.040
8	*ELL*	*elongation factor for RNA polymerase II*	28	0.044
9	*CRLF1*	*cytokine receptor like factor 1*	28	0.047
10	*IFI30*	*IFI30, lysosomal thiol reductase*	28	0.048
11	*ISYNA1*	*inositol-3-phosphate synthase 1*	28	0.050
12	*RAB3A*	*RAB3A, member RAS oncogene family*	28	0.051
13	*CRTC1*	*CREB regulated transcription coactivator 1*	28	0.057
14	*GPR37L1*	*G protein-coupled receptor 37 like 1*	26	0.057
15	*PIK3R2*	*phosphoinositide-3-kinase regulatory subunit 2*	28	0.060
16	*GFRA2*	*GDNF family receptor alpha 2*	22	0.061
17	*EDEM1*	*ER degradation enhancing alpha-mannosidase like protein 1*	12	0.069
18	*FKBP8*	*FK506 binding protein 8*	28	0.069
19	*PDE4C*	*phosphodiesterase 4C*	28	0.083
20	*LGR6*	*leucine rich repeat containing G protein-coupled receptor 6*	26	0.086
21	*HOMER3*	*homer scaffolding protein 3*	28	0.112
22	*LSM4*	*LSM4 homolog, U6 small nuclear RNA and mRNA degradation associated*	28	0.114
23	*COPE*	*coatomer protein complex subunit epsilon*	28	0.149
24	*ARL8B*	*ADP ribosylation factor like GTPase 8B*	12	0.150
25	*PTPN7*	*protein tyrosine phosphatase, non-receptor type 7*	26	0.152
26	*PGPEP1*	*pyroglutamyl-peptidase I*	28	0.156
27	*C19orf60 (also known as REX1BD)*	*chromosome 19 open reading frame 60*	28	0.172

28	*SSBP4*	*single stranded DNA binding protein 4*	28	0.192
29	*UBA52*	*ubiquitin A-52 residue ribosomal protein fusion product 1*	28	0.192
30	*UPF1*	*UPF1, RNA helicase and ATPase*	28	0.192
31	*CERS1*	*ceramide synthase 1*	28	0.198
32	*MPV17L2*	*MPV17 mitochondrial inner membrane protein like 2*	28	0.228
33	*DOK2*	*docking protein 2*	22	0.262
34	*XPO7*	*exportin 7*	22	0.292
35	*ARL8A*	*ADP ribosylation factor like GTPase 8A*	26	0.340
36	*DDX49*	*DEAD-box helicase 49*	28	0.345
37	*SUGP2*	*SURP and G-patch domain containing 2*	28	0.345
38	*KLHL26*	*kelch like family member 26*	28	0.345
39	*KIF21B*	*kinesin family member 21B*	26	0.351
40	*KXD1*	*KxDL motif containing 1*	28	0.351
41	*LRRC25*	*leucine rich repeat containing 25*	28	0.588
42	*TMEM59L*	*transmembrane protein 59 like*	28	0.588
43	*PTPRVP*	*protein tyrosine phosphatase, receptor type, V, pseudogene*	26	1.000

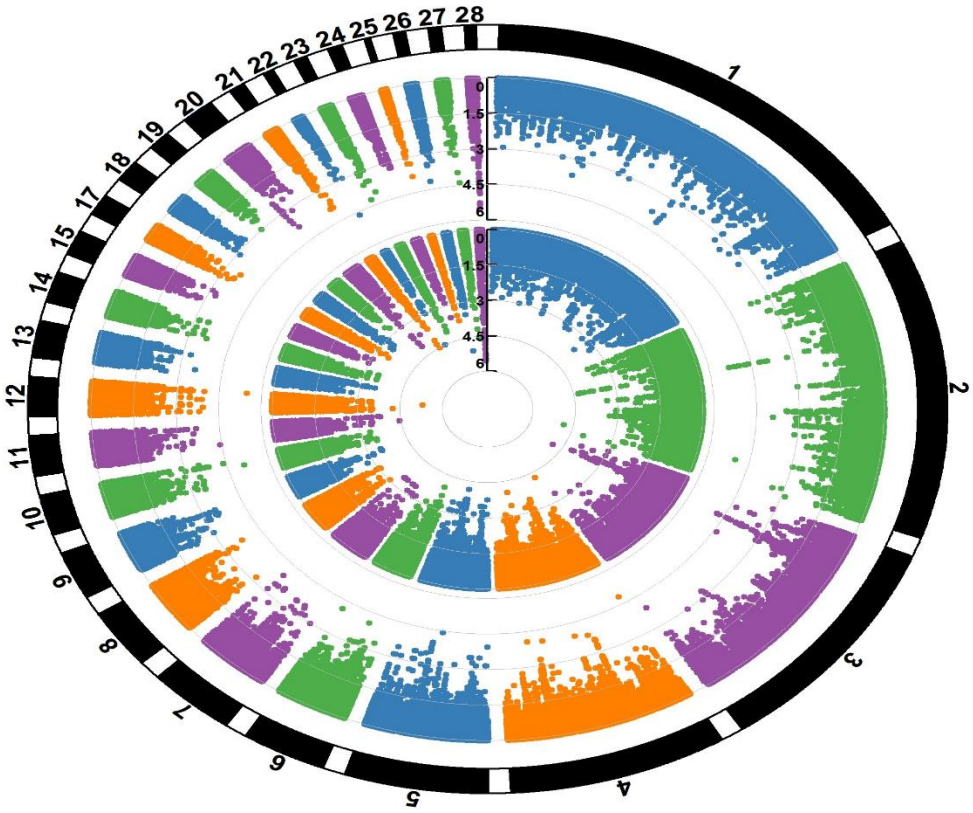


Figure 1. Circular Manhattan plot displaying the chromosome-wide significant associations for EN. The $-\log_{10}$(p-values) of the additive (inner circle) and dominant (outer circle) SNPs are shown across the 28 autosomal chromosomes. This plot was constructed with the CMplot package (https://github.com/YinLiLin/R-CMplot) in R (http://www.r-project.org/).

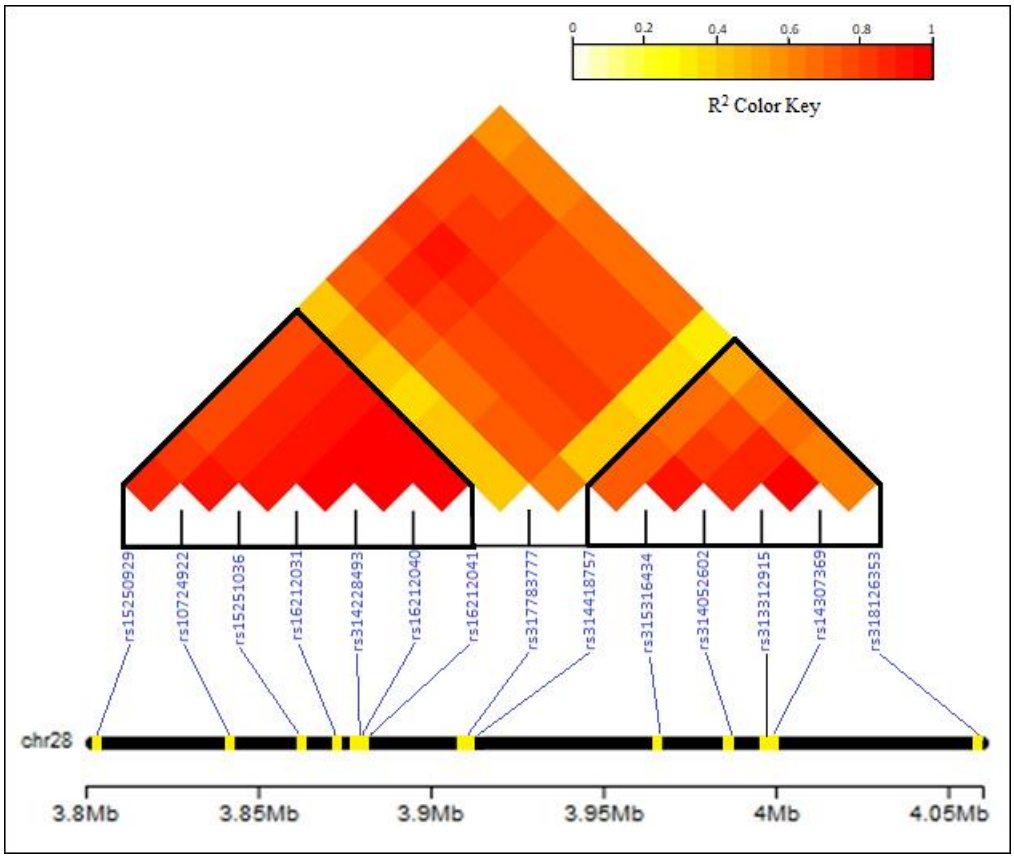


Figure 2. LD heatmap for the 14 SNPs (blue labels) on GGA28. Note the formation of 2 LD blocks (denoted as black lined polygons). LD levels were estimated using the gaston R package and were graphically displayed with use of LDheatmap [48] package in R (http://www.r-project.org/).

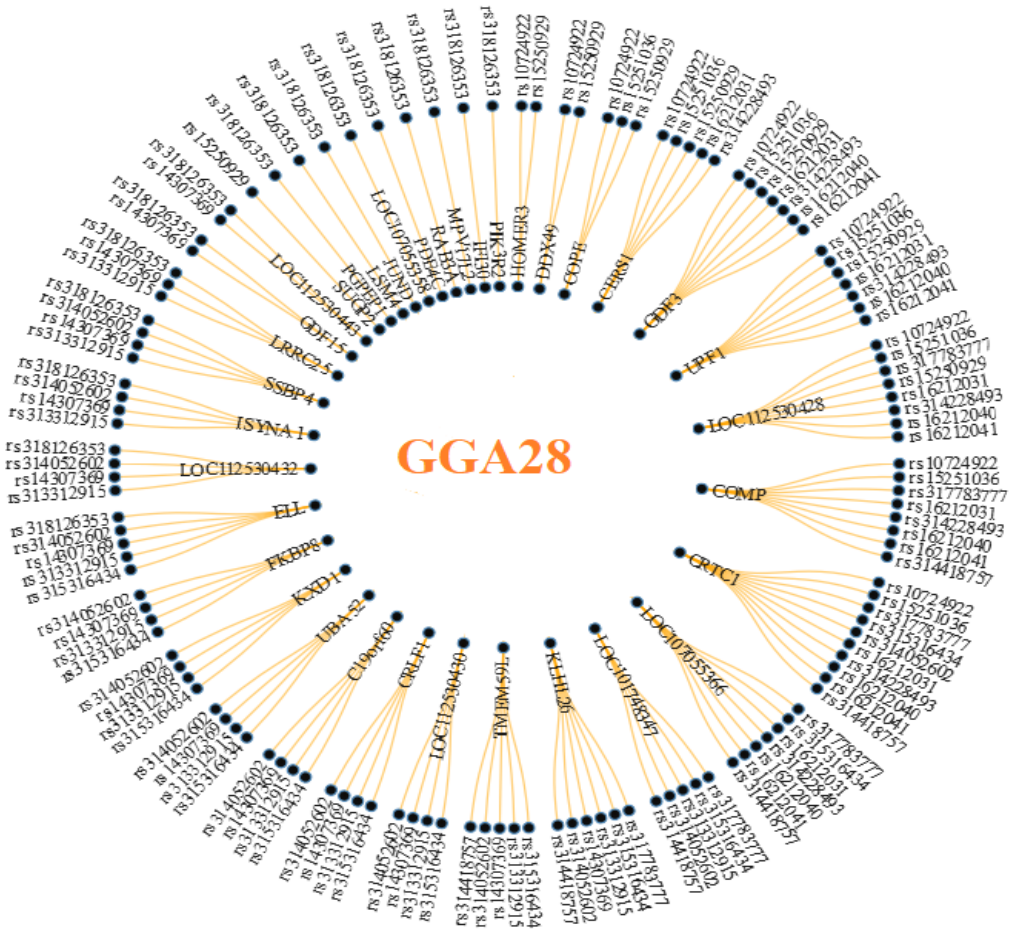


Figure 3. Radial network of significant SNPs associated with positional candidate genes on GGA28. Figure was constructed using the data.tree and networkD3 packages in R (http://www.r-project.org/).

Chapter 5

Detection of pleiotropic loci with antagonistic effects on body weight and egg number in chickens

5.1. Abstract

The present study aimed to identify genetic variants and plausible genes underlying the negative genetic correlation observed between body weight and egg number in female broilers. To this end, bivariate genome-wide association and stepwise conditional-joint analyses were carried out using 52,992 autosomal SNPs and 2,586 female broilers. These analyses pinpointed a total number of 13 independent SNPs exerting cross-phenotype effects with 12 independent markers showing antagonistic effects on the two traits under study. Ten independent SNPs were located within 11 protein coding and/or non-coding genes and twelve growth related QTLs. Examination of the GO slim category summaries of the candidate genes pinpointed *ACVR1* as a true pleiotropic gene with involvement in biological processes relevant to both growth and reproduction. Another plausible pleiotropic gene was *CACNA1H* that exhibited a GO slim category of reproduction and indirect relevance to growth biological processes. Based on literature-based functional evidence, the two aforementioned genes (*ACVR1* and *CACNA1H*) have widespread phenotypic effects on multiple systems (muscle, chondrocytes, bones and oocytes) and for this reason they can be considered as exemplars of horizontal pleiotropy.

5.2. Introduction

Reproductive traits in livestock species often show negative genetic correlation with growth related traits. This antagonistic relationship indicates that the two types of traits share common biological pathways and/or molecular mediators. In chickens, body weight (BW) and egg number (EN) exhibit a clear antagonism as reflected in their negative genetic correlation (r_g) with r_g estimates ranging from -0.05 to -0.55 [1,2,3]. Such a correlation imposes an important constraint in terms of selection response(s) of the individual traits and therefore identifying the genetic factors responsible for this antagonistic relationship is of great interest.

Many Genome-Wide Association Studies (GWAS) have pinpointed the existence of genetic loci harboring variants that are associated with multiple, sometimes seemingly distinct, complex disease or quantitative traits. Such associations are termed cross-phenotype (CP) associations [4] and are potential evidence for pleiotropy. The distinction between a CP association and pleiotropy is important to define. A CP association occurs when a genetic locus is associated with more than one trait regardless of the underlying cause for the observed association [5]. Pleiotropy occurs when a genetic locus truly affects more than one trait and is one possible underlying cause for an observed CP association [4]. In other words, pleiotropy can lead to a CP association but a CP association is not necessarily indicative of a pleiotropic variant.

In the special case where a genetic factor exhibits opposing effects on two different traits i.e. when the same alleles have beneficial effects on one trait and negative effects on a second trait, antagonistic pleiotropy (AP) exists [6,7]. AP was first proposed to explain the evolution of senescence where alleles with positive effects on survival or reproduction at early age decrease fitness in later life. In contrast to AP, synergistic pleiotropy [8,9] (SP) occurs when a

genetic variant simultaneously either increases or decreases performance in two different traits.

Based on the mode of action, pleiotropy between two traits can be distinguished in the following types [10]: 1) biological (or horizontal) pleiotropy when the variant affects directly or indirectly (through an intermediate phenotype) multiple phenotypes, 2) mediated (or vertical) pleiotropy when there is a causal relationship between phenotypes where phenotype B is mediated by phenotype A so as an indirect association occurs between the variant and multiple phenotypes and 3) spurious pleiotropy when the marker is falsely associated with multiple phenotypes due to bias, misclassification or linkage disequilibrium (LD).

As CP associations are indicative of potential pleiotropy, they have been widely explored via multivariate or univariate statistical approaches in GWAS. While multivariate approaches [11] allow for direct identification of CP associations, in the context of univariate analyses, detection of CP associations relies on aggregating results of single traits analyses via meta-analysis techniques [12].

In chicken, CP associations have already been discovered by GWAS for various traits such as daily feed intake and efficiency [13] and for egg weights at different ages [14,15]. To our knowledge, no GWAS has, so far, been reported with the aim to discover genetic variants associated with body weight (BW) and egg number (EN) in chickens.

Given the importance of the two traits from both a biological and an economic point of view, the present study aims to identify genetic variants and genes simultaneously affecting the two traits. To this end we conducted a bivariate GWAS to identify SNP signals associated with both traits. We then applied conditional and joint analysis of the SNP signals detected in the bivariate analysis to obtain independent CP associations. Finally, we examined the GO slim category summaries of the candidate genes underlying the independent CP associations to propose the most relevant genes implicated in the genetic control of the two traits. Our findings are expected to contribute to a better understanding of the genetic mechanism underlying the negative genetic correlation observed between growth and reproduction in broilers.

5.3. Material and methods

5.3.1. Data and quality control

Genotypic and phenotypic data were provided by Aviagen Ltd. The available data consisted of 2,992 female broilers from a grand-grandparent (GGP) commercial line with phenotypic records on body weight (BW) at 35 days of age (average=1822.7g, SD=143.6g) and number of eggs (EN) per hen collected from 28 to 50 weeks of age (average=132.4 eggs, SD=29.8 eggs). Animals were genotyped using the 600k Affymetrix HD SNP array [16] resulting in a total number of 544,927 autosomal SNPs. Quality control (QC) was performed first at a sample and second at a marker level. At a sample level, 406 animals were excluded due to call rate <0.99 and autosomal heterozygosity outside the 1.5 IQR (inter-quartile range: 0.013). At the marker level, 491,935 autosomal SNPs were excluded due to: call rate <0.95, minor allele frequency (MAF)<0.05 and LD r^2>0.70 within windows of 50 SNPs and increments of 5 SNPs. SNP pruning was employed to avoid for spurious CP associations due to strong short-range LD [4,10]. A total of 2,586 samples and 52,992 autosomal SNPs were retained for further analyses. All QC criteria were applied using the SNP & Variation Suite software (http://www.goldenhelix.com), except for LD- based SNP pruning which was performed using PLINK [17] and the '--indep-pairwise' command.

5.3.2. Univariate and bivariate association analyses

First, we performed univariate analyses to detect significant SNP associations for individual traits. The following additive single-locus mixed model was applied:

$$y = W\alpha + x\beta + u + e$$

where y is a n x 1 vector of phenotypic values of BW or EN for n female broilers, W is a n x c matrix of covariates i.e. fixed effects: hatch (36 classes), mating group (17 classes) and a column of 1s, α is a c x 1 vector of the corresponding coefficients including the intercept, x is a n x 1 vector of genotypes for the i^{th} SNP (codes as 0, 1, and 2 according to the number of copies of the minor allele), β is the additive fixed effect of the i^{th} SNP on BW or EN, u is a vector of random polygenic effects, and e is a vector of random residuals. The random effects were assumed to be normally distributed with zero means and the following covariance structure:

$$Var\begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} G\sigma_u^2 & 0 \\ 0 & I\sigma_e^2 \end{bmatrix}$$

where σ_u^2 and σ_e^2 are the polygenic and error variance components, I is the nxn identity matrix, and G is the n x n genetic relatedness matrix (estimated as centered genomic matrix[9,20]. We used the Wald test statistic $F_{Wald} = \frac{\hat{\beta}^2}{Var(\hat{\beta})}$ for each SNP to test for marker-trait associations by comparing the null hypothesis (H_0) where the marker effect sizes are zero, H_0: $\beta = 0$, against the alternative hypothesis H_1 : $\beta \neq 0$.

A bivariate linear mixed model was then applied to test for significant CP associations. The following bivariate linear mixed model was used:

$$Y = WA + x\beta^T + U + E$$

where Y is a n x 2 matrix of phenotypic values for the 2 traits (BW and EN) for n individuals, W is a n × c matrix of covariates i.e. fixed effects: hatch, mating group and a column of 1s, A is a c x 2 matrix of the corresponding coefficients including the intercept, x is a n x 1 vector of genotypes for the i^{th} SNP (codes as 0,1,2 according to the number of copies of the minor allele), β is the additive fixed effect of the i^{th} SNP for the 2 traits, U is a (n x 2) matrix of random effects with $U \sim MVN_{nx2}(0, G, V_g)$ where G is the n x n genetic relatedness matrix (estimated as centered genomic matrix [18,11]) and V_g is a 2 x 2 symmetric matrix of genetic variance component, and E is a (n x 2) matrix of random residuals with $E \sim MVN_{nx2}(0, I, V_e)$ where I is a n x n identity matrix, V_e is a 2 x 2 symmetric positive definite matrix of residual variance component and $MVN_{n\times2}(0, V_1, V_2)$ denotes the n x 2 matrix normal distribution with mean 0, row covariance matrix V_1 (n x n) and column covariance matrix V_2 (2 x 2). Association of each SNP with both traits of BW and EN was obtained by testing the null hypothesis H_0: $\beta = 0$, where 0 is a 2-vector of zeros, against the alternative hypothesis H_1: $\beta \neq 0$. The Wald test statistic was used to infer the significant CP associations. The genetic correlation (r_g) was also estimated between the two traits. All analyses were performed using the GEMMA [19] software (version 0.98.1).

For each association analysis, the estimation of the genomic inflation factor (λ) was used to assess potential systematic bias due to population structure or the analytical approach [20]. If the λ value was greater than 1, it provided evidence for some systematic bias [20]. If the λ value was less than or equal to 1, no adjustment was needed [21]. λ was estimated using the genetic analysis package (gap) in R (http://www.r-project.org/).

5.3.3. Multiple-testing correction

For each of the 3 association analyses (2 univariate and 1 bivariate), the Wald test p-values of the 52,992 SNPs were corrected for multiple comparisons using the false-discovery rate (FDR [22]) correction method in R (http://www.r-project.org/). SNPs with FDR p-values lower than 0.05 were considered as significant.

5.3.4. Selection of independent SNPs

Results obtained from univariate and bivariate analyses were further subject to stepwise conditional and joint (cojo) analysis using the '*cojo-slct'* option and the GCTA [23] tool to select independent SNPs. The cojo-GCTA analysis corrects β and p values of neighboring SNPs (in a sliding window of 10 Mb) based on the LD between the SNPs. This ensures that the SNP with the lowest p value is selected first for conditioning the effect on neighboring loci based on the LD between the neighboring SNPs and the selected SNP. Following LD-based correction of effect, all SNPs that remained significant under the default threshold p-value (5E-8) are run through the same process in a stepwise manner. This process identifies: (i) the number of independent SNP signals in a region and (ii) association signals due to the joint effect of several SNPs. To identify the independent CP associations, we used as input in the cojo-GCTA analysis the summary-level statistics obtained by the bivariate analysis. Specifically, the *b* estimates along with their standard errors were used to estimate t-values for the 52,992 SNPs and t-values were finally converted to p-values using R code (http://www.r-project.org/). For the independent CP associations, the genotypic means and the regression coefficients of the minor allele dose (β) on the phenotypic values of two traits were also estimated using the MANOVA procedure in SAS 9.3 (2012).

5.3.5. Effect prediction of the independent SNPs and identification of positional candidate genes and published QTLs

To predict the consequences of the CP significant SNPs on genes, transcripts, protein sequence and regulatory regions, the Variant Effect Predictor (VEP, https://www.ensembl.org/Tools/VEP, [24]) tool was employed with the latest release (Ensembl release 99, accessed: 28 April 2020).

Physical positions of SNPs were also obtained by the VEP tool using the GRCg6a assembly (https://www.ensembl.org/Gallus_gallus/Info/Annotation, GenBank Assembly ID: GCA_000002315.5, accessed: 28 April 2020). Furthermore, the VEP tool was used to search for positional candidate genes including the independent SNPs. When no positional candidate gene to a given CP association could be assigned, the closest gene to the marker was identified and nominated as candidate. To this end, both Ensembl and NCBI RefSeq transcript databases were used. The VEP tool was also used to assess if the significant SNPs were located within previously reported QTLs. VEP retrieves information for published QTLs via connections with Animal QTL database (Animal QTLdb) and Online Mendelian Inheritance in Animals (OMIA) database.

5.3.6. Functional profile of candidate genes and parent GO terms

Each annotated candidate gene was submitted to g:GOSt of g:Profiler web-based toolset (https://biit.cs.ut.ee/gprofiler, version e99_eg46_p14_f929183) [25] to obtain a full list of associated Gene Ontology Biological Process (GO BP) terms using the "All results" option and the *Gallus gallus* species. Then the CateGorizer tool (https://www.animalgenome.org/cgi-bin/util/gotreei, [26]) was used to group and categorize the GO BP terms into high level summaries. CateGorizer takes input of GO terms, performs step-wise classification against

one of the available GO slim methods (such as GO slim) and finally performs single counting for presence of a term within a GO slim category [26].

5.4. Results

5.4.1. Significant SNPs obtained from univariate and bivariate analyses

Estimations of the genomic inflation factors (univariate analyses: λ_{BW}=0.87, λ_{EN}= 0.96, bivariate analysis: λ=0.85) were less than 1 indicating the absence of population structure or artifacts in the present data. Furthermore, the genomic genetic correlation (r_g) between the two traits was estimated as high as -0.171 ± 0.153 (results not shown). Figure 1 shows the profiles of the SNP p-values (expressed as $-\log_{10}$ values) across the three GWAS. A more detailed view of the statistically significant SNPs is provided in Figure 2. A total number of 58 SNPs across 22 autosomes were found to reach genome-wide significance (FDR p-value <0.05) in the BW univariate analysis (Table S1). A closer inspection of these results revealed the presence of several neighboring SNP signals within distances of less than 500kb on 5 chromosomes (4, 11, 24, 25 and 27, see Figure 2 and Table S1).

No SNP was found to reach significance at the genome-wide level (FDR p-value <0.05) in the EN univariate analysis. However, 5 SNPs located in 2 autosomes (21 and 28) reached chromosome-wide significance (FDR p-value<0.05) (Figure 2 and Table S1) with 4 neighboring SNP signals detected in a region spanning 218 kb on GGA28 (Table S1).

Bivariate analysis identified 51 genome-wide significant CP associations across 21 autosomes (Table S2). As in the univariate cases, several neighboring CP associations were identified within distances of less than 500 kb on 6 chromosomes (4, 11, 14, 25, 27 and 28). Two CP associations (*rs315316434* and *rs316549515*) on GGA28 presented the shortest distance (23,062 bp, Table S2). Of CP associations, marker *rs315329074* (GGA27) was detected in BW univariate analysis with lowest Wald test p-value (6.18E-13, FDR p-value=3.28E-8) (Table S2).

As observed in Figure 2, two autosomes (GGA21 and GGA28) included significant SNPs detected in all three association analyses. Specifically, on GGA21, the CP marker *rs316810914*, which was also identified in BW univariate analysis, lied 659,553bp away from marker *rs316318083* that was detected in bivariate and EN univariate analyses (Table S2). On GGA28, CP marker *rs317501178* was also identified by BW univariate analysis. The latter marker was located 3,045,470 bp away from marker *rs317783777* that was identified by both the bivariate and the EN univariate analyses (Table S2). Such multiple neighboring SNP signals were indicative of long range LD justifying the need to apply a 'cojo' GCTA analysis to obtain independent SNPs.

The number of common SNPs between the bivariate and the two univariate analyses are shown in Figure 3 in form of a Venn diagram. Notably, no significant SNP was common across the three analyses (Figure 3, left Venn diagram). Nevertheless, 40 SNPs were common between the BW univariate analysis and the bivariate analysis and 4 SNPs were common between the EN univariate analysis and the bivariate analysis (Figure 3, left Venn diagram).

5.4.2. Independent SNP signals

A total number of 13 independent CP associations dispersed across 12 autosomes were selected after application of the stepwise conditional-joint analysis (Table 1). From these 13 SNPs, all except for one marker i.e. *rs317501178* (GGA28) were also associated with BW (Figure 3, right Venn diagram) while no independent SNP was associated with EN. Note that in accordance with 'cojo-GCTA' analysis, the markers retained presented the lowest p-values among their neighboring CP association signals. Table 2 shows the regression coefficients (β) of the minor allele dose for the 13 independent CP associations. In all but one marker

(*rs15608447),* estimated β coefficients displayed opposite directions implying antagonistic allelic action i.e. positive effects on one trait and negative effects on the other trait.

5.4.3. Effect prediction of the independent SNPs and identification of positional candidate genes and published QTLs

A total number of 14 positional candidate genes (of which 11 annotated genes) were identified as lying within or in close proximity to the 13 independent SNPs (Table 3). Specifically, ten SNPs were located within nine protein coding genes and two long non-coding (lnc) RNA genes (Table 3). Of these SNPs, one was a missense variant of gene *ZC3H18,* one a synonymous variant of gene *FCRL4,* eight were intron variants of annotated genes *SLAIN2, ACVR1, ST3GAL3, CACNA1H, VPS11, COPA and CACNB1)* and two were intron variants of two lncRNAs. Note that *rs15608447* (GGA4) was an intron variant of both *SLAIN2* and a lncRNA (*LOC107053243*) (Table 3). The rest three SNPs were downstream and upstream variants of one long non-coding RNA and two protein coding genes. Furthermore, a total number of 17 published QTLs were identified within the searched regions (Table 3). Of these, the majority i.e. twelve were related to growth (e.g. duodenum weight, body weight at 9 days, comb weight) while none was related to egg production. Notably, a microRNA (*gga-mir-6646-2*) was also identified at close proximity (1,002 bp) to marker *rs312758346* (GGA25).

5.4.4. GO term profiling of candidate genes and GO slim categories

The full list of GO BP terms identified by the g:GOSt tool per candidate gene for *Gallus gallus* is provided in Table S3. Of the 11 candidate genes, two genes i.e. *ZC3H18* and *FCRL4* did not exhibit any GO BP term (Table S3). With regard to the rest genes, the maximum number (n=351) of GO terms was attained for *ACVR1* and the minimum (n=29) for *ST3GAL3*. Table 4 presents the GO slim categories obtained by CateGOrizer per candidate gene. A detailed description of the GO slim categories per candidate gene is provided in Table S4. Of the nine candidate genes with GO BP terms, only *ACVR1* had GO slim categories relevant to both 'growth' (development, GO:0007275, embryonic development, GO:0009790, morphogenesis, GO:0009653) and 'reproduction' (GO:0000003) thus fully supporting its candidacy as a pleiotropic gene. Of the rest genes, only *CACNA1H* had a 'reproduction' GO slim category (GO:0000003) while its involvement in 'growth' related biological processes could be hypothesized via the GO:0008152 (metabolism) term (Table 4).

5.5. Discussion

The negative genomic estimate of genetic correlation (r_g=-0.17) between BW and EN obtained in the present study is supportive of previous results [1,2,3] confirming a clear antagonism between the two traits in broilers. This genome-wide genetic correlation estimate describes the average CP effects of all implicated causal loci without providing any detailed view on the exact underlying genetic mechanism (e.g. number of causal loci and patterns of pleiotropic effects produced by loci). To this end, we first performed a bivariate analysis to directly identify the individual genomic markers exerting CP effects. Since both pleiotropy and linkage disequilibrium can generate genetic correlation between traits, the next step was to avoid correlation effects due to short and/or long range LD between markers. To this end, first we pruned highly correlated markers to alleviate short range LD and then applied a stepwise conditional and joint analysis to mitigate long range LD. We anticipate that this combination has significantly reduced close linkage between markers so as the resulting CP associations could be attributed to purely QTL pleiotropic effects.

The next step was to consider as most plausible pleiotropic genes those including or being in close proximity to the independent SNPs while also displaying GO slim categories relevant to 'growth' and 'reproduction'. This search strategy resulted in identification of only one gene i.e. *ACVR1* (*serine/threonine-protein kinase receptor or activin receptor type I or activin a receptor, type 1)* fulfilling all the above criteria. *ACVR1* (also known as ALK2) encodes for a bone morphogenetic protein (BMP) type I receptor of the transforming growth factor-beta (TGF-β) superfamily. TGF-β superfamily genes are known to play fundamental role(s) in cell growth while also regulating several reproductive processes (i.e. follicular development, ovulation, oocyte competence, implantation, pregnancy, embryonic development and uterine development) [27]. In mouse embryos, *ACVR1* functions as type I receptor for BMP4 which is necessary for the formation of primordial germ cells (PGCs) [28] and promotes the growth of gonadal PGCs through a Smad1/4 signaling [29]. Furthermore, *ACVR1* has been reported to regulate reproduction via the BMP and anti-Müllerian hormone (AMH) signaling [30]. Specifically, *ACVR1* regulates folliculogenesis through acting as a type I receptor for AMH/MIS (anti-Müllerian hormone/Müllerian inhibiting substance). AMH inhibits FSH (follicle-stimulating hormone) sensitivity of growing follicles and thus could contribute to the accumulation of growing follicles in women with Polycystic ovary syndrome (PCOS) [30]. In chickens, both embryonic ovaries express AMH although ovarian estrogen (from the left ovary) protects the left chick Mullerian duct from AMH action and therefore permits development of the left oviduct [31]. In the same species, AMH is required for the urogenital development and germ cell migration [32], is presented in early follicle development and is expressed in small follicles [33]. So far, in chicken *ACVR1* has been proposed as a positional candidate gene for body weight [34], has a regulatory role in osteogenesis and chondrogenesis during skeletal development [35] and is expressed within the chicken granulosa and thecal layers during ovarian follicle development [36].

Another gene that was associated with the GO slim category of reproduction (GO:0000003) was *CACNA1H (Calcium Voltage-Gated Channel Subunit Alpha1 H*) while its involvement to growth or development processes could be speculated via the general GO slim category of metabolism (GO:0008152). According to QuickGo, the GO:0008152 term describes '*chemical reactions and pathways, including anabolism and catabolism, by which living organisms transform chemical substances. Metabolic processes typically transform small molecules, but also include macromolecular processes such as DNA repair and replication, and protein synthesis and degradation'*. To evaluate the possibilities of *CACNA1H* being another plausible pleiotropic gene, a thorough search of the relevant literature with regard to its functional role was performed. *CACNA1H* (also known as *Cav3.2*) encodes for Cav3.2 channel, a member of the voltage-gated calcium channel family. This gene participates in the T-type Ca^{2+} channels which contribute to signal transduction pathways regulating protein synthesis, cell differentiation, growth, and proliferation[37] that are mainly expressed during embryonic development [37]. Particularly, they are involved in the early stages of muscle differentiation in mice [38] and humans [39]. Cav3.2$^{-/-}$ null mutant female mice presented decreased body weight [40] and reduced litter size [41]. Moreover, Cav3.2 facilitates the Ca^{2+} influx in mouse oocytes and eggs to maintain Ca^{2+} homeostasis during oocyte maturation and *post* fertilization [41]. Cav3.2 may also have a role in reproduction via its control in gonadal endocrine function. Specifically, Cav3.2 participates in the secretion of the Gonadotropin-releasing hormone (GnRH) the pulsatile secretion of which determines the pattern of secretion of follicle stimulating hormone (FSH) and luteinising hormone (LH). Both hormones are known to regulate the endocrine function and gamete maturation of gonads [42]. Persistent, rapid GnRH pulses increase LH which in return stimulates secretion of 17β-estradiol (E2). E2 further upregulates the expression of the T-type Ca^{2+} channel subunits [43]. So far, *CACNA1H* has only been associated with egg quality [44] and body weight [45], in chickens.

Based on the aforementioned functional evidence, it appears that both genes (*ACVR1* and *CACNA1H*) have widespread phenotypic effects on multiple systems (muscle, chondrocytes,

bones and oocytes) and for this reason we hypothesize that they are exemplars of horizontal pleiotropy. Nevertheless, Jordan et al [46] have offered an alternate view of genes' independent effects on multiple traits. Specifically, the authors hypothesized that the pervasive horizontal pleiotropy observed in polygenic traits is, on some level, a logical consequence of widespread polygenicity as 'the more loci are associated with each trait, the more chances there are for associations with multiple traits to overlap'.

Another interesting finding obtained in the present study was the presence of three long non-coding genes (lncRNAs) and a short non-coding gene (*gga-mir-6646-2*) within or close proximity to independent markers. LncRNAs are RNA transcripts greater than 200bp in length that are localized in nucleus and cytoplasm. Nuclear lncRNAs have been reported to act both in-cis and in-trans whereby in-cis acting lncRNAs influence the expression of nearby genes [47]. Although lncRNAs were traditionally thought that they could not encode proteins, some studies found that lncRNAs can encode short peptides [47,48]. LncRNAs can function as molecular decoy for proteins or sponges for other transcripts (such as miRNAs) [47]. They can also regulate a wide range of functions such as epigenetic modification, transcription and post-transcription while playing a key role in tissue development, muscle contraction/relaxation [47] and myogenesis [48]. In chickens, lncRNAs have been reported to regulate muscle development, lipid metabolism, egg production and disease resistance [49]. On the other hand, miRNAs (19-22 nucleotides long) can mediate almost any biological function depending on their targets [47]. As protein-coding genes are regulated by one or more miRNAs, a critical step is to identify genes targeted by the miRNA(s). Several studies used computational tools to predict target candidate genes prior their experimental validation [50]. Following this rationale, we found that *gga-mir-6646-2* was associated with 294 predicted target genes (Table S5) for the species via the miRDB (http://mirdb.org). Although *gga-mir-6646-2* had no documented functions in literature, miRNAs have been reported to be involved in cell growth, cell proliferation, myogenesis and egg production in chickens [51,52].

To conclude, present findings provide a novel insight in the genetic mechanism underlying antagonistic interplay between growth and reproduction in broilers. Further studies are warranted to experimentally validate the functional significance of individual candidate SNPs by using precise mammalian genome editing techniques such as CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats/CRISPR-associated 9) in animal models. CRISPR/Cas9-mediated gene editing may provide evidence which SNP(s) affect the transcriptional activity of the single or nearby genes involved in traits expression (e.g. [53]).

5.6. References

[1] Sang BD, Kong HS, Kim HK, Choi CH, Kim SD, Cho YM, et al. Estimation of genetic parameters for economic traits in Korean native chickens. Asian-Australasian J Anim Sci. Asian-Australasian Association of Animal Production Societies; 2006;19: 319–323. doi:10.5713/ajas.2006.319

[2] Lwelamira J, Kifaro GC, Gwakisa PS. Genetic parameters for body weights, egg traits and antibody response against Newcastle Disease Virus (NDV) vaccine among two Tanzania chicken ecotypes. Trop Anim Health Prod. 2009;41: 51–59. doi:10.1007/s11250-008-9153-2

[3] Tongsiri S, Jeyaruban GM, Hermesch S, van der Werf JHJ, Li L, Chormai T. Genetic parameters and inbreeding effects for production traits of Thai native chickens. Asian-Australasian J Anim Sci. Asian Australasian Association of Animal Production Societies;

2019;32: 930–938. doi:10.5713/ajas.18.0690

[4] Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: Challenges and strategies. Nature Reviews Genetics. 2013. pp. 483–495. doi:10.1038/nrg3461

[5] Tyler AL, Crawford DC, Pendergrass SA. The detection and characterization of pleiotropy: discovery, progress, and promise. Brief Bioinform. 2016;17: 13–22. doi:10.1093/bib/bbv050

[6] Medawar PB. An unsolved problem of biology. London:H.K. Lewis; 1952.

[7] Williams GC. Pleiotropy, Natural Selection, and the Evolution of Senescence. Evolution (N Y). JSTOR; 1957;11: 398. doi:10.2307/2406060

[8] Ostrowski EA, Rozen DE, Lenski RE. Pleiotropic effects of beneficial mutations in Escherichia coli. Evolution. 2005;59: 2343–52. Available: http://www.ncbi.nlm.nih.gov/pubmed/16396175

[9] Leiby N, Marx CJ. Metabolic Erosion Primarily Through Mutation Accumulation, and Not Tradeoffs, Drives Limited Evolution of Substrate Specificity in Escherichia coli. PLoS Biol. Public Library of Science; 2014;12. doi:10.1371/journal.pbio.1001789

[10] van Rheenen W, Peyrot WJ, Schork AJ, Lee SH, Wray NR. Genetic correlations of polygenic disease traits: from theory to practice. Nature Reviews Genetics. Nature Publishing Group; 2019. doi:10.1038/s41576-019-0137-z

[11] Zhou X, Stephens M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat Methods. NIH Public Access; 2014;11: 407–9. doi:10.1038/nmeth.2848

[12] Li X, Zhu X. Cross-phenotype association analysis using summary statistics from GWAS. Methods in Molecular Biology. Humana Press Inc.; 2017. pp. 455–467. doi:10.1007/978-1-4939-7274-6_22

[13] Yuan J, Wang K, Yi G, Ma M, Dou T, Sun C, et al. Genome-wide association studies for feed intake and efficiency in two laying periods of chickens. Genet Sel Evol. BioMed Central; 2015;47: 82. doi:10.1186/s12711-015-0161-1

[14] Liu Z, Sun C, Yan Y, Li G, Wu G, Liu A, et al. Genome-Wide Association Analysis of Age-Dependent Egg Weights in Chickens. Front Genet. Frontiers; 2018;9: 128. doi:10.3389/fgene.2018.00128

[15] Yi G, Shen M, Yuan J, Sun C, Duan Z, Qu L, et al. Genome-wide association study dissects genetic architecture underlying longitudinal egg weights in chickens. BMC Genomics. BioMed Central; 2015;16: 746. doi:10.1186/s12864-015-1945-y

[16] Kranis A, Gheyas AA, Boschiero C, Turner F, Yu L, Smith S, et al. Development of a high density 600K SNP genotyping array for chicken. BMC Genomics. BioMed Central; 2013;14: 59. doi:10.1186/1471-2164-14-59

[17] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. Cell Press; 2007;81: 559–575. doi:10.1086/519795

[18] Zhou X, Carbonetto P, Stephens M. Polygenic Modeling with Bayesian Sparse Linear Mixed

Models. PLoS Genet. 2013;9. doi:10.1371/journal.pgen.1003264

[19] Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012;44: 821–824. doi:10.1038/ng.2310

[20] Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, Willer CJ, et al. Genomic inflation factors under polygenic inheritance. Eur J Hum Genet. Nature Publishing Group; 2011;19: 807–12. doi:10.1038/ejhg.2011.39

[21] Hinrichs AL, Larkin EK, Suarez BK. Population stratification and patterns of linkage disequilibrium. Genetic Epidemiology. 2009. doi:10.1002/gepi.20478

[22] Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing [Internet]. Journal of the Royal Statistical Society. Series B (Methodological). WileyRoyal Statistical Society; 1995. pp. 289–300. doi:10.2307/2346101

[23] Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88: 76–82. doi:10.1016/j.ajhg.2010.11.011

[24] McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. BioMed Central; 2016;17: 122. doi:10.1186/s13059-016-0974-4

[25] Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. 2019;47: 191–198. doi:10.1093/nar/gkz369

[26] Hu, Zhiliang & Bao, Jie & Reecy J. CateGOrizer: A Web-Based Program to Batch Analyze Gene Ontology Classification Categories. In: Online Journal of Bioinformatics. 9. 108-112. [Internet]. 2008 [cited 13 May 2020]. Available: https://www.researchgate.net/publication/284674352_CateGOrizer_A_Web-Based_Program_to_Batch_Analyze_Gene_Ontology_Classification_Categories

[27] Li Q. Transforming growth factor β signaling in uterine development and function. Journal of Animal Science and Biotechnology. BioMed Central Ltd.; 2014. doi:10.1186/2049-1891-5-52

[28] Chuva De Sousa Lopes SM, Roelen BAJ, Monteiro RM, Emmens R, Lin HY, Li E, et al. BMP signaling mediated by ALK2 in the visceral endoderm is necessary for the generation of primordial germ cells in the mouse embryo. Genes Dev. Cold Spring Harbor Laboratory Press; 2004;18: 1838–1849. doi:10.1101/gad.294004

[29] Pesce M, Gioia Klinger F, De Felici M. Derivation in culture of primordial germ cells from cells of the mouse epiblast: Phenotypic induction and growth control by Bmp4 signalling. Mech Dev. 2002;112: 15–24. doi:10.1016/S0925-4773(01)00624-4

[30] Valer, Sánchez-de-Diego, Pimenta-Lopes, Rosa, Ventura. ACVR1 Function in Health and Disease. Cells. 2019;8: 1366. doi:10.3390/cells8111366

[31] Johnson PA, Kent TR, Urick ME, Trevino LS, Giles JR. Expression of anti-Mullerian hormone in hens selected for different ovulation rates. Reproduction. 2009;137: 857–863. doi:10.1530/REP-08-0406

[32] Lambeth LS, Ayers K, Cutting AD, Doran TJ, Sinclair AH, Smith CA. Anti-Müllerian Hormone Is Required for Chicken Embryonic Urogenital System Growth but Not Sexual Differentiation1. Biol Reprod. 2015;93. doi:10.1095/biolreprod.115.131664

[33] Johnson PA, Kent TR, Urick ME, Giles JR. Expression and Regulation of Anti-Mullerian Hormone in an Oviparous Species, the Hen1. Biol Reprod. 2008;78: 13–19. doi:10.1095/biolreprod.107.061879

[34] Moreira GCM, Poleti MD, Pértille F, Boschiero C, Cesar ASM, Godoy TF, et al. Unraveling genomic associations with feed efficiency and body weight traits in chickens through an integrative approach. BMC Genet. BioMed Central Ltd.; 2019;20. doi:10.1186/s12863-019-0783-3

[35] Lin S, Svoboda KKH, Feng JQ, Jiang X. The biological function of type i receptors of bone morphogenetic protein in bone. Bone Res. Sichuan University; 2016;4. doi:10.1038/boneres.2016.5

[36] Lovell TM, Knight PG, Gladwell RT. Differential expression of mRNAs encoding the putative inhibin co-receptor (betaglycan) and activin type-I and type-II receptors in preovulatory and prehierarchical follicles of the laying hen ovary. J Endocrinol. 2006;188: 241–249. doi:10.1677/joe.1.06525

[37] Ertel SI, Ertel EA, Clozel JP. T-type Ca2+ channels and pharmacological blockade: Potential pathophysiological relevance. Cardiovasc Drugs Ther. 1997;11: 723–739. doi:10.1023/A:1007706022381

[38] Berthier C, Monteil A, Lory P, Strube C. α1H mRNA in single skeletal muscle fibres accounts for T-type calcium current transient expression during fetal development in mice. J Physiol. 2002;539: 681–691. doi:10.1113/jphysiol.2001.013246

[39] Carter MT, McMillan HJ, Tomin A, Weiss N. Compound heterozygous CACNA1H mutations associated with severe congenital amyotrophy. Channels (Austin). NLM (Medline); 2019;13: 153–161. doi:10.1080/19336950.2019.1614415

[40] Lundt A, Seidel R, Soós J, Henseler C, Müller R, Bakki M, et al. Cav3.2 T-Type Calcium Channels Are Physiologically Mandatory for the Auditory System. Neuroscience. Elsevier Ltd; 2019;409: 81–100. doi:10.1016/j.neuroscience.2019.04.024

[41] Bernhardt ML, Zhang Y, Erxleben CF, Padilla-Banks E, McDonough CE, Miao YL, et al. CaV3.2 T-type channels mediate Ca2+ entry during oocyte maturation and following fertilization. J Cell Sci. Company of Biologists Ltd; 2015;128: 4442–4452. doi:10.1242/jcs.180026

[42] Marques P, Skorupskaite K, George JT, Anderson RA. Physiology of GNRH and Gonadotropin Secretion [Internet]. Endotext. 2018. Available: http://www.ncbi.nlm.nih.gov/pubmed/25905297

[43] Kelly MJ, Zhang C, Qiu J, Rønnekleiv OK. Pacemaking kisspeptin neurons. Exp Physiol. 2013;98: 1535–1543. doi:10.1113/expphysiol.2013.074559

[44] Zhang Q, Zhu F, Liu L, Zheng CW, Wang DH, Hou ZC, et al. Integrating Transcriptome and Genome Re-Sequencing Data to Identify Key Genes and Mutations Affecting Chicken Eggshell Qualities. Leung FCC, editor. PLoS One. 2015;10: e0125890. doi:10.1371/journal.pone.0125890

[45] Mebratie W, Reyer H, Wimmers K, Bovenhuis H, Jensen J. Genome wide association study of body weight and feed efficiency traits in a commercial broiler chicken population, a re-visitation. Sci Rep. Nature Publishing Group; 2019;9. doi:10.1038/s41598-018-37216-z

[46] Jordan DM, Verbanck M, Do R. HOPS: A quantitative score reveals pervasive horizontal pleiotropy in human genetic variation is driven by extreme polygenicity of human traits and diseases. Genome Biol. BioMed Central Ltd.; 2019;20. doi:10.1186/s13059-019-1844-7

[47] Pasut A, Matsumoto A, Clohessy JG, Pandolfi PP. The pleiotropic role of non-coding genes in development and cancer. Current Opinion in Cell Biology. Elsevier Ltd; 2016. pp. 104–113. doi:10.1016/j.ceb.2016.10.005

[48] Militello G, Hosen MR, Ponomareva Y, Gellert P, Weirick T, John D, et al. A novel long non-coding RNA Myolinc regulates myogenesis through TDP-43 and Filip1. J Mol Cell Biol. Oxford University Press; 2018;10: 102–117. doi:10.1093/jmcb/mjy025

[49] Li Z, Ren T, Li W, Han R. Regulatory Mechanism and Application of lncRNAs in Poultry. Poultry [Working Title]. IntechOpen; 2019. doi:10.5772/intechopen.83800

[50] Liu W, Wang X. Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data. Genome Biol. BioMed Central Ltd.; 2019;20: 18. doi:10.1186/s13059-019-1629-z

[51] Wu N, Zhu Q, Chen B, Gao J, Xu Z, Li D. High-throughput sequencing of pituitary and hypothalamic microRNA transcriptome associated with high rate of egg production. BMC Genomics. BioMed Central; 2017;18: 255. doi:10.1186/s12864-017-3644-3

[52] Wu H, Fan F, Liang C, Zhou Y, Qiao X, Sun Y, et al. Variants of pri-miR-26a-5p polymorphisms are associated with values for chicken egg production variables and affects abundance of mature miRNA. Anim Reprod Sci. Elsevier B.V.; 2019;201: 93–101. doi:10.1016/j.anireprosci.2019.01.002

[53] Wu S, Zhang M, Yang X, Peng F, Zhang J, Tan J, et al. Genome-wide association studies and CRISPR/Cas9-mediated gene editing identify regulatory variants influencing eyebrow thickness in humans. Cooper GM, editor. PLOS Genet. Public Library of Science; 2018;14: e1007640. doi:10.1371/journal.pgen.1007640

[54] Anand L. chromoMap: An R package for Interactive Visualization and Annotation of Chromosomes. bioRxiv. Cold Spring Harbor Laboratory; 2019; 605600. doi:10.1101/605600

[55] Oliveros JC. VENNY. An interactive tool for comparing lists with Venn Diagrams. [Internet]. 2015. Available: http://bioinfogp.cnnb.csic.es/tools/venny/index.html

Tables and Figures of Chapter 5

Table 1. Chromosome position along with Wald and FDR p-values for 13 independent SNPs associated with body weight and egg number in broilers.

SNP	GGA	Position (bp)[1]	Minor allele	MAF	Wald test p-value	FDR p-value
rs13653309	1	52,997,158	A	0.13	2.26E-08	0.0001416
rs313332188	3	99,991,484	B	0.45	5.89E-10	7.804E-06
rs15608447	4	66,459,916	A	0.45	4.34E-08	0.00023
rs313879964	7	36,286,374	A	0.12	9.82E-09	7.431E-05
rs316603825	8	20,680,268	B	0.08	9.46E-09	7.431E-05
rs318098582	11	18,407,493	A	0.13	0.000000024	0.0001416
rs317631529	14	5,738,298	B	0.33	7.69E-08	0.0003396
rs317414603	20	6,729,013	A	0.41	2.51E-09	2.66E-05
rs14291881	24	150,829	A	0.30	4.28E-10	7.563E-06
rs315023079	25	2,292,569	A	0.13	3.48E-10	7.563E-06
rs312758346	25	3,770,684	A	0.44	0.0000377	0.0437699
rs315329074	27	6,920,352	B	0.18	6.18E-13	3.276E-08
rs317501178	28	874,035	A	0.23	0.000000347	0.0013149

[1]Positions are based on GRCg6a assembly

Table 2. Genotypic means (coded as 0, 1, 2 for the dose of the minor allele), regression coefficients (β) for body weight and egg number in female broilers for 13 independent SNPs.

Marker	Genotype		BW		EN	
	Class	n	Mean ± SD	β ± SE	Mean ± SD	β ± SE
rs13653309	0	2197	183.7 ± 13.8a	-6.70 ± 0.42***	131.9 ± 29.9a	0.67 ± 0.94^{NS}
	1	128	187.5 ± 13.6b		135.3 ± 25.6a	
	2	261	171.8 ± 14.8c		135.6 ± 27.7a	
rs313332188	0	1055	183.5 ± 14.9a	-1.14 ± 0.33***	133.2 ± 27.7a	0.63 ± 0.72^{NS}
	1	732	181.0 ± 13.9b		132.3 ± 31.5a	
	2	799	181.7 ± 13.9b		131.7 ± 30.0a	
rs15608447	0	749	180.6 ± 14.4a	1.76 ± 0.39***	131.0 ± 29.8a	1.25 ± 0.84N^{S}
	1	1346	182.5 ± 14.2b		132.6 ± 29.3a	
	2	491	184.1 ± 14.5b		134.0 ± 29.9a	
rs313879964	0	2205	183.7 ± 13.8a	-6.65 ± 0.41***	132.2 ± 29.6a	0.42 ± 0.94NS
	1	119	177.7 ± 12.9b		129.5 ± 30.6a	
	2	262	171.9 ± 14.8c		135.8 ± 28.0a	
rs316603825	0	2366	181.5 ± 14.2a	5.25 ± 0.48***	132.9 ± 29.5a	-3.52 ± 1.06***
	1	9	189.9 ± 8.5b		134.4 ± 24.3a	
	2	211	190.9 ± 13.7b		127.8 ± 29.8b	
rs318098582	0	2202	183.8 ± 13.8a	-6.59 ± 0.41***	131.9 ± 29.9a	1.02 ± 0.92NS
	1	108	177.3 ± 12.9b		135.9 ± 27.3a	
	2	276	172.2 ± 14.5c		135.9 ± 27.6a	
rs317631529	0	1561	180.3 ± 13.9a	3.53 ± 0.30***	132.7 ± 29.2a	-1.00 ± 0.67NS
	1	361	183.2 ± 13.7b		135.5 ± 29.0a	
	2	664	185.4 ± 14.8c		130.4 ± 30.5b	

rs317414603	0	1422	184.6 ± 14.3a	-4.60 ± 0.31***	130.1 ± 30.6a	0.96 ± 0.71NS
	1	207	180.5 ± 14.3b		133.4 ± 29.1b	
	2	957	179.2 ± 13.9b		135.7 ± 27.7b	
rs14291881	0	1669	184.7 ± 13.9a	-4.45 ± 0.31***	131.7 ± 29.9a	0.89 ± 0.69NS
	1	300	179.3 ± 13.7b		130.2 ± 29.5a	
	2	617	177.0 ± 14.3c		135.8 ± 28.4b	
rs315023079	0	2190	183.7 ± 13.9a	-6.65 ± 0.41***	132.0 ± 29.9a	0.28 ± 0.92NS
	1	111	178.0 ± 13.2b		134.8 ± 29.3a	
	2	285	172.9 ± 14.5c		135.3 ± 26.8a	
rs312758346	0	1032	181.7 ± 14.0a	-0.020 ± 0.33NS	132.0 ± 30.8a	0.61 ± 0.71NS
	1	849	183.5 ± 14.9b		132.4 ± 28.5a	
	2	705	181.6 ± 14.2a		133.2 ± 29.0a	
rs315329074	0	2095	180.6 ± 13.8a	5.19 ± 0.35***	132.5 ± 29.5a	-1.64 ± 0.79NS
	1	42	181.9 ± 14.6a		129.0 ± 31.1a	
	2	229	189.9 ± 14.3b		132.4 ± 29.4a	
rs317501178	0	1922	180.7 ± 13.9a	3.93 ± 0.32***	132.5 ± 29.4a	-0.48 ± 0.70NS
	1	121	179.7 ± 13.3a		135.1 ± 24.2a	
	2	543	188.2 ± 14.8b		131.8 ± 31.1a	

[a,b,c] means with different letters as superscipts are significantly different ($p<0.05$)

***$p<0.001$, **$p<0.01$, *$p<0.05$, NS: non significant

Table 3. Positional candidate genes and reported QTLs for the independent SNPs.

SNP	GGA	Consequence (variant)	Positional candidate gene (gene ID)*	Gene description (biotype)	Start position - end position of gene (bp)	Minimum distance from SNP (bp)	QTL(s)
rs13653309	1	intron	*ENSGALG00000051610*	*novel gene (lncRNA)*	52,996,174-53,012,023	0	Duodenum weight (ID: 96627), Body weight 9 days (ID: 96626)
rs313332188	3	downstream gene	*LOC107051696 (107051696)*	*uncharacterized LOC107051696 (lncRNA)*	99,953,839-99,978,579	12905	Comb weight (ID: 127114)
rs15608447	4	intron	*SLAIN2 (ENSGALG00000014115)*	*SLAIN motif family member 2 (protein coding)*	66,459,339-66,485,331	0	none
		intron	*LOC107053243 (107053243)*	*uncharacterized LOC107053243 (lncRNA)*	66,450,553-66,472,818	0	
rs313879964	7	intron	*ACVR1 (ENSGALG00000037301)*	*activin A receptor type 1 (protein coding)*	36,257,915-36,304,135	0	none
rs316603825	8	intron	*ST3GAL3 (ENSGALG00000010083)*	*ST3 beta-galactoside alpha-2,3-sialyltransferase 3 (protein coding)*	20,538,514-20,684,008	0	Feather colour extended black (ID: 157212), Feathered feet (ID: 127123), Body weight 21 days (ID: 95408)
rs318098582	11	missense	*ZC3H18 (ENSGALG00000006118)*	*zinc finger CCCH-type containing 18 (protein coding)*	18,379,252-18,412,636	0	none
rs317631529	14	intron	*CACNA1H (ENSGALG00000005215)*	*calcium voltage-gated channel subunit alpha1 H (protein coding)*	5,667,870-5,837,019	0	Body weight 36 days (ID: 64519), Wattles length (ID: 127121)
rs317414603	20	upstream gene	*ZNFX1 (ENSGALG00000004859)*	*zinc finger NFX1-type containing 1 (protein coding)*	6,715,285-6,727,531	1482	none
rs14291881	24	intron	*VPS11 (ENSGALG00000029536)*	*VPS11, CORVET/HOPS core subunit (protein coding)*	77,314-151,086	0	none
rs315023079	25	intron	*COPA (ENSGALG00000009153)*	*coatomer protein complex subunit alpha (protein coding)*	2,291,057-2,407,537	0	Ileum weight (ID: 96668)
rs312758346	25	synonymous	*FCRL4 (ENSGALG00000010507)*	*Fc receptor like 4 (protein coding)*	3,769,137-3,773,541	0	none

		downstream gene variant	*gga-mir-6646-2 (ENSGALG00000027697)*	*gga-mir-6646-2 (miRNA)*	3,769,573-3,769,682	1002	
rs315329074	27	intron	*CACNB1 (ENSGALG00000025788)*	*calcium voltage-gated channel auxiliary subunit beta 1 (protein coding)*	6,913,922-6,925,806	0	Body weight hatch (ID: 135726), Comb weight (ID: 127127), Femur bone mineral content (ID: 130479), Femur weight (ID: 130480), Proventriculus weight (ID: 96672), Wattles weight (ID: 127120)
rs317501178	28	downstream gene	*RANBP3 (ENSGALG00000000577)*	*RAN binding protein 3 (protein coding)*	832,528-873,641	394	Feather-crested head (ID: 127113), Excreta water content (ID: 96625)

*Note that Ensembl and RefSeq transcript databases were used in VEP tool to identify genes. Gene IDs refer to Ensembl gene IDs ('ENSGAL_') or NCBI gene IDs (numerical).

Table 4. Gene Ontology (GO) slim categories obtained by CateGorizer per candidate gene.

		Gene									
GO Slim category ID	Definition	*SLAIN2*	*ACVR1*	*ST3GAL3*	*CACNA1H*	*ZNFX1*	*VPS11*	*COPA*	*CACNB1*	*RANBP3*	Total
GO:0009058	biosynthesis		✓	✓	✓	✓					4
GO:0009056	catabolism						✓				1
GO:0007154	cell communication		✓		✓				✓		3
GO:0007049	cell cycle		✓								1
GO:0008219	cell death		✓								1
GO:0030154	cell differentiation		✓								1
GO:0016043	cell organization and biogenesis	✓				✓	✓		✓		4
GO:0007267	cell-cell signaling								✓		1
GO:0007010	cytoskeleton organization and biogenesis	✓									1
GO:0007275	development		✓								1
GO:0009790	embryonic development		✓								1
GO:0006811	ion transport				✓				✓		2
GO:0006629	lipid metabolism				✓						1
GO:0008152	metabolism		✓	✓	✓	✓	✓				5
GO:0009653	morphogenesis		✓								1
GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolism		✓			✓					2
GO:0006996	organelle organization and biogenesis	✓				✓	✓				3
GO:0019538	protein metabolism		✓	✓							2
GO:0006464	protein modification		✓	✓							2
GO:0015031	protein transport						✓	✓		✓	3
GO:0040029	regulation of gene expression, epigenetic					✓					1
GO:0000003	reproduction		✓		✓						2

GO:0009719	response to endogenous stimulus		✓		✓				✓		3
GO:0006950	response to stress		✓								1
GO:0007165	signal transduction		✓								1
GO:0006810	transport				✓		✓	✓	✓	✓	5
	Total	3	16	4	8	6	6	2	6	2	53

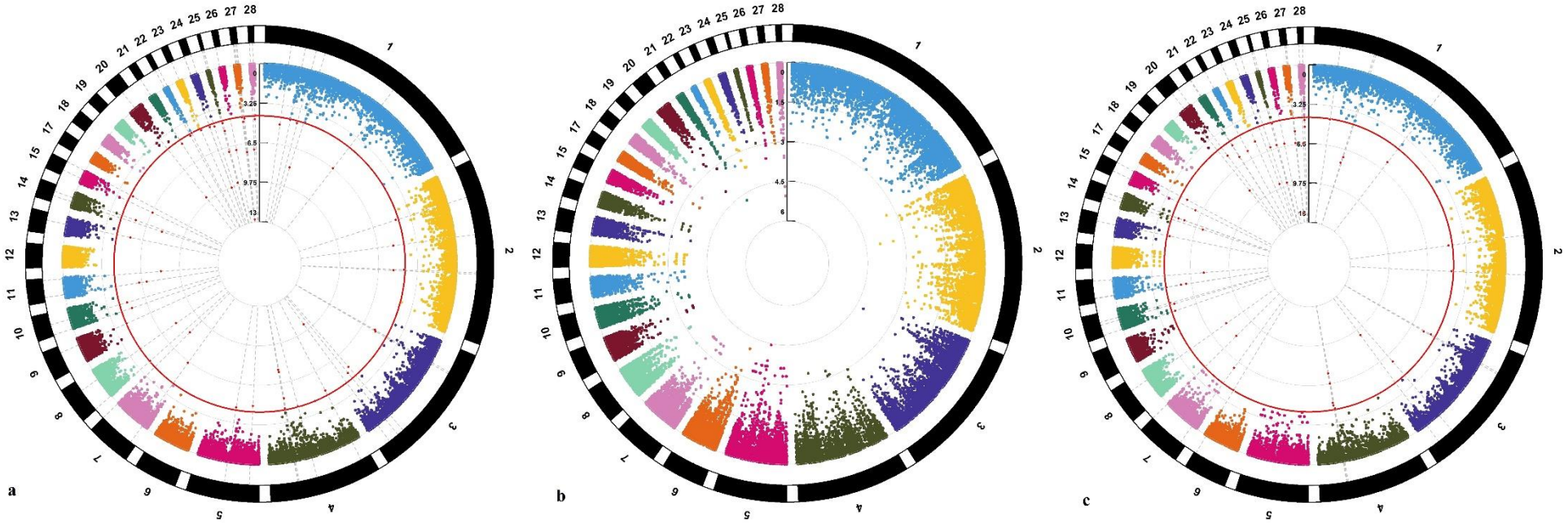


Figure 1. Circular Manhattan plots showing the $-\log_{10}$(p-values) of SNPs across the 28 autosomal chromosomes for body weight (BW) (a), egg number (EN) (b) and both traits (c), respectively. Red dots in the (a) and (c) Manhattan plots denote genome-wide significant SNPs. Plots were constructed using the CMplot package (https://github.com/YinLiLin/R-CMplot) in R (http://www.r-project.org/).

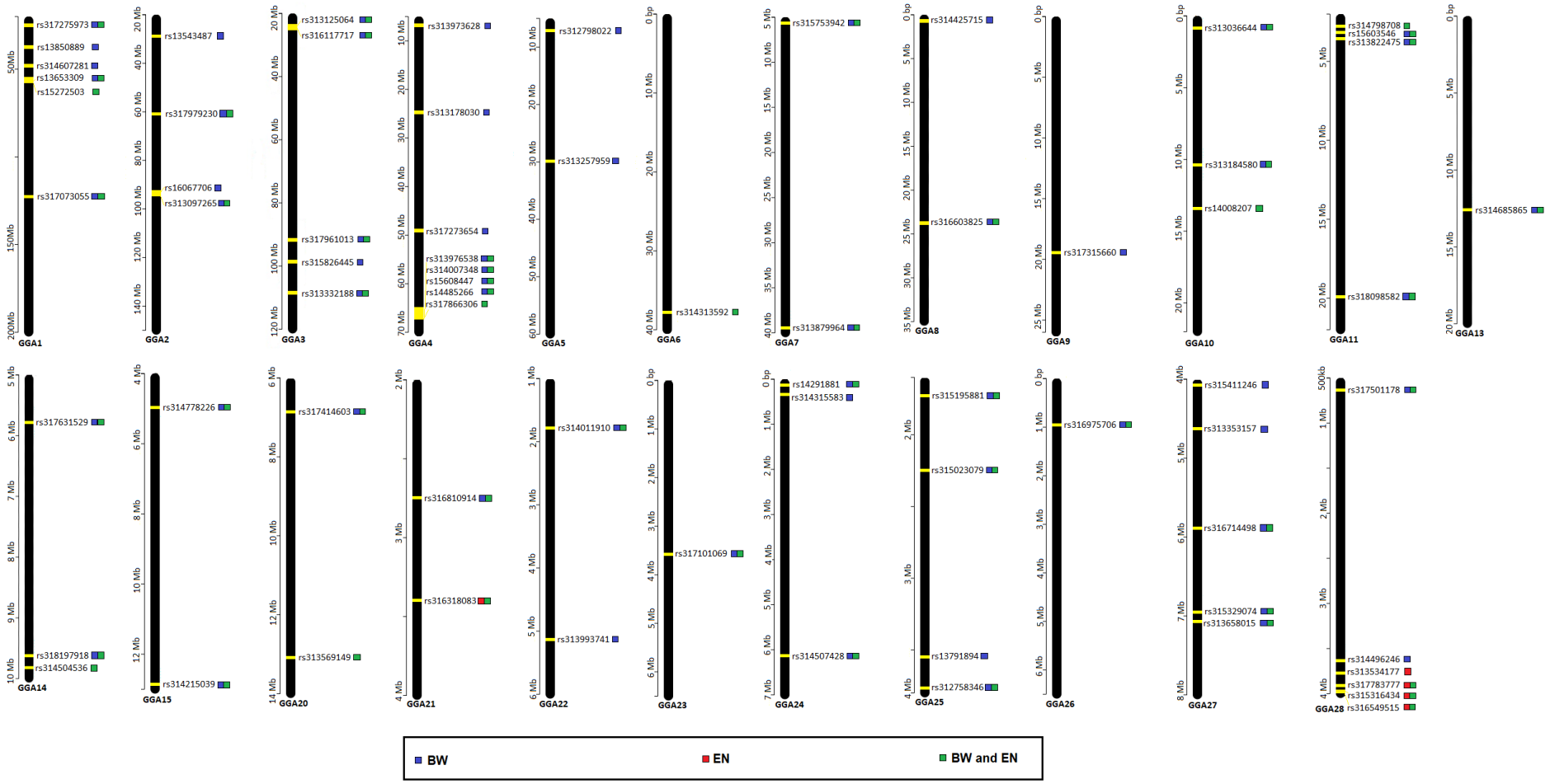


Figure 2. Graphic depiction of the significant SNPs detected by univariate and bivariate analyses across the genome. Yellow color denotes the SNP position(s) per autosome (black color). Each SNP position is linked to at least one colored square representing the associated trait (s) (i.e. body weight (BW): blue square, egg number (EN): red square and both traits: green square). Plot was constructed using the chromoMap package [54] in R(http://www.r-project.org/).

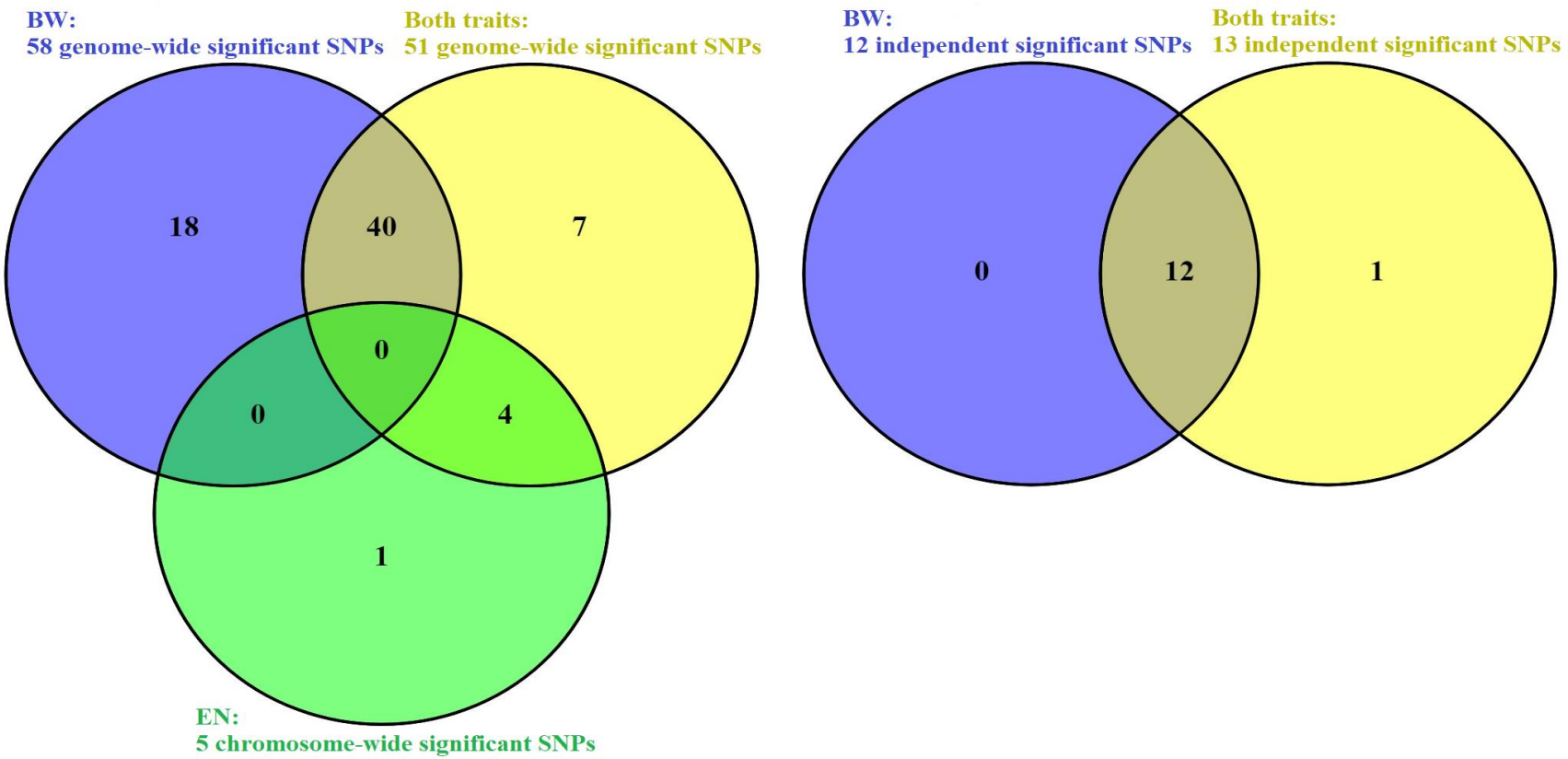


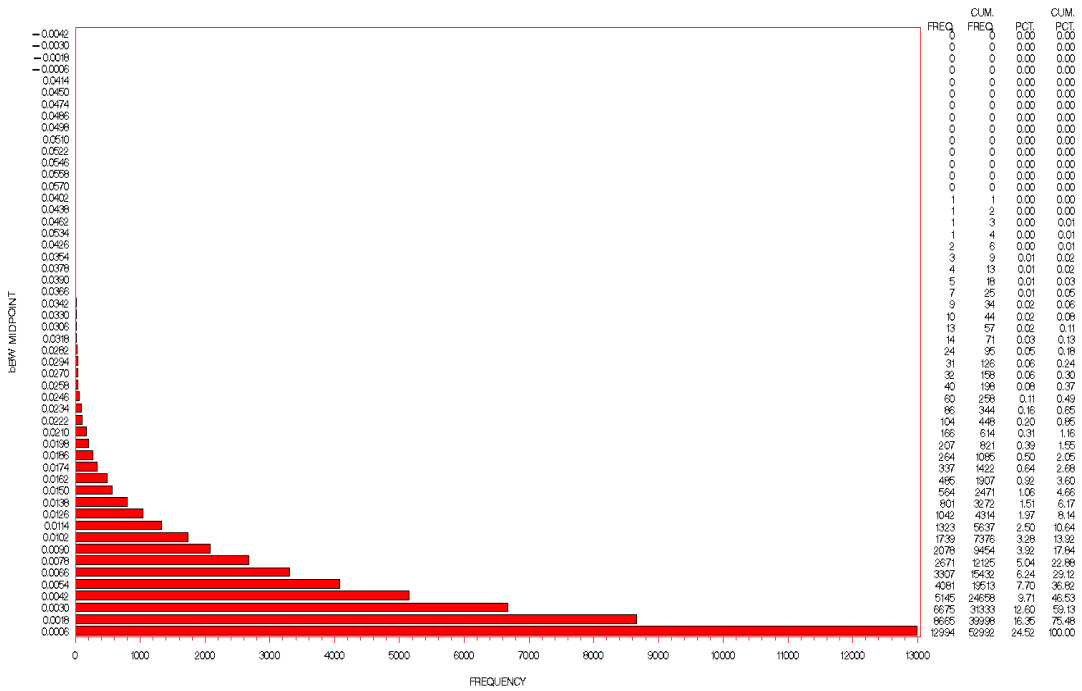
Figure 3. The left Venn diagram presents the number of common significant SNPs between bivariate association analysis (both traits: yellow color) and univariate association analyses(body weight (BW): blue color and egg number (EN): green color) while the right Venn diagram presents the number of common significant SNPs after conditional and joint (cojo) analyses. Venn diagrams were constructed by VENNY 2.1 [55] .

6. General discussion

6.1. Rare alleles and effect size distribution

Most GWASs filter out rare (MAF<0.01) or low-frequency (0.05<MAF<0.01) variants from an analysis to avoid spurious results derived from poor quality of genotype imputation. However, the development of the chicken 600k SNP array has increased the low coverage in previous genotyping technology. For this reason, in contrast to the usual practice of excluding variants with MAF<0.05, a lower MAF threshold was specified (0.01) in two of the four GWAS carried for BW. The use of a lower threshold is based on the premise that also low frequency or rare variants may contribute to phenotypic variation of complex traits such as BW [1]. In accordance to this premise, inclusion of rare or rare and low frequency variants resulted in genomic heritability (h^2) estimates as high as 0.04 and 0.16, respectively, in this population and trait [2]. Such a result implies that low frequency variants explain a significant part of the trait heritability (h^2=0.29).

As most SNPs detected in GWASs segregate at intermediate to high frequencies [3] and the correlation between allele frequencies and locus effect sizes (β) is negative [4], it is rather unlikely that most significant SNPs are tagged to rare causal variants and higher effect sizes or large sample sizes are needed to detect associations involving variants with lower allele frequencies [5]. Based on power calculations [6], the required minimum number of individuals for detecting an association for infrequent variants (MAF=0.001) with low effect sizes (β=0.01 SDs) would exceed 10^8 (80% power, type I error rate=$5x10^{-8}$). This may explain why no low frequency variant could be detected across any analysis here. Furthermore, in both traits, allelic additive effects (β) estimated by GBLUP analysis fit to an exponential distribution (Figure 1) with an abundance of low effect size loci and very few high effect size variants [7].



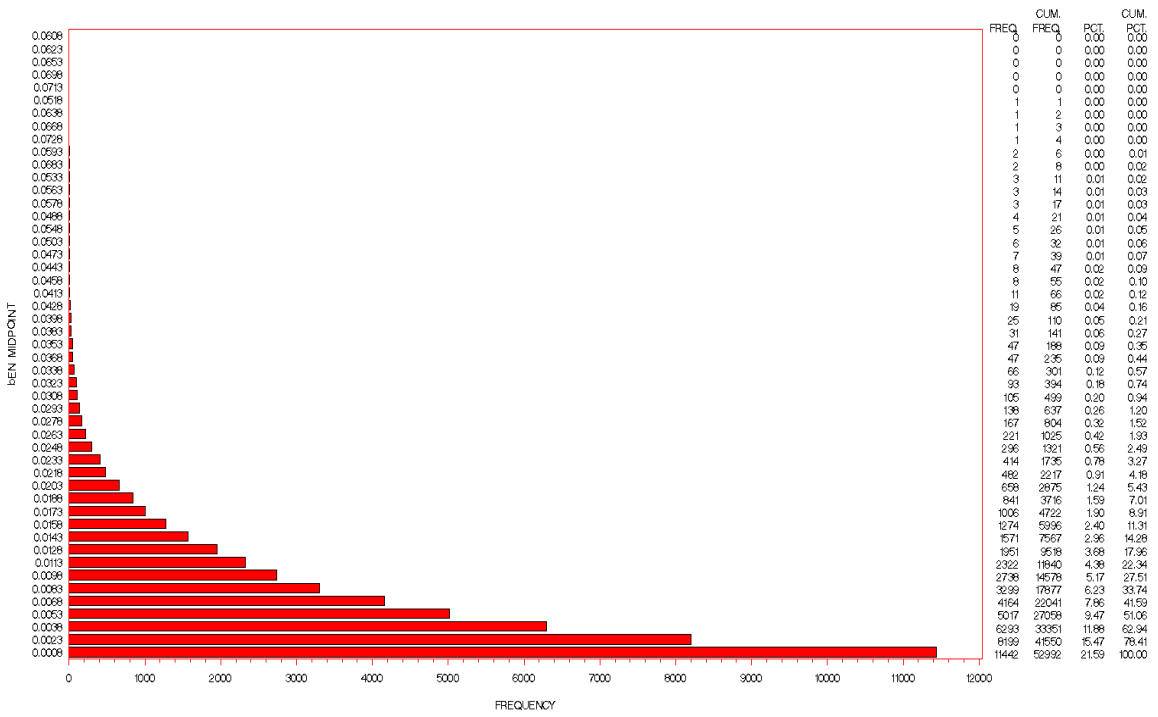


Figure 1. Bar charts depicting absolute allelic effects of 52,992 markers used during bivariate analysis of BW (upper) and EN (lower).

6.2. Sample size and statistical power

While several markers were found to reach genome-wise statistical significance for BW, this was not the case for EN where only chromosome-wise significant markers could be detected. As the yield of GWAS critically depend on the underlying effect-size (β) distribution of the implicated variants [8], this is not a surprising finding and it may be attributed to the low heritability of the trait under study (estimated as high as 0.21).

Albeit a total number of n~2600 animals present an appreciable sample size, a non detectable association implies low statistical power. The required minimum sample size to detect (80% power, type I error rate=$5x10^{-8}$) genome-wise significant markers with MAF=0.10 would, roughly estimated, be over 10^4 and 10^6 for β=0.1 and β=0.01 phenotypic standard deviations (SD), respectively [6]. In general, the expectation is that loci of large effect are likely to be found with low sample sizes, whereas decreasingly lower locus effects would only be found with larger sample sizes. In quantitative traits in humans, the predicted sample sizes needed to identify SNPs which explain 80% of GWAS heritability range between 300k-500k for some of the early growth traits, between 1-2 million for some anthropometric traits and multiple millions for body mass index and some others [9]. In all traits, most of the heritability is explained by loci of large effect (~60%), with those of medium and low effect explaining much lower proportions (~30 and 13%, respectively) of heritability [4].

Whereas such sample sizes are nowadays attainable in human studies due to large-scale consortia, in livestock populations, samples of comparable sizes are unachievable, at least for now, due to insufficient budgets to cover the associated genotyping costs. The latter could be extremely high for large sized populations and the high-throughput genotyping technology. Under a limited budget it is necessary to make an effective allocation of genotyping costs. A useful cost-saving strategy is to genotype only samples of the population with extreme phenotypes (XP) and then perform XP-

GWASs [10]. Results for BW in Chapter 1, have shown that 50% XP-GWAS are useful to detect genomic variants associated with growth related QTLs and plausible causal genes.

6.3. Non-additive gene (inter)action

Most GWASs aim to map additive variants based on the notion that genetic variation in complex traits can be largely explained by additive effects from alleles with small effects on the trait. Nevertheless, non-additive genetic interaction(s) at the same locus (*dominance*) or between different loci (*epistasis*) may also arise resulting in haplotypes with phenotypic values that cannot be attributed to the marginal additive effects of individual alleles. In general, dominance effects in GWASs appear to be minor and efforts to detect epistatic interactions in human GWASs have failed, possibly because they remain severely underpowered [11]. Strikingly, both *dominance* and *epistatic interactions* were detected in analysis of EN (Chapter 4). Specifically, in the case of dominant mode of gene action, the implicated variants spanned a wide spectrum ranging from partial to complete dominance while over-dominance cannot be ruled out for a reproductive trait such as EN. Finally, albeit based on limited number of observations, two locus *epistatic interactions* were also evident in form of additive-by-additive and additive-by-dominance gene interaction. Such results underscore the need to thoroughly exploring the applicability of all possible genetic models in efforts to decipher the genetic architecture of complex traits.

6.4. Pleiotropy

A growing body of genetic research demonstrates the existence of variants that are associated with more than one trait (*pleiotropy*). In polygenic traits, variation affecting a given trait may span a considerable portion of functional genetic variation; as a result, it is bound to overlap with variation affecting other traits [11]. For this reason, it is hypothesized that the pervasive *horizontal pleiotropy* observed in polygenic traits is, at some extent, a logical consequence of widespread polygenicity [12]. This, however, does not mean that every cross phenotype (CP) association detected in GWASs should be attributable to the preceding premise. By identifying genetic variants with antagonistic CP genetic effects, Chapter 5 revealed two genes that simultaneously affect multiple systems offering novel insights in the genetic mechanism underlying the negative co-variation between growth and reproduction in broilers.

In terms of methodology, detection of pleiotropic variants should preferably be based on multivariate approaches as that employed in Chapter 5 that allow for direct identification of CP associations. As results of Chapter 5 have shown, application of bivariate analysis can increase the power to identify additional associations not detected by the ‘standard’ univariate GWAS in the same dataset. Attention should also be paid as to avoiding detection of spurious CP associations arising from short and/or long range LD.

6.5. Importance of non-coding regions

Results obtained herein concur with the general observation that the vast majority of phenotype-associated SNPs identified by GWASs lie within non-coding regions [13]. The tendency to assign potential causal variants preferentially to coding genes within or close to lead variants has contributed to disregarding the role of non-coding elements. Non-coding variants occur in introns, upstream or downstream of coding regions in 3′ and 5′ UTRs, and in intergenic regions and include intronic or promoter regions, small non coding RNAs (ncRNAs) such as miRNAs, long non coding

RNAs (lncRNAs), antisense, and enhancer or insulator regions [14]. Currently, around 2,500 miRNAs and more than 50,000 lncRNAs have been annotated in the human genome [15]. The high proportion of SNPs lying in non-coding regions highlight their potential functional relevance and prompt a better understanding of lncRNA biology as well as regulatory regions such as enhancer to unravel their potential role in complex traits [14]. Many lncRNAs reside in the nucleus conducting key regulatory steps in gene transcription, transcript splicing or chromatin structure while cytoplasmic lncRNAs affect cell homeostasis by modulating translation and stability of mRNA [16]. Implication of miRNAs in growth has also been confirmed in present Chapters (Chapters 2 and 3). Expression quantitate trait loci (eQTL) mapping has been another valuable tool in understanding the function of non-coding variants by establishing genetic association for a given variant with differences in gene expression [17]. Finally, the functional significance of individual candidate SNPs can be experimentally validated by using precise mammalian genome editing techniques such as CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats/CRISPR-associated 9) in animal models. CRISPR/Cas9-mediated gene editing may provide evidence which SNP(s) affect the transcriptional activity of the single or nearby genes involved in traits expression (e.g. [18]).

6.6. Identification of causal genes

6.6.1. Search for candidate genes

A major challenge to the translation of GWAS results into mechanistic understanding is determining the causal variant(s). An intuitive approach is to assign causality to variants preferentially within or closest proximity to lead variants. As demonstrated by the discovery of *MYOM2* gene for BW in Chapter 1, this is an efficient approach to determine causal genes. Nevertheless, the causal gene(s) may not have been directly genotyped, and typically, there are many SNPs in LD with the lead GWAS marker(s). This means that (i) not every gene harboring a significant SNP is a functional candidate and (ii) causal genes may lie tens or hundreds kb apart from the lead markers. Based on the above notions, the search for candidate genes is extended to wider genomic regions around the lead signals, using fixed (e.g. 1Mb, Chapter 2) or variably spanning distances, based on local LD levels (Chapter 3). Often, this search produces numerous positional candidate genes or many candidates with likely functional effects, the experimental validation of which is prohibited because of substantial costs, effort and time. For this reason, assessing gene candidacy in the context of existing biological knowledge and known biological functions is an important step in producing a manageable subset of variants for further validation or exploration [19].

6.6.2. Variant or gene prioritization: fail or success?

'Filtering' the list of candidates to identify a subset of most relevant genes to a specific trait is a complex exercise that needs bioinformatics expertise. Nevertheless, more and more researchers are increasingly motivated to handle this task themselves. The reason for this, is the availability of prominent, often available online, bioinformatics tools that allow for efficient variant/gene annotation and knowledge-driven gene prioritization.

By accepting the challenge of the bioinformatics endeavor, in the present thesis a variety of tools and associated analyses were performed encompassing variant effect prediction (Chapter 1 and 5), functional enrichment analysis (Chapters 2 and 4), gene functional or topological prioritization (Chapters 2 and 4) and detection of functional modules (Chapter 3). Application of these methods has, individually or jointly, demonstrated that extant knowledge can be useful in efforts to prioritize

most likely candidates and in many cases the insights gained from thoroughly interrogating knowledge domains provide enough evidence to implicate a gene. As each analysis uses a different principle foundation to disclose candidate genes, results are, inarguably, method dependent. Thus, careful consideration should be paid to the drawbacks associated with each tool to allow the assessment of variants or genes. Since all have limitations, a combination of tools should be preferably used where possible. Although a candidate that fulfils multiple criteria is most likely to have a genotype-phenotype association, this should not be an absolute criterion for determining its candidacy [19]. Finally, variants or genes should not be discarded as being irrelevant if the knowledge database does not return phenotypic or functional links [19].

6.7. GWAS: present and future

Since the first genome-wide association study for age-related macular degeneration in 2005 [20], 1,181 GWAS with 40,364 genome-wide significant associations have been reported between genetic variants and diseases and traits, in human (NHGRI-EBI Catalog of published GWAS [21]). GWAS have greatly led to the discovery of already known as well as novel genes for traits/diseases (see examples in [22]). Despite the success of GWAS, some studies (e.g. [23,24]) claim that GWAS are just the starting point to find causal genes since follow-up functional studies are required while other studies [25] support that GWAS remains the key strategy to get insight into the biological mechanism of diseases or traits. As a future perspective of GWAS, recent studies suggest the application of next generation sequencing (NGS) techniques for genotyping. The use of whole-genome sequencing (WGS) can increase the power to detect causal variants as every variant (common and rare) can be directly genotyped [26]. To obtain credible GWAS results, a large number of genotyped individuals were required in association analyses but performing WGS in very large sample sizes is currently cost-ineffective (SNP arrays: ~US$40 per sample vs. WGS: >US$1,000 per sample [22]). An efficient imputation strategy in which low-density SNPs were imputed to WGS data at low cost was thus recommended to detect causal loci [27]. Although identifying causal variants might be easier for GWAS using WGS than for GWAS using SNP arrays, functional characterization of the identified genetic variants still remains a challenge.

References

1. Bloom JS, Boocock J, Treusch S, Sadhu MJ, Day L, Oates-Barker H, et al. Rare variants contribute disproportionately to quantitative trait variation in yeast. Elife. eLife Sciences Publications Ltd; 2019;8. doi:10.7554/eLife.49212
2. Kominakis A, Tarsani E, Maniatis G, Olori V, Kranis & A. Estimation of heritability of body weight in broilers using pedigree and dense genome-wide SNP data [Internet]. Available: http://www.wcgalp.org/system/files/proceedings/2018/estimation-heritability-body-weight-broilers-using-pedigree-and-dense-genome-wide-snp-data.pdf
3. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010;42: 565–569. doi:10.1038/ng.608
4. López-Cortegano E, Caballero A. Inferring the nature of missing heritability in human traits using data from the GWAS catalog. Genetics. Genetics Society of America; 2019;212: 891–904. doi:10.1534/genetics.119.302077
5. Bomba L, Walter K, Soranzo N. The impact of rare and low-frequency genetic variants in common disease. Genome Biology. BioMed Central Ltd.; 2017. pp. 1–17. doi:10.1186/s13059-017-1212-4
6. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. American Journal of Human Genetics. Cell Press; 2017. pp. 5–22. doi:10.1016/j.ajhg.2017.06.005
7. King EG, Long AD. The beavis effect in next-generation mapping panels in Drosophila melanogaster.

G3 Genes, Genomes, Genet. Genetics Society of America; 2017;7: 1643–1652. doi:10.1534/g3.117.041426
8. Park JH, Wacholder S, Gail MH, Peters U, Jacobs KB, Chanock SJ, et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. Nat Genet. Nature Publishing Group; 2010;42: 570–575. doi:10.1038/ng.610
9. Zhang Y, Qi G, Park J-H, Chatterjee N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits and implications for the future. bioRxiv. Cold Spring Harbor Laboratory; 2017; 175406. doi:10.1101/175406
10. Yang J, Jiang H, Yeh C-T, Yu J, Jeddeloh JA, Nettleton D, et al. Extreme-phenotype genome-wide association study (XP-GWAS): a method for identifying trait-associated variants by sequencing pools of individuals selected from a diversity panel. Plant J. John Wiley & Sons, Ltd (10.1111); 2015;84: 587–596. doi:10.1111/tpj.13029
11. Sella G, Barton NH. Thinking About the Evolution of Complex Traits in the Era of Genome-Wide Association Studies. Annu Rev Genomics Hum Genet. Annual Reviews; 2019;20: 461–493. doi:10.1146/annurev-genom-083115-022316
12. Jordan DM, Verbanck M, Do R. HOPS: A quantitative score reveals pervasive horizontal pleiotropy in human genetic variation is driven by extreme polygenicity of human traits and diseases. Genome Biol. BioMed Central Ltd.; 2019;20. doi:10.1186/s13059-019-1844-7
13. Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. Nature Publishing Group; 2012;491: 56–65. doi:10.1038/nature11632
14. Giral H, Landmesser U, Kratzer A. Into the Wild: GWAS Exploration of Non-coding RNAs. Frontiers in Cardiovascular Medicine. Frontiers Media S.A.; 2018. p. 181. doi:10.3389/fcvm.2018.00181
15. Volders PJ, Verheggen K, Menschaert G, Vandepoele K, Martens L, Vandesompele J, et al. An update on LNCipedia: a database for annotated human lncRNA sequences. Nucleic Acids Res. 2015;43: 4363–4. doi:10.1093/nar/gkv295
16. Noh JH, Kim KM, McClusky WG, Abdelmohsen K, Gorospe M. Cytoplasmic functions of long noncoding RNAs. Wiley Interdisciplinary Reviews: RNA. Blackwell Publishing Ltd; 2018. p. e1471. doi:10.1002/wrna.1471
17. Osgood JA, Knight JC. Translating GWAS in rheumatic disease: approaches to establishing mechanism and function for genetic associations with ankylosing spondylitis. Brief Funct Genomics. 2018;17: 308–318. doi:10.1093/bfgp/ely015
18. Wu S, Zhang M, Yang X, Peng F, Zhang J, Tan J, et al. Genome-wide association studies and CRISPR/Cas9-mediated gene editing identify regulatory variants influencing eyebrow thickness in humans. Cooper GM, editor. PLOS Genet. Public Library of Science; 2018;14: e1007640. doi:10.1371/journal.pgen.1007640
19. Dashti MJS, Gamieldien J. A practical guide to filtering and prioritizing genetic variants. Biotechniques. Eaton Publishing Company; 2017;62: 18–30. doi:10.2144/000114492
20. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, et al. Complement factor H polymorphism in age-related macular degeneration. Science (80-). 2005;308: 385–389. doi:10.1126/science.1109557
21. GWAS Catalog [Internet]. [cited 1 Apr 2020]. Available: https://www.ebi.ac.uk/gwas/docs/about
22. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. Nature Reviews Genetics. Nature Publishing Group; 2019. pp. 467–484. doi:10.1038/s41576-019-0127-1
23. Ikegawa S. A Short History of the Genome-Wide Association Study: Where We Were and Where We Are Going. Genomics Inform. Korea Genome Organization; 2012;10: 220. doi:10.5808/gi.2012.10.4.220
24. Vandenbroeck K. The era of GWAS is over - Yes. Mult Scler. SAGE Publications Ltd; 2018;24: 256–257. doi:10.1177/1352458517738059
25. De Jager PL. The era of GWAS is over - No. Mult Scler. SAGE Publications Ltd; 2018;24: 258–260. doi:10.1177/1352458517742980
26. Wang Z, Chatterjee N. Increasing mapping precision of genomewide association studies: To genotype and impute, sequence, or both? Genome Biol. BioMed Central Ltd.; 2017;18. doi:10.1186/s13059-017-1255-6
27. Wu P, Wang K, Zhou J, Chen D, Yang Q, Yang X, et al. GWAS on Imputed Whole-Genome Resequencing From Genotyping-by-Sequencing Data for Farrowing Interval of Different Parities in Pigs.

Front Genet. Frontiers Media S.A.; 2019;10: 1012. doi:10.3389/fgene.2019.01012

Appendix – Supplementary Tables and Figures
Chapter 2

Table S1. Positional candidate genes for BW. Note that genes including the significant marker are shown in bold.

Gene ID*	***Description***	**GGA**	**Start Position of the gene (bp)***	**End Position of the gene (bp)***	***Significant SNP(s) associated with each candidate gene***
LOC101750030	*uncharacterized LOC101750030*	1	111729926	111783003	*rs13923872*
CASK	*calcium/calmodulin dependent serine protein kinase*		111777387	111983664	
GPR82	*G protein-coupled receptor 82*		111877626	111886967	
GPR34	*G protein-coupled receptor 34*		111891553	111897268	
NYX	*nyctalopin*		111995838	112016261	
DDX3X	*DEAD-box helicase 3, X-linked*		112021552	112038995	
LOC101750611	*uncharacterized LOC101750611*		112041464	112056885	
LOC107056703	*uncharacterized LOC107056703*		112059923	112065219	
MIR6850	*microRNA 6850*		112065342	112069044	
USP9X	*ubiquitin specific peptidase 9, X-linked*		112078329	112176727	
LOC101750779	*uncharacterized LOC101750779*		112193534	112198521	
LOC107056757	*uncharacterized LOC107056757*		112223451	112228082	
MED14	*mediator complex subunit 14*		112228091	112262695	
LOC107056765	*uncharacterized LOC107056765*		112263096	112268266	
ATP6AP2	*ATPase H+ transporting accessory protein 2*		112273187	112282377	
LOC101750924	*uncharacterized LOC101750924*		112282457	112304013	
LOC107056538	*basic proline-rich protein-like*		112349429	112354018	
BCOR	*BCL6 corepressor*		112378844	112422324	
LOC107057081	*uncharacterized LOC107057081*		112474240	112486871	
LOC101751364	*uncharacterized LOC101751364*		112482075	112497856	
LOC107057076	*uncharacterized LOC107057076*		112498548	112504661	

LOC101751443	*uncharacterized LOC101751443*		112519737	112523119	
MIR6672	*microRNA 6672*		112625566	112625675	
LOC107057061	*uncharacterized LOC107057061*		112635402	112643712	
LOC107057049	*uncharacterized LOC107057049*		112668313	112707049	
MID1IP1	*MID1 interacting protein 1*		112738605	112739753	
TSPAN7	*tetraspanin 7*		112766324	112855838	
LOC101751659	*uncharacterized LOC101751659*		112767424	112769932	
LOC107057044	*uncharacterized LOC107057044*		112853920	112855360	
OTC	*ornithine carbamoyltransferase*		112898625	112924531	
RPGR	*retinitis pigmentosa GTPase regulator*		112926169	112985737	
SRPX	*sushi repeat containing protein, X-linked*		112994814	113035754	
SYTL5	*synaptotagmin like 5*		113039522	113117936	
DYNLT3	*dynein light chain Tctex-type 3*		113133105	113139987	
CYBB	*cytochrome b-245 beta chain*		113149626	113183626	
XK	*X-linked Kx blood group*		113199116	113214148	
LANCL3	*LanC like 3*		113215605	113251683	
LOC107056992	*uncharacterized LOC107056992*		113278216	113282701	
PRRG1	*proline rich and Gla domain 1*		113283937	113314070	
LOC107056924	*maestro heat-like repeat-containing protein family member 2B*		113325651	113339179	
LOC107056990	*maestro heat-like repeat-containing protein family member 2B*		113343513	113348692	
LOC107056981	*maestro heat-like repeat-containing protein family member 2B*		113350799	113354007	
LOC107056965	*maestro heat-like repeat-containing protein family member 2B*		113357864	113369741	
LOC107057105	*uncharacterized LOC107057105*		113387296	113390923	
C1HXORF59	*chromosome 1 open reading frame, human CXORF59*		113403041	113666278	
LOC100857117	*uncharacterized LOC100857117*	4	28172128	28173788	*rs312691174*

LOC107051793	*uncharacterized LOC107051793*		28550125	28563120	
PCDH18	*protocadherin 18*		28880214	28890093	
LOC107051792	*uncharacterized LOC107051792*		28986991	29036383	
SLC7A11	*solute carrier family 7 member 11*		29138177	29196482	
LOC101751121	*uncharacterized LOC101751121*		29205067	29277125	
NOCT	*nocturnin*		29428771	29434135	
ELF2	*E74 like ETS transcription factor 2*		29438718	29467618	
MGARP	*mitochondria localized glutamic acid rich protein*		29471760	29493863	
LOC107051797	*atherin-like*		29494724	29497114	
NAA15	*N(alpha)-acetyltransferase 15, NatA auxiliary subunit*		29496798	29534035	
RAB33B	*RAB33B, member RAS oncogene family*		29539318	29546334	
LOC422442	*uncharacterized LOC422442*		29559136	29561953	
SETD7	*SET domain containing lysine methyltransferase 7*		29561960	29578876	
MGST2	*microsomal glutathione S-transferase 2*		29593387	29603029	
LOC107051791	*microsomal glutathione S-transferase 2-like*		29604032	29611350	
MAML3	*mastermind like transcriptional coactivator 3*		29611253	29815133	
MIR1575	*microRNA 1575*		29768218	29768321	
LOC101751730	*uncharacterized LOC101751730*		29856791	29867266	
SCOC	*short coiled-coil protein*		29862290	29870883	
CLGN	*calmegin*		29872496	29890838	
MGAT4D	*MGAT4 family member D*		29894628	29930674	
ELMOD2	*ELMO domain containing 2*		29930851	29937844	
TBC1D9	*TBC1 domain family member 9*		29945029	29988803	
LOC107051790	*uncharacterized LOC107051790*		29978295	29981680	
RNF150	*ring finger protein 150*		30008764	30129272	
CHIC2	*cysteine rich hydrophobic domain 2*	4	65887673	65915464	*rs15608447*
LNX1	*ligand of numb-protein X 1*		65990792	66133390	
FIP1L1	*factor interacting with PAPOLA and CPSF1*		66133445	66171020	
SCFD2	*sec1 family domain containing 2*		66173796	66368297	

LOC107053245	*uncharacterized LOC107053245*		66257009	66266179	
RASL11B	*RAS like family 11 member B*		66369830	66373534	
LOC107053244	*uncharacterized LOC107053244*		66391577	66398660	
LOC422757	*uncharacterized LOC422757*		66403272	66405708	
USP46	*ubiquitin specific peptidase 46*		66413295	66440521	
SPATA18	*spermatogenesis associated 18*		66567846	66587444	
SGCB	*sarcoglycan beta*		66587612	66593633	
LRRC66	*leucine rich repeat containing 66*		66593723	66602814	
DCUN1D4	*defective in cullin neddylation 1 domain containing 4*		66607031	66645141	
CWH43	*cell wall biogenesis 43 C-terminal homolog*		66651068	66677010	
OCIAD1	*OCIA domain containing 1*		66693766	66708277	
FRYL	*FRY like transcription coactivator*		66708277	66866440	
ZAR1	*zygote arrest 1*		66868239	66871111	
LOC107053243	***uncharacterized LOC107053243***		66870041	66885878	
SLC10A4	*solute carrier family 10 member 4*		66870984	66872626	
SLAIN2	***SLAIN motif family member 2***		66883481	66910807	
TEC	*tec protein tyrosine kinase*		66928003	66969829	
TXK	*TXK tyrosine kinase*		66970277	66990459	
NIPAL1	*NIPA like domain containing 1*		66992447	67002107	
CNGA1	*cyclic nucleotide gated channel alpha 1*		67007806	67015296	
NFXL1	*nuclear transcription factor, X-box binding like 1*		67016510	67061474	
CORIN	*corin, serine peptidase*		67063312	67186506	
ATP10D	*ATPase phospholipid transporting 10D (putative)*		67189360	67228010	
COMMD8	*COMM domain containing 8*		67239252	67243040	
LOC107053239	*uncharacterized LOC107053239*		67246293	67253895	
LOC107053241	*uncharacterized LOC107053241*		67252999	67266059	
GABRB1	*gamma-aminobutyric acid type A receptor beta1 subunit*		67253915	67271613	
LOC107053242	*uncharacterized LOC107053242*		67325419	67334233	
LOC107053240	*uncharacterized LOC107053240*		67330651	67347343	

GABRA4	*gamma-aminobutyric acid type A receptor alpha4 subunit*		67355338	67402602	
GABRA2	*gamma-aminobutyric acid type A receptor alpha2 subunit*		67487682	67549247	
LOC107053237	*uncharacterized LOC107053237*		67489979	67491477	
LOC107053238	*uncharacterized LOC107053238*		67546207	67576158	
GABRG1	*gamma-aminobutyric acid type A receptor gamma1 subunit*		67576299	67633176	
LOC770268	*uncharacterized LOC770268*	10	12530062	12564671	*rs318199727*
FAH	*fumarylacetoacetate hydrolase*		12567084	12577908	
ZFAND6	*zinc finger AN1-type containing 6*		12580889	12630294	
BCL2A1	*BCL2 related protein A1*		12637320	12639335	
MTHFS	*5,10-methenyltetrahydrofolate synthetase (5-formyltetrahydrofolate cyclo-ligase)*		12641387	12667030	
LOC101750317	*uncharacterized LOC101750317*		12707525	12733302	
KIAA1024	*KIAA1024*		12747379	12766338	
LOC107054202	*uncharacterized LOC107054202*		12796949	12800712	
PEX11A	*peroxisomal biogenesis factor 11 alpha*		12815015	12820635	
LOC107054203	*uncharacterized LOC107054203*		12817514	12823049	
PLIN1	*perilipin 1*		12822263	12826843	
KIF7	*kinesin family member 7*		12827027	12836867	
TICRR	*TOPBP1 interacting checkpoint and replication regulator*		12837814	12854365	
RHCG	*Rh family C glycoprotein*		12859333	12867258	
LOC107054204	*uncharacterized LOC107054204*		12902048	12938774	
TRNAR-UCG	*transfer RNA arginine (anticodon UCG)*		12942837	12942909	
POLG	*DNA polymerase gamma, catalytic subunit*		12942988	12953027	
FANCI	*Fanconi anemia complementation group I*		12951763	12975529	
RLBP1	*retinaldehyde binding protein 1*		12976366	12980554	
ABHD2	*abhydrolase domain containing 2*		12981973	13021732	
MFGE8	*milk fat globule-EGF factor 8 protein*		13033323	13040620	

HAPLN3	*hyaluronan and proteoglycan link protein 3*		13042935	13046872	
ACAN	*aggrecan*		13047289	13092387	
AEN	*apoptosis enhancing nuclease*		13131452	13134038	
MIR1720	*microRNA 1720*		13134585	13134649	
MIR7-2	*microRNA 7-2*		13134720	13134818	
MIR3529	*microRNA 3529*		13134724	13134814	
DET1	*de-etiolated homolog 1 (Arabidopsis)*		13149476	13171850	
MRPS11	*mitochondrial ribosomal protein S11*		13171107	13174638	
MRPL46	*mitochondrial ribosomal protein L46*		13174668	13177215	
LOC101751754	*uncharacterized LOC101751754*		13178649	13201172	
LOC101751792	*uncharacterized LOC101751792*		13196390	13209087	
NTRK3	*neurotrophic receptor tyrosine kinase 3*		13227881	13408049	
LOC107054207	*uncharacterized LOC107054207*		13381318	13386137	
LOC107054206	***uncharacterized LOC107054206***		13388442	13616738	
AGBL1	*ATP/GTP binding protein like 1*		13623118	13914436	
KLHL25	*kelch like family member 25*		13954963	13969663	
AKAP13	*A-kinase anchoring protein 13*		13974501	14169240	
LOC107054188	*uncharacterized LOC107054188*		14182433	14183328	
SV2B	*synaptic vesicle glycoprotein 2B*		14187010	14235623	
LOC101747399	*uncharacterized LOC101747399*		14323417	14341914	
SLCO3A1	*solute carrier organic anion transporter family member 3A1*		14345761	14457203	
LOC101747558	*uncharacterized LOC101747558*		14482710	14502486	
ST8SIA2	*ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 2*		14510760	14537564	
LOC107054325	*uncharacterized LOC107054325*	11	17800164	17805321	*rs318098582*
IRF8	*interferon regulatory factor 8*		17821288	17828760	
DUSP22AL	*dual specificity protein phosphatase 22-A-like*		17830204	17840877	
COX4I1	*cytochrome c oxidase subunit 4I1*		17841426	17845527	

EMC8	*ER membrane protein complex subunit 8*		17845679	17853183	
GINS2	*GINS complex subunit 2*		17858071	17867370	
GSE1	*Gse1 coiled-coil protein*		17867595	17954219	
LOC107054326	*uncharacterized LOC107054326*		17912747	17925839	
LOC107054321	*putative protein TPRXL*		17954053	17954808	
KIAA0513	*KIAA0513*		17982718	17994940	
FBXO31	*F-box protein 31*		17998847	18020696	
MAP1LC3B	*microtubule associated protein 1 light chain 3 beta*		18027539	18036281	
ZCCHC14	*zinc finger CCHC-type containing 14*		18036068	18080985	
JPH3	*junctophilin 3*		18127475	18140463	
KLHDC4	*kelch domain containing 4*		18143822	18171651	
SLC7A5	*solute carrier family 7 member 5*		18186088	18218511	
LOC107054327	*uncharacterized LOC107054327*		18218464	18221743	
CA5A	*carbonic anhydrase 5A*		18221264	18236716	
LOC107054328	*uncharacterized LOC107054328*		18236801	18242854	
BANP	*BTG3 associated nuclear protein*		18243614	18386758	
ZNF469	*zinc finger protein 469*		18416335	18579095	
LOC107054329	*uncharacterized LOC107054329*		18541302	18548014	
ZFPM1	*zinc finger protein, FOG family member 1*		18584457	18613764	
CIDEC	*cell death inducing DFFA like effector c*		18614459	18616305	
ZC3H18	***zinc finger CCCH-type containing 18***		18617357	18656718	
MIR1571	*microRNA 1571*		18632364	18632461	
IL17C	*interleukin 17C*		18658034	18663248	
CYBA	*cytochrome b-245 alpha chain*		18663342	18665615	
MVD	*mevalonate diphosphate decarboxylase*		18665712	18668663	
RNF166	*ring finger protein 166*		18670083	18677685	
CTU2	*cytosolic thiouridylase subunit 2*		18677742	18681201	
PIEZO1	*piezo type mechanosensitive ion channel component 1*		18681139	18700645	

LOC107054330	*nascent polypeptide-associated complex subunit alpha, muscle-specific form-like*		18696540	18698959	
CDT1	*chromatin licensing and DNA replication factor 1*		18701940	18705724	
APRT	*adenine phosphoribosyltransferase*		18706797	18709506	
GALNS	*galactosamine (N-acetyl)-6-sulfatase*		18714200	18758480	
TRAPPC2L	*trafficking protein particle complex 2 like*		18758475	18761079	
PABPN1L	*poly(A) binding protein nuclear 1 like, cytoplasmic*		18762170	18765810	
CBFA2T3	*CBFA2/RUNX1 translocation partner 3*		18767307	18788285	
ACSF3	*acyl-CoA synthetase family member 3*		18805694	18846630	
CDH15	*cadherin 15*		18848471	18853003	
SLC22A31	*solute carrier family 22 member 31*		18853081	18856750	
ANKRD11	*ankyrin repeat domain 11*		18857696	18939049	
MIR1560	*microRNA 1560*		18874320	18874423	
MIR1785	*microRNA 1785*		18926659	18926760	
SPG7	*SPG7, paraplegin matrix AAA peptidase subunit*		18946166	18976689	
RPL13	*ribosomal protein L13*		18978071	18982325	
CPNE7	*copine VII*		18983791	18989285	
SULT2B1L1	*sulfotransferase family cytosolic 2B member 1-like 1*		18990705	18993412	
DPEP1	*dipeptidase 1 (renal)*		18993512	18997373	
CHMP1A	*charged multivesicular body protein 1A*		18999512	19003217	
CDK10	*cyclin dependent kinase 10*		19004528	19008784	
MIR6667	*microRNA 6667*		19007943	19008052	
SPATA2L	*spermatogenesis associated 2 like*		19009229	19010941	
VPS9D1	*VPS9 domain containing 1*		19010926	19015340	
ZNF276	*zinc finger protein 276*		19015485	19022670	
FANCA	*Fanconi anemia complementation group A*		19022586	19055008	

LOC769325	*uncharacterized LOC769325*		19056414	19057695	
SPIRE2	*spire-type actin nucleation factor 2*		19057744	19063946	
TCF25	*transcription factor 25*		19064480	19082896	
MC1R	*melanocortin 1 receptor*		19084582	19085526	
LOC107054334	*translation initiation factor IF-2-like*		19086099	19087230	
DEF8	*differentially expressed in FDCP 8 homolog*		19089879	19093651	
DBNDD1	*dysbindin (dystrobrevin binding protein 1) domain containing 1*		19093976	19096929	
GAS8	*growth arrest specific 8*		19097379	19105093	
URAH	*5-hydroxyisourate hydrolase*		19105097	19107786	
CDH3	*cadherin 3*		19107862	19113818	
CDH1	*cadherin 1*		19114383	19123146	
TMCO7	*transmembrane and coiled-coil domains 7*		19123969	19151197	
HAS3	*hyaluronan synthase 3*		19152159	19155850	
CHTF8	*chromosome transmission fidelity factor 8*		19157119	19158772	
UTP4	*UTP4, small subunit processome component*		19158792	19163525	
SNTB2	*syntrophin beta 2*		19163629	19168951	
VPS4A	*vacuolar protein sorting 4 homolog A*		19169788	19172777	
COG8	*component of oligomeric golgi complex 8*		19173208	19174878	
NIP7	*NIP7, nucleolar pre-rRNA processing protein*		19174900	19176268	
TMED6	*transmembrane p24 trafficking protein 6*		19176609	19179130	
TERF2	*telomeric repeat binding factor 2*		19180205	19188036	
CYB5B	*cytochrome b5 type B*		19192249	19206013	
NFAT5	*nuclear factor of activated T-cells 5*		19212134	19277188	
LOC101750188	*envelope glycoprotein gp95-like*		19255421	19262314	
NQO1	*NAD(P)H quinone dehydrogenase 1*		19278734	19280589	
NOB1	*NIN1/PSMD8 binding protein 1 homolog*		19281166	19283712	
WWP2	*WW domain containing E3 ubiquitin protein ligase 2*		19283664	19316100	
MIR140	*microRNA 140*		19310301	19310395	

PSMD7	*proteasome 26S subunit, non-ATPase 7*		19316519	19321831	
ZFHX3	*zinc finger homeobox 3*		19321916	19834198	
LOC107051612	*uncharacterized LOC107051612*		2546549	2572319	
NOC4L	*nucleolar complex associated 4 homolog*		2571851	2582310	
EP400	*E1A binding protein p400*		2582896	2628948	
LOC107051611	*uncharacterized LOC107051611*		2586309	2589222	
PUS1	*pseudouridylate synthase 1*		2629662	2634569	
ULK1	*unc-51 like autophagy activating kinase 1*		2635904	2708526	
MMP17	*matrix metallopeptidase 17*		2722077	2765544	
SFSWAP	*splicing factor SWAP homolog*		2792492	2832182	
STX2	*syntaxin 2*		2926591	3223484	
ADGRD1	*adhesion G protein-coupled receptor D1*		3062231	3191688	
RAN	*RAN, member RAS oncogene family*		3201356	3205602	
RIMBP2	*RIMS binding protein 2*		3223574	3347096	
PIWIL1	*piwi like RNA-mediated gene silencing 1*		3347780	3421290	
FZD10	*frizzled class receptor 10*	15	3432876	3435118	*rs317945754*
LOC107051610	*frizzled-10-like*		3442998	3446063	
TMEM132D	***transmembrane protein 132D***		3526053	3718886	
GLT1D1	*glycosyltransferase 1 domain containing 1*		3736272	3784033	
SLC15A4	*solute carrier family 15 member 4*		3785313	3806720	
TMEM132C	*transmembrane protein 132C*		3830248	4010779	
LOC107051578	*uncharacterized LOC107051578*		4105652	4107324	
LOC107051609	*uncharacterized LOC107051609*		4241859	4320316	
TMEM132B	*transmembrane protein 132B*		4291921	4471778	
AACS	*acetoacetyl-CoA synthetase*		4478298	4515607	
LOC107051608	*uncharacterized LOC107051608*		4510959	4512769	
BRI3BP	*BRI3 binding protein*		4518239	4520623	
DHX37	*DEAH-box helicase 37*		4523892	4535799	
LOC107051607	*uncharacterized LOC107051607*		4530643	4533347	

UBC	*ubiquitin C*		4539277	4540797	
SCARB1	*scavenger receptor class B member 1*		4542170	4558439	
TGFA	*transforming growth factor alpha*		3594207	3611266	
LOC107054987	*leucine-rich repeat extensin-like protein 5*		3607070	3611133	
LOC107054988	*uncharacterized LOC107054988*		3647885	3769840	
LRRTM4	*leucine rich repeat transmembrane neuronal 4*		3854424	4064475	
LOC107054981	*tumor necrosis factor ligand superfamily member 6-like*		4253743	4253898	
LOC107054989	*uncharacterized LOC107054989*		4262027	4273755	
ADRA1A	*adrenoceptor alpha 1A*		4525641	4548511	
LOC101751469	*uncharacterized LOC101751469*		4576077	4577398	
ANXA4	*annexin A4*		4584810	4594055	
SLC20A1	*solute carrier family 20 member 1*		4596702	4604387	
NT5DC4	*5'-nucleotidase domain containing 4*		4604777	4609445	
CKAP2L	*cytoskeleton associated protein 2 like*		4609435	4616294	
LOC107054991	*uncharacterized LOC107054991*	22	4616423	4616974	*rs316794400*
IL1B	*interleukin 1, beta*		4616889	4618625	
OGDH	*oxoglutarate dehydrogenase*		4622977	4643738	
LOC107054982	*zinc finger MIZ domain-containing protein 2-like*		4644533	4658490	
PPIA	*peptidylprolyl isomerase A*		4658952	4660588	
LOC107054992	*nudC domain-containing protein 3 pseudogene*		4661590	4661908	
NUDCD3	*NudC domain containing 3*		4663253	4679543	
CAMK2BL	*calcium/calmodulin-dependent protein kinase II beta-like*		4680833	4697295	
YKT6BL	*synaptobrevin homolog YKT6-B-like*		4699156	4702203	
LOC107054983	*glucokinase-like*		4704014	4710683	
GNLY	*granulysin*		4711546	4712974	
POLD2	*DNA polymerase delta 2, accessory subunit*		4713230	4717098	
AEBP1	*AE binding protein 1*		4717021	4728472	
BGLAP	*bone gamma-carboxyglutamate protein*	25	594	1789	*rs317288536*
SMG5	*SMG5, nonsense mediated mRNA decay factor*		2171	24198	

TMEM79	*transmembrane protein 79*		24606	27213	
GLMP	*glycosylated lysosomal membrane protein*		27661	31956	
CCT3	*chaperonin containing TCP1 subunit 3*		33641	44203	
LOC107055078	*uncharacterized LOC107055078*		46465	47469	
LOC107055080	*nectin-4-like*		54112	60599	
LIM2	*lens intrinsic membrane protein 2*		61134	68960	
LOC107055082	*cytochrome b5 domain-containing protein 1-like*		81045	82719	
LOC107055083	*uncharacterized LOC107055083*		82856	89386	
VPS45	*vacuolar protein sorting 45 homolog*		92295	119429	
PLEKHO1	*pleckstrin homology domain containing O1*		123481	139888	
LOC107055081	*uncharacterized LOC107055081*		142235	150690	
ANP32E	*acidic nuclear phosphoprotein 32 family member E*		162050	174225	
LOC100859767	*cytochrome b5 domain-containing protein 1-like*		204000	205667	
APOA1BP	*apolipoprotein A-I binding protein*		224358	227444	
GPATCH4	*G-patch domain containing 4*		227411	231838	
LOC107055084	*uncharacterized LOC107055084*		235211	236685	
MEX3A	*mex-3 RNA binding family member A*		247371	260425	
LOC107055086	*sperm-associated antigen 4 protein-like*		783429	786252	
LOC100857131	*sperm-associated antigen 4 protein-like*		797890	798435	
UBQLN4	*ubiquilin 4*		804221	815017	
LAMTOR2	*late endosomal/lysosomal adaptor, MAPK and MTOR activator 2*		815073	817920	
RAB25	*RAB25, member RAS oncogene family*		818004	824803	
RAB2B	*RAB2B, member RAS oncogene family*		825467	830852	
LOC101747704	*uncharacterized LOC101747704*		846955	849648	
LOC107055087	*sperm-associated antigen 4 protein-like*		855482	858864	
OTUD7B	*OTU deubiquitinase 7B*		888239	923782	
MTMR11	*myotubularin related protein 11*		925626	933506	
SF3B4	*splicing factor 3b subunit 4*		933590	938879	

SV2A	*synaptic vesicle glycoprotein 2A*		939052	949843	
LOC107055093	*uncharacterized LOC107055093*		950637	951127	
LOC107055108	*feather keratin 3-like*		953490	954648	
LOC107055109	*feather keratin 3-like*		956426	957314	
LOC100859249	*feather keratin 3-like*		959380	960434	
LOC107055107	*feather keratin 1-like*		959446	963950	
LOC100859427	*feather keratin 1-like*		966460	967631	
LOC426914	*feather keratin 1-like*		969985	971047	
F-KER	***feather keratin I***		973471	980575	
LOC429492	*keratin D*		973475	974554	
LOC431325	***feather keratin 1-like***		976657	980857	
LOC431324	*keratin A*		979736	983908	
LOC426913	*feather keratin 1-like*		982753	987425	
LOC431323	*beta-keratin-related protein-like*		991470	993152	
LOC431322	*feather keratin 1-like*		997064	997414	
LOC431321	*keratin*		1002033	1004112	
LOC431320	*feather beta keratin-like*		1008291	1010072	
LOC107055103	*scale keratin-like*		1010842	1012163	
LOC107055106	*uncharacterized LOC107055106*		1014345	1015145	
LOC431317	*scale keratin-like*		1017251	1017835	
LOC431316	*scale keratin-like*		1018813	1019553	
LOC100859586	*scale keratin-like*		1020735	1021650	
LOC100859657	*scale keratin-like*		1020928	1025554	
LOC100859616	*scale keratin-like*		1022275	1023023	
LOC425362	*scale keratin-like*		1026144	1027006	
LOC100857270	*scale keratin-like*		1028628	1029557	
LOC100859756	*scale keratin-like*		1030236	1034932	
LOC100859722	*scale keratin-like*		1030356	1031017	
LOC100857297	*scale keratin-like*		1032585	1044971	

LOC107055105	*uncharacterized LOC107055105*		1032750	1040322	
LOC426912	*scale keratin-like*		1036532	1037365	
LOC100859790	*scale keratin-like*		1037906	1038601	
LOC101751554	*scale keratin-like*		1040076	1040849	
LOC100857367	*scale keratin-like*		1041747	1042135	
LOC107055104	*scale keratin-like*		1044261	1044828	
LOC101750668	*scale keratin-like*		1045428	1046372	
LOC396480	*keratin*		1048622	1050613	
LOC101750550	*scale keratin-like*		1052489	1053802	
LOC396479	*keratin*		1055266	1056859	
LOC431314	*scale keratin-like*		1058333	1060625	
LOC769486	*scale keratin-like*		1064610	1066554	
LOC107055102	*uncharacterized LOC107055102*		1067878	1069936	
LOC408038	*beta-keratin*		1069829	1071422	
LOC431313	*feather beta keratin-like*		1074635	1075569	
LOC107055092	*uncharacterized LOC107055092*		1078270	1079527	
LOC100857468	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		1080919	1087221	
LOC107055101	*uncharacterized LOC107055101*		1093075	1095264	
LOC101751614	*keratin, type I cytoskeletal 9-like*		1098677	1100340	
LOC107055091	*beta-keratin-related protein-like*		1102072	1103363	
LOC107055090	*uncharacterized LOC107055090*		1106254	1107828	
LOC107055100	*uncharacterized LOC107055100*		1109328	1111524	
LOC101751113	*titin-like*		1116120	1122481	
LOC107055099	*uncharacterized LOC107055099*		1116136	1119078	
EDYM2	*epidermal differentiation protein containing Y motif 2*		1122288	1125199	
EDQREP	*epidermal differentiation protein containing glutamine (Q) repeats*		1127324	1130658	
EDPE	*epidermal differentiation protein rich in proline and glutamic acid (E)*		1139306	1142301	

LOC107055098	*epidermal differentiation protein containing glutamine (Q) repeats-like*		1144136	1145825	
EDQCM	*epidermal differentiation protein containing QC motifs*		1148589	1150441	
EDDM	*epidermal differentiation protein containing DPCC motifs*		1154359	1157611	
EDNC	*epidermal differentiation protein encoded by neighbor of cornulin*		1159110	1161256	
CRNN	*cornulin*		1168202	1170459	
SCFN	*scaffoldin*		1173597	1177833	
LOC107055094	*trichohyalin-like*		1178020	1190574	
S100A11	*S100 calcium binding protein A11*		1201165	1202884	
COPA	*coatomer protein complex subunit alpha*		1203325	1221735	
NCSTN	*nicastrin*		1221813	1232683	
NHLH1	*nescient helix-loop-helix 1*		1241431	1246232	
LOC107055095	*uncharacterized LOC107055095*		1246142	1253267	
VANGL2	*VANGL planar cell polarity protein 2*		1250127	1262898	
LY9	*lymphocyte antigen 9*		1266694	1272789	
SLAMF1	*signaling lymphocytic activation molecule family member 1*		1274631	1280327	
CD48	*CD48 molecule*		1282479	1284960	
CD244	*CD244 molecule*		1285835	1293468	
LOC101750757	*uncharacterized LOC101750757*		1296055	1297931	
KIRREL	*kin of IRRE like (Drosophila)*		1301854	1321563	
LOC101750908	*T-lymphocyte surface antigen Ly-9-like*		1323643	1330948	
SLAMF8	*SLAM family member 8*		1331041	1334671	
ETV3	*ETS variant 3*		1334908	1344750	
ETV3L	*ETS variant 3 like*		1353678	1355999	
ARHGEF11	*Rho guanine nucleotide exchange factor 11*		1357481	1378565	
LRRC71	*leucine rich repeat containing 71*		1379498	1382513	
PEAR1	*platelet endothelial aggregation receptor 1*		1382768	1391921	

NTRK1	*neurotrophic receptor tyrosine kinase 1*		1394975	1402575	
INSRR	*insulin receptor related receptor*		1403569	1413068	
LOC100857512	*death-associated protein kinase 2-like*		1413151	1416561	
SH2D2A	*SH2 domain containing 2A*		1416567	1421255	
PRCC	*papillary renal cell carcinoma (translocation-associated)*		1421164	1430083	
HDGF	*hepatoma-derived growth factor*		1432110	1437580	
MRPL24	*mitochondrial ribosomal protein L24*		1438014	1439195	
RRNAD1	*ribosomal RNA adenine dimethylase domain containing 1*		1439293	1442440	
CRABP2	*cellular retinoic acid binding protein 2*		1448446	1451746	
LOC425431	*dnaJ homolog subfamily A member 1-like*		1455339	1458566	
NES	*nestin*		1462876	1470387	
BCAN	*brevican*		1473720	1486466	
HAPLN2	*hyaluronan and proteoglycan link protein 2*		1487083	1489801	
RHBG	*Rh family B glycoprotein*	25	1490958	1496016	*rs317288536, rs312758346*
LOC107055112	*uncharacterized LOC107055112*		1505796	1508625	
LOC107055111	*uncharacterized LOC107055111*		1521253	1544127	
MEF2D	*myocyte enhancer factor 2D*		1557855	1582946	
LOC107055110	*uncharacterized LOC107055110*		1598470	1599898	
LOC101750487	*uncharacterized LOC101750487*		1609167	1620210	
LOC101750716	*uncharacterized LOC101750716*		1620907	1627137	
LOC107055114	*E3 SUMO-protein ligase PIAS3-like*		1636879	1651430	
MIR6662	*microRNA 6662*		1637951	1638060	
INTS3	*integrator complex subunit 3*		1654016	1685697	
LOC107055115	*atrial natriuretic peptide receptor 1-like*		1686320	1696310	
LOC107055116	*atrial natriuretic peptide receptor 1-like*		1696623	1700257	
ILF2	*interleukin enhancer binding factor 2*		1701006	1705564	
SNAPIN	*SNAP associated protein*		1705771	1706895	
IL6R	*interleukin 6 receptor*		1708497	1714085	

SHE	*Src homology 2 domain containing E*		1715301	1720734	
UBE2Q1	*ubiquitin conjugating enzyme E2 Q1*		1722132	1729019	
CHRNB2	*cholinergic receptor nicotinic beta 2 subunit*		1729683	1734479	
ADAR	*adenosine deaminase, RNA specific*		1735473	1747289	
KCNN3	*potassium calcium-activated channel subfamily N member 3*		1758811	1780362	
PMVK	*phosphomevalonate kinase*		1781459	1784289	
PBXIP1	*PBX homeobox interacting protein 1*		1784644	1788641	
PYGO2	*pygopus family PHD finger 2*		1788639	1790498	
SHC1	*SHC adaptor protein 1*		1790787	1800935	
CKS1B	*CDC28 protein kinase regulatory subunit 1B*		1801177	1802314	
FLAD1	*flavin adenine dinucleotide synthetase 1*		1802652	1808111	
ZBTB7B	*zinc finger and BTB domain containing 7B*		1812250	1830425	
DCST2	*DC-STAMP domain containing 2*		1832750	1837666	
SMAD4	*SMAD family member 4*		1838974	1844533	
CHTOP	*chromatin target of PRMT1*		1847078	1853476	
S100A1	*S100 calcium binding protein A1*		1853739	1856049	
S100A13	*S100 calcium binding protein A13*		1857835	1859090	
S100A14	*S100 calcium binding protein A14*		1861081	1863210	
S100A16	*S100 calcium binding protein A16*		1865897	1868591	
S100A4	*S100 calcium binding protein A4*		1869231	1871230	
S100A6	*S100 calcium binding protein A6*		1874323	1875575	
LOC101747386	*protein S100-A9-like*		1877071	1878016	
S100A9	*S100 calcium binding protein A9*		1885186	1886621	
EDKM	*epidermal differentiation protein containing a KKLIQQ motif*		1892914	1895414	
EDQM1	*epidermal differentiation protein containing a glutamine (Q) motif 1*		1895773	1896542	
EDQM2	*epidermal differentiation protein containing a glutamine (Q) motif 2*		1899100	1900320	

EDWM	*epidermal differentiation protein containing WYDP motif*		1906417	1907809	
EDCH5	*epidermal differentiation protein containing cysteine histidine motifs 5*		1909483	1911230	
EDMPN1	*epidermal differentiation protein containing a MPN sequence motif 1*		1912397	1913451	
EDCRP	*epidermal differentiation cysteine-rich protein*		1919906	1922026	
EDCH4	*epidermal differentiation protein containing cysteine histidine motifs 4*		1931045	1932007	
EDGH	*epidermal differentiation protein rich in glycine and histidine*		1940110	1942027	
LOR1	*loricrin 1*		1943612	1951026	
LOR2	*loricrin 2*		1943846	1946131	
LOR3	*loricrin 3*		1952705	1955656	
EDMTF4	*epidermal differentiation protein starting with MTF motif 4*		1960713	1980545	
KRTAP9-1L	*keratin-associated protein 9-1-like*		1982162	1984253	
EDMTF2	*epidermal differentiation protein starting with MTF motif 2*		1987787	1988430	
EDMTF1	*epidermal differentiation protein starting with MTF motif 1*		1996030	1996879	
EDMTF3	*epidermal differentiation protein starting with MTF motif 3*		2001003	2001421	
LOC771066	*claw keratin-like*		2003701	2013388	
LOC771082	*claw keratin-like*		2005571	2006553	*rs312758346*
LOC101748164	*claw keratin-like*		2008222	2013410	
LOC768967	*claw keratin-like*		2009974	2010930	
LOC426217	*claw keratin-like*		2012493	2013457	
LOC100858504	*claw keratin-like*		2014658	2015546	
LOC430658	*claw keratin-like*		2017230	2018201	
KRTAP19-2	*keratin associated protein 19-2*		2019017	2019672	
LOC100858728	*claw keratin-like*		2021632	2026949	

LOC107049025	*claw keratin-like*		2023424	2024386	
LOC395095	*keratin*		2026029	2031235	
LOC107049026	*claw keratin-like*		2027847	2028770	
LOC430661	*claw keratin-like*		2030369	2031296	
LOC426218	*claw keratin-like*		2032190	2033529	
LOC769139	*feather keratin 1-like*		2035478	2057444	
LOC107055121	*uncharacterized LOC107055121*		2035923	2037911	
LOC107055122	*uncharacterized LOC107055122*		2044316	2047342	
LOC107055123	*uncharacterized LOC107055123*		2049190	2052675	
LOC100858797	*feather keratin 1-like*		2059093	2060269	
LOC769121	*feather keratin 1-like*		2062522	2063697	
LOC107055120	*feather keratin 1-like*		2062528	2066628	
LOC107055119	*feather keratin 1-like*		2068407	2069564	
CRTC2	*CREB regulated transcription coactivator 2*		2074821	2076284	
SLC39A1	*solute carrier family 39 member 1*		2076700	2079590	
JTB	*jumping translocation breakpoint*		2079695	2081048	
RPS27	*ribosomal protein S27*		2081137	2082137	
NUP210L	*nucleoporin 210 like*		2082134	2103328	
TPM3	*tropomyosin 3*		2103498	2119531	
C25H1ORF43	*chromosome 25 open reading frame, human C1orf43*		2120011	2126177	
UBAP2L	*ubiquitin associated protein 2 like*		2126287	2148687	
MIR3536	*microRNA 3536*		2139769	2139846	
HAX1	*HCLS1 associated protein X-1*		2148978	2150830	
AQP10	*aquaporin 10*		2150925	2154431	
ATP8B2	*ATPase phospholipid transporting 8B2*		2154823	2162921	
FAM189B	*family with sequence similarity 189 member B*		2163300	2168280	
SCAMP3	*secretory carrier membrane protein 3*		2168436	2172913	
FDPS	*farnesyl diphosphate synthase*		2170966	2195061	
CLK2	*CDC like kinase 2*		2173878	2182893	

HCN3	*hyperpolarization activated cyclic nucleotide gated potassium channel 3*		2183086	2190519	
RUSC1	*RUN and SH3 domain containing 1*		2195161	2201835	
MIR1629	*microRNA 1629*		2201879	2201971	
ASH1L	*ASH1 like histone lysine methyltransferase*		2205058	2244183	
DAP3	*death associated protein 3*		2244309	2252717	
LOC101750628	*uncharacterized LOC101750628*		2245186	2247152	
LOC107055145	*uncharacterized LOC107055145*		2247061	2249304	
MSTO1	*misato 1, mitochondrial distribution and morphology regulator*		2252826	2256364	
GON4L	*gon-4 like*		2256355	2276509	
SYT11	*synaptotagmin 11*		2276722	2287470	
MIR1752	*microRNA 1752*		2277996	2278077	
RIT1	*Ras like without CAAX 1*		2291212	2296654	
KIAA0907	*KIAA0907*		2296753	2306074	
GOLPH3L	*golgi phosphoprotein 3 like*		2306717	2324088	
ENSA	*endosulfine alpha*		2324712	2328959	
MCL1	*BCL2 family apoptosis regulator*		2331554	2337280	
ECM1	*extracellular matrix protein 1*		2338082	2342192	
TARS2	*threonyl-tRNA synthetase 2, mitochondrial (putative)*		2343309	2348907	
RPRD2	*regulation of nuclear pre-mRNA domain containing 2*		2350884	2367837	
PRPF3	*pre-mRNA processing factor 3*		2368227	2381286	
MRPS21	*mitochondrial ribosomal protein S21*		2381556	2383129	
CIART	*circadian associated repressor of transcription*		2384018	2386610	
CA14	*carbonic anhydrase 14*		2387380	2390854	
APH1A	*aph-1 homolog A, gamma-secretase subunit*		2391340	2394697	
LOC101751679	*uncharacterized LOC101751679*		2394867	2398327	
LOC101751615	*uncharacterized LOC101751615*		2398382	2402350	
LOC107055146	*uncharacterized LOC107055146*		2403157	2406137	

LOC770126	*uncharacterized LOC770126*		2406155	2409682	
FCRL4	***Fc receptor like 4***		2411309	2419676	
MIR6646-1	*microRNA 6646-1*		2411755	2411864	
LOC107055147	*uncharacterized LOC107055147*		2433853	2447509	
LOC107055148	*uncharacterized LOC107055148*		2451394	2454004	
LOC107055149	*uncharacterized LOC107055149*		2467577	2473794	
CADM3	*cell adhesion molecule 3*		2483418	2499486	
ACKR1	*atypical chemokine receptor 1 (Duffy blood group)*		2499575	2501683	
CRP	*C-reactive protein*		2502854	2503992	
LOC776376	*C-reactive protein, pentraxin-related*		2504446	2506092	
DUSP23	*dual specificity phosphatase 23*		2506498	2509327	
IGSF9	*immunoglobulin superfamily member 9*		2510347	2528542	
CFAP45	*cilia and flagella associated protein 45*		2529249	2539783	
CFAP126	*cilia and flagella associated protein 126*		2544123	2546852	
B4GALT3	*beta-1,4-galactosyltransferase 3*		2547966	2550190	
LOC100858962	*palmitoyltransferase ZDHHC3-like*		2550458	2553482	
ADAMTS4	*ADAM metallopeptidase with thrombospondin type 1 motif 4*		2553301	2560072	
NDUFS2	*NADH:ubiquinone oxidoreductase core subunit S2*		2562820	2567328	
FCER1G	*Fc fragment of IgE receptor Ig*		2567666	2590029	
LOC107052256	*Fc receptor-like protein 1*		2570713	2573739	
LOC101748032	*Fc receptor-like protein 3-like*		2574091	2575912	
SSR2	*signal sequence receptor subunit 2*		2576778	2580237	
CTSS	*cathepsin S*		2580937	2585626	
CTSK	*cathepsin K*		2586099	2589028	
ARNT	*aryl hydrocarbon receptor nuclear translocator*		2589829	2610524	
SETDB1	*SET domain bifurcated 1*		2611398	2626515	
CERS2	*ceramide synthase 2*		2627398	2631003	
FAM63A	*family with sequence similarity 63 member A*		2631656	2638310	

PRUNE1	*prune exopolyphosphatase*		2638573	2642059	
CDC42SE1	*CDC42 small effector 1*		2642228	2645533	
MLLT11	*myeloid/lymphoid or mixed-lineage leukemia; translocated to, 11*		2646767	2648990	
LOC107052255	*uncharacterized LOC107052255*		2648019	2649834	
GABPB2	*GA binding protein transcription factor beta subunit 2*		2649883	2657535	
SEMA6C	*semaphorin 6C*		2658129	2664668	
LYSMD1	*LysM domain containing 1*		2665269	2666886	
SCNM1	*sodium channel modifier 1*		2666824	2669477	
TMOD4	*tropomodulin 4*		2669480	2673057	
LOC107052254	*uncharacterized LOC107052254*		2672435	2673504	
VPS72	*vacuolar protein sorting-associated protein 72 homolog*		2673468	2675777	
PIP5K1A	*phosphatidylinositol-4-phosphate 5-kinase type 1 alpha*		2676126	2692345	
PSMD4	*proteasome 26S subunit, non-ATPase 4*		2694169	2698198	
ZNF687	*zinc finger protein 687*		2699044	2710407	
PI4KB	*phosphatidylinositol 4-kinase beta*		2710343	2727313	
RFX5	*regulatory factor X5*		2727953	2732051	
SELENBP1L3	*selenium-binding protein 1-like 3*		2732264	2738800	
SELENBP1L2	*selenium-binding protein 1-like 2*		2740745	2745670	
LOC100857190	*selenium-binding protein 1-like*		2751200	2755777	
SELENBP1	*selenium binding protein 1*		2758037	2768146	
POGZ	*pogo transposable element with ZNF domain*		2772407	2798858	
PSMB4	*proteasome subunit beta 4*		2799648	2801947	
MIR6620	*microRNA 6620*		2809804	2809913	
TUFT1	*tuftelin 1*		2816283	2829115	
SNX27	*sorting nexin family member 27*		2829866	2845968	
CELF3	*CUGBP, Elav-like family member 3*		2847235	2856792	
RIIAD1	*regulatory subunit of type II PKA R-subunit (RIIa) domain containing 1*		2857292	2859404	

MRPL9	*mitochondrial ribosomal protein L9*		2859728	2862765	
TDRKH	*tudor and KH domain containing*		2862971	2867558	
RORC	*RAR related orphan receptor C*		2868173	2875867	
THEM4	*thioesterase superfamily member 4*		2877049	2879971	
S100A10	*S100 calcium binding protein A10*		2880417	2882430	
DEDD	*death effector domain containing*		2883775	2893262	
LOC107052253	*uncharacterized LOC107052253*		2889147	2891723	
NIT1	*nitrilase 1*		2893164	2895968	
PFDN2	*prefoldin subunit 2*		2896039	2897647	
ATP6V1AL	*ATPase H+ transporting V1 subunit A-like*		2898067	2904318	
TRNAR-UCU	*transfer RNA arginine (anticodon UCU)*		5519260	5519352	
SLC16A1	*solute carrier family 16 member 1*		3611036	3625979	
MIR1669	*microRNA 1669*		3626430	3626520	
LRIG2	*leucine rich repeats and immunoglobulin like domains 2*		3637024	3655454	
MAGI3	*membrane associated guanylate kinase, WW and PDZ domain containing 3*		3667646	3718891	
PHTF1	*putative homeodomain transcription factor 1*		3718661	3729714	
RSBN1	*round spermatid basic protein 1*		3729822	3746251	
PTPN22	*protein tyrosine phosphatase, non-receptor type 22*		3746522	3760097	
BCL2L15	*BCL2 like 15*	26	3760199	3762959	*rs317627533*
AP4B1	*adaptor related protein complex 4 beta 1 subunit*		3763580	3770851	
DCLRE1B	*DNA cross-link repair 1B*		3770574	3773420	
HIPK1	*homeodomain interacting protein kinase 1*		3777445	3801550	
OLFML3	*olfactomedin like 3*		3801747	3803342	
SYT6	*synaptotagmin 6*		3808772	3837467	
TRIM33	*tripartite motif containing 33*		3844903	3869654	
BCAS2	*breast carcinoma amplified sequence 2*		3869908	3872919	
DENND2C	*DENN domain containing 2C*		3872935	3887414	
AMPD1	*adenosine monophosphate deaminase 1*		3892276	3902376	

NRAS	*neuroblastoma RAS viral oncogene homolog*		3906425	3912827	
CSDE1	*cold shock domain containing E1*		3912971	3930268	
SIKE1	*suppressor of IKBKE 1*		3930830	3935439	
BARL	*bile acid receptor-like*		3941128	3952591	
SYCP1	*synaptonemal complex protein 1*		3951965	3966955	
LOC107049139	*synaptonemal complex protein 1-like*		3966969	3974901	
TSHB	*thyroid stimulating hormone beta*		3974272	3987526	
TSPAN2	*tetraspanin 2*		3985688	4005584	
LOC101747848	*uncharacterized LOC101747848*		4007609	4013783	
NGF	*nerve growth factor*		4027894	4050872	
LOC101747895	*uncharacterized LOC101747895*		4059614	4065666	
LOC101747934	*uncharacterized LOC101747934*		4066019	4077162	
FANCE	*Fanconi anemia complementation group E*		4080342	4084706	
MKRN3	*makorin ring finger protein 3*		4084828	4086964	
PPARD	*peroxisome proliferator activated receptor delta*		4089638	4106338	
DEF6	*DEF6, guanine nucleotide exchange factor*		4108781	4120796	
ZNF76	*zinc finger protein 76*		4121087	4130610	
RPL10A	*ribosomal protein L10a*		4130666	4134401	
SCUBE3	*signal peptide, CUB domain and EGF like domain containing 3*		4135277	4165701	
TCP11	*t-complex 11*		4177656	4188605	
ANKS1A	*ankyrin repeat and sterile alpha motif domain containing 1A*		4186369	4271993	
LOC107055188	*uncharacterized LOC107055188*		4243866	4248941	
TAF11	*TATA-box binding protein associated factor 11*		4272574	4276039	
UHRF1BP1	*UHRF1 binding protein 1*		4276310	4303901	
SNRPC	*small nuclear ribonucleoprotein polypeptide C*		4306380	4310436	
C26H6orf106	*chromosome 26 C6orf106 homolog*		4312154	4344017	
SPDEF	*SAM pointed domain containing ETS transcription factor*		4353136	4359030	

PACSIN1	*protein kinase C and casein kinase substrate in neurons 1*		4359838	4375042	
RPS10	*ribosomal protein S10*		4376915	4382439	
NUDT3	*nudix hydrolase 3*		4384291	4411296	
LOC100858737	*uncharacterized LOC100858737*		4412794	4414947	
HMGA1	*high mobility group AT-hook 1*		4415782	4421656	
LOC107055185	*uncharacterized LOC107055185*		4421016	4430830	
GRM4	*glutamate receptor, metabotropic 4*		4434279	4475476	
LOC101750261	*uncharacterized LOC101750261*		4539410	4571201	
OPN1MSW	*opsin, green sensitive (rhodopsin-like)*		4557195	4562805	
MLN	*motilin*		4573764	4580079	
LEMD2	***LEM domain containing 2***		4584641	4597668	
LOC107055184	***uncharacterized LOC107055184***		4597433	4601335	
IP6K3	*inositol hexakisphosphate kinase 3*		4601290	4614097	
C26H6ORF125	*chromosome 26 open reading frame, human C6orf125*		4615609	4620391	
ITPR3	*inositol 1,4,5-trisphosphate receptor type 3*		4619903	4659888	
LOC768477	*uncharacterized LOC768477*		4669560	4675843	
BAK1	*BCL2 antagonist/killer 1*		4677808	4689669	
LOC107055182	*uncharacterized LOC107055182*		4690003	4698367	
TSPO2	*translocator protein 2*		4698990	4704107	
LOC107055181	*uncharacterized LOC107055181*		4707001	4710843	
APOBEC2	*apolipoprotein B mRNA editing enzyme catalytic subunit 2*		4710803	4718654	
OARD1	*O-acyl-ADP-ribose deacylase 1*		4719478	4722726	
LOC107055164	*glycine-rich protein DOT1-like*		4722793	4723646	
NFYA	*nuclear transcription factor Y subunit alpha*		4723085	4737464	
LOC100858470	*uncharacterized LOC100858470*		4737878	4774348	
TREM-B1	*triggering receptor expressed on myeloid cells B1*		4741376	4747862	
TREM2	*triggering receptor expressed on myeloid cells 2*		4749175	4753562	
TREM-B2	*triggering receptor expressed on myeloid cells B2*		4755477	4761898	

LOC107055180	*uncharacterized LOC107055180*		4782696	4797766	
LOC107055165	*uncharacterized LOC107055165*		4802076	4806903	
FOXP4L	*forkhead box protein P4-like*		4844332	4887023	
MDFI	*MyoD family inhibitor*		4891121	4906184	
TFEB	*transcription factor EB*		4925985	4940293	
GASTL	*gastricsin-like*		4942094	4945050	
PGC	*progastricsin (pepsinogen C)*		4946918	4950762	
FRS3	*fibroblast growth factor receptor substrate 3*		4952522	4965746	
PRICKLE4	*prickle planar cell polarity protein 4*		4966313	4973328	
LOC101749017	*platelet binding protein GspB-like*		4973531	4985366	
TOMM6	*translocase of outer mitochondrial membrane 6*		4985455	4986384	
USP49	*ubiquitin specific peptidase 49*		4986408	5019177	
LOC107055176	*uncharacterized LOC107055176*		5019458	5022472	
MED20	*mediator complex subunit 20*		5022173	5026778	
BYSL	*bystin like*		5026764	5031080	
CCND3	*cyclin D3*		5030661	5069048	
TAF8	*TATA-box binding protein associated factor 8*		5069068	5076732	
PIFO	*primary cilia formation*		5078237	5081253	
CHIA-M31	*chitinase-M31, acidic*		5081386	5086043	
CHIA	*chitinase, acidic*		5088647	5092767	
LOC768786	*acidic mammalian chitinase-like*		5095653	5100518	
LOC107055174	*uncharacterized LOC107055174*		5107408	5111368	
LOC107055171	*uncharacterized LOC107055171*		5120914	5123009	
BTG2	*BTG anti-proliferation factor 2*		5123154	5127216	
LOC107055173	*uncharacterized LOC107055173*		5128663	5129912	
LOC107055172	*uncharacterized LOC107055172*		5129960	5131921	
FMOD	*fibromodulin*		5133650	5140211	
LOC107055169	*uncharacterized LOC107055169*		5153629	5170330	
PRELP	*proline and arginine rich end leucine rich repeat protein*		5163841	5174857	

OPTC	*opticin*		5176849	5180331	
ATP2B4	*ATPase plasma membrane Ca2+ transporting 4*		5210403	5247582	
LOC107055168	*uncharacterized LOC107055168*		5245482	5259467	
MIR7454	*microRNA 7454*		5270406	5270459	
LOC107055210	*uncharacterized LOC107055210*		506	1277	
LOC107049042	*olfactory receptor 4M1-like*		7537	8651	
LOC768958	*olfactory receptor 6B1-like*		19365	20677	
MROH8	*maestro heat like repeat family member 8*		28315	37726	
LOC107055211	*uncharacterized LOC107055211*		47117	47583	
LOC101751094	*uncharacterized LOC101751094*		58462	62958	
LOC107049117	*uncharacterized LOC107049117*		66251	68943	
LOH11CR2A	*loss of heterozygosity, 11, chromosomal region 2, gene A*		160440	171210	
LOC107055212	*uncharacterized LOC107055212*		160850	166792	
DAD1	*defender against cell death 1*		172289	175051	
IGHVL	*Ig heavy chain Mem5-like*		181398	731094	
LOC101750797	*immunoglobulin omega chain-like*		224224	584730	
LOC101748259	*uncharacterized LOC101748259*	27	254315	268854	*rs314452928*
LOC101748117	*Ig heavy chain V region C3-like*		275201	277549	
LOC107055220	*uncharacterized LOC107055220*		294205	319579	
LOC107055218	*Ig heavy chain V region C3-like*		330149	331828	
IGHVC3L	*Ig heavy chain V region C3-like*		354463	409472	
LOC107055215	*uncharacterized LOC107055215*		371255	377432	
LOC101750872	*Ig heavy chain V region C3-like*		378419	382678	
LOC107055221	*uncharacterized LOC107055221*		390237	392841	
LOC101747602	*Ig heavy chain V region C3-like*		437483	475144	
MIR6644-2	*microRNA 6644-2*		442675	442784	
LOC101747562	*Ig heavy chain V region G4-like*		444766	446020	
MIR6644-1	*microRNA 6644-1*		450920	451029	
LOC100857271	*Ig heavy chain V region G4-like*		481943	483748	

LOC107055242	*uncharacterized LOC107055242*		484823	490120	
TCRA	*T-cell receptor V alpha*		490904	494158	
LOC101752222	*uncharacterized LOC101752222*		499474	507239	
LOC101752169	*uncharacterized LOC101752169*		507517	511586	
LOC107055244	*uncharacterized LOC107055244*		529451	539743	
LOC107055227	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		535417	536596	
LOC107055236	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		542651	543854	
LOC107055225	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		542810	562235	
LOC107055214	*uncharacterized LOC107055214*		549390	555126	
LOC107055238	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		549746	550086	
LOC107055217	*uncharacterized LOC107055217*		562035	563898	
LOC107055240	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		568703	569448	
LOC107055233	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		576637	577512	
LOC107055248	*uncharacterized LOC107055248*		577639	581534	
LOC769422	*T-cell receptor alpha chain V region 2B4-like*		588220	592260	
LOC107055223	*Ig lambda chain V-I region EPS-like*		592321	593573	
LOC770434	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		599735	600920	
TRAV2B4L	*T-cell receptor alpha chain V region 2B4-like*		604190	607247	
LOC107055231	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		613699	614616	
LOC107055247	*uncharacterized LOC107055247*		614641	622721	
LOC101751813	*uncharacterized LOC101751813*		624438	626417	
LOC429206	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		633011	633442	
LOC107055229	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		642194	643374	
LOC107055245	*uncharacterized LOC107055245*		643179	644193	
LOC100859830	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		650260	651359	
LOC107055216	*uncharacterized LOC107055216*		653782	685699	
LOC107055224	*Ig lambda chain V-II region BO-like*		658586	660512	
LOC107055243	*uncharacterized LOC107055243*		666775	670901	
LOC107055232	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		673821	674992	

LOC107055219	*uncharacterized LOC107055219*		684935	686381	
LOC107055234	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		707411	708631	
LOC107055213	*uncharacterized LOC107055213*		711599	713493	
LOC107055230	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		721439	722626	
LOC107055239	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		726214	727417	
LOC769926	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		742659	743831	
LOC425854	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		760032	760903	
LOC107055222	*T-cell receptor alpha chain V region 2B4-like*		764282	766324	
LOC107055228	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		774697	775550	
LOC107055241	*uncharacterized LOC107055241*		782172	799004	
LOC107055237	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		783652	784840	
LOC101749259	*immunoglobulin omega chain-like*		787662	788394	
LOC107055246	*uncharacterized LOC107055246*		788821	790251	
LOC107055235	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		794600	795702	
LOC101750937	*immunoglobulin iota chain-like*		807203	810694	
LOC107055251	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		816809	817994	
LOC107049059	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		821387	822603	
LOC427060	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		826840	828009	
LOC101749128	*immunoglobulin omega chain-like*		827802	829273	
LOC107055250	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		832169	833270	
LOC107049028	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		836471	837684	
LOC107055249	*Ig lambda chain V-V region DEL-like*		839810	841237	
LOC107055252	*uncharacterized LOC107055252*		841664	844422	
LOC107055253	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		848755	849857	
LOC107055254	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		858372	859593	
LOC107055255	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		869063	870093	
LOC107055258	*feather keratin Cos1-2-like*		873292	874517	
LOC425497	*T-cell receptor alpha chain V region 2B4-like*		877700	880735	
FKCOSL	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		886974	888161	

LOC770684	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		891650	892662	
TCRAV2B4L	*T-cell receptor alpha chain V region 2B4-like*		894657	916906	
LOC107055256	*uncharacterized LOC107055256*		898181	899322	
FK21	*feather keratin 21*		906726	907819	
LOC107055257	*uncharacterized LOC107055257*		911214	911732	
LOC430902	*T-cell receptor alpha chain V region CTL-L17-like*		921699	923034	
LOC107055261	*uncharacterized LOC107055261*		932708	933596	
LOC107055263	*feather keratin Cos1-2-like*		950730	954092	
LOC428299	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		958207	959601	
LOC428298	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		967848	968775	
LOC428297	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		972474	973601	
LOC100858427	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		980257	981451	
LOC107055265	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		986134	987373	
LOC107055266	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		989578	990839	
LOC107055264	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		992970	993892	
LOC107055269	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		996105	997366	
LOC428295	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		999181	1000420	
LOC107055268	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		1002631	1003892	
LOC100859466	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		1005986	1013849	
LOC107055267	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		1006024	1006946	
LOC107055270	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		1009162	1010150	
LOC428293	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		1021096	1022276	
LOC107055262	*uncharacterized LOC107055262*		1030898	1032082	
LOC428291	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*		1046023	1047177	
LOC107055272	*feather keratin 3-like*		1058299	1069666	
LOC100859500	*feather keratin 3-like*		1062921	1063981	
LOC428289	*feather keratin 3-like*		1068928	1069257	
TRNAI-UAU	*transfer RNA isoleucine (anticodon UAU)*	27	4520423	4520513	*rs315329074*
TRNAQ-UUG	*transfer RNA glutamine (anticodon UUG)*		3640823	3640894	

MEOX1	*mesenchyme homeobox 1*		3530597	3536176	
ETV4	*ETS variant 4*		3548537	3564258	
DHX8	*DEAH-box helicase 8*		3567743	3579115	
PHB	*prohibitin*		3580883	3585030	
LOC101750197	*uncharacterized LOC101750197*		3582478	3592058	
ZNF652	*zinc finger protein 652*		3592238	3620525	
PHOSPHO1	*phosphoethanolamine/phosphocholine phosphatase*		3623753	3631815	
ABI3	*ABI family member 3*		3631245	3637401	
GNGT2	*G protein subunit gamma transducin 2*		3637559	3639810	
IGF2BP1	*insulin like growth factor 2 mRNA binding protein 1*		3648588	3672763	
GIP	*gastric inhibitory polypeptide*		3682114	3689763	
SNF8	*SNF8, ESCRT-II complex subunit*		3690040	3693337	
UBE2Z	*ubiquitin conjugating enzyme E2 Z*		3693620	3705485	
ATP5G1	*ATP synthase, H+ transporting, mitochondrial Fo complex subunit C1 (subunit 9)*		3707200	3709533	
CALCOCO2	*calcium binding and coiled-coil domain 2*		3709693	3719291	
HOXB13	*homeobox B13*		3742876	3745452	
LOC107055286	*uncharacterized LOC107055286*		3764713	3778499	
MIR196A1	*microRNA 196a-1*		3775036	3775130	
HOXB9	*homeobox B9*		3781877	3788281	
HOXB8	*homeobox B8*		3795352	3796999	
HOXB7	*homeobox B7*		3798149	3803481	
LOC107055284	*uncharacterized LOC107055284*		3810034	3821988	
HOXB6	*homeobox B6*		3812716	3815203	
HOXB5	*homeobox B5*		3818138	3820563	
MIR10A	*microRNA 10a*		3834170	3834243	
HOXB4	*homeobox B4*		3835207	3840489	
LOC107055283	*uncharacterized LOC107055283*		3837488	3843542	
HOXB3	*homeobox B3*		3840562	3864857	

LOC107055285	*uncharacterized LOC107055285*		3852258	3856707	
HOXB2	*homeobox B2*		3867101	3871833	
LOC107055282	*uncharacterized LOC107055282*		3868775	3870699	
HOXB1	*homeobox B1*		3879443	3882191	
LOC419994	*src kinase-associated phosphoprotein 1-like*		3927113	3969088	
LOC101751838	*uncharacterized LOC101751838*		3980545	3984567	
TBKBP1	*TBK1 binding protein 1*		4013103	4024426	
KPNB1	*karyopherin subunit beta 1*		4029274	4049551	
NPEPPS	*aminopeptidase puromycin sensitive*		4052247	4080818	
MRPL45	*mitochondrial ribosomal protein L45*		4083516	4087147	
GPR179	*G protein-coupled receptor 179*		4089962	4097067	
SOCS7	*suppressor of cytokine signaling 7*		4098027	4106014	
SKAP1	*src kinase associated phosphoprotein 1*		4120634	4160530	
SNX11	*sorting nexin 11*		4162424	4168386	
CBX1	*chromobox 1*		4168538	4176552	
NFE2L1	*nuclear factor, erythroid 2 like 1*		4179406	4188312	
LOC107055292	*uncharacterized LOC107055292*		4188341	4189844	
CDK5RAP3	*CDK5 regulatory subunit associated protein 3*		4192066	4195013	
PRR15L	*proline rich 15 like*		4195737	4196844	
PNPO	*pyridoxamine 5'-phosphate oxidase*		4199228	4201514	
SP2	*Sp2 transcription factor*		4202904	4212077	
SP6	*Sp6 transcription factor*		4219693	4222841	
SCRN2	*secernin 2*		4223661	4226254	
LRRC46	*leucine rich repeat containing 46*		4226229	4229327	
MRPL10	*mitochondrial ribosomal protein L10*		4229306	4231308	
OSBPL7	*oxysterol binding protein like 7*		4231743	4238314	
TBX21	*T-box 21*		4238676	4245894	
ARHGAP23	*Rho GTPase activating protein 23*		4260935	4274769	
SRCIN1	*SRC kinase signaling inhibitor 1*		4278768	4318568	

LOC107055293	*SKI/DACH domain-containing protein 1-like*		4356552	4360423	
MIR6663	*microRNA 6663*		4371072	4371181	
MLLT6	*MLLT6, PHD finger domain containing*		4374478	4403841	
MIR1735	*microRNA 1735*		4391229	4391307	
LOC107055294	*polycomb group RING finger protein 2-like*		4405220	4407907	
LOC107055296	*protein AF-17-like*		4410156	4419961	
CISD3	*CDGSH iron sulfur domain 3*		4421720	4422397	
PCGF2	*polycomb group ring finger 2*		4422951	4428429	
LOC107055298	*POU domain, class 3, transcription factor 3-like*		4425831	4427935	
PSMB3	*proteasome subunit beta 3*		4430285	4433174	
PIP4K2B	*phosphatidylinositol-5-phosphate 4-kinase type 2 beta*		4434946	4450427	
CWC25	*CWC25 spliceosome associated protein homolog*		4451015	4459377	
RPL23	*ribosomal protein L23*		4463064	4465066	
LASP1	*LIM and SH3 protein 1*		4466693	4486609	
FBXO47	*F-box protein 47*		4490226	4500566	
LOC101749109	*uncharacterized LOC101749109*		4498009	4506617	
PLXDC1	*plexin domain containing 1*		4505312	4518001	
LOC100858629	*dickkopf-related protein 1-like*		4519315	4521946	
CACNB1	***calcium voltage-gated channel auxiliary subunit beta 1***		4521859	4534179	
RPL19	*ribosomal protein L19*		4534596	4536944	
FBXL20	*F-box and leucine rich repeat protein 20*		4537208	4558569	
MED1	*mediator complex subunit 1*		4558906	4572538	
CDK12	*cyclin dependent kinase 12*		4572918	4595934	
NEUROD2	*neuronal differentiation 2*		4610541	4613182	
PPP1R1B	*protein phosphatase 1 regulatory inhibitor subunit 1B*		4621254	4626595	
STARD3	*StAR related lipid transfer domain containing 3*		4626814	4643949	
TCAP	*titin-cap*		4644741	4646231	
PNMT	*phenylethanolamine N-methyltransferase*		4647235	4649104	
PGAP3	*post-GPI attachment to proteins 3*		4649381	4652406	

ERBB2	*erb-b2 receptor tyrosine kinase 2*		4653394	4662322	
MIR6547	*microRNA 6547*		4660683	4660802	
MIEN1	*migration and invasion enhancer 1*		4663033	4664878	
GRB7	*growth factor receptor bound protein 7*		4665681	4671472	
LOC100858293	*retinol dehydrogenase 8-like*		4671617	4674052	
IKZF3	*IKAROS family zinc finger 3*		4676390	4700355	
ZPBP2	*zona pellucida binding protein 2*		4700373	4706427	
LRRC3C	*leucine rich repeat containing 3C*		4710386	4714334	
LOC107055297	*basic proline-rich protein-like*		4713858	4715843	
ORMDL3	*ORMDL sphingolipid biosynthesis regulator 3*		4717225	4725376	
GSDMA	*gasdermin A*		4725470	4730952	
PSMD3	*proteasome 26S subunit, non-ATPase 3*		4731389	4735157	
CSF3	*colony stimulating factor 3*		4737417	4739289	
MIR6884	*microRNA 6884*		4741104	4757219	
THRA	*thyroid hormone receptor, alpha*		4764287	4775504	
NR1D1	*nuclear receptor subfamily 1 group D member 1*		4776457	4783436	
MSL1	*male specific lethal 1 homolog*		4787718	4792939	
CASC3	*cancer susceptibility 3*		4793935	4803203	
RAPGEFL1	*Rap guanine nucleotide exchange factor like 1*		4804082	4809927	
WIPF2	*WAS/WASL interacting protein family member 2*		4810433	4822471	
CDC6	*cell division cycle 6*		4822688	4827078	
RARA	*retinoic acid receptor alpha*		4833202	4836052	
GJD3	*gap junction protein delta 3*		4836854	4839029	
TOP2A	*topoisomerase (DNA) II alpha*		4839029	4856730	
LOC101747522	*collagen alpha-1(XVIII) chain-like*		4857355	4863016	
IGFBP4	*insulin like growth factor binding protein 4*		4865485	4870779	
TNS4	*tensin 4*		4870793	4882163	
LOC107055315	*uncharacterized LOC107055315*		4882026	4907719	
CCR7	*C-C motif chemokine receptor 7*		4891171	4899901	

SMARCE1	*SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily e, member 1*		4907121	4921663	
KRT222	*keratin 222*		4920500	4929780	
LOC107055316	*uncharacterized LOC107055316*		4922980	4937276	
KRT12	*keratin 12*		4932629	4939390	
KRT20	*keratin 20*		4939479	4943592	
KRT23	*keratin 23*		4953002	4962692	
KRT15	*keratin 15*		4965385	4969810	
KRT19	*keratin 19*		4972722	4977384	
LOC420043	*keratin 16-like*		4983598	4987746	
LOC100857659	*keratin, type I cytoskeletal 42-like*		4992053	4995428	
KRT10	*keratin, type I cytoskeletal 10-like*		4997475	5002485	
KRT9L	*keratin, type I cytoskeletal 9-like*		5008298	5012957	
LOC772080	*keratin, type I cytoskeletal 17-like*		5016347	5021212	
KRTC42L	*keratin, type I cytoskeletal 42-like*		5025520	5030185	
LOC771995	*keratin, type I cytoskeletal 42-like*		5035336	5038163	
LOC107055312	*uncharacterized LOC107055312*		5040540	5042410	
KRT14	*keratin 14*		5042180	5046007	
KRT17	*keratin 17*		5048947	5052819	
LOC107055313	*uncharacterized LOC107055313*		5051305	5052029	
EIF1	*eukaryotic translation initiation factor 1*		5064906	5066718	
LOC396365	*preprogastrin*		5069166	5069814	
HAP1	*huntingtin associated protein 1*		5069870	5078795	
JUP	*junction plakoglobin*		5081253	5096734	
P3H4	*prolyl 3-hydroxylase family member 4 (non-enzymatic)*		5097792	5102079	
FKBP10	*FK506 binding protein 10*		5102602	5108343	
NT5C3B	*5'-nucleotidase, cytosolic IIIB*		5108609	5113795	
KLHL10	*kelch like family member 10*		5113816	5117579	
KLHL11	*kelch like family member 11*		5117750	5123733	

ACLY	*ATP citrate lyase*		5124218	5152667	
TTC25	*tetratricopeptide repeat domain 25*		5149242	5156676	
CNP	*2',3'-cyclic nucleotide 3' phosphodiesterase*		5156868	5161795	
DNAJC7	*DnaJ heat shock protein family (Hsp40) member C7*		5161978	5181291	
LOC107055311	*uncharacterized LOC107055311*		5172889	5180535	
NKIRAS2	*NFKB inhibitor interacting Ras like 2*		5181433	5183584	
ZNF385C	*zinc finger protein 385C*		5183603	5237098	
ZNF862L	*zinc finger protein 862-like*		5198667	5206571	
LOC107055310	*uncharacterized LOC107055310*		5223122	5224406	
DHX58	*DExH-box helicase 58*		5242809	5249758	
KAT2A	*lysine acetyltransferase 2A*		5249877	5256053	
LOC772158	*heat shock protein 30C-like*		5256460	5257349	
HSPB9	*heat shock protein family B (small) member 9*		5258006	5258998	
RAB5C	*RAB5C, member RAS oncogene family*		5260322	5270725	
KCNH4	*potassium voltage-gated channel subfamily H member 4*		5271367	5280067	
HCRT	*hypocretin neuropeptide precursor*		5280135	5281806	
PIBPPDD4L	*1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase delta-4-like*		5281905	5292750	
GHDC	*GH3 domain containing*		5294204	5298441	
STAT5B	*signal transducer and activator of transcription 5B*		5297863	5311670	
STAT3	*signal transducer and activator of transcription 3*		5319501	5334933	
PTRF	*polymerase I and transcript release factor*		5335537	5350085	
ATP6V0A1	*ATPase H+ transporting V0 subunit a1*		5352689	5381019	
NAGLU	*N-acetylglucosaminidase, alpha*		5381264	5384665	
HSD17B1	*hydroxysteroid 17-beta dehydrogenase 1*		5384880	5386410	
MLX	*MLX, MAX dimerization protein*		5390278	5394161	
PSMC3IP	*PSMC3 interacting protein*		5393341	5397274	
FAM134C	*family with sequence similarity 134 member C*		5397367	5405515	
TUBG1	*tubulin gamma 1*		5405589	5413030	

LOC107055309	*uncharacterized LOC107055309*		5407102	5408044	
PLEKHH3	*pleckstrin homology, MyTH4 and FERM domain containing H3*		5413790	5421146	
CCR10	*C-C motif chemokine receptor 10*		5422621	5427142	
CNTNAP1	*contactin associated protein 1*		5427307	5435737	
EZH1	*enhancer of zeste 1 polycomb repressive complex 2 subunit*		5435970	5451101	
RAMP2	*receptor activity modifying protein 2*		5453137	5454994	
VPS25	*vacuolar protein sorting 25 homolog*		5455180	5457660	
WNK4	*WNK lysine deficient protein kinase 4*		5461875	5473473	
COA3	*cytochrome c oxidase assembly factor 3*		5473589	5474238	
CNTD1	*cyclin N-terminal domain containing 1*		5474297	5478907	
BECN1	*beclin 1*		5478573	5483381	
PSME3	*proteasome activator subunit 3*		5483531	5490236	
AOC3	*amine oxidase, copper containing 3*		5490347	5504360	
G6PC	*glucose-6-phosphatase catalytic subunit*		5506845	5510258	
AARSD1	*alanyl-tRNA synthetase domain containing 1*		5511314	5515677	
PTGES3L	*prostaglandin E synthase 3 (cytosolic)-like*		5515779	5517799	
RUNDC1	*RUN domain containing 1*		5518274	5520926	
RPL27	*ribosomal protein L27*		5521996	5524344	
IFI35	*interferon-induced protein 35*		5524907	5535384	

* Note: Positions are based on Gallus gallus 5.0 genome assembly.

Table S2. Growth related QTL/Associations reported within 1Mb distances from significant SNPs.

SNP ID	GGA	Number of QTL/associations	QTL/association (bp)*	QTL/association type	QTL/association IDs*
rs13923872	1	22	37278942 -128288555	Chest width	16706
			18054807 -171631116	Visceral fat weight	17319
			18054807 -171717298	Total white fat weight	17332
			25724479 -171631116	Subcutaneous neck fat weight	17325
			18054807 -196202543	Body weight	1797
			37278942 -133528161	Body weight (140 days)	55919
			18054807 - 171631116	Body weight (140 days)	17076
			18054807 -171631116	Carcass weight	17110
			37278942 -111865771	Breast muscle percentage	9407
			100051042 -123004362	Breast muscle weight	9410
			2420814 -171631116	Shank length	9409
			106349346 -123007835	Carcass fat content	17119
			113159926 -128288555	Shank weight	14341
			113159926 -128288555	Femur weight	14342
			18054807 -172427968	Spleen weight	1851
			18054807 -171631116	Growth (70-105 days)	55937
			18054807 -168151247	Abdominal fat weight	6858
			18054807 -171631116	Shank length	9294

			113159926 -115848566	Breast muscle weight	13385
			111368640 -164599096	Body weight (35 days)	14355
			6580919 -171631116	Subcutaneous fat thickness	14359
			37278942 -111865771	Abdominal fat percentage	14362
rs312691174	4	37			
			17148380 -81264760	Body weight (168 days)	24875
			17148380 -81264760	Body weight (21 days)	24842
			17148380 -81264760	Body weight (336 days)	24883
			17148380 -81264760	Body weight (42 days)	24855
			17148380 -81264760	Average daily gain	24899, 24905,24911
			17148380 -81264760	Body weight (84 days)	24866
			17148380 -81264760	Body weight (504 days)	24890
			4964691 -87025255	Visceral fat weight	17321
			10768639 -91268419	Average daily gain	24914
			18357474 -31942137	Shank length	9295
			18354191 -37883325	Thigh muscle weight	9395
			18354191 -47647218	Drumstick and thigh muscle weight	13404

			17148380 -81264760	Body weight (day of first egg)	14457, 14464, 14470
			17148380 -81264760	Head percentage	15571
rs15608447	4	14	67546750 -67546790	Body weight (28 days)	65710
			17148380 -81264760	Body weight (168 days)	24875
			47647218 -89464128	Body weight	200,820,152,016
			17148380 -81264760	Body weight (21 days)	24842
			17148380 -81264760	Body weight (336 days)	24883
			47647218 -89464128	Carcass weight	2012
			17148380 -81264760	Body weight (42 days)	24855
			17148380 -81264760	Average daily gain	24899, 24905, 24911
			17148380 -81264760	Body weight (84 days)	24866
			66031512 -66100420	Body weight	3339
			47647218 -89464128	Liver weight	2017
			52604411 -82619142	Growth (14-28 days)	12499
			17148380 -81264760	Body weight (504 days)	24890
			48804413 -85154534	Growth (28-42 days)	12500
			52191247 -89464128	Abdominal fat percentage	9421
			48404949 -82619142	Growth (0-14 days)	12498
			52535768 -70787114	Shank length	9286

			49665708 -85877678	Total white fat weight	17334
			30906204 -83247658	Tibia width	2035
			61970484 -83247658	Tibia width	2038
			62452715 -89022456	Growth (42-56 days)	12501
			4964691 -87025255	Visceral fat weight	17321
			62331035 -82550230	Pectoralis major weight	2041
			31561525 -89318267	Body weight (35 days)	55905
			32974594 -89318267	Growth (0-35 days)	55930
			47647218 -89464128	Drumstick muscle weight	2057
			10768639 -91268419	Average daily gain	24914
			47647218 -89464128	Drumstick weight	2059
			47647218 -89464128	Wing weight	2060
			47647218 -89464128	Body weight (42 days)	9759
			47647218 -89464128	Body weight (63 days)	9760
			47647218 -89464128	Growth (21-42 days)	9761
			47647218 -89464128	Growth (42-63 days)	9762
			62331035 -82299229	Shank length	11795
			47647218 -89464128	Skin fat weight	12636
			47647218 -89464128	Drumstick and thigh muscle weight	13395

			17148380 -81264760	Body weight (day of first egg)	14457, 14464, 14470
rs318199727	10	13	13330009 -13330034	Dressing percentage	57550
			13330009 -13330034	Breast muscle percentage	57551
			12736750 -12736790	Spleen weight	21767, 21787, 21788, 21742, 21743
			12736750 -12736790	Abdominal fat percentage	14489, 14488
			13329989 -13330029	Breast muscle percentage	57547
			13329989 -13330029	Drumstick and thigh muscle percentage	57548
			13329989 -13330029	Abdominal fat percentage	57549
			692555 -20423025	Carcass weight	17113
			1541735 -16171711	Body weight (140 days)	55923
			2357400 -17864188	Body weight (35 days)	55907
			692555 -20423025	Body weight (70 days)	55911
			2552841 -18059263	Growth (0-35 days)	55931
			4410690 -18434155	Body weight (105 days)	55917
rs318098582	11	14			
			1133281 -19983730	Body weight (140 days)	55924

			13523302 -18192592	Body weight (46 days)	6636
			13523302 -18192592	Body weight (112 days)	6637
			13523302 -18192592	Growth (8-46 days)	6638
			13523302 -18192592	Thigh muscle weight	6735
			18193544 -20208550	Thigh meat-to-bone ratio	6736
			18193544 -20208550	Body weight (40 days)	6737
			9752914 -18192592	Tibia width	9337
			953174 -20208550	Body weight (140 days)	17080
			6823128 -20208550	Carcass weight	17114
			12510855 -20208550	Carcass weight	17088
			6910612 -20208550	Spleen weight	2287
			18193544 -18870770	Body weight	2284, 2285
			18642683 -18686657	Growth (8-46 days)	9519
rs317945754	15	21	3731712 -3769767	Spleen weight	2349
			3731712 -3769767	Body weight (42 days)	9727
			3731712 -3769767	Carcass weight	9728
			3731712 -3769767	Spleen percentage	12588
			2812987 -10689472	Body weight (336 days)	24887
			4236686 -4265310	Abdominal fat weight	11995
			1931502 -7215657	Visceral fat weight	17323

			2519182 -7215657	Subcutaneous neck fat weight	17331
			3749008 -7973093	Drumstick and thigh percentage	15586
			3749008 -7973093	Abdominal fat weight	2337
			3749008 -7973093	Abdominal fat percentage	2339, 2340
			3749008 -8228905	Body weight (35 days)	3355
			1931502 -7215657	Total white fat weight	17337
			2812987 -10689472	Liver weight	2348
			2403639 -10689472	Abdominal fat percentage	9450
			1931502 -10689472	Abdominal fat weight	9451
			2812987 -10689472	Abdominal fat weight	2347, 12631
			3749008 -10689472	Body weight (46 days)	6648
			3749008 -10689472	Growth (8-46 days)	6649
			2812987 -10689472	Fat distribution	12645
			1931502 -9638429	Breast muscle weight	9449
rs316794400	22	1	-	Breast muscle percentage	95429
rs317627533	26	6			
			3118976 -4116802	Body weight (28 days)	95418
			1263919 -4918464	Body weight (63 days)	9453

			2499704 -4918464	Shank weight	2383
			4610791 -4624276	Liver percentage	2385
			-	Abdominal fat weight	30883
			4873346 -4886832	Breast muscle weight	6957
rs314452928	27	4	-	Growth (105-140 days)	55944
			404762 -4520058	Wing weight	17109
			54597 -4520058	Growth (0-35 days)	55932
			54597 -4520058	Body weight (35 days)	55906
rs315329074	27	65	3834510 -3834550	Shank length	66068, 66069, 66070
			3363708 -3363748	Shank length	66067
			3971422 -3971462	Shank circumference	66063
			3564173 -3564213	Shank circumference	66065
			3624903 -3624943	Shank circumference	66064, 66066
			3869461 -3869501	Shank length	66071
			3456748 -3456788	Abdominal fat weight	66072
			1798380 -3707375	Abdominal fat weight	11817, 11809
			1798380 -3707375	Abdominal fat percentage	11820
			1798380 -3707375	Carcass fat content	17135, 17126
			1798380 -3707375	Head percentage	15599
			1798380 -3707375	Body weight	2406, 2407

			1798380 -3707375	Body weight (1 day)	7178
			1798380 -3707375	Body weight (41 days)	7186
			1365641 -4520058	Humerus length	2397
			3522988 -3534446	Body weight (112 days)	9521
			3522988 -3534446	Body weight (200 days)	9522
			3522988 -3534446	Growth (46-112 days)	9523
			1365641 -4520058	Body weight	2410
			1809980 -3707375	Body weight (35 days)	3356
			1809980 -3707375	Abdominal fat percentage	3354
			1798380 -3707375	Carcass protein content, dry matter basis	17124
			1798380 -3707375	Carcass fat content, dry matter basis	17125
			3701574 -3713173	Body weight (42 days)	9775
			3701574 -3713173	Growth (21-42 days)	9776
			3701574 -3713173	Body weight (day of first egg)	14459, 14466, 14473
			3701574 -3713173	Body weight (168 days)	24878
			3701574 -3713173	Body weight (336 days)	24888

			3701574 -3713173	Body weight (504 days)	24892
			3701574 -3713173	Average daily gain	24907
			3707375 -3968049	Body weight	2404, 2405
			3707375 -4520058	Thigh weight	2411
			3707375 -4520058	Wing weight	2412
			3788374 -3889766	Drumstick and thigh weight	11920
			3788374 -3889766	Drumstick and thigh percentage	11921
			3788374 -3889766	Abdominal fat percentage	11934
			3788374 -3889766	Pectoralis major percent	11950
			3204318 -4520058	Shank weight	2413
			1798380 -3707375	Body weight (112 days)	6652
			1798380 -3707375	Body weight (200 days)	6653
			1798380 -3707375	Growth (46-112 days)	6654
			1798380 -3379175	Femur length	6778
			1798380 -3707375	Shank weight percentage	15567
			2263107 -4520058	Carcass weight	17116
			404762 -4520058	Wing weight	17109
			2780009 -4520058	Body weight (105 days)	55918
			54597 -4520058	Growth (0-35 days)	55932
			1365641 -4520058	Body weight	2409

			54597 -4520058	Body weight (35 days)	55906
			2454458 -4520058	Body weight (140 days)	55926
			2639460 -4520058	Body weight (70 days)	55912
			3707375 -4520058	Body weight (35 days)	7159
			1850810 -4520058	Carcass weight	17090
			2639460 -4520058	Growth (35-70 days)	55936
			4377710 -4389305	Shank length	9288
			3707375 -4520058	Shank weight percentage	15595
			2390652 -4520058	Breast muscle weight	17096
			3597175 -4520058	Drumstick and thigh weight	17105
			3707375 -4520058	Intramuscular fat	3360
			2141304 -4520058	Body weight (140 days)	17084
			3707375 -4520058	Body weight	2408
			3788374 -5629582	Thigh percentage	30886
			5159872 -5171472	Body weight (56 days)	12395
			5159872 -5171472	Body weight (hatch)	16623
			5159872 -5171472	Body weight (300 days)	16624

*Note: The positions of QTL/associations were mapped to *Gallus gallus_5.0* assembly and thus some regions could not be remapped ("-"). QTL/associations IDs are available in ChickenQTLdb.

Table S3. Prioritized genes by guilt by association gene prioritization analysis.

Rank	Gene ID	Overall p-value	GGA	Markers
1	*SMAD4*	0,000365	25	rs317288536,rs312758346
2	*CHRNB2*	0,000541	25	rs317288536,rs312758346
3	*CDH1*	0,000558	11	rs318098582
4	*NTRK1*	0,000754	25	rs317288536
5	*RARA*	0,00103	27	rs315329074
6	*STAT5B*	0,00113	27	rs315329074
7	*SCARB1*	0,00117	15	rs317945754
8	*NR1D1*	0,00151	27	rs315329074
9	*SHC1*	0,00166	25	rs317288536,rs312758346
10	*CYBB*	0,00171	1	rs13923872
11	*PHB*	0,00173	27	rs315329074
12	*CYBA*	0,00226	11	rs318098582
13	*MC1R*	0,0023	11	rs318098582
14	*CNGA1*	0,00234	4	rs15608447
15	*MED1*	0,00242	27	rs315329074
16	*NGF*	0,00257	26	rs317627533
17	*G6PC*	0,00293	27	rs315329074
18	*NRAS*	0,00295	26	rs317627533
19	*JUP*	0,00314	27	rs315329074
20	*GABRB1*	0,00348	4	rs15608447
21	*LAMTOR2*	0,00371	25	rs317288536
22	*GABRA2*	0,00382	4	rs15608447
23	*ARNT*	0,00391	25	rs312758346
24	*ATP2B4*	0,00427	26	rs317627533
25	*NCSTN*	0,0057	25	rs317288536
26	*TGFA*	0,00578	22	rs316794400
27	*PIP5K1A*	0,00597	25	rs312758346
28	*NFYA*	0,00636	26	rs317627533
29	*CACNB1*	0,00648	27	rs315329074
30	*GABRA4*	0,00656	4	rs15608447

31	*POLG*	0,00664	10	rs318199727
32	*ADRA1A*	0,00669	22	rs316794400
33	*KAT2A*	0,00684	27	rs315329074
34	*MEF2D*	0,00715	25	rs317288536,rs312758346
35	*ATP6V0A1*	0,00776	27	rs315329074
36	*HAX1*	0,00791	25	rs312758346
37	*MEOX1*	0,00809	27	rs315329074
38	*CASK*	0,00817	1	rs13923872
39	*SMARCE1*	0,00838	27	rs315329074
40	*ULK1*	0,00841	15	rs317945754
41	*DDX3X*	0,00841	1	rs13923872
42	*UBC*	0,00855	15	rs317945754
43	*PIEZO1*	0,00881	11	rs318098582
44	*HCN3*	0,00898	25	rs312758346
45	*BECN1*	0,00933	27	rs315329074
46	*MCL1*	0,00968	25	rs312758346
47	*ZFPM1*	0,00977	11	rs318098582
48	*RAN*	0,00986	15	rs317945754
49	*GRM4*	0,0101	26	rs317627533
50	*USP9X*	0,0106	1	rs13923872
51	*PIP4K2B*	0,0106	27	rs315329074
52	*SLC16A1*	0,0112	26	rs317627533
53	*IRF8*	0,0113	11	rs318098582
54	*NTRK3*	0,0114	10	rs318199727
55	*ADAR*	0,0115	25	rs317288536,rs312758346
56	*PPP1R1B*	0,012	27	rs315329074
57	*MED14*	0,012	1	rs13923872
58	*INSRR*	0,0121	25	rs317288536,rs312758346
59	*GIP*	0,0122	27	rs315329074
60	*KPNB1*	0,0123	27	rs315329074
61	*KRT17*	0,0124	27	rs315329074
62	*DPEP1*	0,0125	11	rs318098582

63	*CTSS*	0,0126	25	rs312758346
64	*TXK*	0,0129	4	rs15608447
65	*TPM3*	0,0131	25	rs312758346
66	*FCER1G*	0,0132	25	rs312758346
67	*KRT14*	0,0134	27	rs315329074
68	*CCR7*	0,0136	27	rs315329074
69	*SV2A*	0,0137	25	rs317288536
70	*GABRG1*	0,0138	4	rs15608447
71	*PTRF*	0,01408	27	rs315329074
72	*CDH3*	0,0141	11	rs318098582
73	*CSF3*	0,0143	27	rs315329074
74	*ATP6AP2*	0,0143	1	rs13923872
75	*CCND3*	0,0144	26	rs317627533
76	*THEM4*	0,0145	25	rs312758346
77	*SYT11*	0,0146	25	rs312758346
78	*S100A9*	0,0146	25	rs317288536,rs312758346
79	*STARD3*	0,0147	27	rs315329074
80	*SGCB*	0,0149	4	rs15608447
81	*ACKR1*	0,0152	25	rs312758346
82	*PCGF2*	0,0152	27	rs315329074
83	*KCNH4*	0,0152	27	rs315329074
84	*ECM1*	0,0155	25	rs312758346
85	*ETV4*	0,0157	27	rs315329074
86	*SRCIN1*	0,0158	27	rs315329074
87	*IKZF3*	0,016	27	rs315329074
88	*BCL2A1*	0,0161	10	rs318199727
89	*KCNN3*	0,0161	25	rs317288536,rs312758346
90	*NFAT5*	0,0162	11	rs318098582
91	*CORIN*	0,0164	4	rs15608447
92	*PMVK*	0,0164	25	rs317288536,rs312758346
93	*PI4KB*	0,0165	25	rs312758346
94	*RHCG*	0,0165	10	rs318199727

95	*BAK1*	0,0169	26	rs317627533
96	*VPS4A*	0,0172	11	rs318098582
97	*LRRTM4*	0,0173	22	rs316794400
98	*BCOR*	0,0174	1	rs13923872
99	*IGFBP4*	0,0179	27	rs315329074
100	*GRB7*	0,018	27	rs315329074
101	*WNK4*	0,0181	27	rs315329074
102	*CCR10*	0,0181	27	rs315329074
103	*FMOD*	0,0182	26	rs317627533
104	*SKAP1*	0,0182	27	rs315329074
105	*RLBP1*	0,0185	10	rs318199727
106	*NDUFS2*	0,0185	25	rs312758346
107	*APH1A*	0,0188	25	rs312758346
108	*BCAN*	0,0188	25	rs317288536,rs312758346
109	*RAMP2*	0,0188	27	rs315329074
110	*SLC7A5*	0,019	11	rs318098582
111	*COPA*	0,0191	25	rs317288536
112	*HAP1*	0,0194	27	rs315329074
113	*IGF2BP1*	0,0197	27	rs315329074
114	*FRS3*	0,0197	26	rs317627533
115	*MFGE8*	0,0199	10	rs318199727
116	*AEBP1*	0,02	22	rs316794400
117	*PSME3*	0,0201	27	rs315329074
118	*CHMP1A*	0,0211	11	rs318098582
119	*SETDB1*	0,0211	25	rs312758346
120	*KIF7*	0,0215	10	rs318199727
121	*TBX21*	0,0217	27	rs315329074
122	*PACSIN1*	0,0218	26	rs317627533
123	*MVD*	0,0222	11	rs318098582
124	*ZBTB7B*	0,0224	25	rs317288536,rs312758346
125	*GLMP*	0,0225	25	rs317288536
126	*PTPN22*	0,0228	26	rs317627533

127	*XK*	0,0229	1	rs13923872
128	*TRIM33*	0,023	26	rs317627533
129	*CRTC2*	0,0233	25	rs312758346
130	*PSMD7*	0,0233	11	rs318098582
131	*TERF2*	0,0235	11	rs318098582
132	*S100A6*	0,0236	25	rs317288536,rs312758346
133	*PUS1*	0,0237	15	rs317945754
134	*TOP2A*	0,0238	27	rs315329074
135	*PSMD4*	0,024	25	rs312758346
136	*CDH15*	0,0241	11	rs318098582
137	*RPGR*	0,0242	1	rs13923872
138	*PNMT*	0,0242	27	rs315329074
139	*MED20*	0,0244	26	rs317627533
140	*AOC3*	0,0246	27	rs315329074
141	*ANXA4*	0,025	22	rs316794400
142	*GALNS*	0,0251	11	rs318098582
143	*SV2B*	0,0252	10	rs318199727
144	*RFX5*	0,0255	25	rs312758346
145	*SPG7*	0,0258	11	rs318098582
146	*PPIA*	0,026	22	rs316794400
147	*PGAP3*	0,0262	27	rs315329074
148	*RAB5C*	0,0262	27	rs315329074
149	*HDGF*	0,0264	25	rs317288536,rs312758346
150	*HOXB3*	0,0265	27	rs315329074
151	*FKBP10*	0,0266	27	rs315329074
152	*SLC20A1*	0,0268	22	rs316794400
153	*ANKRD11*	0,0269	11	rs318098582
154	*FZD10*	0,027	15	rs317945754
155	*RUSC1*	0,0271	25	rs312758346
156	*CBFA2T3*	0,0272	11	rs318098582
157	*PSMB4*	0,0273	25	rs312758346
158	*VANGL2*	0,0275	25	rs317288536

159	*CNP*	0,0276	27	rs315329074
160	*GNGT2*	0,0276	27	rs315329074
161	*AKAP13*	0,0276	10	rs318199727
162	*COX4I1*	0,0276	11	rs318098582
163	*HIPK1*	0,0278	26	rs317627533
164	*PRELP*	0,0279	26	rs317627533
165	*MLX*	0,0286	27	rs315329074
166	*HSD17B1*	0,0286	27	rs315329074
167	*CIDEC*	0,0286	11	rs318098582
168	*NAGLU*	0,0287	27	rs315329074
169	*AP4B1*	0,0291	26	rs317627533
170	*S100A1*	0,0294	25	rs317288536,rs312758346
171	*FANCA*	0,0294	11	rs318098582
172	*SOCS7*	0,0298	27	rs315329074
173	*MMP17*	0,03	15	rs317945754
174	*S100A10*	0,0304	25	rs312758346
175	*CDC6*	0,0304	27	rs315329074
176	*PSMB3*	0,0306	27	rs315329074
177	*KRT10*	0,0306	27	rs315329074
178	*CNTNAP1*	0,0315	27	rs315329074
179	*SLAMF1*	0,0316	25	rs317288536
180	*KRT19*	0,0317	27	rs315329074
181	*HOXB5*	0,0318	27	rs315329074
182	*PSMD3*	0,032	27	rs315329074
183	*RIT1*	0,0322	25	rs312758346
184	*NEUROD2*	0,0322	27	rs315329074
185	*PNPO*	0,0324	27	rs315329074
186	*SLC7A11*	0,0327	4	rs312691174
187	*STX2*	0,0327	15	rs317945754
188	*CKS1B*	0,0329	25	rs317288536,rs312758346
189	*IL17C*	0,0331	11	rs318098582
190	*SLC39A1*	0,0333	25	rs312758346

191	*NFXL1*	0,0343	4	rs15608447
192	*ST8SIA2*	0,0343	10	rs318199727
193	*MGST2*	0,0349	4	rs312691174
194	*ARHGEF11*	0,0351	25	rs317288536
195	*RHBG*	0,0352	25	rs317288536,rs312758346
196	*OSBPL7*	0,0356	27	rs315329074
197	*CDT1*	0,0358	11	rs318098582
198	*ACLY*	0,0358	27	rs315329074
199	*KRT12*	0,0364	27	rs315329074
200	*CIART*	0,0368	25	rs312758346
201	*ADAMTS4*	0,0369	25	rs312758346
202	*S100A4*	0,0371	25	rs317288536,rs312758346
203	*TREM2*	0,0378	26	rs317627533
204	*MTHFS*	0,0384	10	rs318199727
205	*TFEB*	0,039	26	rs317627533
206	*BTG2*	0,039	26	rs317627533
207	*SNAPIN*	0,0392	25	rs317288536,rs312758346
208	*MKRN3*	0,0392	26	rs317627533
209	*SLCO3A1*	0,0396	10	rs318199727
210	*NQO1*	0,0398	11	rs318098582
211	*ACSF3*	0,0401	11	rs318098582
212	*OTC*	0,0403	1	rs13923872
213	*MAP1LC3B*	0,0408	11	rs318098582
214	*CERS2*	0,0411	25	rs312758346
215	*PEX11A*	0,0411	10	rs318199727
216	*MAML3*	0,0412	4	rs312691174
217	*VPS25*	0,0415	27	rs315329074
218	*NHLH1*	0,0418	25	rs317288536
219	*S100A13*	0,0422	25	rs317288536,rs312758346
220	*CLK2*	0,0422	25	rs312758346
221	*SH2D2A*	0,0424	25	rs317288536,rs312758346
222	*HOXB2*	0,0429	27	rs315329074

223	*POGZ*	0,0431	25	rs312758346
224	*TCAP*	0,0432	27	rs315329074
225	*S100A14*	0,0434	25	rs317288536,rs312758346
226	*ZNF469*	0,0436	11	rs318098582
227	*TUBG1*	0,0441	27	rs315329074
228	*SYT6*	0,0441	26	rs317627533
229	*SNX27*	0,0444	25	rs312758346
230	*ELMOD2*	0,045	4	rs312691174
231	*ATP8B2*	0,0454	25	rs312758346
232	*TARS2*	0,0456	25	rs312758346
233	*LASP1*	0,0457	27	rs315329074
234	*FDPS*	0,0458	25	rs312758346
235	*KLHL10*	0,0461	27	rs315329074
236	*SLC15A4*	0,0464	15	rs317945754
237	*WWP2*	0,0467	11	rs318098582
238	*KIRREL*	0,04684	25	rs317288536
239	*TSHB*	0,0471	26	rs317627533
240	*CHIC2*	0,0471	4	rs15608447
241	*SLC10A4*	0,0473	4	rs15608447
242	*EP400*	0,0476	15	rs317945754
243	*HOXB4*	0,048	27	rs315329074
244	*TMEM79*	0,0483	25	rs317288536
245	*SMG5*	0,0486	25	rs317288536
246	*PHOSPHO1*	0,0487	27	rs315329074
247	*CDK12*	0,04878	27	rs315329074
248	*SETD7*	0,0495	4	rs312691174

Marker minimum distance (bp)*: distances are based on Gallus gallus 5.0 assembly

Table S4. Results of topological gene network analysis.

Gene ID	**Node degree**
UBC	68
STAT3	40
SMAD4	34
SHC1	30
ERBB2	28
NRAS, PSMD4	26
CDC6	23
PSMD7	22
RARA, RPL10A	21
PSMB4	20
CDH1	19
RPL19, STAT5B	18
MED1, PSMD3	17
RPS27A, RPL13, MRPL24	16
UBA52, CDT1	15
MRPS11	14
MCL1, BYSL	13
NTRK1, EP300, KAT2A, THRA	12
MEF2D, CCND3, UBB, PSME3	11

BECN1, NFYA, SETD7	10
TUBG1, HSP90AA1, HDAC2, RAC1, RPS27, MAGOH, CREBBP, CDK1, CKS1B, USP9X, LCK, GRB2,MED14	9
PIP5K1A, SRC , NNB1, HDAC1, SMARCE1, FYN, CDK2, MRPL3, PSMC5, NIP7, TEC	8
AR, LK1, ERBB4, VPS4A, NCSTN, ESR1, HDAC3, PPARD, MRPL15, AKAP13, ARHGEF11, EGFR, RPS3, NPEPPS, PSMC2, IKZF3, GRB7, NOC4L	7
EIF4A3, DHX8, CUL3, CCT3, AKT1, NCOR1, PDGFRB, IMP3, SMURF1, ZFPM1, MAPK1, PSMD8, LYN, PLCG1, PTPN11, DAD1, KIT, MRPL4, MED20, PS6, JAK1, PSMC4, RPS5, PSMD2	6
FANCI, DDX3X, TERF2, KPNB1, CDKN1A, CBFA2T3, TP53, NR4A1, CAV1, PIK3CG, CCND1, FANCE, FANCA, TRIM33, KAT2B, MYC, PCNA, UBQLN4, VPS25, SYK,RPL5, SEC61A1, HNF4A, CUL1, EZH1, TAF8, IMP4, MRPL46, DHX37, UTP4, PSMA4	5
CDK5, PBX1, UHRF1BP1, MTOR, TOP2A,TBP, RHOC, TAF10, GNA12, SMAD2, ITPR3, PFDN2, CDC42, EFTUD2, MYOD1, ACTB, MAPK8, PIK3CA, NGFR, NIFK, NOB1, MED15, BCL2L1, ABL1, E2F1, MRPS2, CCNA1, SETDB1, CDK19, MRTO4, SMARCD3, BCAS2, TNKS, SSR2, ERBB3, VPS4B, TAF11, CHMP4B	4
NGF, VPS45, ADAR, PPP1R1B, BCAN, OGDH, MCM5, VANGL2, DMD, TBK1, BCOR, COPA, MAGI3, SNRPC, PRPF19, NFKB1, ACLY, PSEN1, NR1D1, NOTCH1, YWHAQ, BAK1, SQSTM1, BTRC, PTEN, UPF1, PLCB2, LNX1, RBBP4, RPN1, YWHAZ, YWHAB, TAF9, IGF2BP1, BCL6, SNU13, RAC2, CHEK1, HSPA8, SNF8, NR3C1, CASC3, NSA2, FOXO3, SKP2, RUNX1, NCOA2, DDOST, RND1, SREBF2, HES1, TBX21, RBX1, NCOA3, ZAP70, IRF8, WIPF2, PPARG, FBXO31, SH2D2A, TSR1,CTNND1, ETV4, TUBA4A, DEF6, NR0B2, CDH15, FTSJ3, TXK, INSRR, CRTC2, CDH3, PTPN22, SKAP1,MAP1LC3B, UBE4B, TGFA, MRPL9, SMAD3, NTRK3, CHMP1A	3

MMP7, STX6, WDR48, TPM3, CCL2, CNTNAP1, DICER1, SPDEF, STX2, USP46, PRICKLE1, NCBP1, GINS2, PDHX, ARNT, DVL2, MMP2, TGFB1, MAD2L2, NFYB, PIWIL1, SNAP23, PHLPP1, RBBP5, GABARAP, HGS, CCR10, APRT, CKAP5, FOXO1, PRKACA, LEP, TSHR, DDX20, KCNA1, RUNX2, ATG16L1, NKX3-1, UBE2Z, ILF2, PRPF8, SSU72, TCF7, SMG1, DYNLL1, HOXB5, HOXB7, STRN, DYNC1H1, NCBP2, LDHA, NUMB, S100A1, MAPK14, NES, ATP2B4, EIF1, BIK, CSNK1D, STK11, SNRPD3, PPP1CC, SUPT7L, SRPRA, TRIM28, CDC5L, SEC61B, RFWD2, SPI1, ATG3, GATA4, LCP2,WASL, EZH2, SMAD7, HAX1, NOTCH3, RPTOR, SNW1, STUB1, TRAF6, RORC, GSK3B, PRKCA, ASXL1, FOXM1, WDR12, XRCC6, TOP2B, ARVCF, GATA6, APP, PPARA, RAB5A, HDAC5, CSK, DNMT3A, PRPF3, RAD1, HIPK1, GATA3, CDKN1B, FIP1L1, ADRA1A, ELP3, ATR, ARF1, HIST3H3, ATM, EED, TRIM24, ATP2A2, DOK1, MRPL44, NR1D2, DYNLT3, PABPC1, SIKE1, MAML3, NEDD8, SMG5, HRAS, FANCD2, EIF2S1, MRPL27, WRN, CALM2, MRPS10, SLAMF1, RHOV, BCL2A1, PBX2, FCER1G, CYBA, CYBB, IL6R	2
KRT14, FZD10, CCR7, WNK4, G6PC, CRP, TSHB, SLC7A5, SCARB1, IL1B, TCF25, CHIC2, SLAIN2, KCNH4, KRT10, S100A11, LC20A1, SLC7A11, MRPL10, RFX5, GABRG1, NAA15, PLIN1, RAB33B, CYB5B, AACS, HAP1, SNTB2, FMOD, TBKBP1, KRT19, HOXB1, ADAMTS4, CD244,, ETV3, BGLAP, SNAPIN, SNX27, SP2, SGCB, NFE2L1, ANXA4, ZNF76, PPIA, PRCC, TDRKH, FDPS, DCLRE1B, AMPD1, KLHL11, KLHL10, KLHL25, PYGO2, ZNF652, THEM4, APH1A, DNAJC7, RUSC1, PLEKHO1	1

Chapter 3

Table S1. Positional candidate genes for BW.

SNP ID	GGA	Position of the marker (bp)	*Gene ID**	*Description*	Start Position of the gene (bp)*	End Position of the gene (bp)*	Orientation of the gene	Minimum distance from gene (bp)
rs13923872	1	112,741,685	*USP9X*	*ubiquitin specific peptidase 9, X-linked*	112,078,329	112,176,727	minus	564,958
			LOC101750779	*uncharacterized LOC101750779*	112,193,534	112,198,521	plus	543,164
			LOC107056757	*uncharacterized LOC107056757*	112,223,451	112,228,082	minus	513,603
			MED14	*mediator complex subunit 14*	112,228,091	112,262,695	plus	478,99
			LOC107056765	*uncharacterized LOC107056765*	112,263,096	112,268,266	plus	473,419
			ATP6AP2	*ATPase H+ transporting accessory protein 2*	112,273,187	112,282,377	minus	459,308
			LOC101750924	*uncharacterized LOC101750924*	112,282,457	112,304,013	plus	437,672
			LOC107056538	*basic proline-rich protein-like*	112,349,429	112,354,018	minus	387,667
			BCOR	*BCL6 corepressor*	112,378,844	112,422,324	plus	319,361
			LOC107057081	*uncharacterized LOC107057081*	112,474,240	112,486,871	minus	254,814
			LOC101751364	*uncharacterized LOC101751364*	112,482,075	112,497,856	plus	243,829
			LOC107057076	*uncharacterized LOC107057076*	112,498,548	112,504,661	plus	237,024
			LOC101751443	*uncharacterized LOC101751443*	112,519,737	112,523,119	plus	218,566
			MIR6672	*microRNA 6672*	112,625,566	112,625,675	plus	116,01
			LOC107057061	*uncharacterized LOC107057061*	112,635,402	112,643,712	plus	97,973
			LOC107057049	*uncharacterized*	112,668,313	112,707,049	minus	34,636

				LOC107057049				
			MID1IP1	*MID1 interacting protein 1*	112,738,605	112,739,753	minus	1,932
			TSPAN7	*tetraspanin 7*	112,766,324	112,855,838	minus	24,639
			LOC101751659	*uncharacterized LOC101751659*	112,767,424	112,769,932	plus	25,739
			LOC107057044	*uncharacterized LOC107057044*	112,853,920	112,855,360	minus	112,235
			OTC	*ornithine carbamoyltransferase*	112,898,625	112,924,531	minus	156,94
			RPGR	*retinitis pigmentosa GTPase regulator*	112,926,169	112,985,737	plus	184,484
			SRPX	*sushi repeat containing protein, X-linked*	112,994,814	113,035,754	plus	253,129
			SYTL5	*synaptotagmin like 5*	113,039,522	113,117,936	minus	297,837
			DYNLT3	*dynein light chain Tctex-type 3*	113,133,105	113,139,987	plus	391,42
			CYBB	*cytochrome b-245 beta chain*	113,149,626	113,183,626	minus	407,941
			XK	*X-linked Kx blood group*	113,199,116	113,214,148	minus	457,431
			LANCL3	*LanC like 3*	113,215,605	113,251,683	minus	473,92
			LOC107056992	*uncharacterized LOC107056992*	113,278,216	113,282,701	minus	536,531
			PRRG1	*proline rich and Gla domain 1*	113,283,937	113,314,070	minus	542,252
			LOC107056924	*maestro heat-like repeat-containing protein family member 2B*	113,325,651	113,339,179	plus	583,966
			LOC107056990	*maestro heat-like repeat-containing protein family member 2B*	113,343,513	113,348,692	plus	601,828
			LOC107056981	*maestro heat-like*	113,350,799	113,354,007	plus	609,114

				repeat-containing protein family member 2B				
rs312691174	4	29,074,989	*LOC107051793*	*uncharacterized LOC107051793*	28,550,125	28,563,120	minus	511,869
			PCDH18	*protocadherin 18*	28,880,214	28,890,093	minus	184,896
			LOC107051792	*uncharacterized LOC107051792*	28,986,991	29,036,383	plus	38,606
			SLC7A11	*solute carrier family 7 member 11*	29,138,177	29,196,482	minus	63,188
			LOC101751121	*uncharacterized LOC101751121*	29,205,067	29,277,125	minus	130,078
			NOCT	*nocturnin*	29,428,771	29,434,135	plus	353,782
			ELF2	*E74 like ETS transcription factor 2*	29,438,718	29,467,618	minus	363,729
			MGARP	*mitochondria localized glutamic acid rich protein*	29,471,760	29,493,863	minus	396,771
			LOC107051797	*atherin-like*	29,494,724	29,497,114	minus	419,735
			NAA15	*N(alpha)-acetyltransferase 15, NatA auxiliary subunit*	29,496,798	29,534,035	plus	421,809
			RAB33B	*RAB33B, member RAS oncogene family*	29,539,318	29,546,334	plus	464,329
			LOC422442	*uncharacterized LOC422442*	29,559,136	29,561,953	minus	484,147
			SETD7	*SET domain containing lysine methyltransferase 7*	29,561,960	29,578,876	minus	486,971
			MGST2	*microsomal glutathione S-transferase 2*	29,593,387	29,603,029	plus	518,398
			LOC107051791	*microsomal glutathione S-transferase 2-like*	29,604,032	29,611,350	plus	529,043

			MAML3	*mastermind like transcriptional coactivator 3*	29,611,253	29,815,133	minus	536,264
rs15608447	4	66,885,210	*FIP1L1*	*factor interacting with PAPOLA and CPSF1*	66,133,445	66,171,020	minus	714,19
			SCFD2	*sec1 family domain containing 2*	66,173,796	66,368,297	plus	516,913
			LOC107053245	*uncharacterized LOC107053245*	66,257,009	66,266,179	minus	619,031
			RASL11B	*RAS like family 11 member B*	66,369,830	66,373,534	minus	511,676
			LOC107053244	*uncharacterized LOC107053244*	66,391,577	66,398,660	minus	486,55
			LOC422757	*uncharacterized LOC422757*	66,403,272	66,405,708	minus	479,502
			USP46	*ubiquitin specific peptidase 46*	66,413,295	66,440,521	plus	444,689
			SPATA18	*spermatogenesis associated 18*	66,567,846	66,587,444	minus	297,766
			SGCB	*sarcoglycan beta*	66,587,612	66,593,633	plus	291,577
			LRRC66	*leucine rich repeat containing 66*	66,593,723	66,602,814	plus	282,396
			DCUN1D4	*defective in cullin neddylation 1 domain containing 4*	66,607,031	66,645,141	minus	240,069
			CWH43	*cell wall biogenesis 43 C-terminal homolog*	66,651,068	66,677,010	minus	208,2
			OCIAD1	*OCIA domain containing 1*	66,693,766	66,708,277	minus	176,933
			FRYL	*FRY like transcription coactivator*	66,708,277	66,866,440	plus	18,77
			ZAR1	*zygote arrest 1*	66,868,239	66,871,111	minus	14,099
			LOC107053243	***uncharacterized LOC107053243***	66,870,041	66,885,878	plus	0

			SLC10A4	*solute carrier family 10 member 4*	66,870,984	66,872,626	minus	12,584
			SLAIN2	***SLAIN motif family member 2***	66,883,481	66,910,807	minus	0
			TEC	*tec protein tyrosine kinase*	66,928,003	66,969,829	plus	42,793
			TXK	*TXK tyrosine kinase*	66,970,277	66,990,459	plus	85,067
			NIPAL1	*NIPA like domain containing 1*	66,992,447	67,002,107	minus	107,237
			CNGA1	*cyclic nucleotide gated channel alpha 1*	67,007,806	67,015,296	plus	122,596
			NFXL1	*nuclear transcription factor, X-box binding like 1*	67,016,510	67,061,474	plus	131,3
			CORIN	*corin, serine peptidase*	67,063,312	67,186,506	plus	178,102
			ATP10D	*ATPase phospholipid transporting 10D (putative)*	67,189,360	67,228,010	minus	304,15
			COMMD8	*COMM domain containing 8*	67,239,252	67,243,040	plus	354,042
			LOC107053239	*uncharacterized LOC107053239*	67,246,293	67,253,895	minus	361,083
			LOC107053241	*uncharacterized LOC107053241*	67,252,999	67,266,059	plus	367,789
			GABRB1	*gamma-aminobutyric acid type A receptor beta1 subunit*	67,253,915	67,271,613	minus	368,705
			LOC107053242	*uncharacterized LOC107053242*	67,325,419	67,334,233	plus	440,209
			LOC107053240	*uncharacterized LOC107053240*	67,330,651	67,347,343	minus	445,441
			GABRA4	*gamma-aminobutyric acid type A receptor alpha4 subunit*	67,355,338	67,402,602	plus	470,128
			GABRA2	*gamma-aminobutyric*	67,487,682	67,549,247	plus	602,472

				acid type A receptor alpha2 subunit				
			LOC107053237	*uncharacterized LOC107053237*	67,489,979	67,491,477	minus	604,769
			LOC107053238	*uncharacterized LOC107053238*	67,546,207	67,576,158	minus	660,997
			GABRG1	*gamma-aminobutyric acid type A receptor gamma1 subunit*	67,576,299	67,633,176	plus	691,089
rs318199727	10	13,536,548	*LOC107054202*	*uncharacterized LOC107054202*	12,796,949	12,800,712	minus	735,836
			PEX11A	*peroxisomal biogenesis factor 11 alpha*	12,815,015	12,820,635	plus	715,913
			LOC107054203	*uncharacterized LOC107054203*	12,817,514	12,823,049	minus	713,499
			PLIN1	*perilipin 1*	12,822,263	12,826,843	plus	709,705
			KIF7	*kinesin family member 7*	12,827,027	12,836,867	plus	699,681
			TICRR	*TOPBP1 interacting checkpoint and replication regulator*	12,837,814	12,854,365	minus	682,183
			RHCG	*Rh family C glycoprotein*	12,859,333	12,867,258	plus	669,29
			LOC107054204	*uncharacterized LOC107054204*	12,902,048	12,938,774	minus	597,774
			TRNAR-UCG	*transfer RNA arginine (anticodon UCG)*	12,942,837	12,942,909	minus	593,639
			POLG	*DNA polymerase gamma, catalytic subunit*	12,942,988	12,953,027	plus	583,521
			FANCI	*Fanconi anemia complementation group I*	12,951,763	12,975,529	minus	561,019
			RLBP1	*retinaldehyde binding protein 1*	12,976,366	12,980,554	plus	555,994

			ABHD2	*abhydrolase domain containing 2*	12,981,973	13,021,732	minus	514,816
			MFGE8	*milk fat globule-EGF factor 8 protein*	13,033,323	13,040,620	plus	495,928
			HAPLN3	*hyaluronan and proteoglycan link protein 3*	13,042,935	13,046,872	plus	489,676
			ACAN	*aggrecan*	13,047,289	13,092,387	minus	444,161
			AEN	*apoptosis enhancing nuclease*	13,131,452	13,134,038	minus	402,51
			MIR1720	*microRNA 1720*	13,134,585	13,134,649	minus	401,899
			MIR7-2	*microRNA 7-2*	13,134,720	13,134,818	minus	401,73
			MIR3529	*microRNA 3529*	13,134,724	13,134,814	plus	401,734
			DET1	*de-etiolated homolog 1 (Arabidopsis)*	13,149,476	13,171,850	plus	364,698
			MRPS11	*mitochondrial ribosomal protein S11*	13,171,107	13,174,638	minus	361,91
			MRPL46	*mitochondrial ribosomal protein L46*	13,174,668	13,177,215	plus	359,333
			LOC101751754	*uncharacterized LOC101751754*	13,178,649	13,201,172	plus	335,376
			LOC101751792	*uncharacterized LOC101751792*	13,196,390	13,209,087	minus	327,461
			NTRK3	*neurotrophic receptor tyrosine kinase 3*	13,227,881	13,408,049	plus	128,499
			LOC107054207	*uncharacterized LOC107054207*	13,381,318	13,386,137	minus	150,411
			LOC107054206	***uncharacterized LOC107054206***	13,388,442	13,616,738	minus	0
			AGBL1	*ATP/GTP binding protein like 1*	13,623,118	13,914,436	minus	86,57
			KLHL25	*kelch like family member 25*	13,954,963	13,969,663	plus	418,415

			AKAP13	*A-kinase anchoring protein 13*	13,974,501	14,169,240	minus	437,953
			LOC107054188	*uncharacterized LOC107054188*	14,182,433	14,183,328	minus	645,885
			SV2B	*synaptic vesicle glycoprotein 2B*	14,187,010	14,235,623	plus	650,462
rs318098582	11	18,651,449	*BANP*	*BTG3 associated nuclear protein*	18,243,614	18,386,758	plus	264,691
			ZNF469	*zinc finger protein 469*	18,416,335	18,579,095	plus	72,354
			LOC107054329	*uncharacterized LOC107054329*	18,541,302	18,548,014	minus	103,435
			ZFPM1	*zinc finger protein, FOG family member 1*	18,584,457	18,613,764	plus	37,685
			CIDEC	*cell death inducing DFFA like effector c*	18,614,459	18,616,305	minus	35,144
			ZC3H18	***zinc finger CCCH-type containing 18***	18,617,357	18,656,718	plus	0
			MIR1571	*microRNA 1571*	18,632,364	18,632,461	plus	18,988
			IL17C	*interleukin 17C*	18,658,034	18,663,248	plus	6,585
			CYBA	*cytochrome b-245 alpha chain*	18,663,342	18,665,615	minus	11,893
			MVD	*mevalonate diphosphate decarboxylase*	18,665,712	18,668,663	minus	14,263
			RNF166	*ring finger protein 166*	18,670,083	18,677,685	minus	18,634
			CTU2	*cytosolic thiouridylase subunit 2*	18,677,742	18,681,201	plus	26,293
			PIEZO1	*piezo type mechanosensitive ion channel component 1*	18,681,139	18,700,645	minus	29,69
			LOC107054330	*nascent polypeptide-associated complex subunit alpha,*	18,696,540	18,698,959	plus	45,091

				muscle-specific form-like				
			CDT1	*chromatin licensing and DNA replication factor 1*	18,701,940	18,705,724	plus	50,491
			APRT	*adenine phosphoribosyltransferase*	18,706,797	18,709,506	minus	55,348
			GALNS	*galactosamine (N-acetyl)-6-sulfatase*	18,714,200	18,758,480	minus	62,751
			TRAPPC2L	*trafficking protein particle complex 2 like*	18,758,475	18,761,079	plus	107,026
			PABPN1L	*poly(A) binding protein nuclear 1 like, cytoplasmic*	18,762,170	18,765,810	minus	110,721
			CBFA2T3	*CBFA2/RUNX1 translocation partner 3*	18,767,307	18,788,285	minus	115,858
			ACSF3	*acyl-CoA synthetase family member 3*	18,805,694	18,846,630	plus	154,245
			CDH15	*cadherin 15*	18,848,471	18,853,003	plus	197,022
			SLC22A31	*solute carrier family 22 member 31*	18,853,081	18,856,750	minus	201,632
			ANKRD11	*ankyrin repeat domain 11*	18,857,696	18,939,049	minus	206,247
			MIR1560	*microRNA 1560*	18,874,320	18,874,423	minus	222,871
			MIR1785	*microRNA 1785*	18,926,659	18,926,760	minus	275,21
			SPG7	*SPG7, paraplegin matrix AAA peptidase subunit*	18,946,166	18,976,689	plus	294,717
rs317945754	15	3,557,083	*EP400*	*E1A binding protein p400*	2,582,896	2,628,948	minus	928,135
			PUS1	*pseudouridylate synthase 1*	2,629,662	2,634,569	minus	922,514
			ULK1	*unc-51 like autophagy activating*	2,635,904	2,708,526	minus	848,557

				kinase 1				
			MMP17	*matrix metallopeptidase 17*	2,722,077	2,765,544	minus	791,539
			SFSWAP	*splicing factor SWAP homolog*	2,792,492	2,832,182	minus	724,901
			STX2	*syntaxin 2*	2,926,591	3,223,484	plus	333,599
			ADGRD1	*adhesion G protein-coupled receptor D1*	3,062,231	3,191,688	minus	365,395
			RAN	*RAN, member RAS oncogene family*	3,201,356	3,205,602	minus	351,481
			RIMBP2	*RIMS binding protein 2*	3,223,574	3,347,096	plus	209,987
			PIWIL1	*piwi like RNA-mediated gene silencing 1*	3,347,780	3,421,290	minus	135,793
			FZD10	*frizzled class receptor 10*	3,432,876	3,435,118	minus	121,965
			LOC107051610	*frizzled-10-like*	3,442,998	3,446,063	minus	111,02
			TMEM132D	***transmembrane protein 132D***	3,526,053	3,718,886	plus	0
			GLT1D1	*glycosyltransferase 1 domain containing 1*	3,736,272	3,784,033	minus	179,189
			SLC15A4	*solute carrier family 15 member 4*	3,785,313	3,806,720	plus	228,23
			TMEM132C	*transmembrane protein 132C*	3,830,248	4,010,779	minus	273,165
			LOC107051578	*uncharacterized LOC107051578*	4,105,652	4,107,324	minus	548,569
			LOC107051609	*uncharacterized LOC107051609*	4,241,859	4,320,316	plus	684,776
			TMEM132B	*transmembrane protein 132B*	4,291,921	4,471,778	minus	734,838
			AACS	*acetoacetyl-CoA synthetase*	4,478,298	4,515,607	minus	921,215
rs316794400	22	4,594,855	*LOC101751469*	*uncharacterized LOC101751469*	4,576,077	4,577,398	minus	17,457

			ANXA4	*annexin A4*	4,584,810	4,594,055	plus	800
			SLC20A1	*solute carrier family 20 member 1*	4,596,702	4,604,387	plus	1,847
			NT5DC4	*5'-nucleotidase domain containing 4*	4,604,777	4,609,445	plus	9,922
			CKAP2L	*cytoskeleton associated protein 2 like*	4,609,435	4,616,294	minus	14,58
			LOC107054991	*uncharacterized LOC107054991*	4,616,423	4,616,974	plus	21,568
			IL1B	*interleukin 1, beta*	4,616,889	4,618,625	minus	22,034
rs317288536	25	976,833	*BGLAP*	*bone gamma-carboxyglutamate protein*	594	1,789	plus	975,044
			SMG5	*SMG5, nonsense mediated mRNA decay factor*	2,171	24,198	minus	952,635
			TMEM79	*transmembrane protein 79*	24,606	27,213	plus	949,62
			GLMP	*glycosylated lysosomal membrane protein*	27,661	31,956	minus	944,877
			CCT3	*chaperonin containing TCP1 subunit 3*	33,641	44,203	minus	932,63
			LOC107055078	*uncharacterized LOC107055078*	46,465	47,469	minus	929,364
			LOC107055080	*nectin-4-like*	54,112	60,599	minus	916,234
			LIM2	*lens intrinsic membrane protein 2*	61,134	68,96	minus	907,873
			LOC107055082	*cytochrome b5 domain-containing protein 1-like*	81,045	82,719	minus	894,114
			LOC107055083	*uncharacterized LOC107055083*	82,856	89,386	plus	887,447
			VPS45	*vacuolar protein sorting 45 homolog*	92,295	119,429	plus	857,404

			PLEKHO1	*pleckstrin homology domain containing O1*	123,481	139,888	plus	836,945
			LOC107055081	*uncharacterized LOC107055081*	142,235	150,69	plus	826,143
			ANP32E	*acidic nuclear phosphoprotein 32 family member E*	162,05	174,225	minus	802,608
			LOC100859767	*cytochrome b5 domain-containing protein 1-like*	204	205,667	plus	771,166
			APOA1BP	*apolipoprotein A-I binding protein*	224,358	227,444	plus	749,389
			GPATCH4	*G-patch domain containing 4*	227,411	231,838	minus	744,995
			LOC107055084	*uncharacterized LOC107055084*	235,211	236,685	plus	740,148
			MEX3A	*mex-3 RNA binding family member A*	247,371	260,425	minus	716,408
			LOC107055086	*sperm-associated antigen 4 protein-like*	783,429	786,252	minus	190,581
			LOC100857131	*sperm-associated antigen 4 protein-like*	797,89	798,435	minus	178,398
			UBQLN4	*ubiquilin 4*	804,221	815,017	minus	161,816
			LAMTOR2	*late endosomal/lysosomal adaptor, MAPK and MTOR activator 2*	815,073	817,92	plus	158,913
			RAB25	*RAB25, member RAS oncogene family*	818,004	824,803	plus	152,03
			RAB2B	*RAB2B, member RAS oncogene family*	825,467	830,852	plus	145,981
			LOC101747704	*uncharacterized LOC101747704*	846,955	849,648	minus	127,185
			LOC107055087	*sperm-associated antigen 4 protein-like*	855,482	858,864	minus	117,969
			OTUD7B	*OTU deubiquitinase*	888,239	923,782	plus	53,051

				7B				
			MTMR11	*myotubularin related protein 11*	925,626	933,506	plus	43,327
			SF3B4	*splicing factor 3b subunit 4*	933,59	938,879	plus	37,954
			SV2A	*synaptic vesicle glycoprotein 2A*	939,052	949,843	plus	26,99
			LOC107055093	*uncharacterized LOC107055093*	950,637	951,127	plus	25,706
			LOC107055108	*feather keratin 3-like*	953,49	954,648	plus	22,185
			LOC107055109	*feather keratin 3-like*	956,426	957,314	plus	19,519
			LOC100859249	*feather keratin 3-like*	959,38	960,434	plus	16,399
			LOC107055107	*feather keratin 1-like*	959,446	963,95	plus	12,883
			LOC100859427	*feather keratin 1-like*	966,46	967,631	plus	9,202
			LOC426914	*feather keratin 1-like*	969,985	971,047	plus	5,786
			F-KER	***feather keratin I***	973,471	980,575	plus	0
			LOC429492	*keratin D*	973,475	974,554	plus	2,279
			LOC431325	***feather keratin 1-like***	976,657	980,857	plus	0
			LOC431324	*keratin A*	979,736	983,908	plus	2,903
			LOC426913	*feather keratin 1-like*	982,753	987,425	plus	5,92
			LOC431323	*beta-keratin-related protein-like*	991,47	993,152	plus	14,637
			LOC431322	*feather keratin 1-like*	997,064	997,414	plus	20,231
			LOC431321	*keratin*	1,002,033	1,004,112	plus	25,2
			LOC431320	*feather beta keratin-like*	1,008,291	1,010,072	plus	31,458
			LOC107055103	*scale keratin-like*	1,010,842	1,012,163	minus	34,009
			LOC107055106	*uncharacterized LOC107055106*	1,014,345	1,015,145	plus	37,512
			LOC431317	*scale keratin-like*	1,017,251	1,017,835	minus	40,418
			LOC431316	*scale keratin-like*	1,018,813	1,019,553	plus	41,98
			LOC100859586	*scale keratin-like*	1,020,735	1,021,650	minus	43,902

			LOC100859657	*scale keratin-like*	1,020,928	1,025,554	minus	44,095
			LOC100859616	*scale keratin-like*	1,022,275	1,023,023	plus	45,442
			LOC425362	*scale keratin-like*	1,026,144	1,027,006	plus	49,311
			LOC100857270	*scale keratin-like*	1,028,628	1,029,557	minus	51,795
			LOC100859756	*scale keratin-like*	1,030,236	1,034,932	plus	53,403
			LOC100859722	*scale keratin-like*	1,030,356	1,031,017	plus	53,523
			LOC100857297	*scale keratin-like*	1,032,585	1,044,971	minus	55,752
			LOC107055105	*uncharacterized LOC107055105*	1,032,750	1,040,322	plus	55,917
			LOC426912	*scale keratin-like*	1,036,532	1,037,365	minus	59,699
			LOC100859790	*scale keratin-like*	1,037,906	1,038,601	plus	61,073
			LOC101751554	*scale keratin-like*	1,040,076	1,040,849	minus	63,243
			LOC100857367	*scale keratin-like*	1,041,747	1,042,135	plus	64,914
			LOC107055104	*scale keratin-like*	1,044,261	1,044,828	minus	67,428
			LOC101750668	*scale keratin-like*	1,045,428	1,046,372	plus	68,595
			LOC396480	*keratin*	1,048,622	1,050,613	minus	71,789
			LOC101750550	*scale keratin-like*	1,052,489	1,053,802	plus	75,656
			LOC396479	*keratin*	1,055,266	1,056,859	minus	78,433
			LOC431314	*scale keratin-like*	1,058,333	1,060,625	plus	81,5
			LOC769486	*scale keratin-like*	1,064,610	1,066,554	plus	87,777
			LOC107055102	*uncharacterized LOC107055102*	1,067,878	1,069,936	plus	91,045
			LOC408038	*beta-keratin*	1,069,829	1,071,422	minus	92,996
			LOC431313	*feather beta keratin-like*	1,074,635	1,075,569	plus	97,802
			LOC107055092	*uncharacterized LOC107055092*	1,078,270	1,079,527	minus	101,437
			LOC100857468	*feather keratin Cos1-1/Cos1-3/Cos2-1-like*	1,080,919	1,087,221	minus	104,086
			LOC107055101	*uncharacterized LOC107055101*	1,093,075	1,095,264	minus	116,242
			LOC101751614	*keratin, type I*	1,098,677	1,100,340	minus	121,844

				cytoskeletal 9-like				
			LOC107055091	*beta-keratin-related protein-like*	1,102,072	1,103,363	plus	125,239
			LOC107055090	*uncharacterized LOC107055090*	1,106,254	1,107,828	plus	129,421
			LOC107055100	*uncharacterized LOC107055100*	1,109,328	1,111,524	minus	132,495
			LOC101751113	*titin-like*	1,116,120	1,122,481	plus	139,287
			LOC107055099	*uncharacterized LOC107055099*	1,116,136	1,119,078	minus	139,303
			EDYM2	*epidermal differentiation protein containing Y motif 2*	1,122,288	1,125,199	minus	145,455
			EDQREP	*epidermal differentiation protein containing glutamine (Q) repeats*	1,127,324	1,130,658	minus	150,491
			EDPE	*epidermal differentiation protein rich in proline and glutamic acid (E)*	1,139,306	1,142,301	plus	162,473
			LOC107055098	*epidermal differentiation protein containing glutamine (Q) repeats-like*	1,144,136	1,145,825	minus	167,303
			EDQCM	*epidermal differentiation protein containing QC motifs*	1,148,589	1,150,441	minus	171,756
			EDDM	*epidermal differentiation protein containing DPCC motifs*	1,154,359	1,157,611	minus	177,526
			EDNC	*epidermal*	1,159,110	1,161,256	plus	182,277

				differentiation protein encoded by neighbor of cornulin				
			CRNN	*cornulin*	1,168,202	1,170,459	plus	191,369
			SCFN	*scaffoldin*	1,173,597	1,177,833	plus	196,764
			LOC107055094	*trichohyalin-like*	1,178,020	1,190,574	plus	201,187
			S100A11	*S100 calcium binding protein A11*	1,201,165	1,202,884	plus	224,332
			COPA	*coatomer protein complex subunit alpha*	1,203,325	1,221,735	minus	226,492
			NCSTN	*nicastrin*	1,221,813	1,232,683	plus	244,98
			NHLH1	*nescient helix-loop-helix 1*	1,241,431	1,246,232	plus	264,598
			LOC107055095	*uncharacterized LOC107055095*	1,246,142	1,253,267	minus	269,309
			VANGL2	*VANGL planar cell polarity protein 2*	1,250,127	1,262,898	plus	273,294
			LY9	*lymphocyte antigen 9*	1,266,694	1,272,789	minus	289,861
			SLAMF1	*signaling lymphocytic activation molecule family member 1*	1,274,631	1,280,327	minus	297,798
			CD48	*CD48 molecule*	1,282,479	1,284,960	minus	305,646
			CD244	*CD244 molecule*	1,285,835	1,293,468	minus	309,002
			LOC101750757	*uncharacterized LOC101750757*	1,296,055	1,297,931	plus	319,222
			KIRREL	*kin of IRRE like (Drosophila)*	1,301,854	1,321,563	plus	325,021
			LOC101750908	*T-lymphocyte surface antigen Ly-9-like*	1,323,643	1,330,948	plus	346,81
			SLAMF8	*SLAM family member 8*	1,331,041	1,334,671	plus	354,208
			ETV3	*ETS variant 3*	1,334,908	1,344,750	plus	358,075
			ETV3L	*ETS variant 3 like*	1,353,678	1,355,999	plus	376,845
			ARHGEF11	*Rho guanine*	1,357,481	1,378,565	plus	380,648

				nucleotide exchange factor 11				
			LRRC71	*leucine rich repeat containing 71*	1,379,498	1,382,513	minus	402,665
			PEAR1	*platelet endothelial aggregation receptor 1*	1,382,768	1,391,921	minus	405,935
			NTRK1	*neurotrophic receptor tyrosine kinase 1*	1,394,975	1,402,575	minus	418,142
			INSRR	*insulin receptor related receptor*	1,403,569	1,413,068	minus	426,736
			LOC100857512	*death-associated protein kinase 2-like*	1,413,151	1,416,561	minus	436,318
			SH2D2A	*SH2 domain containing 2A*	1,416,567	1,421,255	plus	439,734
			PRCC	*papillary renal cell carcinoma (translocation-associated)*	1,421,164	1,430,083	minus	444,331
			HDGF	*hepatoma-derived growth factor*	1,432,110	1,437,580	plus	455,277
			MRPL24	*mitochondrial ribosomal protein L24*	1,438,014	1,439,195	plus	461,181
			RRNAD1	*ribosomal RNA adenine dimethylase domain containing 1*	1,439,293	1,442,440	minus	462,46
			CRABP2	*cellular retinoic acid binding protein 2*	1,448,446	1,451,746	plus	471,613
			LOC425431	*dnaJ homolog subfamily A member 1-like*	1,455,339	1,458,566	minus	478,506
			NES	*nestin*	1,462,876	1,470,387	minus	486,043
			BCAN	*brevican*	1,473,720	1,486,466	plus	496,887
			HAPLN2	*hyaluronan and proteoglycan link*	1,487,083	1,489,801	minus	510,25

				protein 2				
			RHBG	*Rh family B glycoprotein*	1,490,958	1,496,016	plus	514,125
			LOC107055112	*uncharacterized LOC107055112*	1,505,796	1,508,625	minus	528,963
			LOC107055111	*uncharacterized LOC107055111*	1,521,253	1,544,127	minus	544,42
			MEF2D	*myocyte enhancer factor 2D*	1,557,855	1,582,946	minus	581,022
			LOC107055110	*uncharacterized LOC107055110*	1,598,470	1,599,898	plus	621,637
			LOC101750487	*uncharacterized LOC101750487*	1,609,167	1,620,210	minus	632,334
			LOC101750716	*uncharacterized LOC101750716*	1,620,907	1,627,137	minus	644,074
			LOC107055114	*E3 SUMO-protein ligase PIAS3-like*	1,636,879	1,651,430	plus	660,046
			MIR6662	*microRNA 6662*	1,637,951	1,638,060	minus	661,118
			INTS3	*integrator complex subunit 3*	1,654,016	1,685,697	minus	677,183
			LOC107055115	*atrial natriuretic peptide receptor 1-like*	1,686,320	1,696,310	minus	709,487
			LOC107055116	*atrial natriuretic peptide receptor 1-like*	1,696,623	1,700,257	minus	719,79
			ILF2	*interleukin enhancer binding factor 2*	1,701,006	1,705,564	plus	724,173
			SNAPIN	*SNAP associated protein*	1,705,771	1,706,895	minus	728,938
			IL6R	*interleukin 6 receptor*	1,708,497	1,714,085	plus	731,664
			SHE	*Src homology 2 domain containing E*	1,715,301	1,720,734	minus	738,468
			UBE2Q1	*ubiquitin conjugating enzyme E2 Q1*	1,722,132	1,729,019	minus	745,299
			CHRNB2	*cholinergic receptor nicotinic beta 2*	1,729,683	1,734,479	plus	752,85

				subunit				
			ADAR	*adenosine deaminase, RNA specific*	1,735,473	1,747,289	minus	758,64
			KCNN3	*potassium calcium-activated channel subfamily N member 3*	1,758,811	1,780,362	minus	781,978
			PMVK	*phosphomevalonate kinase*	1,781,459	1,784,289	minus	804,626
			PBXIP1	*PBX homeobox interacting protein 1*	1,784,644	1,788,641	minus	807,811
			PYGO2	*pygopus family PHD finger 2*	1,788,639	1,790,498	minus	811,806
			SHC1	*SHC adaptor protein 1*	1,790,787	1,800,935	minus	813,954
			CKS1B	*CDC28 protein kinase regulatory subunit 1B*	1,801,177	1,802,314	plus	824,344
			FLAD1	*flavin adenine dinucleotide synthetase 1*	1,802,652	1,808,111	plus	825,819
			ZBTB7B	*zinc finger and BTB domain containing 7B*	1,812,250	1,830,425	plus	835,417
			DCST2	*DC-STAMP domain containing 2*	1,832,750	1,837,666	minus	855,917
			SMAD4	*SMAD family member 4*	1,838,974	1,844,533	minus	862,141
			CHTOP	*chromatin target of PRMT1*	1,847,078	1,853,476	minus	870,245
			S100A1	*S100 calcium binding protein A1*	1,853,739	1,856,049	minus	876,906
			S100A13	*S100 calcium binding protein A13*	1,857,835	1,859,090	plus	881,002
			S100A14	*S100 calcium binding protein A14*	1,861,081	1,863,210	plus	884,248
			S100A16	*S100 calcium binding*	1,865,897	1,868,591	plus	889,064

				protein A16				
			S100A4	*S100 calcium binding protein A4*	1,869,231	1,871,230	plus	892,398
			S100A6	*S100 calcium binding protein A6*	1,874,323	1,875,575	plus	897,49
			LOC101747386	*protein S100-A9-like*	1,877,071	1,878,016	plus	900,238
			S100A9	*S100 calcium binding protein A9*	1,885,186	1,886,621	plus	908,353
			EDKM	*epidermal differentiation protein containing a KKLIQQ motif*	1,892,914	1,895,414	plus	916,081
			EDQM1	*epidermal differentiation protein containing a glutamine (Q) motif 1*	1,895,773	1,896,542	minus	918,94
			EDQM2	*epidermal differentiation protein containing a glutamine (Q) motif 2*	1,899,100	1,900,320	minus	922,267
			EDWM	*epidermal differentiation protein containing WYDP motif*	1,906,417	1,907,809	minus	929,584
			EDCH5	*epidermal differentiation protein containing cysteine histidine motifs 5*	1,909,483	1,911,230	minus	932,65
			EDMPN1	*epidermal differentiation protein containing a MPN sequence motif 1*	1,912,397	1,913,451	minus	935,564
			EDCRP	*epidermal differentiation cysteine-rich protein*	1,919,906	1,922,026	minus	943,073

			EDCH4	*epidermal differentiation protein containing cysteine histidine motifs 4*	1,931,045	1,932,007	minus	954,212
			EDGH	*epidermal differentiation protein rich in glycine and histidine*	1,940,110	1,942,027	minus	963,277
			LOR1	*loricrin 1*	1,943,612	1,951,026	minus	966,779
			LOR2	*loricrin 2*	1,943,846	1,946,131	minus	967,013
			LOR3	*loricrin 3*	1,952,705	1,955,656	minus	975,872
			EDMTF4	*epidermal differentiation protein starting with MTF motif 4*	1,960,713	1,980,545	plus	983,88
rs317627533	26	4,597,439	*SYT6*	*synaptotagmin 6*	3,808,772	3,837,467	minus	759,972
			TRIM33	*tripartite motif containing 33*	3,844,903	3,869,654	minus	727,785
			BCAS2	*breast carcinoma amplified sequence 2*	3,869,908	3,872,919	minus	724,52
			DENND2C	*DENN domain containing 2C*	3,872,935	3,887,414	minus	710,025
			AMPD1	*adenosine monophosphate deaminase 1*	3,892,276	3,902,376	minus	695,063
			NRAS	*neuroblastoma RAS viral oncogene homolog*	3,906,425	3,912,827	minus	684,612
			CSDE1	*cold shock domain containing E1*	3,912,971	3,930,268	minus	667,171
			SIKE1	*suppressor of IKBKE 1*	3,930,830	3,935,439	minus	662
			BARL	*bile acid receptor-like*	3,941,128	3,952,591	plus	644,848
			SYCP1	*synaptonemal complex protein 1*	3,951,965	3,966,955	plus	630,484

			LOC107049139	*synaptonemal complex protein 1-like*	3,966,969	3,974,901	plus	622,538
			TSHB	*thyroid stimulating hormone beta*	3,974,272	3,987,526	plus	609,913
			TSPAN2	*tetraspanin 2*	3,985,688	4,005,584	minus	591,855
			LOC101747848	*uncharacterized LOC101747848*	4,007,609	4,013,783	plus	583,656
			NGF	*nerve growth factor*	4,027,894	4,050,872	minus	546,567
			LOC101747895	*uncharacterized LOC101747895*	4,059,614	4,065,666	minus	531,773
			LOC101747934	*uncharacterized LOC101747934*	4,066,019	4,077,162	minus	520,277
			FANCE	*Fanconi anemia complementation group E*	4,080,342	4,084,706	minus	512,733
			MKRN3	*makorin ring finger protein 3*	4,084,828	4,086,964	minus	510,475
			PPARD	*peroxisome proliferator activated receptor delta*	4,089,638	4,106,338	minus	491,101
			DEF6	*DEF6, guanine nucleotide exchange factor*	4,108,781	4,120,796	minus	476,643
			ZNF76	*zinc finger protein 76*	4,121,087	4,130,610	minus	466,829
			RPL10A	*ribosomal protein L10a*	4,130,666	4,134,401	plus	463,038
			SCUBE3	*signal peptide, CUB domain and EGF like domain containing 3*	4,135,277	4,165,701	minus	431,738
			TCP11	*t-complex 11*	4,177,656	4,188,605	plus	408,834
			ANKS1A	*ankyrin repeat and sterile alpha motif domain containing 1A*	4,186,369	4,271,993	minus	325,446
			LOC107055188	*uncharacterized*	4,243,866	4,248,941	plus	348,498

				LOC107055188				
			TAF11	*TATA-box binding protein associated factor 11*	4,272,574	4,276,039	plus	321,4
			UHRF1BP1	*UHRF1 binding protein 1*	4,276,310	4,303,901	minus	293,538
			SNRPC	*small nuclear ribonucleoprotein polypeptide C*	4,306,380	4,310,436	minus	287,003
			C26H6orf106	*chromosome 26 C6orf106 homolog*	4,312,154	4,344,017	plus	253,422
			SPDEF	*SAM pointed domain containing ETS transcription factor*	4,353,136	4,359,030	plus	238,409
			PACSIN1	*protein kinase C and casein kinase substrate in neurons 1*	4,359,838	4,375,042	minus	222,397
			RPS10	*ribosomal protein S10*	4,376,915	4,382,439	plus	215
			NUDT3	*nudix hydrolase 3*	4,384,291	4,411,296	plus	186,143
			LOC100858737	*uncharacterized LOC100858737*	4,412,794	4,414,947	plus	182,492
			HMGA1	*high mobility group AT-hook 1*	4,415,782	4,421,656	minus	175,783
			LOC107055185	*uncharacterized LOC107055185*	4,421,016	4,430,830	plus	166,609
			GRM4	*glutamate receptor, metabotropic 4*	4,434,279	4,475,476	plus	121,963
			LOC101750261	*uncharacterized LOC101750261*	4,539,410	4,571,201	plus	26,238
			OPN1MSW	*opsin, green sensitive (rhodopsin-like)*	4,557,195	4,562,805	minus	34,634
			MLN	*motilin*	4,573,764	4,580,079	plus	17,36
			LEMD2	***LEM domain containing 2***	4,584,641	4,597,668	plus	0

			LOC107055184	***uncharacterized LOC107055184***	4,597,433	4,601,335	minus	0
			IP6K3	*inositol hexakisphosphate kinase 3*	4,601,290	4,614,097	plus	3,851
			C26H6ORF125	*chromosome 26 open reading frame, human C6orf125*	4,615,609	4,620,391	plus	18,17
			ITPR3	*inositol 1,4,5-trisphosphate receptor type 3*	4,619,903	4,659,888	minus	22,464
			LOC768477	*uncharacterized LOC768477*	4,669,560	4,675,843	minus	72,121
			BAK1	*BCL2 antagonist/killer 1*	4,677,808	4,689,669	plus	80,369
			LOC107055182	*uncharacterized LOC107055182*	4,690,003	4,698,367	minus	92,564
			TSPO2	*translocator protein 2*	4,698,990	4,704,107	plus	101,551
			LOC107055181	*uncharacterized LOC107055181*	4,707,001	4,710,843	minus	109,562
			APOBEC2	*apolipoprotein B mRNA editing enzyme catalytic subunit 2*	4,710,803	4,718,654	plus	113,364
			OARD1	*O-acyl-ADP-ribose deacylase 1*	4,719,478	4,722,726	minus	122,039
			LOC107055164	*glycine-rich protein DOT1-like*	4,722,793	4,723,646	plus	125,354
			NFYA	*nuclear transcription factor Y subunit alpha*	4,723,085	4,737,464	plus	125,646
			LOC100858470	*uncharacterized LOC100858470*	4,737,878	4,774,348	plus	140,439
			TREM-B1	*triggering receptor expressed on myeloid cells B1*	4,741,376	4,747,862	minus	143,937
			TREM2	*triggering receptor expressed on myeloid*	4,749,175	4,753,562	minus	151,736

				cells 2				
			TREM-B2	*triggering receptor expressed on myeloid cells B2*	4,755,477	4,761,898	minus	158,038
			LOC107055180	*uncharacterized LOC107055180*	4,782,696	4,797,766	plus	185,257
			LOC107055165	*uncharacterized LOC107055165*	4,802,076	4,806,903	plus	204,637
			FOXP4L	*forkhead box protein P4-like*	4,844,332	4,887,023	plus	246,893
			MDFI	*MyoD family inhibitor*	4,891,121	4,906,184	plus	293,682
			TFEB	*transcription factor EB*	4,925,985	4,940,293	minus	328,546
			GASTL	*gastricsin-like*	4,942,094	4,945,050	minus	344,655
			PGC	*progastricsin (pepsinogen C)*	4,946,918	4,950,762	minus	349,479
			FRS3	*fibroblast growth factor receptor substrate 3*	4,952,522	4,965,746	minus	355,083
			PRICKLE4	*prickle planar cell polarity protein 4*	4,966,313	4,973,328	plus	368,874
			LOC101749017	*platelet binding protein GspB-like*	4,973,531	4,985,366	minus	376,092
			TOMM6	*translocase of outer mitochondrial membrane 6*	4,985,455	4,986,384	plus	388,016
			USP49	*ubiquitin specific peptidase 49*	4,986,408	5,019,177	minus	388,969
			LOC107055176	*uncharacterized LOC107055176*	5,019,458	5,022,472	plus	422,019
			MED20	*mediator complex subunit 20*	5,022,173	5,026,778	minus	424,734
			BYSL	*bystin like*	5,026,764	5,031,080	plus	429,325
			CCND3	*cyclin D3*	5,030,661	5,069,048	minus	433,222
			TAF8	*TATA-box binding*	5,069,068	5,076,732	plus	471,629

				protein associated factor 8				
			PIFO	*primary cilia formation*	5,078,237	5,081,253	minus	480,798
			CHIA-M31	*chitinase-M31, acidic*	5,081,386	5,086,043	minus	483,947
			CHIA	*chitinase, acidic*	5,088,647	5,092,767	minus	491,208
			LOC768786	*acidic mammalian chitinase-like*	5,095,653	5,100,518	minus	498,214
			LOC107055174	*uncharacterized LOC107055174*	5,107,408	5,111,368	plus	509,969
			LOC107055171	*uncharacterized LOC107055171*	5,120,914	5,123,009	minus	523,475
			BTG2	*BTG anti-proliferation factor 2*	5,123,154	5,127,216	plus	525,715
			LOC107055173	*uncharacterized LOC107055173*	5,128,663	5,129,912	minus	531,224
			LOC107055172	*uncharacterized LOC107055172*	5,129,960	5,131,921	plus	532,521
			FMOD	*fibromodulin*	5,133,650	5,140,211	minus	536,211
			LOC107055169	*uncharacterized LOC107055169*	5,153,629	5,170,330	minus	556,19
			PRELP	*proline and arginine rich end leucine rich repeat protein*	5,163,841	5,174,857	plus	566,402
			OPTC	*opticin*	5,176,849	5,180,331	plus	579,41
			ATP2B4	*ATPase plasma membrane Ca2+ transporting 4*	5,210,403	5,247,582	plus	612,964
			LOC107055168	*uncharacterized LOC107055168*	5,245,482	5,259,467	minus	648,043
			MIR7454	*microRNA 7454*	5,270,406	5,270,459	minus	672,967
rs314452928	27	104,022	*LOC107055210*	*uncharacterized LOC107055210*	506	1,277	plus	102,745
			LOC107049042	*olfactory receptor 4M1-like*	7,537	8,651	minus	95,371
			LOC768958	*olfactory receptor*	19,365	20,677	minus	83,345

				6B1-like				
			MROH8	*maestro heat like repeat family member 8*	28,315	37,726	minus	66,296
			LOC107055211	*uncharacterized LOC107055211*	47,117	47,583	minus	56,439
			LOC101751094	*uncharacterized LOC101751094*	58,462	62,958	minus	41,064
			LOC107049117	*uncharacterized LOC107049117*	66,251	68,943	minus	35,079
			LOH11CR2A	*loss of heterozygosity, 11, chromosomal region 2, gene A*	160,44	171,21	plus	56,418
			LOC107055212	*uncharacterized LOC107055212*	160,85	166,792	minus	56,828
			DAD1	*defender against cell death 1*	172,289	175,051	plus	68,267
			IGHVL	*Ig heavy chain Mem5-like*	181,398	731,094	minus	77,376
			LOC101750797	*immunoglobulin omega chain-like*	224,224	584,73	minus	120,202
rs315329074	27	4,528,275	*TRNAI-UAU*	*transfer RNA isoleucine (anticodon UAU)*	4,520,423	4,520,513	plus	7,762
			TRNAQ-UUG	*transfer RNA glutamine (anticodon UUG)*	3,640,823	3,640,894	minus	887,381
			MEOX1	*mesenchyme homeobox 1*	3,530,597	3,536,176	plus	992,099
			ETV4	*ETS variant 4*	3,548,537	3,564,258	plus	964,017
			DHX8	*DEAH-box helicase 8*	3,567,743	3,579,115	minus	949,16
			PHB	*prohibitin*	3,580,883	3,585,030	plus	943,245
			LOC101750197	*uncharacterized LOC101750197*	3,582,478	3,592,058	minus	936,217
			ZNF652	*zinc finger protein*	3,592,238	3,620,525	plus	907,75

				652				
			PHOSPHO1	*phosphoethanolamine /phosphocholine phosphatase*	3,623,753	3,631,815	plus	896,46
			ABI3	*ABI family member 3*	3,631,245	3,637,401	minus	890,874
			GNGT2	*G protein subunit gamma transducin 2*	3,637,559	3,639,810	plus	888,465
			IGF2BP1	*insulin like growth factor 2 mRNA binding protein 1*	3,648,588	3,672,763	minus	855,512
			GIP	*gastric inhibitory polypeptide*	3,682,114	3,689,763	plus	838,512
			SNF8	*SNF8, ESCRT-II complex subunit*	3,690,040	3,693,337	plus	834,938
			UBE2Z	*ubiquitin conjugating enzyme E2 Z*	3,693,620	3,705,485	minus	822,79
			ATP5G1	*ATP synthase, H+ transporting, mitochondrial Fo complex subunit C1 (subunit 9)*	3,707,200	3,709,533	minus	818,742
			CALCOCO2	*calcium binding and coiled-coil domain 2*	3,709,693	3,719,291	plus	808,984
			HOXB13	*homeobox B13*	3,742,876	3,745,452	plus	782,823
			LOC107055286	*uncharacterized LOC107055286*	3,764,713	3,778,499	plus	749,776
			MIR196A1	*microRNA 196a-1*	3,775,036	3,775,130	plus	753,145
			HOXB9	*homeobox B9*	3,781,877	3,788,281	plus	739,994
			HOXB8	*homeobox B8*	3,795,352	3,796,999	plus	731,276
			HOXB7	*homeobox B7*	3,798,149	3,803,481	plus	724,794
			LOC107055284	*uncharacterized LOC107055284*	3,810,034	3,821,988	minus	706,287
			HOXB6	*homeobox B6*	3,812,716	3,815,203	plus	713,072
			HOXB5	*homeobox B5*	3,818,138	3,820,563	plus	707,712
			MIR10A	*microRNA 10a*	3,834,170	3,834,243	plus	694,032

			HOXB4	*homeobox B4*	3,835,207	3,840,489	plus	687,786
			LOC107055283	*uncharacterized LOC107055283*	3,837,488	3,843,542	minus	684,733
			HOXB3	*homeobox B3*	3,840,562	3,864,857	plus	663,418
			LOC107055285	*uncharacterized LOC107055285*	3,852,258	3,856,707	minus	671,568
			HOXB2	*homeobox B2*	3,867,101	3,871,833	plus	656,442
			LOC107055282	*uncharacterized LOC107055282*	3,868,775	3,870,699	minus	657,576
			HOXB1	*homeobox B1*	3,879,443	3,882,191	plus	646,084
			LOC419994	*src kinase-associated phosphoprotein 1-like*	3,927,113	3,969,088	plus	559,187
			LOC101751838	*uncharacterized LOC101751838*	3,980,545	3,984,567	minus	543,708
			TBKBP1	*TBK1 binding protein 1*	4,013,103	4,024,426	minus	503,849
			KPNB1	*karyopherin subunit beta 1*	4,029,274	4,049,551	minus	478,724
			NPEPPS	*aminopeptidase puromycin sensitive*	4,052,247	4,080,818	minus	447,457
			MRPL45	*mitochondrial ribosomal protein L45*	4,083,516	4,087,147	plus	441,128
			GPR179	*G protein-coupled receptor 179*	4,089,962	4,097,067	minus	431,208
			SOCS7	*suppressor of cytokine signaling 7*	4,098,027	4,106,014	plus	422,261
			SKAP1	*src kinase associated phosphoprotein 1*	4,120,634	4,160,530	plus	367,745
			SNX11	*sorting nexin 11*	4,162,424	4,168,386	minus	359,889
			CBX1	*chromobox 1*	4,168,538	4,176,552	plus	351,723
			NFE2L1	*nuclear factor, erythroid 2 like 1*	4,179,406	4,188,312	minus	339,963
			LOC107055292	*uncharacterized LOC107055292*	4,188,341	4,189,844	plus	338,431

			CDK5RAP3	*CDK5 regulatory subunit associated protein 3*	4,192,066	4,195,013	minus	333,262
			PRR15L	*proline rich 15 like*	4,195,737	4,196,844	plus	331,431
			PNPO	*pyridoxamine 5'-phosphate oxidase*	4,199,228	4,201,514	minus	326,761
			SP2	*Sp2 transcription factor*	4,202,904	4,212,077	minus	316,198
			SP6	*Sp6 transcription factor*	4,219,693	4,222,841	plus	305,434
			SCRN2	*secernin 2*	4,223,661	4,226,254	plus	302,021
			LRRC46	*leucine rich repeat containing 46*	4,226,229	4,229,327	minus	298,948
			MRPL10	*mitochondrial ribosomal protein L10*	4,229,306	4,231,308	plus	296,967
			OSBPL7	*oxysterol binding protein like 7*	4,231,743	4,238,314	plus	289,961
			TBX21	*T-box 21*	4,238,676	4,245,894	minus	282,381
			ARHGAP23	*Rho GTPase activating protein 23*	4,260,935	4,274,769	plus	253,506
			SRCIN1	*SRC kinase signaling inhibitor 1*	4,278,768	4,318,568	minus	209,707
			LOC107055293	*SKI/DACH domain-containing protein 1-like*	4,356,552	4,360,423	minus	167,852
			MIR6663	*microRNA 6663*	4,371,072	4,371,181	minus	157,094
			MLLT6	*MLLT6, PHD finger domain containing*	4,374,478	4,403,841	plus	124,434
			MIR1735	*microRNA 1735*	4,391,229	4,391,307	plus	136,968
			LOC107055294	*polycomb group RING finger protein 2-like*	4,405,220	4,407,907	minus	120,368
			LOC107055296	*protein AF-17-like*	4,410,156	4,419,961	plus	108,314
			CISD3	*CDGSH iron sulfur domain 3*	4,421,720	4,422,397	plus	105,878

			PCGF2	*polycomb group ring finger 2*	4,422,951	4,428,429	minus	99,846
			LOC107055298	*POU domain, class 3, transcription factor 3-like*	4,425,831	4,427,935	plus	100,34
			PSMB3	*proteasome subunit beta 3*	4,430,285	4,433,174	plus	95,101
			PIP4K2B	*phosphatidylinositol-5-phosphate 4-kinase type 2 beta*	4,434,946	4,450,427	minus	77,848
			CWC25	*CWC25 spliceosome associated protein homolog*	4,451,015	4,459,377	minus	68,898
			RPL23	*ribosomal protein L23*	4,463,064	4,465,066	minus	63,209
			LASP1	*LIM and SH3 protein 1*	4,466,693	4,486,609	plus	41,666
			FBXO47	*F-box protein 47*	4,490,226	4,500,566	minus	27,709
			LOC101749109	*uncharacterized LOC101749109*	4,498,009	4,506,617	plus	21,658
			PLXDC1	*plexin domain containing 1*	4,505,312	4,518,001	minus	10,274
			LOC100858629	*dickkopf-related protein 1-like*	4,519,315	4,521,946	plus	6,329
			CACNB1	***calcium voltage-gated channel auxiliary subunit beta 1***	4,521,859	4,534,179	minus	0
			RPL19	*ribosomal protein L19*	4,534,596	4,536,944	plus	6,321
			FBXL20	*F-box and leucine rich repeat protein 20*	4,537,208	4,558,569	minus	8,933
			MED1	*mediator complex subunit 1*	4,558,906	4,572,538	minus	30,631
			CDK12	*cyclin dependent kinase 12*	4,572,918	4,595,934	plus	44,643
			NEUROD2	*neuronal*	4,610,541	4,613,182	minus	82,266

				differentiation 2				
			PPP1R1B	*protein phosphatase 1 regulatory inhibitor subunit 1B*	4,621,254	4,626,595	plus	92,979
			STARD3	*StAR related lipid transfer domain containing 3*	4,626,814	4,643,949	plus	98,539
			TCAP	*titin-cap*	4,644,741	4,646,231	plus	116,466
			PNMT	*phenylethanolamine N-methyltransferase*	4,647,235	4,649,104	plus	118,96
			PGAP3	*post-GPI attachment to proteins 3*	4,649,381	4,652,406	minus	121,106
			ERBB2	*erb-b2 receptor tyrosine kinase 2*	4,653,394	4,662,322	plus	125,119
			MIR6547	*microRNA 6547*	4,660,683	4,660,802	plus	132,408
			MIEN1	*migration and invasion enhancer 1*	4,663,033	4,664,878	minus	134,758
			GRB7	*growth factor receptor bound protein 7*	4,665,681	4,671,472	plus	137,406
			LOC100858293	*retinol dehydrogenase 8-like*	4,671,617	4,674,052	plus	143,342
			IKZF3	*IKAROS family zinc finger 3*	4,676,390	4,700,355	minus	148,115
			ZPBP2	*zona pellucida binding protein 2*	4,700,373	4,706,427	plus	172,098
			LRRC3C	*leucine rich repeat containing 3C*	4,710,386	4,714,334	minus	182,111
			LOC107055297	*basic proline-rich protein-like*	4,713,858	4,715,843	plus	185,583
			ORMDL3	*ORMDL sphingolipid biosynthesis regulator 3*	4,717,225	4,725,376	plus	188,95
			GSDMA	*gasdermin A*	4,725,470	4,730,952	plus	197,195
			PSMD3	*proteasome 26S subunit, non-ATPase*	4,731,389	4,735,157	plus	203,114

				3				
			CSF3	*colony stimulating factor 3*	4,737,417	4,739,289	plus	209,142
			MIR6884	*microRNA 6884*	4,741,104	4,757,219	minus	212,829
			THRA	*thyroid hormone receptor, alpha*	4,764,287	4,775,504	plus	236,012
			NR1D1	*nuclear receptor subfamily 1 group D member 1*	4,776,457	4,783,436	minus	248,182
			MSL1	*male specific lethal 1 homolog*	4,787,718	4,792,939	plus	259,443
			CASC3	*cancer susceptibility 3*	4,793,935	4,803,203	plus	265,66
			RAPGEFL1	*Rap guanine nucleotide exchange factor like 1*	4,804,082	4,809,927	plus	275,807
			WIPF2	*WAS/WASL interacting protein family member 2*	4,810,433	4,822,471	plus	282,158
			CDC6	*cell division cycle 6*	4,822,688	4,827,078	plus	294,413
			RARA	*retinoic acid receptor alpha*	4,833,202	4,836,052	plus	304,927
			GJD3	*gap junction protein delta 3*	4,836,854	4,839,029	minus	308,579
			TOP2A	*topoisomerase (DNA) II alpha*	4,839,029	4,856,730	minus	310,754
			LOC101747522	*collagen alpha-1(XVIII) chain-like*	4,857,355	4,863,016	minus	329,08
			IGFBP4	*insulin like growth factor binding protein 4*	4,865,485	4,870,779	plus	337,21
			TNS4	*tensin 4*	4,870,793	4,882,163	minus	342,518
			LOC107055315	*uncharacterized LOC107055315*	4,882,026	4,907,719	plus	353,751
			CCR7	*C-C motif chemokine receptor 7*	4,891,171	4,899,901	minus	362,896

			SMARCE1	*SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily e, member 1*	4,907,121	4,921,663	minus	378,846
			KRT222	*keratin 222*	4,920,500	4,929,780	minus	392,225
			LOC107055316	*uncharacterized LOC107055316*	4,922,980	4,937,276	plus	394,705
			KRT12	*keratin 12*	4,932,629	4,939,390	minus	404,354
			KRT20	*keratin 20*	4,939,479	4,943,592	minus	411,204
			KRT23	*keratin 23*	4,953,002	4,962,692	minus	424,727
			KRT15	*keratin 15*	4,965,385	4,969,810	minus	437,11
			KRT19	*keratin 19*	4,972,722	4,977,384	minus	444,447
			LOC420043	*keratin 16-like*	4,983,598	4,987,746	minus	455,323
			LOC100857659	*keratin, type I cytoskeletal 42-like*	4,992,053	4,995,428	minus	463,778
			KRT10	*keratin, type I cytoskeletal 10-like*	4,997,475	5,002,485	minus	469,2
			KRT9L	*keratin, type I cytoskeletal 9-like*	5,008,298	5,012,957	minus	480,023
			LOC772080	*keratin, type I cytoskeletal 17-like*	5,016,347	5,021,212	minus	488,072
			KRTC42L	*keratin, type I cytoskeletal 42-like*	5,025,520	5,030,185	minus	497,245
			LOC771995	*keratin, type I cytoskeletal 42-like*	5,035,336	5,038,163	minus	507,061
			LOC107055312	*uncharacterized LOC107055312*	5,040,540	5,042,410	plus	512,265
			KRT14	*keratin 14*	5,042,180	5,046,007	minus	513,905
			KRT17	*keratin 17*	5,048,947	5,052,819	minus	520,672
			LOC107055313	*uncharacterized LOC107055313*	5,051,305	5,052,029	plus	523,03
			EIF1	*eukaryotic translation initiation*	5,064,906	5,066,718	plus	536,631

				factor 1				
			LOC396365	*preprogastrin*	5,069,166	5,069,814	plus	540,891
			HAP1	*huntingtin associated protein 1*	5,069,870	5,078,795	minus	541,595
			JUP	*junction plakoglobin*	5,081,253	5,096,734	minus	552,978
			P3H4	*prolyl 3-hydroxylase family member 4 (non-enzymatic)*	5,097,792	5,102,079	minus	569,517
			FKBP10	*FK506 binding protein 10*	5,102,602	5,108,343	plus	574,327
			NT5C3B	*5'-nucleotidase, cytosolic IIIB*	5,108,609	5,113,795	minus	580,334
			KLHL10	*kelch like family member 10*	5,113,816	5,117,579	plus	585,541
			KLHL11	*kelch like family member 11*	5,117,750	5,123,733	minus	589,475
			ACLY	*ATP citrate lyase*	5,124,218	5,152,667	minus	595,943
			TTC25	*tetratricopeptide repeat domain 25*	5,149,242	5,156,676	plus	620,967
			CNP	*2',3'-cyclic nucleotide 3' phosphodiesterase*	5,156,868	5,161,795	plus	628,593
			DNAJC7	*DnaJ heat shock protein family (Hsp40) member C7*	5,161,978	5,181,291	minus	633,703
			LOC107055311	*uncharacterized LOC107055311*	5,172,889	5,180,535	plus	644,614
			NKIRAS2	*NFKB inhibitor interacting Ras like 2*	5,181,433	5,183,584	plus	653,158
			ZNF385C	*zinc finger protein 385C*	5,183,603	5,237,098	minus	655,328
			ZNF862L	*zinc finger protein 862-like*	5,198,667	5,206,571	minus	670,392
			LOC107055310	*uncharacterized LOC107055310*	5,223,122	5,224,406	minus	694,847
			DHX58	*DExH-box helicase 58*	5,242,809	5,249,758	minus	714,534

			KAT2A	*lysine acetyltransferase 2A*	5,249,877	5,256,053	minus	721,602
			LOC772158	*heat shock protein 30C-like*	5,256,460	5,257,349	minus	728,185
			HSPB9	*heat shock protein family B (small) member 9*	5,258,006	5,258,998	plus	729,731
			RAB5C	*RAB5C, member RAS oncogene family*	5,260,322	5,270,725	minus	732,047
			KCNH4	*potassium voltage-gated channel subfamily H member 4*	5,271,367	5,280,067	minus	743,092
			HCRT	*hypocretin neuropeptide precursor*	5,280,135	5,281,806	minus	751,86
			PIBPPDD4L	*1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase delta-4-like*	5,281,905	5,292,750	minus	753,63
			GHDC	*GH3 domain containing*	5,294,204	5,298,441	minus	765,929
			STAT5B	*signal transducer and activator of transcription 5B*	5,297,863	5,311,670	minus	769,588
			STAT3	*signal transducer and activator of transcription 3*	5,319,501	5,334,933	minus	791,226
			PTRF	*polymerase I and transcript release factor*	5,335,537	5,350,085	minus	807,262
			ATP6V0A1	*ATPase H+ transporting V0 subunit a1*	5,352,689	5,381,019	plus	824,414
			NAGLU	*N-acetylglucosaminidase, alpha*	5,381,264	5,384,665	plus	852,989

			HSD17B1	*hydroxysteroid 17-beta dehydrogenase 1*	5,384,880	5,386,410	plus	856,605
			MLX	*MLX, MAX dimerization protein*	5,390,278	5,394,161	plus	862,003
			PSMC3IP	*PSMC3 interacting protein*	5,393,341	5,397,274	minus	865,066
			FAM134C	*family with sequence similarity 134 member C*	5,397,367	5,405,515	minus	869,092
			TUBG1	*tubulin gamma 1*	5,405,589	5,413,030	plus	877,314
			LOC107055309	*uncharacterized LOC107055309*	5,407,102	5,408,044	minus	878,827
			PLEKHH3	*pleckstrin homology, MyTH4 and FERM domain containing H3*	5,413,790	5,421,146	minus	885,515
			CCR10	*C-C motif chemokine receptor 10*	5,422,621	5,427,142	minus	894,346
			CNTNAP1	*contactin associated protein 1*	5,427,307	5,435,737	plus	899,032
			EZH1	*enhancer of zeste 1 polycomb repressive complex 2 subunit*	5,435,970	5,451,101	minus	907,695
			RAMP2	*receptor activity modifying protein 2*	5,453,137	5,454,994	plus	924,862
			VPS25	*vacuolar protein sorting 25 homolog*	5,455,180	5,457,660	plus	926,905
			WNK4	*WNK lysine deficient protein kinase 4*	5,461,875	5,473,473	plus	933,6
			COA3	*cytochrome c oxidase assembly factor 3*	5,473,589	5,474,238	minus	945,314
			CNTD1	*cyclin N-terminal domain containing 1*	5,474,297	5,478,907	plus	946,022
			BECN1	*beclin 1*	5,478,573	5,483,381	minus	950,298
			PSME3	*proteasome activator subunit 3*	5,483,531	5,490,236	plus	955,256
			AOC3	*amine oxidase,*	5,490,347	5,504,360	plus	962,072

				copper containing 3				
			G6PC	*glucose-6-phosphatase catalytic subunit*	5,506,845	5,510,258	plus	978,57
			AARSD1	*alanyl-tRNA synthetase domain containing 1*	5,511,314	5,515,677	minus	983,039
			PTGES3L	*prostaglandin E synthase 3 (cytosolic)-like*	5,515,779	5,517,799	minus	987,504
			RUNDC1	*RUN domain containing 1*	5,518,274	5,520,926	plus	989,999
			RPL27	*ribosomal protein L27*	5,521,996	5,524,344	plus	993,721
			IFI35	*interferon-induced protein 35*	5,524,907	5,535,384	plus	996,632

* Note: Positions are based on Gallus gallus 5.0 genome assembly. Genes including the significant marker are shown in bold.

Table S2. Published QTL/associations related to growth traits in the searched genomic regions. In bold are shown the QTL that included all candidate genes in the predefined distances.

SNP ID	GGA	Distance (bp)	Number of QTL	QTL (bp)*	QTL type	QTL IDs*
rs13923872	1	613,054	20	**37,278,942 -128,288,555**	**Chest width**	**16706**
				18,054,807 -171,631,116	**Visceral fat weight**	**17319**
				18,054,807 -171,717,298	**Total white fat weight**	**17332**
				25,724,479 -171,631,116	**Subcutaneous neck fat weight**	**17325**
				18,054,807 -196,202,543	**Body weight**	**1797**
				37,278,942 -133,528,161	**Body weight**	**55919**
					(140 days)	
				18,054,807 - 171,631,116	**Body weight**	**17076**

					(140 days)	
				18,054,807 -171,631,116	**Carcass weight**	**17110**
				100,051,042 -123,004,362	**Breast muscle weight**	**9410**
				2,420,814 -171,631,116	**Shank length**	**9409**
				106,349,346 -123,007,835	**Carcass fat content**	**17119**
				113,159,926 -128,288,555	Shank weight	14341
				113,159,926 -128,288,555	Femur weight	14342
				18,054,807 -172,427,968	**Spleen weight**	**1851**
				18,054,807 -171,631,116	**Growth (70-105 days)**	**55937**
				18,054,807 -168,151,247	**Abdominal fat weight**	**6858**
				18,054,807 -171,631,116	**Shank length**	**9294**
				113,159,926 -115,848,566	Breast muscle weight	13385
				111,368,640 -164,599,096	**Body weight (35 days)**	**14355**
				6,580,919 -171,631,116	**Subcutaneous fat thickness**	**14359**
rs312691174	4	650,472	14	**17,148,380 -81,264,760**	**Body weight (168 days)**	**24875**
				17,148,380 -81,264,760	**Body weight (21 days)**	**24842**
				17,148,380 -81,264,760	**Body weight (336 days)**	**24883**
				17,148,380 -81,264,760	**Body weight (42 days)**	**24855**
				17,148,380 -81,264,760	**Average daily gain**	**24899, 24905,24911**
				17,148,380 -81,264,760	**Body weight (84 days)**	**24866**
				17,148,380 -81,264,760	**Body weight (504 days)**	**24890**
				4,964,691 -87,025,255	**Visceral fat weight**	**17321**
				10,768,639 -91,268,419	**Average daily gain**	**24914**
				18,357,474 -31,942,137	**Shank length**	**9295**
				18,354,191 -37,883,325	**Thigh muscle weight**	**9395**
				18,354,191 -47,647,218	**Drumstick and thigh muscle weight**	**13404**
				17,148,380 -81,264,760	**Body weight (day of first egg)**	**14457, 14464, 14470**
				17,148,380 -81,264,760	**Head percentage**	**15571**

rs15608447	4	718,407	36	67,546,750 -67,546,790	Body weight (28 days)	65710
				17,148,380 -81,264,760	**Body weight (168 days)**	**24875**
				47,647,218 -89,464,128	**Body weight**	**200.820.152.016**
				17,148,380 -81,264,760	**Body weight (21 days)**	**24842**
				17,148,380 -81,264,760	**Body weight (336 days)**	**24883**
				47,647,218 -89,464,128	**Carcass weight**	**2012**
				17,148,380 -81,264,760	**Body weight (42 days)**	**24855**
				17,148,380 -81,264,760	**Average daily gain**	**24899, 24905, 24911**
				17,148,380 -81,264,760	**Body weight (84 days)**	**24866**
				47,647,218 -89,464,128	**Liver weight**	**2017**
				52,604,411 -82,619,142	**Growth (14-28 days)**	**12499**
				17,148,380 -81,264,760	**Body weight (504 days)**	**24890**
				48,804,413 -85,154,534	**Growth (28-42 days)**	**12500**
				52,191,247 -89,464,128	**Abdominal fat percentage**	**9421**
				48,404,949 -82,619,142	**Growth (0-14 days)**	**12498**
				52,535,768 -70,787,114	**Shank length**	**9286**
				49,665,708 -85,877,678	**Total white fat weight**	**17334**
				30,906,204 -83,247,658	**Tibia width**	**2035**
				61,970,484 -83,247,658	**Tibia width**	**2038**
				62,452,715 -89,022,456	**Growth (42-56 days)**	**12501**
				4,964,691 -87,025,255	**Visceral fat weight**	**17321**
				62,331,035 -82,550,230	**Pectoralis major weight**	**2041**
				31,561,525 -89,318,267	**Body weight (35 days)**	**55905**
				32,974,594 -89,318,267	**Growth (0-35 days)**	**55930**
				47,647,218 -89,464,128	**Drumstick muscle weight**	**2057**
				10,768,639 -91,268,419	**Average daily gain**	**24914**
				47,647,218 -89,464,128	**Drumstick weight**	**2059**
				47,647,218 -89,464,128	**Wing weight**	**2060**

				47,647,218 -89,464,128	**Body weight (42 days)**	**9759**
				47,647,218 -89,464,128	**Body weight (63 days)**	**9760**
				47,647,218 -89,464,128	**Growth (21-42 days)**	**9761**
				47,647,218 -89,464,128	**Growth (42-63 days)**	**9762**
				62,331,035 -82,299,229	**Shank length**	**11795**
				47,647,218 -89,464,128	**Skin fat weight**	**12636**
				47,647,218 -89,464,128	**Drumstick and thigh muscle weight**	**13395**
				17,148,380 -81,264,760	**Body weight (day of first egg)**	**14457, 14464, 14470**
rs318199727	10	737,906	11	13,330,009 -13,330,034	Dressing percentage	57550
				13,330,009 -13,330,034	Breast muscle percentage	57551
				13,329,989 -13,330,029	Breast muscle percentage	57547
				13,329,989 -13,330,029	Drumstick and thigh muscle percentage	57548
				13,329,989 -13,330,029	Abdominal fat percentage	57549
				692,555 -20,423,025	**Carcass weight**	**17113**
				1,541,735 -16,171,711	**Body weight (140 days)**	**55923**
				2,357,400 -17,864,188	**Body weight (35 days)**	**55907**
				692,555 -20,423,025	**Body weight (70 days)**	**55911**
				2,552,841 -18,059,263	**Growth (0-35 days)**	**55931**
				4,410,690 -18,434,155	**Body weight (105 days)**	**55917**
rs318098582	11	300,257	9	**1,133,281 -19,983,730**	**Body weight (140 days)**	**55924**
				18,193,544 -20,208,550	**Thigh meat-to-bone ratio**	**6736**
				18,193,544 -20,208,550	**Body weight (40 days)**	**6737**
				953,174 -20,208,550	**Body weight (140 days)**	**17080**
				6,823,128 -20,208,550	**Carcass weight**	**17114**
				12,510,855 -20,208,550	**Carcass weight**	**17088**
				6,910,612 -20,208,550	**Spleen weight**	**2287**
				18,193,544 -18,870,770	Body weight	2284, 2285
				18,642,683 -18,686,657	Growth (8-46 days)	9519

rs317945754	15	935,183	21	3,731,712 -3,769,767	Spleen weight	2349
				3,731,712 -3,769,767	Body weight (42 days)	9727
				3,731,712 -3,769,767	Carcass weight	9728
				3,731,712 -3,769,767	Spleen percentage	12588
				2,812,987 -10,689,472	Body weight (336 days)	24887
				4,236,686 -4,265,310	Abdominal fat weight	11995
				1,931,502 -7,215,657	**Visceral fat weight**	**17323**
				2,519,182 -7,215,657	**Subcutaneous neck fat weight**	**17331**
				3,749,008 -7,973,093	Drumstick and thigh percentage	15586
				3,749,008 -7,973,093	Abdominal fat weight	2337
				3,749,008 -7,973,093	Abdominal fat percentage	2339, 2340
				3,749,008 -8,228,905	Body weight (35 days)	3355
				1,931,502 -7,215,657	**Total white fat weight**	**17337**
				2,812,987 -10,689,472	Liver weight	2348
				2,403,639 -10,689,472	**Abdominal fat percentage**	**9450**
				1,931,502 -10,689,472	**Abdominal fat weight**	**9451**
				2,812,987 -10,689,472	Abdominal fat weight	2347, 12631
				3,749,008 -10,689,472	Body weight (46 days)	6648
				3,749,008 -10,689,472	Growth (8-46 days)	6649
				2,812,987 -10,689,472	Fat distribution	12645
				1,931,502 -9,638,429	**Breast muscle weight**	**9449**
rs316794400	22	26,589	1	QTL: 3625173-4599266 bp could not be remapped from Gallus gallus 4 to Gallus gallus 5.0 assembly by NCBI Genome Remapping Service	Breast muscle percentage	95429
rs317627533	26	773,988	6	3,118,976 -4,116,802	Body weight (28 days)	95418

				1,263,919 -4,918,464	Body weight (63 days)	9453
				2,499,704 -4,918,464	Shank weight	2383
				4,610,791 -4,624,276	Liver percentage	2385
				QTL: 4715796-108192374 bp could not be remapped from Gallus gallus 4 to Gallus gallus 5.0 assembly by NCBI Genome Remapping Service	Abdominal fat weight	30883
				4,873,346 -4,886,832	Breast muscle weight	6957
rs314452928	27	140,067	3	QTL: 81131-81301 bp could not be remapped from Gallus gallus 4 to Gallus gallus 5.0 assembly by NCBI Genome Remapping Service	Growth (105-140 days)	55944
				54,597 -4,520,058	Growth (0-35 days)	55932
				54,597 -4,520,058	Body weight (35 days)	55906
rs315329074	27	998,553	65	3,834,510 -3,834,550	Shank length	66068, 66069, 66070
				3,363,708 -3,363,748	Shank length	66067
				3,971,422 -3,971,462	Shank circumference	66063
				3,564,173 -3,564,213	Shank circumference	66065
				3,624,903 -3,624,943	Shank circumference	66064, 66066
				3,869,461 -3,869,501	Shank length	66071
				3,456,748 -3,456,788	Abdominal fat weight	66072
				1,798,380 -3,707,375	Abdominal fat weight	11817, 11809
				1,798,380 -3,707,375	Abdominal fat percentage	11820
				1,798,380 -3,707,375	Carcass fat content	17135, 17126
				1,798,380 -3,707,375	Head percentage	15599

1,798,380 -3,707,375	Body weight	2406, 2407
1,798,380 -3,707,375	Body weight (1 day)	7178
1,798,380 -3,707,375	Body weight (41 days)	7186
1,365,641 -4,520,058	Humerus length	2397
3,522,988 -3,534,446	Body weight (112 days)	9521
3,522,988 -3,534,446	Body weight (200 days)	9522
3,522,988 -3,534,446	Growth (46-112 days)	9523
1,365,641 -4,520,058	Body weight	2410
1,809,980 -3,707,375	Body weight (35 days)	3356
1,809,980 -3,707,375	Abdominal fat percentage	3354
1,798,380 -3,707,375	Carcass protein content, dry matter basis	17124
1,798,380 -3,707,375	Carcass fat content, dry matter basis	17125
3,701,574 -3,713,173	Body weight (42 days)	9775
3,701,574 -3,713,173	Growth (21-42 days)	9776
3,701,574 -3,713,173	Body weight (day of first egg)	14459, 14466, 14473
3,701,574 -3,713,173	Body weight (168 days)	24878
3,701,574 -3,713,173	Body weight (336 days)	24888
3,701,574 -3,713,173	Body weight (504 days)	24892
3,701,574 -3,713,173	Average daily gain	24907
3,707,375 -3,968,049	Body weight	2404, 2405
3,707,375 -4,520,058	Thigh weight	2411
3,707,375 -4,520,058	Wing weight	2412
3,788,374 -3,889,766	Drumstick and thigh weight	11920
3,788,374 -3,889,766	Drumstick and thigh percentage	11921
3,788,374 -3,889,766	Abdominal fat percentage	11934
3,788,374 -3,889,766	Pectoralis major percent	11950
3,204,318 -4,520,058	Shank weight	2413
1,798,380 -3,707,375	Body weight (112 days)	6652

				1,798,380 -3,707,375	Body weight (200 days)	6653
				1,798,380 -3,707,375	Growth (46-112 days)	6654
				1,798,380 -3,379,175	Femur length	6778
				1,798,380 -3,707,375	Shank weight percentage	15567
				2,263,107 -4,520,058	Carcass weight	17116
				404,762 -4,520,058	Wing weight	17109
				2,780,009 -4,520,058	Body weight (105 days)	55918
				54,597 -4,520,058	Growth (0-35 days)	55932
				1,365,641 -4,520,058	Body weight	2409
				54,597 -4,520,058	Body weight (35 days)	55906
				2,454,458 -4,520,058	Body weight (140 days)	55926
				2,639,460 -4,520,058	Body weight (70 days)	55912
				3,707,375 -4,520,058	Body weight (35 days)	7159
				1,850,810 -4,520,058	Carcass weight	17090
				2,639,460 -4,520,058	Growth (35-70 days)	55936
				4,377,710 -4,389,305	Shank length	9288
				3,707,375 -4,520,058	Shank weight percentage	15595
				2,390,652 -4,520,058	Breast muscle weight	17096
				3,597,175 -4,520,058	Drumstick and thigh weight	17105
				3,707,375 -4,520,058	Intramuscular fat	3360
				2,141,304 -4,520,058	Body weight (140 days)	17084
				3,707,375 -4,520,058	Body weight	2408
				3,788,374 -5,629,582	**Thigh percentage**	**30886**
				5,159,872 -5,171,472	Body weight (56 days)	12395
				5,159,872 -5,171,472	Body weight (hatch)	16623
				5,159,872 -5,171,472	Body weight (300 days)	16624

*Note: The positions of QTL were remapped to Gallus_gallus-5.0 genome assembly using NCBI database (https://www.ncbi.nlm.nih.gov/genome/tools/remap).

Table S3. Genes and chromosomes per module.

Module_ ID	GGA	Gene ID
module_1	4	*NIPAL1*
module_2	1	*SYTL5, SRPX, PRRG1, XK, OTC, CYBB, RPGR, MED14, TSPAN7*
	4	*ZAR1, PCDH18, LRRC66, NOCT, GABRA4, GABRG1, MGARP, CORIN, SPATA18, RASL11B, CWH43, SLC10A4, TEC, GABRB1, SLC7A11, SGCB, CNGA1, GABRA2, TXK*
	10	*TICRR, RHCG, SV2B, PLIN1, MFGE8, ACAN, RLBP1, HAPLN3, NTRK3*
	11	*CIDEC, SLC22A31, ZFPM1, IL17C, ZNF469, CBFA2T3, CDH15, CYBA*
	15	*ADGRD1, TMEM132C, PIWIL1, TMEM132D, TMEM132B, STX2, FZD10, RIMBP2, SLC15A4, GLT1D1, MMP17*
	22	*IL1B, ANXA4*
	25	*NES, LIM2, BCAN, SLAMF8, TMEM79, RHBG, INSRR, S100A14, HAPLN2, S100A16, LRRC71, PEAR1, DCST2, SV2A, SLAMF1, CD48, NHLH1, CRABP2, RAB25, ZBTB7B, CHRNB2, S100A13, PLEKHO1, OTUD7B, S100A11, ETV3, S100A9, MTMR11, SH2D2A, CD244, S100A6, KCNN3, S100A1, S100A4, LY9, CRNN*
	26	*DENND2C, OPTC, IP6K3, PIFO, PACSIN1, TCP11, FRS3, CHIA, SCUBE3, SYT6, USP49, TSPO2, GRM4, PRELP, SPDEF, SYCP1, MKRN3, MLN, MDFI, FMOD, TSHB, AMPD1, NGF, PGC, TREM2, TSPAN2, ATP2B4, BTG2*

	27	*GPR179, ARHGAP23, KRT222, LRRC46, KLHL10, LRRC3C, TNS4, RAPGEFL1, HOXB8, HOXB4, GNGT2, FKBP10, GSDMA, KCNH4, HOXB9, CCR10, TTC25, ZPBP2, GJD3, TBKBP1, HCRT, G6PC, THRA, MEOX1, RAMP2, HOXB6, HAP1, NEUROD2, HOXB2, AOC3, P3H4, GIP, KRT15, HSD17B1, HOXB13, CNTNAP1, KRT17, IGFBP4, KRT20, PNMT, KRT12, HOXB1, HOXB7, HOXB5, PTRF, PRR15L, CACNB1, PPP1R1B, KRT23, WNK4, CCR7, KLHL11, SKAP1, IKZF3, HOXB3, PLXDC1, ORMDL3, TBX21, PGAP3, OSBPL7, KRT14, KRT19, CSF3, NR1D1, ABI3*
module_3	1	*BCOR*
	4	*NFXL1, ATP10D*
	10	*KIF7*
	11	*SPG7*
	22	*NT5DC4*
	25	*SHE, KIRREL*
	26	*ANKS1A*
	27	*FBXO47, AARSD1, PHOSPHO1, PLEKHH3, SRCIN1, PCGF2, FBXL20, NT5C3B, RUNDC1, MLLT6, STAT5B, ERBB2, SOCS7*
module_4	1	*DYNLT3,ATP6AP2*
	4	*RAB33B, FRYL, USP46, ELF2*
	10	*AGBL1*
	11	*CTU2, GALNS*
	25	*ETV3L, RAB2B*
	26	*PPARD, UHRF1BP1, SIKE1, NRAS*
	27	*MROH8, VWA5A, ETV4*
module_5	1	*USP9X, MID1IP1*
	4	*SCFD2, COMMD8, FIP1L1, NAA15, OCIAD1, SLAIN2, DCUN1D4*
	10	*DET1, MRPL46, AEN, KLHL25, FANCI, MRPS11, POLG, AKAP13*
	11	*PABPN1L, ACSF3, MVD, ANKRD11, TRAPPC2L, BANP, ZC3H18, APRT, PIEZO1, CDT1*

	15	*AACS, ULK1, EP400, RAN, SFSWAP, PUS1*
	22	*CKAP2L, SLC20A1*
	25	*VANGL2, MEX3A, RRNAD1, PYGO2, BGLAP, NTRK1, ARHGEF11, CHTOP, FLAD1, MEF2D, CCT3, MRPL24, ILF2, ADAR, PMVK, VPS45, PRCC, COPA, PBXIP1, GPATCH4, NCSTN, ANP32E, SMG5, NAXE, UBE2Q1, INTS3, LAMTOR2, SNAPIN, SHC1, SMAD4, CKS1B, SF3B4, HDGF, IL6R, UBQLN4*
	26	*TOMM6, FANCE, BCAS2, APOBEC2, BYSL, PRICKLE4, ZNF76, NFYA, TAF11, LEMD2, RPS10, RPL10A, SNRPC, CSDE1, NUDT3, MED20, OARD1, TRIM33, TFEB, TAF8, CCND3, HMGA1, BAK1*
	27	*PTGES3L, HSPB9, IGF2BP1, PSMC3IP, TCAP, TUBG1, MRPL45, GRB7, RPL23, STARD3, RARA, SP2, SMARCE1, WIPF2, CASC3, RPL19, CBX1, ATP5G1, PSMB3, EIF1, PSMD3, ATP6V0A1, ACLY, NFE2L1, PHB, MSL1, BECN1, NPEPPS, EZH1, MED1, MRPL10, UBE2Z, LASP1, KRT10, FAM134C, PIP4K2B, CNP, DNAJC7, MLX, SNF8, SNX11, NKIRAS2, COA3, CISD3, MIEN1, RPL27, TOP2A, DAD1, CALCOCO2, KAT2A, JUP, CDC6, PSME3, KPNB1, STAT3, DHX8, VPS25, RAB5C, CDK12*
module_6	1	*LANCL3*
	4	*SETD7, MGST2, MAML3*
	10	*PEX11A, ABHD2*
	11	*RNF166*
	25	*GLMP*
	26	*ITPR3, DEF6*
	27	*ZNF385C, SP6, CNTD1, PNPO, DHX58, GHDC, NAGLU, CWC25, CDK5RAP3, SCRN2, IFI35, ZNF652*

Table S4. Significantly enriched GO biological processes (BPs) per each module. GO BP terms in bold are associated with development. None enriched GO BP was found for module_6.

Module_ ID	GO_ID	GO BP term	Considered relevant to BW35 as a subclass or child term of the developmental process or growth parent term	P-value	Number of Genes	Associated Genes Found
module_2	**GO:0048704**	**embryonic skeletal system morphogenesis**	yes	0.00000000483	9	*HOXB3*, *MDFI*, *HOXB4*, *HOXB1*, *HOXB2*, *HOXB7*, *HOXB8*, *HOXB5*, *HOXB6*
	GO:0009952	**anterior/posterior pattern specification**	yes	0.000000125	10	*HOXB3*, *HOXB4*, *HOXB1*, *HOXB2*, *BTG2*, *HOXB7*, *HOXB8*, *HOXB5*, *HOXB6*, *HOXB9*
	GO:0007417	**central nervous system development**	yes	0.001	7	*HAPLN2*, *NES*, *HAPLN3*, *GABRA4*, *ACAN*, *BCAN*, *NHLH1*
	GO:0007275	**multicellular organism development**	yes	0.003	14	*CSF3*, *ZBTB7B*, *ZAR1*, *TBX21*, *HOXB3*, *FZD10*, *TCP11*, *HOXB1*, *HOXB2*, *HOXB7*, *HOXB8*, *MEOX1*, *PIWIL1*, *SPDEF*

	GO:0030851	**granulocyte differentiation**	yes	0.005	3	*CSF3, ZFPM1, CBFA2T3*
	GO:0042340	keratan sulfate catabolic process	no	0.006	3	*FMOD, ACAN, PRELP*
	GO:0006811	ion transport	no	0.099	6	*CYBB, GABRA2, WNK4, GABRB1, CHRNB2, SLC15A4*
	GO:0008544	**epidermis development**	yes	0.011	5	*KRT17, KRT15, CRABP2, KRT14, HOXB13*
	GO:0001501	**skeletal system development**	yes	0.013	6	*HAPLN2, HAPLN3, ACAN, BCAN, IGFBP4, PRELP*
	GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	no	0.017	5	*NTRK3, TXK, TEC, NGF, INSRR*
	GO:0007155	cell adhesion	no	0.020	11	*HAPLN2, CDH15, SRPX, HAPLN3, ACAN, BCAN, CNTNAP1, MFGE8, SLAMF1, AOC3, PCDH18*
	GO:0021570	**rhombomere 4 development**	yes	0.020	2	*HOXB1, HOXB2*
	GO:0007214	gamma-aminobutyric acid signaling pathway	no	0.021	3	*GABRG1, GABRA2, GABRA4*
	GO:0008306	associative learning	no	0.025	3	*BTG2, NEUROD2, CHRNB2*

	GO:0007165	signal transduction	no	0.029	20	*S100A6, CD244, GIP, STX2, GABRB1, CRABP2, S100A9, S100A11, ARHGAP23, ANXA4, SLAMF1, CD48, SH2D2A, KRT17, PPP1R1B, IL1B, CHRNB2, CNTNAP1, FRS3, IGFBP4*
	GO:1904044	response to aldosterone	no	0.030	2	*CYBA, CYBB*
	GO:0070634	transepithelial ammonium transport	no	0.030	2	*RHCG, RHBG*
	GO:1904845	cellular response to L-glutamine	no	0.030	2	*CYBA, CYBB*
	GO:0018146	keratan sulfate biosynthetic process	no	0.033	3	*FMOD, ACAN, PRELP*
	GO:0006954	inflammatory response	no	0.042	9	*IL17C, CYBA, CYBB, CCR7, TSPAN2, S100A9, IL1B, IGFBP4, AOC3*
	GO:0030154	**cell differentiation**	yes	0.049	10	*ZBTB7B, SH2D2A, TCP11, NR1D1, STX2, SPDEF, TXK, NHLH1, TEC, ETV3*
module_3	GO:0016311	dephosphorylation	no	0.005	3	*NT5C3B, PHOSPHO1, NT5DC4*

	GO:0071364	cellular response to epidermal growth factor stimulus	no	0.039	2	*ERBB2, STAT5B*
module_4	**GO:0030154**	**cell differentiation**	yes	0.005	4	*PPARD, ELF2, ETV3L, ETV4*
	GO:0007264	small GTPase mediated signal transduction	no	0.015	3	*NRAS, RAB2B, RAB33B*
	GO:0006357	regulation of transcription from RNA polymerase II promoter	no	0.044	3	*PPARD, ELF2, ETV3L*
module_5	GO:0045893	positive regulation of transcription, DNA-templated	no	0.0000287	16	*CKS1B, RAN, TAF8, PHB, SMAD4, TFEB, NAA15, BANP, NFYA, HMGA1, STAT3, ARHGEF11, ILF2, RARA, NFE2L1, MED1*
	GO:0045471	response to ethanol	no	0.000256	7	*BAK1, BGLAP, NTRK1, RARA, RPL10A, AACS, STAT3*
	GO:0000184	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	no	0.000502	7	*RPL19, RPL23, SMG5, RPL27, RPS10, CASC3, RPL10A*

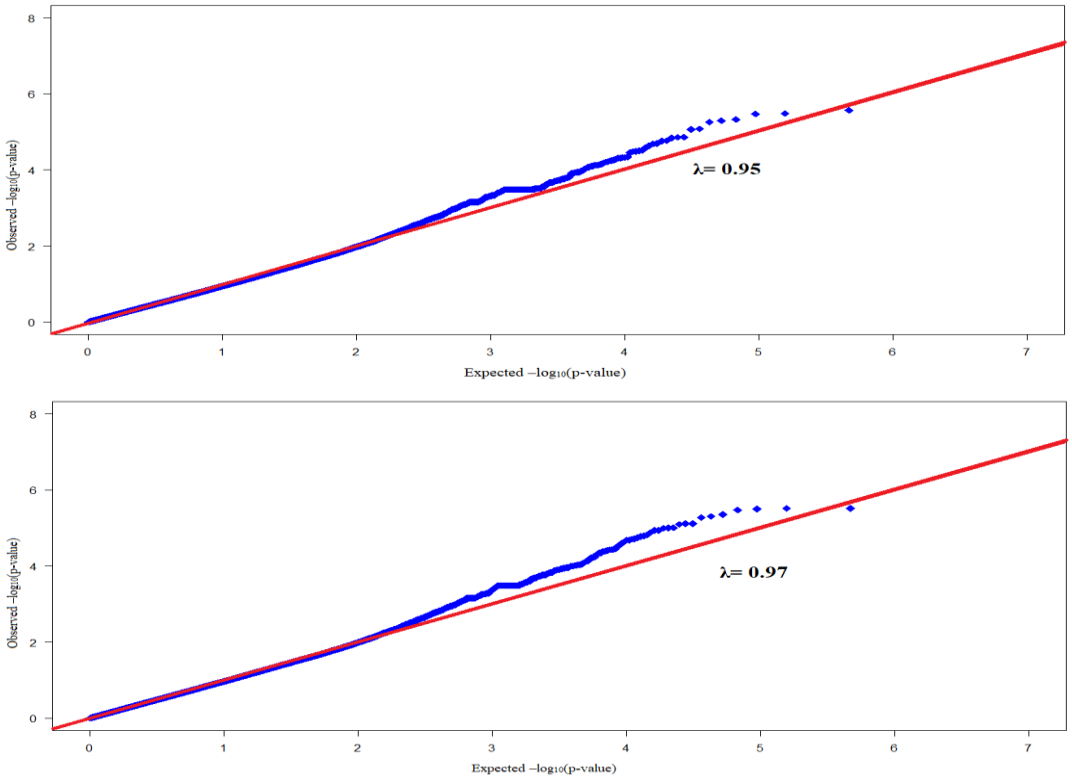
	GO:0070102	interleukin-6-mediated signaling pathway	no	0.002	3	*SMAD4, IL6R, STAT3*
	GO:0046902	regulation of mitochondrial membrane permeability	no	0.003	3	*BAK1, CNP, STAT3*
	GO:0006412	translation	no	0.006	8	*MRPL24, MRPL10, RPL19, RPL23, MRPS11, RPL27, RPS10, RPL10A*
	GO:0070125	mitochondrial translational elongation	no	0.006	5	*MRPL24, MRPL10, MRPS11, MRPL45, MRPL46*
	GO:0070126	mitochondrial translational termination	no	0.006	5	*MRPL24, MRPL10, MRPS11, MRPL45, MRPL46*
	GO:0006413	translational initiation	no	0.006	6	*RPL19, RPL23, RPL27, RPS10, EIF1, RPL10A*
	GO:0006614	SRP-dependent cotranslational protein targeting to membrane	no	0.008	5	*RPL19, RPL23, RPL27, RPS10, RPL10A*
	GO:0006606	protein import into nucleus	no	0.013	4	*RAN, KPNB1, STAT3, ADAR*
	GO:0019083	viral transcription	no	0.015	5	*RPL19, RPL23, RPL27, RPS10, RPL10A*

	GO:0045737	positive regulation of cyclin-dependent protein serine/threonine kinase activity	no	0.017	3	*CKS1B, CDC6, CCND3*
	GO:0006397	mRNA processing	no	0.018	6	*APOBEC2, BCAS2, CDK12, CASC3, SF3B4, ADAR*
	GO:0008285	negative regulation of cell proliferation	no	0.019	9	*CDC6, BAK1, BECN1, NTRK1, PHB, SMAD4, RARA, HMGA1, STAT3*
	GO:0019287	isopentenyl diphosphate biosynthetic process, mevalonate pathway	no	0.025	2	*MVD, PMVK*
	GO:0006914	autophagy	no	0.026	5	*BECN1, SNF8, TFEB, UBQLN4, VPS25*
	GO:0016236	macroautophagy	no	0.026	4	*TOMM6, BECN1, LAMTOR2, ULK1*
	GO:0035264	**multicellular organism growth**	yes	0.030	4	*KAT2A, SP2, ANKRD11, RARA*

	GO:0043328	protein targeting to vacuole involved in ubiquitin-dependent protein catabolic process via the multivesicular body sorting pathway	no	0.033	2	*SNF8, VPS25*
	GO:0006364	rRNA processing	no	0.035	6	*RPL19, RPL23, BYSL, RPL27, RPS10, RPL10A*
	GO:0000398	mRNA splicing, via spliceosome	no	0.039	6	*BCAS2, DHX8, FIP1L1, CASC3, SNRPC, SF3B4*
	GO:0006695	cholesterol biosynthetic process	no	0.040	3	*MVD, ACLY, PMVK*
	GO:0016032	viral process	no	0.041	7	*KAT2A, RAN, PSMB3, CALCOCO2, SNAPIN, SHC1, STAT3*
	GO:0042493	response to drug	no	0.043	7	*BAK1, BGLAP, BECN1, NTRK1, DAD1, AACS, STAT3*
	GO:0010508	positive regulation of autophagy	no	0.044	3	*BECN1, ULK1, TFEB*
	GO:0060348	**bone development**	yes	0.048	3	*BGLAP, ANKRD11, AKAP13*

	GO:0006886	intracellular protein transport	no	0.049	6	*COPA, RAN, VPS45, SNAPIN, KPNB1, SNX11*
	GO:0000395	mRNA 5'-splice site recognition	no	0.049	2	*SFSWAP, SNRPC*

Chapter 4



Supplementary Figure 1. Quantile- quantile (Q-Q) plots of the additive (top) and dominant (bottom) SNP effects for EN. Blue dots denote the $-\log_{10}$(p-value) obtained from the additive (λ= 0.95) and dominant (λ= 0.97) genetic models and the red lines represent the expected values for the null hypothesis under no association. Q-Q plots were constructed with the qqman package [49] in R (http://www.r-project.org/).

Supplementary Table 1. List of training genes retrieved by the NCBI database for the 'reproduction' and 'egg production' queried terms in *Gallus gallus*.

Gene ID	***Description***	**Queried term**
GNRH1	*gonadotropin releasing hormone 1*	reproduction
IGF1	*insulin like growth factor 1*	
MSTN	*myostatin*	
TGFB2	*transforming growth factor beta 2*	
GHRL	*ghrelin, preproghrelin*	
PCK1	*phosphoenolpyruvate carboxykinase 1*	
IGF2	*insulin like growth factor 2*	
LITAF	*lipopolysaccharide induced TNF factor*	
STAT3	*signal transducer and activator of transcription 3*	
TSHR	*thyroid stimulating hormone receptor*	
ACTA2	*actin, alpha 2, smooth muscle, aorta*	
LAMA1	*laminin subunit alpha 1*	
STAT5A	*signal transducer and activator of transcription 5A*	
ZNF764L	*zinc finger protein 764-like*	
PRL	*prolactin*	egg production
GH	*growth hormone*	
PPARG	*peroxisome proliferator-activated receptor gamma*	
PRLR	*prolactin receptor*	
HMGCR	*3-hydroxy-3-methylglutaryl-CoA reductase*	
GDF9	*growth differentiation factor 9*	
POSTN	*periostin*	
NPY	*neuropeptide Y*	
GNRHR	*gonadotropin-releasing hormone receptor*	
FOXL2	*forkhead box L2*	
RARRES1	*retinoic acid receptor responder 1*	
THRSP	*thyroid hormone responsive*	
BMP15	*bone morphogenetic protein 15*	
CETP	*cholesteryl ester transfer protein*	
PDGFRL	*platelet derived growth factor receptor like*	

C9ORF152	*chromosome 2 open reading frame, human C9orf152*	
NCOA1	*nuclear receptor coactivator 1*	

Supplementary Table 2. Positional candidate genes for EN.

Gene ID	***Description***	***GGA***	**Start position of the gene (bp)***	**End position of the gene (bp)***	**Significant SNP(s) associated with each candidate gene**
ITPR1	*inositol 1,4,5-trisphosphate receptor type 1*	12	18802491	18953176	*rs313298834*
LOC112533364	*uncharacterized LOC112533364*		18964030	18970288	
BHLHE40	*basic helix-loop-helix family member e40*		18986975	18991321	
LOC107054403	*uncharacterized LOC107054403*		18991726	18994904	
LOC107054404	*uncharacterized LOC107054404*		18996262	19001149	
ARL8B	*ADP ribosylation factor like GTPase 8B*		19004930	19021010	
EDEM1	*ER degradation enhancing alpha-mannosidase like protein 1*		19022229	19032275	
XPO7	*exportin 7*	22	1659759	1691710	*rs314011910*
DOK2	*docking protein 2*		1692828	1698227	
LOC101749127	*uncharacterized LOC101749127*		1698237	1704849	
GFRA2	*GDNF family receptor alpha 2*		1714010	1734922	
LOC107054969	*uncharacterized LOC107054969*		1725523	1731382	
LOC107054968	*uncharacterized LOC107054968*		1748674	1758867	
KIF21B	*kinesin family member 21B*	26	289715	321414	*rs313045367*
ELF3	*E74 like ETS transcription factor 3*		323816	327820	
GPR37L1	*G protein-coupled receptor 37 like 1*		337772	353528	
ARL8A	*ADP ribosylation factor like GTPase 8A*		355155	365068	
PTPN7	*protein tyrosine phosphatase, non-receptor type 7*		367516	377636	
PTPRVP	*protein tyrosine phosphatase, receptor type, V, pseudogene*		376921	408510	
LOC112530334	*uncharacterized LOC112530334*		410270	413273	

LGR6	*leucine rich repeat containing G protein-coupled receptor 6*		410381	484213	
SUGP2	*SURP and G-patch domain containing 2*		3761200	3776727	*rs15250929*
HOMER3	*homer scaffolding protein 3*		3774743	3814238	*rs10724922, rs15250929*
DDX49	*DEAD-box helicase 49*		3818734	3824430	*rs10724922, rs15250929*
COPE	*coatomer protein complex subunit epsilon*		3824578	3829382	*rs10724922, rs15251036, rs15250929*
CERS1	*ceramide synthase 1*		3831598	3842053	*rs10724922, rs15251036, rs15250929,rs16212031, rs314228493*
GDF3	*growth differentiation factor 3*		3843493	3845584	*rs10724922, rs15251036, rs15250929, rs16212031, rs314228493, rs16212040, rs16212041*
UPF1	*UPF1, RNA helicase and ATPase*		3847285	3866737	*rs10724922, rs15251036, rs15250929, rs16212031, rs314228493, rs16212040, rs16212041*
LOC112530428	*uncharacterized LOC112530428*		3866889	3871355	*rs10724922, rs15251036, rs317783777, rs15250929, rs16212031, rs314228493, rs16212040, rs16212041*
COMP	*cartilage oligomeric matrix protein*	28	3885739	3901177	*rs10724922, rs15251036, rs317783777, rs16212031, rs314228493, rs16212040, rs16212041, rs314418757*
CRTC1	*CREB regulated transcription coactivator 1*		3903025	3943104	*rs10724922, rs15251036, rs317783777, rs315316434, rs314052602, rs16212031, rs314228493, rs16212040, rs16212041, rs314418757*
LOC107055366	*uncharacterized LOC107055366*		3929316	3930938	*rs317783777, rs315316434, rs16212031, rs314228493, rs16212040, rs16212041, rs314418757*
LOC101748347	*uncharacterized LOC101748347*		3943292	3949939	*rs317783777, rs315316434, rs313312915, rs314052602, rs314418757*
KLHL26	*kelch like family member 26*		3951133	3966079	*rs317783777, rs315316434, rs313312915, rs14307369, rs314052602, rs314418757*
TMEM59L	*transmembrane protein 59 like*		3971283	3976388	*rs315316434, rs313312915, rs14307369, rs314052602, rs314418757*
LOC112530430	*uncharacterized LOC112530430*		3976324	3977914	*rs315316434, rs313312915, rs14307369, rs314052602*

CRLF1	*cytokine receptor like factor 1*		3979732	3985638	*rs315316434, rs313312915, rs14307369, rs314052602*
C19orf60	*chromosome 19 open reading frame 60*		3985818	3987504	*rs315316434, rs313312915, rs14307369, rs314052602*
UBA52	*ubiquitin A-52 residue ribosomal protein fusion product 1*		3987712	3988725	*rs315316434, rs313312915, rs14307369, rs314052602*
KXD1	*KxDL motif containing 1*		3989908	3991381	*rs315316434, rs313312915, rs14307369, rs314052602*
FKBP8	*FK506 binding protein 8*		3991437	3997184	*rs315316434, rs313312915, rs14307369, rs314052602*
ELL	*elongation factor for RNA polymerase II*		3997782	4031850	*rs315316434, rs313312915, rs14307369, rs314052602, rs318126353*
LOC112530432	*translation initiation factor IF-2-like*		4031011	4033491	*rs313312915, rs14307369, rs314052602, rs318126353*
ISYNA1	*inositol-3-phosphate synthase 1*		4033985	4037537	*rs313312915, rs14307369, rs314052602, rs318126353*
SSBP4	*single stranded DNA binding protein 4*		4037993	4047230	*rs313312915, rs14307369, rs314052602, rs318126353*
LRRC25	*leucine rich repeat containing 25*		4047671	4051078	*rs313312915, rs14307369, rs318126353*
GDF15	*growth differentiation factor 15*		4050997	4053678	*rs14307369, rs318126353*
LOC112530443	*translation initiation factor IF-2-like*		4053708	4055686	*rs14307369, rs318126353*
PGPEP1	*pyroglutamyl-peptidase I*		4057878	4068206	*rs318126353*
LSM4	*LSM4 homolog, U6 small nuclear RNA and mRNA degradation associated*		4068726	4075072	*rs318126353*
JUND	*JunD proto-oncogene, AP-1 transcription factor subunit*		4074009	4079212	*rs318126353*
LOC107055358	*proline-rich receptor-like protein kinase PERK2*		4079280	4082415	*rs318126353*
PDE4C	*phosphodiesterase 4C*		4084333	4091257	*rs318126353*
RAB3A	*RAB3A, member RAS oncogene family*		4092352	4096084	*rs318126353*
MPV17L2	*MPV17 mitochondrial inner membrane protein like 2*		4096022	4098287	*rs318126353*
IFI30	*IFI30, lysosomal thiol reductase*		4098284	4100575	*rs318126353*
PIK3R2	*phosphoinositide-3-kinase regulatory subunit 2*		4100899	4119455	*rs318126353*

* Positions were based on GRCg6a assembly

Supplementary Table 3. Genes obtained by previous GWA studies for egg and reproductive traits in *Gallus gallus.*

Genes	**Trait**	**Reference**
AJAP1,TNFRSF9,C1ORF174,CAMTA1,CEP104,PDAI1,SDHB,DJ-1,PADI3,MRPS16	eggshell blueness	Darwish et al. 2019 [1]
ZNF704,CA2,RFT1 ,PRKCD,PDGFD,DYNC2H1,DCUN1D5,FGF9,KIAA1468,PHLPP1,ZCCHC2,TLL1	yield of extraembryonic fluid, age at first egg, egg weight	Kudinov et al. 2019 [2]
NCAPG,FGFPB1,KPNA3,CDC25A,WDR48,BST1,THSD7B	egg albumen quality	Qu et al. 2019 [3]
DLEU7,MIR15A,CECR2,MEIS1,SPRED2,RNASEH2B,KCNRG,SPRYD7,CECR1,CECR5	egg weight	Liu et al. 2018 [4]
MSX2,DRD1,RHOA,SDF4,TNFRSF4,TTLL10,LOC419425,MIR429	egg quality	Liu et al. 2018 [5]
FAM184B,HTT,KCNH7,CDC42BPA,KCNIP4,GJA5,CBFB,GPC6,COG6,PPARGC1A,CNKSR2,MED30,STK31	age at first egg , body weight at first egg, egg weight, egg number	Fan et al. 2017 [6]
IGFBP3,GORAB,CCKAR	oviduct development	Shen et al. 2017 [7]
RGS3,AMH,DCLK1,NBEA,SMAD9	follicle number	Shen et al. 2017 [8]
ADIPOR2,LRTM2,SOX5,CMAS,ERC1,CHUNK-1,STRAP,PLCZ1,AEBP2,SLCO1A2,IAPP,KCNJ8,GYS2,SPX,KRAS,IQSEC3,SLC6A12,WNK1,CACNA2D4,DCP1B,CACNA1C,BCL2L13,MICAL3,BPGM,CALD1,RERG,PTPRO,EPS8,DERA,MGST1,LMO3,RERGL,PLEKHA5,SLCO1C1,PYROXD1,BCAT1,C2CD5,ST8SIA1,LDHB,LRMP,LYRM5,RAD52,WNT5B,AKR,CAPZA3,TMTC1,TUBAT,BICD1,DENND5B,SLC6A13,B4GALNT3,RASSF8,SLC15A5,WASH1,CYB5R3,ITM2C,KNDC1,GPR123(ADGRA1)	eggshell ultrastructure	Duan et al. 2016 [9]

DGKQ,SHROOM3,RBPJ,TRPC6,C4,TRPM8,MCCC2,PTPRD,AUTS2,ITGA6,PLCB1,CCDC82,LGSN,EPHX2,ANKMY1,APEM1,POLA1,DAPK1,SMYD3,ARL8A,LOC10751519,LOC100857741,C5H15ORF29,LOC101751216,MEF2A,ELFN1,LAMC3,HAS2,PRL,PXDN,SLC8A1,KATNBL1	total egg numbers , egg weight, eggshell thickness, eggshell color	Liao et al. 2016 [10]
DOTL1,GALC,RPA2,ACTL9,ZAP70,BRCA2,RFX2,ONECUT3,ZAR1,REXO1,NRTN,EML4,KCNG3,ADAMTS10,SCAMP4,ABHD17A,ELL,ENSGALG00000028314,RGS3F,OAZ,C19orf35,AP3D1,MYO1F,MUC16,SPG20,STARD13,NEK5,NEK3,CKAP2,DHRS12,ACER1b,ACSBG2,ACER1	yolk and ovary weights	Sun et al. 2015 [11]
ITPR2,PIK3C2G,RECQL,ABCC9,CASC1	eggshell quality	Sun et al. 2015 [12]
CAB39L,FOXO1,CDADC1	egg weight	Yi et al. 2015 [13]

GTF2A1,STON2,CLSPN,FARSB,KIAA1549,CALM1,Gprotein,GNB2L1,TRIM27,GRIK3,B-G,TSN,THYN1,SEL1L,OPCML,SPPL2B,TCF3,TAP2,PLXNB2, RPS29,METAP1D, TLK1,ZAK ,PPP1R9B, TRA2B ,ZC3H12A, RRAGC ,STMN1 ,PNISR ,KIAA0776, ATG5 ,NRXN3,MOGAT1, MIR1741, SMG6 ,ADORA2B, SVOPL ,HIPK2 ,TAB1 ,PFKP ,PPP4R1L, C3AR1 ,PDE3A ,ARNTL2 ,PPFIBP1 ,HTR2C ,LOC101749001, DMC1 ,ZC3H14, KCNK10 ,TTC7B ,DIO2 ,C5H14ORF159, TTC8 ,LOC423393, NEURL1B ,CDKAL1 ,SOX4 ,MIR216B, MIR7463 ,RTN4IP1 ,MIR6583 ,CCDC88C ,POU3F1	egg number , egg laying rate and age at first egg	Yuan et al. 2015 [14]
GRB14,GALNT1,ODZ2,ZNF536,ATM,LOC418918,ENOX1,BLK	egg production and quality	Liu et al. 2011 [15]

References of Supplementary Table 3

1. Darwish HYA, Dalirsefat SB, Dong X, Hua G, Chen J, Zhang Y, et al. Genome-wide association study and a post replication analysis revealed a promising genomic region and candidate genes for chicken eggshell blueness. PLoS One. 2019;14:e0209181. doi:10.1371/journal.pone.0209181.

2. Kudinov AA, Dementieva N V., Mitrofanova O V., Stanishevskaya OI, Fedorova ES, Larkina TA, et al. Genome-wide association studies targeting the yield of extraembryonic fluid and production traits in Russian White chickens. BMC Genomics. 2019;20:270. doi:10.1186/s12864-019-5605-5.

3. Qu L, Shen M, Guo J, Wang X, Dou T, Hu Y, et al. Identification of potential genomic regions and candidate genes for egg albumen quality by a genome-wide association study. Arch Anim Breed. 2019;62:113–23. doi:10.5194/aab-62-113-2019.

4. Liu Z, Sun C, Yan Y, Li G, Wu G, Liu A, et al. Genome-Wide Association Analysis of Age-Dependent Egg Weights in Chickens. Front Genet. 2018;9:128. doi:10.3389/fgene.2018.00128.

5. Liu Z, Sun C, Yan Y, Li G, Shi F, Wu G, et al. Genetic variations for egg quality of chickens at late laying period revealed by genome-wide association study. Sci Rep. 2018;8:10832. doi:10.1038/s41598-018-29162-7.

6. Fan QC, Wu PF, Dai GJ, Zhang GX, Zhang T, Xue Q, et al. Identification of 19 loci for reproductive traits in a local Chinese chicken by genome-wide study. Genet Mol Res. 2017;16. doi:10.4238/gmr16019431.

7. Shen M, Qu L, Ma M, Dou T, Lu J, Guo J, et al. A genome-wide study to identify genes responsible for oviduct development in chickens. PLoS One. 2017;12:e0189955. doi:10.1371/journal.pone.0189955.

8. Shen M, Sun H, Qu L, Ma M, Dou T, Lu J, et al. Genetic Architecture and Candidate Genes Identified for Follicle Number in Chicken. Sci Rep. 2017;7:16412. doi:10.1038/s41598-017-16557-1.

9. Duan Z, Sun C, Shen M, Wang K, Yang N, Zheng J, et al. Genetic architecture dissection by genome-wide association analysis reveals avian eggshell ultrastructure traits. Sci Rep. 2016;6:28836. doi:10.1038/srep28836.

10. Liao R, Zhang X, Chen Q, Wang Z, Wang Q, Yang C, et al. Genome-wide association study reveals novel variants for growth and egg traits in Dongxiang blue-shelled and White Leghorn chickens. Anim Genet. 2016;47:588–96. doi:10.1111/age.12456.

11. Sun C, Lu J, Yi G, Yuan J, Duan Z, Qu L, et al. Promising Loci and Genes for Yolk and Ovary Weight in Chickens Revealed by a Genome-Wide Association Study. PLoS One. 2015;10:e0137145. doi:10.1371/journal.pone.0137145.

12. Sun C, Qu L, Yi G, Yuan J, Duan Z, Shen M, et al. Genome-wide association study revealed a promising region and candidate genes for eggshell quality in an F2 resource population. BMC Genomics. 2015;16:565. doi:10.1186/s12864-015-1795-7.

13. Yi G, Shen M, Yuan J, Sun C, Duan Z, Qu L, et al. Genome-wide association study dissects genetic architecture underlying longitudinal egg weights in chickens. BMC Genomics. 2015;16:746. doi:10.1186/s12864-015-1945-y.

14. Yuan J, Sun C, Dou T, Yi G, Qu L, Qu L, et al. Identification of Promising Mutants Associated with Egg Production Traits Revealed by Genome-Wide Association Study. PLoS One. 2015;10:e0140615. doi:10.1371/journal.pone.0140615.

15. Liu W, Li D, Liu J, Chen S, Qu L, Zheng J, et al. A Genome-Wide SNP Scan Reveals Novel Loci for Egg Production and Quality Traits in White Leghorn and Brown-Egg Dwarf Layers. PLoS One. 2011;6:e28600. doi:10.1371/journal.pone.0028600.

Chapter 5

Table S1. Genome-wide significant SNPs for body weight (BW) and chromosome-wide significant SNPs for egg number (EN).

SNP ID	GGA	Position (bp)*	Wald test p-value	FDR p-value	Distance (bp)	Trait
rs317275973	1	23082139	1,28E-05	0,020476944	12592975	BW
rs13850889	1	35675114	4,46E-05	0,04534326	9338826	
rs314607281	1	45013940	3,27E-05	0,038514056	7983218	
rs13653309	1	52997158	7,12E-09	4,72E-05	68237936	
rs317073055	1	121235094	3,22E-07	0,000899406		
rs13543487	2	27103453	3,15E-05	0,038350828	32365880	
rs317979230	2	59469333	6,79E-06	0,011793922	32389099	
rs16067706	2	91858432	4,28E-05	0,044442073	1282396	
rs313097265	2	93140828	5,81E-06	0,011397788		
rs313125064	3	21986853	4,05E-06	0,008938395	1460073	
rs316117717	3	23446926	5,08E-06	0,010759583	61658923	
rs317961013	3	85105849	6,24E-06	0,011525504	5797015	
rs315826445	3	90902864	1,43E-05	0,022363434	9088620	
rs313332188	3	99991484	5,30E-11	1,40E-06		
rs313973628	4	8970286	2,06E-05	0,029432058	17560376	
rs313178030	4	26530662	5,17E-05	0,047275168	23130710	
rs317273654	4	49661372	4,70E-05	0,04534326	15783372	
rs313976538	4	65444744	5,73E-06	0,011397788	781079	
rs314007348	4	66225823	3,77E-08	0,00016656	234093	
rs15608447	4	66459916	2,42E-08	0,000116691	238906	
rs14485266	4	66698822	3,83E-05	0,040584485		
rs312798022	5	8828819	1,67E-05	0,025277547	21829468	
rs313257959	5	30658287	3,11E-05	0,038350828		
rs315753942	7	5203886	4,25E-07	0,001124835	31082488	
rs313879964	7	36286374	4,25E-09	3,75E-05		
rs314425715	8	770143	4,82E-05	0,045576328	19910125	

rs316603825	8	20680268	1,14E-08	6,70E-05		
rs317315660	9	17942760	4,71E-05	0,04534326		
rs313036644	10	1446308	6,31E-06	0,011525504	8991974	
rs313184580	10	10438282	1,12E-05	0,018565284		
rs15603546	11	3817134	5,21E-07	0,001315423	219957	
rs313822475	11	4037091	1,19E-07	0,000391989	14370402	
rs318098582	11	18407493	6,24E-09	4,72E-05		
rs314685865	13	12041828	3,53E-06	0,008125337		
rs317631529	14	5738298	1,29E-08	6,85E-05	3878760	
rs318197918	14	9617058	3,12E-06	0,007525859		
rs314778226	15	4845973	2,63E-07	0,000774041	8010884	
rs314215039	15	12856857	2,97E-05	0,038350828		
rs317414603	20	6729013	3,95E-10	4,19E-06		
rs316810914	21	2722503	2,22E-05	0,030998493		
rs314011910	22	1711605	3,51E-05	0,039411178	3430481	
rs313993741	22	5142086	5,02E-05	0,046681406		
rs317101069	23	3379059	3,64E-05	0,039411178		
rs14291881	24	150829	1,67E-10	2,95E-06	213139	
rs314315583	24	363968	3,37E-05	0,038776251	5785443	
rs314507428	24	6149411	2,67E-05	0,036313203		
rs315195881	25	1794356	1,19E-07	0,000391989	498213	
rs315023079	25	2292569	2,78E-10	3,68E-06	1273013	
rs13791894	25	3565582	3,18E-05	0,038350828	205102	
rs312758346	25	3770684	6,90E-06	0,011793922		
rs316975706	26	805527	1,26E-07	0,000391989		
rs315411246	27	4002212	4,65E-05	0,04534326	597349	
rs313353157	27	4599561	1,93E-05	0,028396588	1254027	
rs316714498	27	5853588	6,09E-08	0,000248242	1066764	
rs315329074	27	6920352	1,59E-13	8,41E-09	144105	

rs313658015	27	7064457	3,59E-05	0,039411178		
rs317501178	28	874035	8,84E-08	0,00033453	2787008	
rs314496246	28	3661043	2,79E-05	0,036944764		
rs316318083	21	3382056	1,76E-05	0,021637347		
rs313534177	28	3777407	0,000252	0,049357166	142098	EN
rs317783777	28	3919505	2,01E-05	0,006873961	52423	
rs315316434	28	3971928	8,90E-06	0,006079608	23062	
rs316549515	28	3994990	0,000289	0,049357166		
* Note: Positions are based on GRCg6a assembly						

Table S2. Genome-wide significant cross-phenotype (CP) associations for BW and EN. Note that 44 SNPs were also identified by univariate association analyses.

SNP ID	GGA	Position (bp)*	Wald test p-value	FDR p-value	Distance (bp)	Statistical analysis
rs317275973	1	23082139	4,22E-05	0,045922629	29915019	Bivariate analysis and Univariate analysis for BW
rs13653309	1	52997158	2,26E-08	0,000141565	1429802	Bivariate analysis and Univariate analysis for BW
rs15272503	1	54426960	6,99E-08	0,000336529	66808134	Bivariate analysis
rs317073055	1	121235094	1,73E-06	0,003665709		Bivariate analysis and Univariate analysis for BW
rs317979230	2	59469333	2,27E-05	0,029832557	33671495	Bivariate analysis and Univariate analysis for BW
rs313097265	2	93140828	3,37E-05	0,040579648		Bivariate analysis and Univariate analysis for BW
rs313125064	3	21986853	1,85E-05	0,026379236	1460073	Bivariate analysis and Univariate analysis for BW
rs316117717	3	23446926	4,34E-06	0,007661716	61658923	Bivariate analysis and Univariate analysis for BW
rs317961013	3	85105849	1,55E-06	0,003562813	14885635	Bivariate analysis and Univariate analysis for BW
rs313332188	3	99991484	5,89E-10	7,80E-06		Bivariate analysis and Univariate analysis for BW
rs313976538	4	65444744	3,80E-05	0,043769917	781079	Bivariate analysis and Univariate analysis for BW
rs314007348	4	66225823	3,23E-07	0,001314906	234093	Bivariate analysis and Univariate analysis for BW
rs15608447	4	66459916	4,34E-08	0,000230024	238906	Bivariate analysis and Univariate analysis for BW
rs14485266	4	66698822	2,12E-05	0,028770838	234382	Bivariate analysis and Univariate analysis for BW
rs317866306	4	66933204	1,26E-05	0,019711107		Bivariate analysis
rs314313592	6	34543637	4,67E-05	0,048866635		Bivariate analysis

rs315753942	7	5203886	1,30E-06	0,003277951	31082488	Bivariate analysis and Univariate analysis for BW
rs313879964	7	36286374	9,82E-09	7,43E-05		Bivariate analysis and Univariate analysis for BW
rs316603825	8	20680268	9,46E-09	7,43E-05		Bivariate analysis and Univariate analysis for BW
rs313036644	10	1446308	2,31E-05	0,029832557	8991974	Bivariate analysis and Univariate analysis for BW
rs313184580	10	10438282	1,89E-05	0,026379236	2675362	Bivariate analysis and Univariate analysis for BW
rs14008207	10	13113644	4,00E-05	0,045113352		Bivariate analysis
rs314798708	11	3373672	4,25E-05	0,045922629	443462	Bivariate analysis
rs15603546	11	3817134	2,40E-06	0,004900058	219957	Bivariate analysis and Univariate analysis for BW
rs313822475	11	4037091	8,70E-07	0,002427468	14370402	Bivariate analysis and Univariate analysis for BW
rs318098582	11	18407493	2,40E-08	0,000141565		Bivariate analysis and Univariate analysis for BW
rs314685865	13	12041828	2,74E-05	0,034622475		Bivariate analysis and Univariate analysis for BW
rs317631529	14	5738298	7,69E-08	0,000339649	3878760	Bivariate analysis and Univariate analysis for BW
rs318197918	14	9617058	1,30E-05	0,019737688	205140	Bivariate analysis and Univariate analysis for BW
rs314504536	14	9822198	3,95E-06	0,007468611		Bivariate analysis
rs314778226	15	4845973	1,13E-06	0,002983047	8010884	Bivariate analysis and Univariate analysis for BW
rs314215039	15	12856857	1,12E-05	0,018005895		Bivariate analysis and Univariate analysis for BW
rs317414603	20	6729013	2,51E-09	2,66E-05	6353624	Bivariate analysis and Univariate analysis for BW
rs313569149	20	13082637	6,62E-06	0,011309187		Bivariate analysis
rs316810914	21	2722503	1,40E-06	0,003370082	659553	Bivariate analysis and Univariate analysis for BW
rs316318083	21	3382056	1,81E-05	0,026379236		Bivariate analysis and Univariate analysis for EN
rs314011910	22	1711605	1,64E-06	0,003624311		Bivariate analysis and Univariate analysis for BW
rs317101069	23	3379059	4,33E-06	0,007661716		Bivariate analysis and Univariate analysis for BW
rs14291881	24	150829	4,28E-10	7,56E-06	5998582	Bivariate analysis and Univariate analysis for BW
rs314507428	24	6149411	5,17E-07	0,00160387		Bivariate analysis and Univariate analysis for BW
rs315195881	25	1794356	5,45E-07	0,00160387	498213	Bivariate analysis and Univariate analysis for BW
rs315023079	25	2292569	3,48E-10	7,56E-06	1478115	Bivariate analysis and Univariate analysis for BW
rs312758346	25	3770684	3,77E-05	0,043769917		Bivariate analysis and Univariate analysis for BW
rs316975706	26	805527	5,13E-07	0,00160387		Bivariate analysis and Univariate analysis for BW
rs316714498	27	5853588	4,12E-07	0,001454562	1066764	Bivariate analysis and Univariate analysis for BW
rs315329074	27	6920352	6,18E-13	3,28E-08	144105	Bivariate analysis and Univariate analysis for BW

rs313658015	27	7064457	3,76E-06	0,007372002		Bivariate analysis and Univariate analysis for BW
rs317501178	28	874035	3,47E-07	0,001314906	3045470	Bivariate analysis and Univariate analysis for BW
rs317783777	28	3919505	2,89E-05	0,035590746	52423	Bivariate analysis and Univariate analysis for EN
rs315316434	28	3971928	7,95E-06	0,013157669	23062	Bivariate analysis and Univariate analysis for EN
rs316549515	28	3994990	4,70E-05	0,048866635		Bivariate analysis and Univariate analysis for EN
** Note: Positions are based on GRCg6a assembly*						

Table S3. GO biological processes (BPs) per gene.

GO_ID	GO BP term	P-value	Gene involved
GO:0007020	microtubule nucleation	0.012645755	
GO:0031116	positive regulation of microtubule polymerization	0.012645755	
GO:0031112	positive regulation of microtubule polymerization or depolymerization	0.012645755	
GO:0031122	cytoplasmic microtubule organization	0.012903831	
GO:0046785	microtubule polymerization	0.012903831	
GO:0031113	regulation of microtubule polymerization	0.012903831	
GO:0031110	regulation of microtubule polymerization or depolymerization	0.012903831	
GO:0031109	microtubule polymerization or depolymerization	0.016258827	
GO:0032273	positive regulation of protein polymerization	0.018667543	
GO:0032271	regulation of protein polymerization	0.019269721	
GO:0031334	positive regulation of protein-containing complex assembly	0.019269721	
GO:1902905	positive regulation of supramolecular fiber organization	0.019269721	
GO:0051495	positive regulation of cytoskeleton organization	0.019269721	
GO:0070507	regulation of microtubule cytoskeleton organization	0.019269721	*SLAIN2*
GO:0032886	regulation of microtubule-based process	0.019269721	
GO:0051258	protein polymerization	0.020436443	
GO:1902903	regulation of supramolecular fiber organization	0.025397776	
GO:0043254	regulation of protein-containing complex assembly	0.026295141	
GO:0044089	positive regulation of cellular component biogenesis	0.032993059	
GO:0051493	regulation of cytoskeleton organization	0.034233864	
GO:0010638	positive regulation of organelle organization	0.034324191	
GO:0000226	microtubule cytoskeleton organization	0.035966497	
GO:0097435	supramolecular fiber organization	0.037937264	
GO:0044087	regulation of cellular component biogenesis	0.044224011	
GO:0007017	microtubule-based process	0.044224011	
GO:0034622	cellular protein-containing complex assembly	0.044746516	
GO:0051130	positive regulation of cellular component organization	0.04944557	

GO:0033043	regulation of organelle organization	0.054518687	
GO:0065003	protein-containing complex assembly	0.061858297	
GO:0043933	protein-containing complex subunit organization	0.065093585	
GO:0007010	cytoskeleton organization	0.065093585	
GO:0051128	regulation of cellular component organization	0.087842831	
GO:0022607	cellular component assembly	0.101549242	
GO:0044085	cellular component biogenesis	0.107595181	
GO:0006996	organelle organization	0.141890529	
GO:0048522	positive regulation of cellular process	0.172323498	
GO:0048518	positive regulation of biological process	0.18978397	
GO:0016043	cellular component organization	0.201523887	
GO:0071840	cellular component organization or biogenesis	0.20284161	
GO:0050794	regulation of cellular process	0.31871818	
GO:0050789	regulation of biological process	0.330640219	
GO:0065007	biological regulation	0.346338831	
GO:0009987	cellular process	0.472010144	
GO:0008150	biological process	0.527139103	
GO:0060957	endocardial cell fate commitment	0.003602808	*ACVR1*
GO:0062042	regulation of cardiac epithelial to mesenchymal transition	0.003602808	
GO:0062043	positive regulation of cardiac epithelial to mesenchymal transition	0.003602808	
GO:0003289	atrial septum primum morphogenesis	0.003602808	
GO:0003274	endocardial cushion fusion	0.003602808	
GO:0061445	endocardial cushion cell fate commitment	0.003602808	
GO:0061443	endocardial cushion cell differentiation	0.003602808	
GO:0003284	septum primum development	0.003602808	
GO:1905005	regulation of epithelial to mesenchymal transition involved in endocardial cushion formation	0.003602808	
GO:1905007	positive regulation of epithelial to mesenchymal transition involved in endocardial cushion formation	0.003602808	
GO:2000015	regulation of determination of dorsal identity	0.003602808	
GO:2000017	positive regulation of determination of dorsal identity	0.003602808	

GO:0061343	cell adhesion involved in heart morphogenesis	0.004117495	
GO:0061312	BMP signaling pathway involved in heart development	0.004117495	
GO:0060839	endothelial cell fate commitment	0.004803744	
GO:0060923	cardiac muscle cell fate commitment	0.004803744	
GO:0060956	endocardial cell differentiation	0.004803744	
GO:0003348	cardiac endothelial cell differentiation	0.004803744	
GO:0032926	negative regulation of activin receptor signaling pathway	0.005404213	
GO:0003174	mitral valve development	0.005404213	
GO:0060911	cardiac cell fate commitment	0.005404213	
GO:0003183	mitral valve morphogenesis	0.005404213	
GO:0048262	determination of dorsal/ventral asymmetry	0.005404213	
GO:0048263	determination of dorsal identity	0.005404213	
GO:0042693	muscle cell fate commitment	0.005661556	
GO:0072148	epithelial cell fate commitment	0.005661556	
GO:0003157	endocardium development	0.005661556	
GO:0003198	epithelial to mesenchymal transition involved in endocardial cushion formation	0.005661556	
GO:1901213	regulation of transcription from RNA polymerase II promoter involved in heart development	0.005963269	
GO:0003272	endocardial cushion formation	0.006304915	
GO:0060413	atrial septum morphogenesis	0.006304915	
GO:0032925	regulation of activin receptor signaling pathway	0.006304915	
GO:0061311	cell surface receptor signaling pathway involved in heart development	0.007411491	
GO:0003283	atrial septum development	0.007411491	
GO:0003181	atrioventricular valve morphogenesis	0.007411491	
GO:0001702	gastrulation with mouth forming second	0.007605929	
GO:0003171	atrioventricular valve development	0.007789856	
GO:0060037	pharyngeal system development	0.007964103	
GO:0003203	endocardial cushion morphogenesis	0.008129414	
GO:0003209	cardiac atrium morphogenesis	0.00864674	
GO:0002526	acute inflammatory response	0.00896699	

GO:0001569	branching involved in blood vessel morphogenesis	0.00896699	
GO:2000826	regulation of heart morphogenesis	0.00896699	
GO:0032924	activin receptor signaling pathway	0.00896699	
GO:0003230	cardiac atrium development	0.00896699	
GO:0030501	positive regulation of bone mineralization	0.009042343	
GO:0060317	cardiac epithelial to mesenchymal transition	0.009042343	
GO:0010718	positive regulation of epithelial to mesenchymal transition	0.009042343	
GO:0060412	ventricular septum morphogenesis	0.009042343	
GO:0072132	mesenchyme morphogenesis	0.009042343	
GO:0003197	endocardial cushion development	0.009042343	
GO:0010862	positive regulation of pathway-restricted SMAD protein phosphorylation	0.00914559	
GO:0110151	positive regulation of biomineralization	0.009170785	
GO:0003179	heart valve morphogenesis	0.009170785	
GO:0070169	positive regulation of biomineral tissue development	0.009170785	
GO:0003170	heart valve development	0.010036395	
GO:0051145	smooth muscle cell differentiation	0.010365975	
GO:0060393	regulation of pathway-restricted SMAD protein phosphorylation	0.010435721	
GO:0045669	positive regulation of osteoblast differentiation	0.011048612	
GO:0060389	pathway-restricted SMAD protein phosphorylation	0.011048612	
GO:0003143	embryonic heart tube morphogenesis	0.011339987	
GO:0001707	mesoderm formation	0.011750698	
GO:0001755	neural crest cell migration	0.011750698	
GO:0030500	regulation of bone mineralization	0.011750698	
GO:0003281	ventricular septum development	0.011750698	
GO:0110110	positive regulation of animal organ morphogenesis	0.011791009	
GO:0048332	mesoderm morphogenesis	0.011830117	
GO:0035050	embryonic heart tube development	0.011868075	
GO:0060411	cardiac septum morphogenesis	0.012146611	
GO:0010717	regulation of epithelial to mesenchymal transition	0.012146611	

GO:2001237	negative regulation of extrinsic apoptotic signaling pathway	0.012297586	
GO:0045778	positive regulation of ossification	0.012297586	
GO:0110149	regulation of biomineralization	0.012297586	
GO:0070167	regulation of biomineral tissue development	0.012297586	
GO:0045446	endothelial cell differentiation	0.012297586	
GO:0055007	cardiac muscle cell differentiation	0.012515018	
GO:0003158	endothelium development	0.012563639	
GO:0014032	neural crest cell development	0.012563639	
GO:0018107	peptidyl-threonine phosphorylation	0.012587027	
GO:0048864	stem cell development	0.012829513	
GO:0090100	positive regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	0.012829513	
GO:0014031	mesenchymal cell development	0.012829513	
GO:0014033	neural crest cell differentiation	0.01284857	
GO:0018210	peptidyl-threonine modification	0.013038735	
GO:0009953	dorsal/ventral pattern formation	0.013054882	
GO:0030282	bone mineralization	0.013238226	
GO:0007498	mesoderm development	0.013251709	
GO:0045667	regulation of osteoblast differentiation	0.013264885	
GO:0003279	cardiac septum development	0.01329036	
GO:0007368	determination of left/right symmetry	0.01329036	
GO:0035051	cardiocyte differentiation	0.013797989	
GO:0001704	formation of primary germ layer	0.013797989	
GO:0009855	determination of bilateral symmetry	0.013797989	
GO:0003231	cardiac ventricle development	0.013797989	
GO:0009799	specification of symmetry	0.013804445	
GO:0090101	negative regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	0.013810765	
GO:0031214	biomineral tissue development	0.01382302	
GO:0110148	biomineralization	0.01382302	
GO:0003206	cardiac chamber morphogenesis	0.013974529	

GO:2001236	regulation of extrinsic apoptotic signaling pathway	0.014699458	
GO:0001837	epithelial to mesenchymal transition	0.014835093	
GO:0000082	G1/S transition of mitotic cell cycle	0.014835093	
GO:0048754	branching morphogenesis of an epithelial tube	0.014970893	
GO:0050731	positive regulation of peptidyl-tyrosine phosphorylation	0.015104081	
GO:0044843	cell cycle G1/S phase transition	0.015646482	
GO:0007179	transforming growth factor beta receptor signaling pathway	0.015770784	
GO:0030509	BMP signaling pathway	0.015892762	
GO:0003205	cardiac chamber development	0.016412794	
GO:0007369	gastrulation	0.016579295	
GO:0071772	response to BMP	0.016579295	
GO:2000027	regulation of animal organ morphogenesis	0.016579295	
GO:0071773	cellular response to BMP stimulus	0.016579295	
GO:0061138	morphogenesis of a branching epithelium	0.016579295	
GO:0001649	osteoblast differentiation	0.016666905	
GO:0030278	regulation of ossification	0.016666905	
GO:0048738	cardiac muscle tissue development	0.016895929	
GO:2001234	negative regulation of apoptotic signaling pathway	0.017098073	
GO:0001763	morphogenesis of a branching structure	0.017098073	
GO:0071560	cellular response to transforming growth factor beta stimulus	0.017173386	
GO:0097191	extrinsic apoptotic signaling pathway	0.017173386	
GO:0007281	germ cell development	0.01724623	
GO:0071559	response to transforming growth factor beta	0.01724623	
GO:0048863	stem cell differentiation	0.017781602	
GO:0050730	regulation of peptidyl-tyrosine phosphorylation	0.017781602	
GO:0048762	mesenchymal cell differentiation	0.019368698	
GO:0098742	cell-cell adhesion via plasma-membrane adhesion molecules	0.019786852	
GO:0090092	regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	0.019971473	
GO:0051146	striated muscle cell differentiation	0.020828736	

GO:0003007	heart morphogenesis	0.021002418	
GO:0022412	cellular process involved in reproduction in multicellular organism	0.021838561	
GO:0060485	mesenchyme development	0.022053554	
GO:0045165	cell fate commitment	0.022053554	
GO:0060562	epithelial tube morphogenesis	0.024163196	
GO:0048562	embryonic organ morphogenesis	0.024659222	
GO:0042692	muscle cell differentiation	0.024659222	
GO:0007178	transmembrane receptor protein serine/threonine kinase signaling pathway	0.025322596	
GO:0018212	peptidyl-tyrosine modification	0.025322596	
GO:2001233	regulation of apoptotic signaling pathway	0.025322596	
GO:0018108	peptidyl-tyrosine phosphorylation	0.025322596	
GO:0001503	ossification	0.025322596	
GO:0044772	mitotic cell cycle phase transition	0.02565404	
GO:0014706	striated muscle tissue development	0.026488253	
GO:0043009	chordate embryonic development	0.026605354	
GO:0009792	embryo development ending in birth or egg hatching	0.027341861	
GO:0060537	muscle tissue development	0.027341861	
GO:0001667	ameboidal-type cell migration	0.027341861	
GO:0044770	cell cycle phase transition	0.027351933	
GO:0003002	regionalization	0.027371671	
GO:0048568	embryonic organ development	0.027371671	
GO:0030335	positive regulation of cell migration	0.029110691	
GO:0001525	angiogenesis	0.02929652	
GO:0006954	inflammatory response	0.02929652	
GO:2000147	positive regulation of cell motility	0.029481804	
GO:0051272	positive regulation of cellular component movement	0.029758261	
GO:0007276	gamete generation	0.029938175	
GO:0002009	morphogenesis of an epithelium	0.030300542	
GO:0040017	positive regulation of locomotion	0.030382919	

GO:0007389	pattern specification process	0.031011515	
GO:0048609	multicellular organismal reproductive process	0.032425275	
GO:0032504	multicellular organism reproduction	0.032425275	
GO:0097190	apoptotic signaling pathway	0.033118984	
GO:0048514	blood vessel morphogenesis	0.03318142	
GO:0030855	epithelial cell differentiation	0.033419915	
GO:0007507	heart development	0.033919122	
GO:0019953	sexual reproduction	0.034324938	
GO:0061061	muscle structure development	0.034725864	
GO:0044703	multi-organism reproductive process	0.035121988	
GO:0003006	developmental process involved in reproduction	0.035729626	
GO:0048729	tissue morphogenesis	0.035729626	
GO:0071363	cellular response to growth factor stimulus	0.035773768	
GO:0001568	blood vessel development	0.035985945	
GO:0070848	response to growth factor	0.036195656	
GO:0001944	vasculature development	0.036939138	
GO:0048598	embryonic morphogenesis	0.036939138	
GO:0072358	cardiovascular system development	0.037139807	
GO:1903047	mitotic cell cycle process	0.039081311	
GO:0043066	negative regulation of apoptotic process	0.039081311	
GO:0098609	cell-cell adhesion	0.039347525	
GO:0043069	negative regulation of programmed cell death	0.039691274	
GO:0060548	negative regulation of cell death	0.04275598	
GO:0030334	regulation of cell migration	0.04275598	
GO:0035239	tube morphogenesis	0.0432337	
GO:2000145	regulation of cell motility	0.044572449	
GO:0045597	positive regulation of cell differentiation	0.04527007	
GO:0009790	embryo development	0.045336961	
GO:0000278	mitotic cell cycle	0.047274945	

GO:0051270	regulation of cellular component movement	0.047274945	
GO:0001934	positive regulation of protein phosphorylation	0.047274945	
GO:0040012	regulation of locomotion	0.047274945	
GO:0007167	enzyme linked receptor protein signaling pathway	0.048012555	
GO:0022414	reproductive process	0.048012555	
GO:0022603	regulation of anatomical structure morphogenesis	0.048012555	
GO:0000003	reproduction	0.048012555	
GO:0042327	positive regulation of phosphorylation	0.04951552	
GO:0006952	defense response	0.04951552	
GO:0035295	tube development	0.049998157	
GO:0072359	circulatory system development	0.050546863	
GO:0060429	epithelium development	0.050546863	
GO:0045937	positive regulation of phosphate metabolic process	0.050546863	
GO:0009887	animal organ morphogenesis	0.050546863	
GO:0010562	positive regulation of phosphorus metabolic process	0.050546863	
GO:0048646	anatomical structure formation involved in morphogenesis	0.050653345	
GO:0031401	positive regulation of protein modification process	0.054663299	
GO:0009968	negative regulation of signal transduction	0.055172418	
GO:0007155	cell adhesion	0.055817167	
GO:0022610	biological adhesion	0.055895997	
GO:0042981	regulation of apoptotic process	0.056391783	
GO:0022402	cell cycle process	0.056605662	
GO:0043067	regulation of programmed cell death	0.056886448	
GO:0010648	negative regulation of cell communication	0.057301809	
GO:0023057	negative regulation of signaling	0.057303435	
GO:0018193	peptidyl-amino acid modification	0.057441001	
GO:0051094	positive regulation of developmental process	0.057441958	
GO:0071495	cellular response to endogenous stimulus	0.057644933	
GO:0045944	positive regulation of transcription by RNA polymerase II	0.057711962	

GO:0010941	regulation of cell death	0.059579775	
GO:0016477	cell migration	0.06042141	
GO:0001932	regulation of protein phosphorylation	0.06042141	
GO:0009719	response to endogenous stimulus	0.060606146	
GO:0009967	positive regulation of signal transduction	0.062950888	
GO:0006915	apoptotic process	0.06364418	
GO:0048585	negative regulation of response to stimulus	0.064085352	
GO:0012501	programmed cell death	0.064085352	
GO:0048870	cell motility	0.064085352	
GO:0032270	positive regulation of cellular protein metabolic process	0.064085352	
GO:0051674	localization of cell	0.064085352	
GO:0042325	regulation of phosphorylation	0.064501379	
GO:0051240	positive regulation of multicellular organismal process	0.065419415	
GO:0010647	positive regulation of cell communication	0.066628559	
GO:0051247	positive regulation of protein metabolic process	0.066628559	
GO:0023056	positive regulation of signaling	0.066628559	
GO:0007049	cell cycle	0.066651954	
GO:0008219	cell death	0.066759731	
GO:0045595	regulation of cell differentiation	0.066759731	
GO:0045893	positive regulation of transcription, DNA-templated	0.066843593	
GO:0019220	regulation of phosphate metabolic process	0.068513405	
GO:0040011	locomotion	0.068513405	
GO:1903508	positive regulation of nucleic acid-templated transcription	0.068513405	
GO:1902680	positive regulation of RNA biosynthetic process	0.068513405	
GO:0051174	regulation of phosphorus metabolic process	0.068513405	
GO:0051704	multi-organism process	0.06918588	
GO:0031399	regulation of protein modification process	0.070756793	
GO:0009888	tissue development	0.070756793	
GO:0051254	positive regulation of RNA metabolic process	0.070756793	

GO:2000026	regulation of multicellular organismal development	0.073232594	
GO:0010557	positive regulation of macromolecule biosynthetic process	0.075556871	
GO:0045935	positive regulation of nucleobase-containing compound metabolic process	0.075556871	
GO:0006468	protein phosphorylation	0.075644506	
GO:0006928	movement of cell or subcellular component	0.075644506	
GO:0048584	positive regulation of response to stimulus	0.075976022	
GO:0031328	positive regulation of cellular biosynthetic process	0.076247482	
GO:0009891	positive regulation of biosynthetic process	0.076802724	
GO:0010628	positive regulation of gene expression	0.078378882	
GO:0048468	cell development	0.078467464	
GO:0006357	regulation of transcription by RNA polymerase II	0.080815936	
GO:0006366	transcription by RNA polymerase II	0.082752004	
GO:0071310	cellular response to organic substance	0.082934685	
GO:0050793	regulation of developmental process	0.08919385	
GO:0016310	phosphorylation	0.08919385	
GO:0032879	regulation of localization	0.093258901	
GO:0032268	regulation of cellular protein metabolic process	0.093258901	
GO:0007166	cell surface receptor signaling pathway	0.094083984	
GO:0010033	response to organic substance	0.094083984	
GO:0009653	anatomical structure morphogenesis	0.094655601	
GO:0051246	regulation of protein metabolic process	0.098268297	
GO:0051239	regulation of multicellular organismal process	0.100555994	
GO:0009966	regulation of signal transduction	0.100555994	
GO:0070887	cellular response to chemical stimulus	0.100555994	
GO:0048513	animal organ development	0.108111037	
GO:0006950	response to stress	0.111233372	
GO:0006355	regulation of transcription, DNA-templated	0.11204601	
GO:0010646	regulation of cell communication	0.112209205	
GO:1903506	regulation of nucleic acid-templated transcription	0.112209205	

GO:0023051	regulation of signaling	0.112209205	
GO:2001141	regulation of RNA biosynthetic process	0.112209205	
GO:0051173	positive regulation of nitrogen compound metabolic process	0.112209205	
GO:0032774	RNA biosynthetic process	0.114012275	
GO:0031325	positive regulation of cellular metabolic process	0.114012275	
GO:0006793	phosphorus metabolic process	0.114012275	
GO:0006796	phosphate-containing compound metabolic process	0.114012275	
GO:0097659	nucleic acid-templated transcription	0.114012275	
GO:0006351	transcription, DNA-templated	0.114012275	
GO:0010604	positive regulation of macromolecule metabolic process	0.114067713	
GO:0051252	regulation of RNA metabolic process	0.118537469	
GO:2000112	regulation of cellular macromolecule biosynthetic process	0.121256623	
GO:0030154	cell differentiation	0.121437316	
GO:0048583	regulation of response to stimulus	0.122045133	
GO:0009893	positive regulation of metabolic process	0.122045133	
GO:0010556	regulation of macromolecule biosynthetic process	0.123019075	
GO:0019219	regulation of nucleobase-containing compound metabolic process	0.123737831	
GO:0042221	response to chemical	0.124402125	
GO:0048869	cellular developmental process	0.124848338	
GO:0034654	nucleobase-containing compound biosynthetic process	0.124848338	
GO:0031326	regulation of cellular biosynthetic process	0.124848338	
GO:0009889	regulation of biosynthetic process	0.125499859	
GO:0018130	heterocycle biosynthetic process	0.125964855	
GO:0019438	aromatic compound biosynthetic process	0.125964855	
GO:1901362	organic cyclic compound biosynthetic process	0.129120782	
GO:0036211	protein modification process	0.130945924	
GO:0006464	cellular protein modification process	0.130945924	
GO:0010468	regulation of gene expression	0.130945924	
GO:0048731	system development	0.13528426	

GO:0043412	macromolecule modification	0.136930498	
GO:0048523	negative regulation of cellular process	0.142121308	
GO:0016070	RNA metabolic process	0.142652372	
GO:0007275	multicellular organism development	0.142652372	
GO:0044271	cellular nitrogen compound biosynthetic process	0.14270407	
GO:0034645	cellular macromolecule biosynthetic process	0.143971965	
GO:0009059	macromolecule biosynthetic process	0.14695727	
GO:0048519	negative regulation of biological process	0.154200197	
GO:0090304	nucleic acid metabolic process	0.156675973	
GO:0048856	anatomical structure development	0.156675973	
GO:0048522	positive regulation of cellular process	0.158109187	
GO:0010467	gene expression	0.159209816	
GO:0044267	cellular protein metabolic process	0.159209816	
GO:0032502	developmental process	0.167690206	
GO:0051171	regulation of nitrogen compound metabolic process	0.16832682	
GO:0007165	signal transduction	0.16832682	
GO:0006139	nucleobase-containing compound metabolic process	0.16832682	
GO:0044249	cellular biosynthetic process	0.168971711	
GO:0046483	heterocycle metabolic process	0.169900857	
GO:0080090	regulation of primary metabolic process	0.169900857	
GO:1901576	organic substance biosynthetic process	0.169900857	
GO:0006725	cellular aromatic compound metabolic process	0.170488449	
GO:0009058	biosynthetic process	0.171826244	
GO:0048518	positive regulation of biological process	0.171829645	
GO:0060255	regulation of macromolecule metabolic process	0.172229664	
GO:1901360	organic cyclic compound metabolic process	0.173154484	
GO:0031323	regulation of cellular metabolic process	0.173460438	
GO:0019538	protein metabolic process	0.173502516	
GO:0023052	signaling	0.175155262	

GO:0007154	cell communication	0.176233757	
GO:0051179	localization	0.181341353	
GO:0032501	multicellular organismal process	0.181341353	
GO:0034641	cellular nitrogen compound metabolic process	0.181341353	
GO:0019222	regulation of metabolic process	0.181341353	
GO:1901564	organonitrogen compound metabolic process	0.195684878	
GO:0051716	cellular response to stimulus	0.200307617	
GO:0050896	response to stimulus	0.232237663	
GO:0044260	cellular macromolecule metabolic process	0.238336368	
GO:0043170	macromolecule metabolic process	0.273855696	
GO:0006807	nitrogen compound metabolic process	0.289868054	
GO:0050794	regulation of cellular process	0.296501673	
GO:0044238	primary metabolic process	0.304185946	
GO:0044237	cellular metabolic process	0.308400394	
GO:0050789	regulation of biological process	0.31254883	
GO:0071704	organic substance metabolic process	0.3141815	
GO:0065007	biological regulation	0.332490691	
GO:0008152	metabolic process	0.332490691	
GO:0009987	cellular process	0.462600591	
GO:0008150	biological process	0.527139103	
GO:0097503	sialylation	0.028576121	*ST3GAL3*
GO:0006486	protein glycosylation	0.053580227	
GO:0070085	glycosylation	0.053580227	
GO:0043413	macromolecule glycosylation	0.053580227	
GO:0009100	glycoprotein metabolic process	0.054572453	
GO:0009101	glycoprotein biosynthetic process	0.054572453	
GO:1901137	carbohydrate derivative biosynthetic process	0.074161838	
GO:1901135	carbohydrate derivative metabolic process	0.101207095	
GO:1901566	organonitrogen compound biosynthetic process	0.151744494	

GO:0006464	cellular protein modification process	0.24897572	
GO:0044267	cellular protein metabolic process	0.24897572	
GO:0044249	cellular biosynthetic process	0.24897572	
GO:0043412	macromolecule modification	0.24897572	
GO:0034645	cellular macromolecule biosynthetic process	0.24897572	
GO:0019538	protein metabolic process	0.24897572	
GO:0009059	macromolecule biosynthetic process	0.24897572	
GO:0009058	biosynthetic process	0.24897572	
GO:0036211	protein modification process	0.24897572	
GO:1901576	organic substance biosynthetic process	0.24897572	
GO:1901564	organonitrogen compound metabolic process	0.272425686	
GO:0044260	cellular macromolecule metabolic process	0.318816523	
GO:0043170	macromolecule metabolic process	0.346439671	
GO:0044237	cellular metabolic process	0.346439671	
GO:0044238	primary metabolic process	0.346439671	
GO:0071704	organic substance metabolic process	0.346439671	
GO:0006807	nitrogen compound metabolic process	0.346439671	
GO:0008152	metabolic process	0.354820168	
GO:0009987	cellular process	0.477757021	
GO:0008150	biological process	0.527139103	
GO:2000344	positive regulation of acrosome reaction	0.003117815	*CACNA1H*
GO:0032342	aldosterone biosynthetic process	0.003117815	
GO:0032341	aldosterone metabolic process	0.003117815	
GO:0034650	cortisol metabolic process	0.003117815	
GO:0034651	cortisol biosynthetic process	0.003117815	
GO:0035864	response to potassium ion	0.003117815	
GO:0035865	cellular response to potassium ion	0.003117815	
GO:0045956	positive regulation of calcium ion-dependent exocytosis	0.003117815	
GO:0034309	primary alcohol biosynthetic process	0.003117815	

GO:0008212	mineralocorticoid metabolic process	0.003117815	
GO:1902644	tertiary alcohol metabolic process	0.003117815	
GO:0060046	regulation of acrosome reaction	0.003117815	
GO:1905516	positive regulation of fertilization	0.003117815	
GO:1902645	tertiary alcohol biosynthetic process	0.003117815	
GO:0006705	mineralocorticoid biosynthetic process	0.003117815	
GO:0006704	glucocorticoid biosynthetic process	0.003117815	
GO:0006700	C21-steroid hormone biosynthetic process	0.00326046	
GO:0046184	aldehyde biosynthetic process	0.003387256	
GO:0080154	regulation of fertilization	0.003500704	
GO:0008211	glucocorticoid metabolic process	0.003879947	
GO:0008207	C21-steroid hormone metabolic process	0.00395913	
GO:0019228	neuronal action potential	0.004786948	
GO:0120178	steroid hormone biosynthetic process	0.004849934	
GO:0007340	acrosome reaction	0.004849934	
GO:0086010	membrane depolarization during action potential	0.005321071	
GO:0017158	regulation of calcium ion-dependent exocytosis	0.005329598	
GO:1903307	positive regulation of regulated secretory pathway	0.005938695	
GO:0042181	ketone biosynthetic process	0.005938695	
GO:0042446	hormone biosynthetic process	0.006973177	
GO:0006081	cellular aldehyde metabolic process	0.006973177	
GO:0034308	primary alcohol metabolic process	0.006973177	
GO:0051899	membrane depolarization	0.007173012	
GO:0019226	transmission of nerve impulse	0.007173012	
GO:0017156	calcium-ion regulated exocytosis	0.007173012	
GO:0045921	positive regulation of exocytosis	0.007640051	
GO:0070509	calcium ion import	0.007640051	
GO:2000243	positive regulation of reproductive process	0.007640051	
GO:0034754	cellular hormone metabolic process	0.008460036	

GO:0007338	single fertilization	0.008953725	
GO:0001508	action potential	0.010669855	
GO:0006694	steroid biosynthetic process	0.010959592	
GO:0009566	fertilization	0.010959592	
GO:0071248	cellular response to metal ion	0.010959592	
GO:0035637	multicellular organismal signaling	0.010959592	
GO:0046165	alcohol biosynthetic process	0.0113118	
GO:0035725	sodium ion transmembrane transport	0.0113118	
GO:0071241	cellular response to inorganic substance	0.0113118	
GO:2000241	regulation of reproductive process	0.0113118	
GO:1903305	regulation of regulated secretory pathway	0.0113118	
GO:0042180	cellular ketone metabolic process	0.012748399	
GO:0006814	sodium ion transport	0.012933158	
GO:0042445	hormone metabolic process	0.013217403	
GO:0017157	regulation of exocytosis	0.014013826	
GO:0010038	response to metal ion	0.014209678	
GO:0045055	regulated exocytosis	0.014209678	
GO:0008202	steroid metabolic process	0.014252868	
GO:1901617	organic hydroxy compound biosynthetic process	0.014780752	
GO:1903532	positive regulation of secretion by cell	0.017583998	
GO:0051047	positive regulation of secretion	0.017921662	
GO:0006066	alcohol metabolic process	0.017921662	
GO:0070588	calcium ion transmembrane transport	0.018718248	
GO:0006887	exocytosis	0.019757336	
GO:0010035	response to inorganic substance	0.020147573	
GO:0043902	positive regulation of multi-organism process	0.020179191	
GO:0006816	calcium ion transport	0.022341675	
GO:0042391	regulation of membrane potential	0.022423073	
GO:0017144	drug metabolic process	0.023376081	

GO:1901615	organic hydroxy compound metabolic process	0.023376081	
GO:0010817	regulation of hormone levels	0.023376081	
GO:0070838	divalent metal ion transport	0.023498273	
GO:0072511	divalent inorganic cation transport	0.023498273	
GO:0060627	regulation of vesicle-mediated transport	0.024018722	
GO:0015672	monovalent inorganic cation transport	0.025766446	
GO:0043900	regulation of multi-organism process	0.025766446	
GO:0008610	lipid biosynthetic process	0.026235835	
GO:0019953	sexual reproduction	0.028198904	
GO:1903530	regulation of secretion by cell	0.028198904	
GO:0044283	small molecule biosynthetic process	0.028198904	
GO:0044703	multi-organism reproductive process	0.028198904	
GO:0032870	cellular response to hormone stimulus	0.028198904	
GO:0051046	regulation of secretion	0.029082498	
GO:0009725	response to hormone	0.030688086	
GO:0051050	positive regulation of transport	0.03292279	
GO:0098662	inorganic cation transmembrane transport	0.03490633	
GO:0030001	metal ion transport	0.03586506	
GO:0032940	secretion by cell	0.036028083	
GO:0098660	inorganic ion transmembrane transport	0.037413779	
GO:0140352	export from cell	0.037413779	
GO:0098655	cation transmembrane transport	0.03817669	
GO:0046903	secretion	0.038429956	
GO:0000003	reproduction	0.038739227	
GO:0022414	reproductive process	0.038739227	
GO:0050877	nervous system process	0.041541066	
GO:0006812	cation transport	0.045639503	
GO:0034220	ion transmembrane transport	0.0473762	
GO:0006629	lipid metabolic process	0.047691021	

GO:0071495	cellular response to endogenous stimulus	0.048913623	
GO:0009719	response to endogenous stimulus	0.05209084	
GO:0016192	vesicle-mediated transport	0.055707759	
GO:0051049	regulation of transport	0.05825464	
GO:0003008	system process	0.059763264	
GO:0055085	transmembrane transport	0.060916789	
GO:0044281	small molecule metabolic process	0.060916789	
GO:0006811	ion transport	0.06107618	
GO:0051704	multi-organism process	0.06107618	
GO:0071310	cellular response to organic substance	0.077337497	
GO:0032879	regulation of localization	0.087234065	
GO:0010033	response to organic substance	0.08811997	
GO:0070887	cellular response to chemical stimulus	0.094481551	
GO:0065008	regulation of biological quality	0.125436654	
GO:0042221	response to chemical	0.125436654	
GO:1901362	organic cyclic compound biosynthetic process	0.132135965	
GO:0006810	transport	0.152990595	
GO:0051234	establishment of localization	0.156170316	
GO:0048522	positive regulation of cellular process	0.165512292	
GO:0044249	cellular biosynthetic process	0.17928033	
GO:1901576	organic substance biosynthetic process	0.180306227	
GO:0048518	positive regulation of biological process	0.181048688	
GO:0009058	biosynthetic process	0.181048688	
GO:1901360	organic cyclic compound metabolic process	0.182034201	
GO:0023052	signaling	0.184286052	
GO:0007154	cell communication	0.184456518	
GO:0032501	multicellular organismal process	0.188856889	
GO:0051179	localization	0.188856889	
GO:0051716	cellular response to stimulus	0.208319921	

GO:0050896	response to stimulus	0.240319194	
GO:0050794	regulation of cellular process	0.307995378	
GO:0044238	primary metabolic process	0.314422973	
GO:0044237	cellular metabolic process	0.317227597	
GO:0050789	regulation of biological process	0.319946436	
GO:0071704	organic substance metabolic process	0.320085086	
GO:0008152	metabolic process	0.335567528	
GO:0065007	biological regulation	0.335567528	
GO:0009987	cellular process	0.464725049	
GO:0008150	biological process	0.527139103	
GO:0031048	chromatin silencing by small RNA	0.006404993	
GO:0030702	chromatin silencing at centromere	0.006404993	
GO:0006342	chromatin silencing	0.035227459	
GO:0034401	chromatin organization involved in regulation of transcription	0.035227459	
GO:0031047	gene silencing by RNA	0.035227459	
GO:0097549	chromatin organization involved in negative regulation of transcription	0.035227459	
GO:0045814	negative regulation of gene expression, epigenetic	0.035227459	
GO:0016458	gene silencing	0.055243061	
GO:0040029	regulation of gene expression, epigenetic	0.055509936	
GO:0006325	chromatin organization	0.155321071	*ZNFX1*
GO:0051276	chromosome organization	0.158683692	
GO:0045934	negative regulation of nucleobase-containing compound metabolic process	0.158683692	
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	0.158683692	
GO:0045892	negative regulation of transcription, DNA-templated	0.158683692	
GO:0051253	negative regulation of RNA metabolic process	0.158683692	
GO:0031327	negative regulation of cellular biosynthetic process	0.158683692	
GO:0010558	negative regulation of macromolecule biosynthetic process	0.158683692	
GO:0009890	negative regulation of biosynthetic process	0.158683692	
GO:1902679	negative regulation of RNA biosynthetic process	0.158683692	

GO:1903507	negative regulation of nucleic acid-templated transcription	0.158683692	
GO:0010629	negative regulation of gene expression	0.171104803	
GO:0048519	negative regulation of biological process	0.204546538	
GO:0044271	cellular nitrogen compound biosynthetic process	0.204546538	
GO:0046483	heterocycle metabolic process	0.204546538	
GO:0044249	cellular biosynthetic process	0.204546538	
GO:0006139	nucleobase-containing compound metabolic process	0.204546538	
GO:0051171	regulation of nitrogen compound metabolic process	0.204546538	
GO:0051172	negative regulation of nitrogen compound metabolic process	0.204546538	
GO:0051252	regulation of RNA metabolic process	0.204546538	
GO:0060255	regulation of macromolecule metabolic process	0.204546538	
GO:0080090	regulation of primary metabolic process	0.204546538	
GO:0090304	nucleic acid metabolic process	0.204546538	
GO:0097659	nucleic acid-templated transcription	0.204546538	
GO:1901360	organic cyclic compound metabolic process	0.204546538	
GO:1901362	organic cyclic compound biosynthetic process	0.204546538	
GO:1901576	organic substance biosynthetic process	0.204546538	
GO:1903506	regulation of nucleic acid-templated transcription	0.204546538	
GO:2000112	regulation of cellular macromolecule biosynthetic process	0.204546538	
GO:0048523	negative regulation of cellular process	0.204546538	
GO:0034654	nucleobase-containing compound biosynthetic process	0.204546538	
GO:2001141	regulation of RNA biosynthetic process	0.204546538	
GO:0032774	RNA biosynthetic process	0.204546538	
GO:0006351	transcription, DNA-templated	0.204546538	
GO:0006355	regulation of transcription, DNA-templated	0.204546538	
GO:0006725	cellular aromatic compound metabolic process	0.204546538	
GO:0006996	organelle organization	0.204546538	
GO:0009058	biosynthetic process	0.204546538	
GO:0009059	macromolecule biosynthetic process	0.204546538	

GO:0009889	regulation of biosynthetic process	0.204546538	
GO:0009892	negative regulation of metabolic process	0.204546538	
GO:0010467	gene expression	0.204546538	
GO:0034645	cellular macromolecule biosynthetic process	0.204546538	
GO:0010556	regulation of macromolecule biosynthetic process	0.204546538	
GO:0010605	negative regulation of macromolecule metabolic process	0.204546538	
GO:0010468	regulation of gene expression	0.204546538	
GO:0031323	regulation of cellular metabolic process	0.204546538	
GO:0031326	regulation of cellular biosynthetic process	0.204546538	
GO:0016070	RNA metabolic process	0.204546538	
GO:0018130	heterocycle biosynthetic process	0.204546538	
GO:0019219	regulation of nucleobase-containing compound metabolic process	0.204546538	
GO:0031324	negative regulation of cellular metabolic process	0.204546538	
GO:0019438	aromatic compound biosynthetic process	0.204546538	
GO:0034641	cellular nitrogen compound metabolic process	0.208852028	
GO:0019222	regulation of metabolic process	0.208852028	
GO:0016043	cellular component organization	0.208852028	
GO:0071840	cellular component organization or biogenesis	0.212480778	
GO:0044260	cellular macromolecule metabolic process	0.268770697	
GO:0043170	macromolecule metabolic process	0.30517906	
GO:0006807	nitrogen compound metabolic process	0.319274958	
GO:0050794	regulation of cellular process	0.322857377	
GO:0071704	organic substance metabolic process	0.327390332	
GO:0050789	regulation of biological process	0.327390332	
GO:0044238	primary metabolic process	0.327390332	
GO:0044237	cellular metabolic process	0.327390332	
GO:0008152	metabolic process	0.339296056	
GO:0065007	biological regulation	0.339296056	
GO:0009987	cellular process	0.467273324	

GO:0008150	biological process	0.527139103	
GO:1901998	toxin transport	0.015716866	*VPS11*
GO:0140056	organelle localization by membrane tethering	0.015716866	
GO:0031338	regulation of vesicle fusion	0.015716866	
GO:0006904	vesicle docking involved in exocytosis	0.015716866	
GO:0140029	exocytic process	0.015716866	
GO:0007032	endosome organization	0.015716866	
GO:0022406	membrane docking	0.015716866	
GO:0035493	SNARE complex assembly	0.015716866	
GO:0048278	vesicle docking	0.015716866	
GO:0035542	regulation of SNARE complex assembly	0.015716866	
GO:0006906	vesicle fusion	0.016710461	
GO:0090174	organelle membrane fusion	0.016710461	
GO:0061025	membrane fusion	0.018968632	
GO:0048284	organelle fusion	0.018968632	
GO:0007033	vacuole organization	0.022581705	
GO:0016050	vesicle organization	0.033703194	
GO:0006887	exocytosis	0.035227459	
GO:0061919	process utilizing autophagic mechanism	0.03549844	
GO:0043254	regulation of protein-containing complex assembly	0.03549844	
GO:0006914	autophagy	0.03549844	
GO:0010256	endomembrane system organization	0.036130728	
GO:0060627	regulation of vesicle-mediated transport	0.038429956	
GO:0051640	organelle localization	0.041000521	
GO:0061024	membrane organization	0.04821194	
GO:0032940	secretion by cell	0.06017635	
GO:0046903	secretion	0.06017635	
GO:0034622	cellular protein-containing complex assembly	0.06017635	
GO:0140352	export from cell	0.06017635	

GO:0044087	regulation of cellular component biogenesis	0.06017635	
GO:0006886	intracellular protein transport	0.061783544	
GO:0033043	regulation of organelle organization	0.073864028	
GO:0065003	protein-containing complex assembly	0.078313352	
GO:0051049	regulation of transport	0.078313352	
GO:0046907	intracellular transport	0.078313352	
GO:0045184	establishment of protein localization	0.078313352	
GO:0016192	vesicle-mediated transport	0.078313352	
GO:0043933	protein-containing complex subunit organization	0.078313352	
GO:0042886	amide transport	0.078313352	
GO:0015833	peptide transport	0.078313352	
GO:0015031	protein transport	0.078313352	
GO:0034613	cellular protein localization	0.081294137	
GO:0070727	cellular macromolecule localization	0.081294137	
GO:0044248	cellular catabolic process	0.082587453	
GO:0071705	nitrogen compound transport	0.082587453	
GO:0009056	catabolic process	0.091662085	
GO:0051128	regulation of cellular component organization	0.091662085	
GO:0071702	organic substance transport	0.093405304	
GO:0032879	regulation of localization	0.093405304	
GO:0008104	protein localization	0.093405304	
GO:0051641	cellular localization	0.095981278	
GO:0022607	cellular component assembly	0.098562499	
GO:0033036	macromolecule localization	0.100731774	
GO:0044085	cellular component biogenesis	0.103534986	
GO:0006996	organelle organization	0.137949125	
GO:0006810	transport	0.153670553	
GO:0051234	establishment of localization	0.155426648	
GO:0051179	localization	0.198049337	

GO:0016043	cellular component organization	0.198049337	
GO:0071840	cellular component organization or biogenesis	0.201122613	
GO:0050794	regulation of cellular process	0.31871818	
GO:0050789	regulation of biological process	0.327973765	
GO:0044237	cellular metabolic process	0.327973765	
GO:0065007	biological regulation	0.340927287	
GO:0008152	metabolic process	0.340927287	
GO:0009987	cellular process	0.468379297	
GO:0008150	biological process	0.527139103	
GO:1902463	protein localization to cell leading edge	0.003592544	*COPA*
GO:0030157	pancreatic juice secretion	0.003592544	
GO:0099612	protein localization to axon	0.003832047	
GO:0006891	intra-Golgi vesicle-mediated transport	0.00574807	
GO:0007589	body fluid secretion	0.00574807	
GO:0032941	secretion by tissue	0.00574807	
GO:0006890	retrograde vesicle-mediated transport, Golgi to endoplasmic reticulum	0.0069798	
GO:0022600	digestive system process	0.008083224	
GO:0007586	digestion	0.008462437	
GO:0006888	endoplasmic reticulum to Golgi vesicle-mediated transport	0.012502053	
GO:0050878	regulation of body fluid levels	0.020771436	
GO:0048193	Golgi vesicle transport	0.02454905	
GO:0046903	secretion	0.068976843	
GO:0006886	intracellular protein transport	0.070208573	
GO:0046907	intracellular transport	0.078723571	
GO:0045184	establishment of protein localization	0.078723571	
GO:0070727	cellular macromolecule localization	0.078723571	
GO:0042886	amide transport	0.078723571	
GO:0003008	system process	0.078723571	
GO:0016192	vesicle-mediated transport	0.078723571	

GO:0015833	peptide transport	0.078723571	
GO:0015031	protein transport	0.078723571	
GO:0034613	cellular protein localization	0.078723571	
GO:0071705	nitrogen compound transport	0.080293357	
GO:0008104	protein localization	0.093350872	
GO:0071702	organic substance transport	0.093350872	
GO:0051641	cellular localization	0.094257708	
GO:0033036	macromolecule localization	0.099205535	
GO:0065008	regulation of biological quality	0.123781721	
GO:0051234	establishment of localization	0.148893563	
GO:0006810	transport	0.148893563	
GO:0032501	multicellular organismal process	0.183981795	
GO:0051179	localization	0.183981795	
GO:0065007	biological regulation	0.340319573	
GO:0008150	biological process	0.527139103	
GO:0061577	calcium ion transmembrane transport via high voltage-gated calcium channel	0.006774511	*CACNB1*
GO:1902514	regulation of calcium ion transmembrane transport via high voltage-gated calcium channel	0.006774511	
GO:1901385	regulation of voltage-gated calcium channel activity	0.01192314	
GO:1904646	cellular response to amyloid-beta	0.01192314	
GO:1904645	response to amyloid-beta	0.01192314	
GO:1901019	regulation of calcium ion transmembrane transporter activity	0.015484598	
GO:0007528	neuromuscular junction development	0.015484598	
GO:1903169	regulation of calcium ion transmembrane transport	0.025291509	
GO:2001257	regulation of cation channel activity	0.025291509	
GO:0051924	regulation of calcium ion transport	0.029182511	
GO:0022898	regulation of transmembrane transporter activity	0.029182511	
GO:0032409	regulation of transporter activity	0.029182511	
GO:0032412	regulation of ion transmembrane transporter activity	0.029182511	
GO:1901653	cellular response to peptide	0.032698308	

GO:1904062	regulation of cation transmembrane transport	0.032698308	
GO:0070588	calcium ion transmembrane transport	0.032836455	
GO:0010959	regulation of metal ion transport	0.032836455	
GO:1901652	response to peptide	0.033119834	
GO:0006816	calcium ion transport	0.034969383	
GO:0050808	synapse organization	0.034969383	
GO:0034765	regulation of ion transmembrane transport	0.034969383	
GO:0072511	divalent inorganic cation transport	0.035463095	
GO:0070838	divalent metal ion transport	0.035463095	
GO:0034762	regulation of transmembrane transport	0.03590491	
GO:0071417	cellular response to organonitrogen compound	0.036961734	
GO:1901699	cellular response to nitrogen compound	0.037833041	
GO:0043269	regulation of ion transport	0.037836901	
GO:0099537	trans-synaptic signaling	0.039073634	
GO:0007268	chemical synaptic transmission	0.039073634	
GO:0099536	synaptic signaling	0.039073634	
GO:0098916	anterograde trans-synaptic signaling	0.039073634	
GO:1901698	response to nitrogen compound	0.042582643	
GO:0098662	inorganic cation transmembrane transport	0.042582643	
GO:0030001	metal ion transport	0.042582643	
GO:0010243	response to organonitrogen compound	0.042582643	
GO:1901701	cellular response to oxygen-containing compound	0.043713426	
GO:0098655	cation transmembrane transport	0.043713426	
GO:0098660	inorganic ion transmembrane transport	0.043713426	
GO:1901700	response to oxygen-containing compound	0.05351864	
GO:0006812	cation transport	0.05351864	
GO:0034220	ion transmembrane transport	0.053667349	
GO:0071495	cellular response to endogenous stimulus	0.055228398	
GO:0007267	cell-cell signaling	0.056721136	

GO:0009719	response to endogenous stimulus	0.056721136	
GO:0051049	regulation of transport	0.063288991	
GO:0055085	transmembrane transport	0.066073107	
GO:0006811	ion transport	0.066073107	
GO:0071310	cellular response to organic substance	0.083495853	
GO:0032879	regulation of localization	0.093054689	
GO:0010033	response to organic substance	0.093054689	
GO:0070887	cellular response to chemical stimulus	0.098021969	
GO:0065009	regulation of molecular function	0.098021969	
GO:0042221	response to chemical	0.128434511	
GO:0006810	transport	0.156516304	
GO:0051234	establishment of localization	0.158252587	
GO:0023052	signaling	0.19301415	
GO:0007154	cell communication	0.19301415	
GO:0051179	localization	0.194692569	
GO:0016043	cellular component organization	0.194692569	
GO:0071840	cellular component organization or biogenesis	0.19777057	
GO:0051716	cellular response to stimulus	0.208699375	
GO:0050896	response to stimulus	0.238768747	
GO:0050789	regulation of biological process	0.322767833	
GO:0065007	biological regulation	0.340927287	
GO:0009987	cellular process	0.468379297	
GO:0008150	biological process	0.527139103	
GO:0006611	protein export from nucleus	0.06589752	*RANBP3*
GO:0006913	nucleocytoplasmic transport	0.06589752	
GO:0051169	nuclear transport	0.06589752	
GO:0051168	nuclear export	0.06589752	
GO:0043547	positive regulation of GTPase activity	0.06589752	
GO:0043087	regulation of GTPase activity	0.06589752	

GO:0051345	positive regulation of hydrolase activity	0.080766253	
GO:0070727	cellular macromolecule localization	0.086221054	
GO:0051336	regulation of hydrolase activity	0.086221054	
GO:0046907	intracellular transport	0.086221054	
GO:0045184	establishment of protein localization	0.086221054	
GO:0044093	positive regulation of molecular function	0.086221054	
GO:0042886	amide transport	0.086221054	
GO:0034613	cellular protein localization	0.086221054	
GO:0006886	intracellular protein transport	0.086221054	
GO:0043085	positive regulation of catalytic activity	0.086221054	
GO:0015833	peptide transport	0.086221054	
GO:0015031	protein transport	0.086221054	
GO:0071705	nitrogen compound transport	0.086934161	
GO:0050790	regulation of catalytic activity	0.091209558	
GO:0008104	protein localization	0.094563221	
GO:0071702	organic substance transport	0.094563221	
GO:0051641	cellular localization	0.09484316	
GO:0033036	macromolecule localization	0.095237313	
GO:0065009	regulation of molecular function	0.095237313	
GO:0051234	establishment of localization	0.146530173	
GO:0006810	transport	0.146530173	
GO:0051179	localization	0.185859161	
GO:0065007	biological regulation	0.341996024	
GO:0008150	biological process	0.527139103	

Table S4. Gene Ontology (GO) slim categories obtained by CateGOrizer per candidate gene.

GO slim category ID	Definition	Counted terms	Fraction	Gene
GO:0008150	biological_process	44	42.31%	*SLAIN2*
GO:0016043	cell organization and biogenesis	31	29.81%	
GO:0006996	organelle organization and biogenesis	16	15.38%	
GO:0007010	cytoskeleton organization and biogenesis	13	12.50%	
GO:0008150	biological_process	345	41.37%	*ACVR1*
GO:0007275	development	101	12.11%	
GO:0008152	metabolism	96	11.51%	
GO:0009653	morphogenesis	43	5.16%	
GO:0030154	cell differentiation	40	4.80%	
GO:0009058	biosynthesis	32	3.84%	
GO:0007154	cell communication	30	3.60%	
GO:0007165	signal transduction	26	3.12%	
GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolism	23	2.76%	
GO:0019538	protein metabolism	23	2.76%	
GO:0006464	protein modification	16	1.92%	
GO:0008219	cell death	15	1.80%	
GO:0009790	embryonic development	13	1.56%	
GO:0000003	reproduction	10	1.20%	
GO:0009719	response to endogenous stimulus	9	1.08%	
GO:0007049	cell cycle	8	0.96%	
GO:0006950	response to stress	4	0.48%	
GO:0008150	biological_process	29	40.28%	*ST3GAL3*
GO:0008152	metabolism	27	37.50%	
GO:0009058	biosynthesis	9	12.50%	
GO:0019538	protein metabolism	5	6.94%	
GO:0006464	protein modification	2	2.78%	

GO:0008150	biological_process	133	47.84%	
GO:0008152	metabolism	41	14.75%	
GO:0006810	transport	36	12.95%	
GO:0009058	biosynthesis	19	6.83%	
GO:0006811	ion transport	15	5.40%	*CACNA1H*
GO:0006629	lipid metabolism	14	5.04%	
GO:0000003	reproduction	13	4.68%	
GO:0009719	response to endogenous stimulus	4	1.44%	
GO:0007154	cell communication	3	1.08%	
GO:0008150	biological_process	78	37.50%	
GO:0008152	metabolism	64	30.77%	
GO:0009058	biosynthesis	30	14.42%	
GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolism	20	9.62%	*ZNFX1*
GO:0016043	cell organization and biogenesis	9	4.33%	
GO:0040029	regulation of gene expression, epigenetic	5	2.40%	
GO:0006996	organelle organization and biogenesis	2	0.96%	
GO:0008150	biological_process	62	50.41%	
GO:0006810	transport	21	17.07%	
GO:0016043	cell organization and biogenesis	21	17.07%	
GO:0006996	organelle organization and biogenesis	11	8.94%	*VPS11*
GO:0008152	metabolism	4	3.25%	
GO:0009056	catabolism	2	1.63%	
GO:0015031	protein transport	2	1.63%	
GO:0008150	biological_process	35	64.81%	
GO:0006810	transport	17	31.48%	*COPA*
GO:0015031	protein transport	2	3.70%	
GO:0008150	biological_process	66	47.14%	
GO:0006810	transport	29	20.71%	*CACNB1*
GO:0006811	ion transport	23	16.43%	

GO:0009719	response to endogenous stimulus	8	5.71%	
GO:0007154	cell communication	6	4.29%	
GO:0007267	cell-cell signaling	5	3.57%	
GO:0016043	cell organization and biogenesis	3	2.14%	
GO:0008150	biological_process	30	66.67%	*RANBP3*
GO:0006810	transport	12	26.67%	
GO:0015031	protein transport	3	6.67%	

Table S5. Predicted target genes for *gga-mir-6646-2.*

Gene Symbol	***Gene Description***	**Target Rank**	**Target Score**
ZNF555	*zinc finger protein 555*	1	98
TSTD2	*thiosulfate sulfurtransferase like domain containing 2*	2	96
XKR6	*XK related 6*	3	95
ITM2B	*integral membrane protein 2B*	4	
SETBP1	*SET binding protein 1*	5	
MYH10	*myosin, heavy chain 10, non-muscle*	6	93
LOC107053409	*mucin-5AC-like*	7	
SMG1	*SMG1, nonsense mediated mRNA decay associated PI3K related kinase*	8	92
LCAT	*lecithin-cholesterol acyltransferase*	9	
ENDOUL	*endonuclease, polyU-specific-like*	10	91
TMLHE	*trimethyllysine hydroxylase, epsilon*	11	
SEPT7	*septin 7*	12	
CALN1	*calneuron 1*	13	
CCBE1	*collagen and calcium binding EGF domains 1*	14	90
EPHA7	*EPH receptor A7*	15	
SGIP1	*SH3 domain GRB2 like endophilin interacting protein 1*	16	
TPD52	*tumor protein D52*	17	
ACTN1	*actinin, alpha 1*	18	89
GRB2	*growth factor receptor bound protein 2*	19	
C13H5orf15	*chromosome 13 C5orf15 homolog*	20	
KPNA3	*karyopherin subunit alpha 3*	21	88
KPNA4	*karyopherin subunit alpha 4*	22	87

KHDRBS3	*KH RNA binding domain containing, signal transduction associated 3*	23	
GDAP1	*ganglioside induced differentiation associated protein 1*	24	86
AZI2	*5-azacytidine induced 2*	25	
CDAN1	*codanin 1*	26	
NCOA3	*nuclear receptor coactivator 3*	27	
GRSF1	*G-rich RNA sequence binding factor 1*	28	
LZTS3	*leucine zipper tumor suppressor family member 3*	29	
CREB5	*cAMP responsive element binding protein 5*	30	85
LRRK2	*leucine rich repeat kinase 2*	31	
C1HXORF59	*chromosome 1 open reading frame, human CXORF59*	32	
B4GALT5	*beta-1,4-galactosyltransferase 5*	33	84
TBL3	*transducin beta like 3*	34	
CABLES1	*Cdk5 and Abl enzyme substrate 1*	35	
EPT1L	*ethanolaminephosphotransferase 1-like*	36	
LOC771066	*claw keratin-like*	37	83
PSMA1	*proteasome subunit alpha 1*	38	
FBXO41	*F-box protein 41*	39	
TCF12	*transcription factor 12*	40	
RAB5B	*RAB5B, member RAS oncogene family*	41	
REEP2	*receptor accessory protein 2*	42	82
BAZ2B	*bromodomain adjacent to zinc finger domain 2B*	43	
LMX1B	*LIM homeobox transcription factor 1 beta*	44	81
PYCRL	*pyrroline-5-carboxylate reductase-like*	45	
TMEM98	*transmembrane protein 98*	46	
MVB12B	*multivesicular body subunit 12B*	47	
SGCZ	*sarcoglycan zeta*	48	
SLC16A8	*solute carrier family 16 member 8*	49	
PDIK1L	*PDLIM1 interacting kinase 1 like*	50	
L3MBTL3	*L3MBTL3, histone methyl-lysine binding protein*	51	80
IGSF3	*immunoglobulin superfamily member 3*	52	
CCDC71	*coiled-coil domain containing 71*	53	
SYNPO2L	*synaptopodin 2 like*	54	
TRPS1	*transcriptional repressor GATA binding 1*	55	79
HNRNPK	*heterogeneous nuclear ribonucleoprotein K*	56	

TRABD2B	*TraB domain containing 2B*	57	
CENPM	*centromere protein M*	58	
DGKG	*diacylglycerol kinase gamma*	59	
C2H8ORF37	*chromosome 2 open reading frame, human C8orf37*	60	
BBX	*BBX, HMG-box containing*	61	
CHMP6	*charged multivesicular body protein 6*	62	
KNSTRN	*kinetochore-localized astrin/SPAG5 binding protein*	63	78
MIF4GD	*MIF4G domain containing*	64	
GRHL2	*grainyhead like transcription factor 2*	65	
TUB	*tubby bipartite transcription factor*	66	
SYNE3	*spectrin repeat containing nuclear envelope family member 3*	67	
OSBPL9	*oxysterol binding protein like 9*	68	
LOC426218	*claw keratin-like*	69	
CDC14B	*cell division cycle 14B*	70	77
HAO2	*hydroxyacid oxidase 2*	71	
CAP2	*cyclase associated actin cytoskeleton regulatory protein 2*	72	
LOC101748210	*uncharacterized LOC101748210*	73	
TMEM135	*transmembrane protein 135*	74	
OLFML1	*olfactomedin like 1*	75	
GLRA4	*glycine receptor alpha 4*	76	
LOC101750514	*LIM homeobox transcription factor 1-alpha-like*	77	
NFATC3	*nuclear factor of activated T-cells 3*	78	
PAK3	*p21 (RAC1) activated kinase 3*	79	76
PSEN1	*presenilin 1*	80	
RTCA	*RNA 3'-terminal phosphate cyclase*	81	
ANKRD16	*ankyrin repeat domain 16*	82	
RNF103	*ring finger protein 103*	83	
PAPPA	*pappalysin 1*	84	
RFLNA	*refilin A*	85	
PTPN2	*protein tyrosine phosphatase, non-receptor type 2*	86	
RAB3C	*RAB3C, member RAS oncogene family*	87	75
HS3ST4	*heparan sulfate-glucosamine 3-sulfotransferase 4*	88	
EPHB6	*EPH receptor B6*	89	
FLRT3	*fibronectin leucine rich transmembrane protein 3*	90	

AGK	*acylglycerol kinase*	91	
ARHGEF12	*Rho guanine nucleotide exchange factor 12*	92	
ADAMTS7	*ADAM metallopeptidase with thrombospondin type 1 motif 7*	93	74
OMP	*olfactory marker protein*	94	
SNCB	*synuclein beta*	95	
SLC7A1	*solute carrier family 7 member 1*	96	
RBBP4	*RB binding protein 4, chromatin remodeling factor*	97	
BNIP3L	*BCL2 interacting protein 3 like*	98	
HOXB1	*homeobox B1*	99	73
CLC2DL4	*C-type lectin domain family 2 member D-like 4*	100	
GPATCH2L	*G-patch domain containing 2 like*	101	
MCHR1	*melanin-concentrating hormone receptor 1*	102	
SLC38A1	*solute carrier family 38 member 1*	103	
TMEM136	*transmembrane protein 136*	104	72
STK38	*serine/threonine kinase 38*	105	
IRS4	*insulin receptor substrate 4*	106	
SMAD7B	*TGF-beta signal pathway antagonist Smad7*	107	
EHF	*ETS homologous factor*	108	
BTAF1	*B-TFIID TATA-box binding protein associated factor 1*	109	71
ELMSAN1	*ELM2 and Myb/SANT domain containing 1*	110	
FAM189A1	*family with sequence similarity 189 member A1*	111	
ECT2	*epithelial cell transforming 2*	112	
NAXD	*NAD(P)HX dehydratase*	113	
LARP1	*La ribonucleoprotein domain family member 1*	114	
VPS45	*vacuolar protein sorting 45 homolog*	115	70
FOXO4	*forkhead box O4*	116	
MAGI1	*membrane associated guanylate kinase, WW and PDZ domain containing 1*	117	
CD44	*CD44 molecule (Indian blood group)*	118	
CHTF8	*chromosome transmission fidelity factor 8*	119	
FSIP1	*fibrous sheath interacting protein 1*	120	69
GNA12	*G protein subunit alpha 12*	121	
ZNF516	*zinc finger protein 516*	122	
CUEDC1	*CUE domain containing 1*	123	68
CAPN7	*calpain 7*	124	

DUSP23	*dual specificity phosphatase 23*	125	
LRRC66	*leucine rich repeat containing 66*	126	
DVL2	*dishevelled segment polarity protein 2*	127	
DUSP3	*dual specificity phosphatase 3*	128	
PDE7B	*phosphodiesterase 7B*	129	67
SSTR2	*somatostatin receptor 2*	130	
FBXO3	*F-box protein 3*	131	
SKI	*SKI proto-oncogene*	132	
CD83	*CD83 molecule*	133	
PLEKHG3	*pleckstrin homology and RhoGEF domain containing G3*	134	
KMT2C	*lysine methyltransferase 2C*	135	
IGSF9B	*immunoglobulin superfamily member 9B*	136	
IQSEC3	*IQ motif and Sec7 domain 3*	137	66
MUC4	*mucin 4, cell surface associated*	138	
OARD1	*O-acyl-ADP-ribose deacylase 1*	139	
TNFRSF8	*TNF receptor superfamily member 8*	140	
ZDHHC2	*zinc finger DHHC-type containing 2*	141	
MYEF2	*myelin expression factor 2*	142	
VANGL1	*VANGL planar cell polarity protein 1*	143	
RANBP10	*RAN binding protein 10*	144	
KLHDC10	*kelch domain containing 10*	145	
VPS36	*vacuolar protein sorting 36 homolog*	146	65
TLE4Z1	*transducin like enhancer of split 4-Z 1*	147	
BRAF	*B-Raf proto-oncogene, serine/threonine kinase*	148	
LOC428541	*tetraspanin-18-like*	149	
CLIP2	*CAP-GLY domain containing linker protein 2*	150	
C3H1ORF131	*chromosome 3 open reading frame, human C1orf131*	151	
FNDC3A	*fibronectin type III domain containing 3A*	152	
GBE	*eye-globin*	153	
CPE	*carboxypeptidase E*	154	
PALM	*paralemmin*	155	
LOC422090	*uncharacterized LOC422090*	156	64
NVL	*nuclear VCP-like*	157	
IVD	*isovaleryl-CoA dehydrogenase*	158	

ABI2	*abl interactor 2*	159	
PDE8A	*phosphodiesterase 8A*	160	
UTP15	*UTP15, small subunit processome component*	161	
TMEM74B	*transmembrane protein 74B*	162	
SNX29	*sorting nexin 29*	163	
ATG16L1	*autophagy related 16 like 1*	164	63
TMEM266	*transmembrane protein 266*	165	
EXOC2	*exocyst complex component 2*	166	
MGAT4D	*MGAT4 family member D*	167	
HOMER3	*homer scaffold protein 3*	168	
LARGE1	*LARGE xylosyl- and glucuronyltransferase 1*	169	
GREM2	*gremlin 2, DAN family BMP antagonist*	170	62
KRT75L2	*keratin, type II cytoskeletal 75-like 2*	171	
SPR	*sepiapterin reductase (7,8-dihydrobiopterin:NADP+ oxidoreductase)*	172	
B4GALNT3	*beta-1,4-N-acetyl-galactosaminyltransferase 3*	173	
CLIC5	*chloride intracellular channel 5*	174	
MEF2A	*myocyte enhancer factor 2A*	175	
CACNA2D4	*calcium voltage-gated channel auxiliary subunit alpha2delta 4*	176	
ARSA	*arylsulfatase A*	177	61
FADS1	*fatty acid desaturase 1*	178	
RBPMS2	*RNA binding protein, mRNA processing factor 2*	179	
KCNK16	*potassium two pore domain channel subfamily K member 16*	180	
SPOCK1	*SPARC/osteonectin, cwcv and kazal like domains proteoglycan 1*	181	
RIMS4	*regulating synaptic membrane exocytosis 4*	182	
GHRH	*growth hormone releasing hormone*	183	
CRISP2	*cysteine rich secretory protein 2*	184	60
KIAA1958	*KIAA1958*	185	
FGF18	*fibroblast growth factor 18*	186	
RNF38	*ring finger protein 38*	187	
MICAL1	*microtubule associated monooxygenase, calponin and LIM domain containing 1*	188	
SIRT1	*sirtuin 1*	189	
KCNJ12	*potassium voltage-gated channel subfamily J member 12*	190	
ASB1	*ankyrin repeat and SOCS box containing 1*	191	
WASF2	*WAS protein family member 2*	192	

BTBD3	*BTB domain containing 3*	193	
PAG1	*phosphoprotein membrane anchor with glycosphingolipid microdomains 1*	194	
SPRY3	*sprouty RTK signaling antagonist 3*	195	
GRIK1	*glutamate ionotropic receptor kainate type subunit 1*	196	59
STAU1	*staufen double-stranded RNA binding protein 1*	197	
LOC107054473	*uncharacterized LOC107054473*	198	
ADAMTS15	*ADAM metallopeptidase with thrombospondin type 1 motif, 15*	199	
KIF3B	*kinesin family member 3B*	200	
LMOD3	*leiomodin 3*	201	
TECTB	*tectorin beta*	202	
MXD1	*MAX dimerization protein 1*	203	58
ZDHHC1	*zinc finger DHHC-type containing 1*	204	
SLC9A4	*solute carrier family 9 member A4*	205	
SLC31A1	*solute carrier family 31 member 1*	206	
GLP2R	*glucagon-like peptide 2 receptor*	207	
TTYH3	*tweety family member 3*	208	
TXNDC11	*thioredoxin domain containing 11*	209	
SRD5A2	*steroid 5 alpha-reductase 2*	210	
NIPAL2	*NIPA like domain containing 2*	211	57
KIAA0232	*KIAA0232*	212	
CDRT1	*CMT1A duplicated region transcript 1*	213	
RLTPR	*RGD motif, leucine rich repeats, tropomodulin domain and proline-rich containing*	214	
PRDM11	*PR/SET domain 11*	215	
NPFFR1	*neuropeptide FF receptor 1*	216	
PCNX4	*pecanex homolog 4 (Drosophila)*	217	56
LOC107049932	*deleted in malignant brain tumors 1 protein-like*	218	
NT5C2	*5'-nucleotidase, cytosolic II*	219	
ZNF608	*zinc finger protein 608*	220	
TGFBR1	*transforming growth factor beta receptor 1*	221	
RREB1	*ras responsive element binding protein 1*	222	
WDR37	*WD repeat domain 37*	223	
PLEKHG4	*pleckstrin homology and RhoGEF domain containing G4*	224	
PPME1	*protein phosphatase methylesterase 1*	225	
NR6A1	*nuclear receptor subfamily 6 group A member 1*	226	

ADAMTSL2L	*ADAMTS-like protein 2-like*	227	
AVL9	*AVL9 cell migration associated*	228	55
TMEM273	*transmembrane protein 273*	229	
UBAP1L	*ubiquitin associated protein 1 like*	230	
LRRCC1	*leucine rich repeat and coiled-coil centrosomal protein 1*	231	
AMN1	*antagonist of mitotic exit network 1 homolog*	232	
MAPK1	*mitogen-activated protein kinase 1*	233	
COMMD10	*COMM domain containing 10*	234	
RORA	*RAR related orphan receptor A*	235	
NBEA	*neurobeachin*	236	
JAG1	*jagged 1*	237	
SESN3	*sestrin 3*	238	
ATG14	*autophagy related 14*	239	
CP	*ceruloplasmin*	240	
IL11RA	*interleukin 11 receptor subunit alpha*	241	54
ADHFE1	*alcohol dehydrogenase, iron containing 1*	242	
TMEM8C	*transmembrane protein 8C*	243	
IGDCC3	*immunoglobulin superfamily DCC subclass member 3*	244	
NFAT5	*nuclear factor of activated T-cells 5*	245	
FADS1L2	*fatty acid desaturase 1-like 2*	246	
BDKRB1	*bradykinin receptor B1*	247	
USP45	*ubiquitin specific peptidase 45*	248	
CACNA1B	*calcium voltage-gated channel subunit alpha1 B*	249	
ZIC3	*Zic family member 3*	250	
TIMM17A	*translocase of inner mitochondrial membrane 17 homolog A (yeast)*	251	53
TTC1	*tetratricopeptide repeat domain 1*	252	
SFRP1	*secreted frizzled related protein 1*	253	
POMT2	*protein O-mannosyltransferase 2*	254	
TMCO3	*transmembrane and coiled-coil domains 3*	255	
TOP2A	*topoisomerase (DNA) II alpha*	256	
MED26	*mediator complex subunit 26*	257	
KCNK10	*potassium two pore domain channel subfamily K member 10*	258	
FLT4	*fms related tyrosine kinase 4*	259	
UHRF1BP1	*UHRF1 binding protein 1*	260	

TRIM27.1	*tripartite motif containing 27.1*	261	
SLC30A9	*solute carrier family 30 member 9*	262	
PIANP	*PILR alpha associated neural protein*	263	
CPPED1	*calcineurin like phosphoesterase domain containing 1*	264	
PTHLH	*parathyroid hormone like hormone*	265	
IL20RB	*interleukin 20 receptor subunit beta*	266	52
PRELID1	*PRELI domain containing 1*	267	
CDC42EP1	*CDC42 effector protein 1*	268	
SLC13A4	*solute carrier family 13 member 4*	269	
GAN	*gigaxonin*	270	
REPS1	*RALBP1 associated Eps domain containing 1*	271	
PRAM1	*PML-RARA regulated adaptor molecule 1*	272	
LOC101751874	*uncharacterized LOC101751874*	273	
SSPN	*sarcospan*	274	
ANGPTL7	*angiopoietin like 7*	275	
FARP1	*FERM, ARH/RhoGEF and pleckstrin domain protein 1*	276	
PDE6H	*phosphodiesterase 6H*	277	
TMEM88B	*transmembrane protein 88B*	278	51
NHLRC4	*NHL repeat containing 4*	279	
TMED3	*transmembrane p24 trafficking protein 3*	280	
DCX	*doublecortin*	281	
SYNM	*synemin*	282	
PSKH2	*protein serine kinase H2*	283	
ABRA	*actin binding Rho activating protein*	284	
TMEM45A	*transmembrane protein 45A*	285	
AKT3	*AKT serine/threonine kinase 3*	286	
CREB3L1	*cAMP responsive element binding protein 3 like 1*	287	
C5H15orf52	*chromosome 5 C15orf52 homolog*	288	
RAD9B	*RAD9 checkpoint clamp component B*	289	50
MRPL38	*mitochondrial ribosomal protein L38*	290	
PBRM1	*polybromo 1*	291	
CCNL2	*cyclin L2*	292	
AKAP11	*A-kinase anchoring protein 11*	293	
TTC26	*tetratricopeptide repeat domain 26*	294	