



HAL
open science

Re-sequencing data for refining candidate genes and polymorphisms in QTL regions affecting adiposity in chicken

Pierre-François Roux, Morgane Boutin, Colette Désert, Anis Djari, Diane Esquerre, Christophe Klopp, Sandrine Lagarrigue, Olivier Demeure

► **To cite this version:**

Pierre-François Roux, Morgane Boutin, Colette Désert, Anis Djari, Diane Esquerre, et al.. Re-sequencing data for refining candidate genes and polymorphisms in QTL regions affecting adiposity in chicken. PLoS ONE, 2014, 9 (10), pp.e111299. <10.1371/journal.pone.0111299>. <hal-01207758>

HAL Id: hal-01207758

<https://hal.science/hal-01207758v1>

Submitted on 27 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization



Re-Sequencing Data for Refining Candidate Genes and Polymorphisms in QTL Regions Affecting Adiposity in Chicken

Pierre-François Roux^{1,2,3}, Morgane Boutin^{1,2,3}, Colette Désert^{1,2,3}, Anis Djari⁴, Diane Esquerré⁵, Christophe Klopp⁴, Sandrine Lagarrigue^{1,2,3,9}, Olivier Demeure^{1,2,*,9}

1 INRA, UMR1348 PEGASE, Saint-Gilles, France, 2 Agrocampus Ouest, UMR1348 PEGASE, Rennes, France, 3 Université Européenne de Bretagne, Rennes, France, 4 INRA, SIGENAE, Castanet-Tolosan, France, 5 INRA, UMR1388 GenPhySE, GeT-PlaGe, Castanet-Tolosan, France

Abstract

In this study, we propose an approach aiming at fine-mapping adiposity QTL in chicken, integrating whole genome re-sequencing data. First, two QTL regions for adiposity were identified by performing a classical linkage analysis on 1362 offspring in 11 sire families obtained by crossing two meat-type chicken lines divergently selected for abdominal fat weight. Those regions, located on chromosome 7 and 19, contained a total of 77 and 84 genes, respectively. Then, SNPs and indels in these regions were identified by re-sequencing sires. Considering issues related to polymorphism annotations for regulatory regions, we focused on the 120 and 104 polymorphisms having an impact on protein sequence, and located in coding regions of 35 and 42 genes situated in the two QTL regions. Subsequently, a filter was applied on SNPs considering their potential impact on the protein function based on conservation criteria. For the two regions, we identified 42 and 34 functional polymorphisms carried by 18 and 24 genes, and likely to deeply impact protein, including 3 coding indels and 4 nonsense SNPs. Finally, using gene functional annotation, a short list of 17 and 4 polymorphisms in 6 and 4 functional genes has been defined. Even if we cannot exclude that the causal polymorphisms may be located in regulatory regions, this strategy gives a complete overview of the candidate polymorphisms in coding regions and prioritize them on conservation- and functional-based arguments.

Citation: Roux P-F, Boutin M, Désert C, Djari A, Esquerré D, et al. (2014) Re-Sequencing Data for Refining Candidate Genes and Polymorphisms in QTL Regions Affecting Adiposity in Chicken. PLoS ONE 9(10): e111299. doi:10.1371/journal.pone.0111299

Editor: Hikmet Budak, Sabanci University, Turkey

Received: May 27, 2014; **Accepted:** September 22, 2014; **Published:** October 21, 2014

Copyright: © 2014 Roux et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All DNA re-sequencing related data are available from the NCBI-SRA database (accession number SRP042034).

Funding: This work was funded by INRA, in the scope of program ChickSeq (2010). Pierre-François Roux's thesis grant was co-funded by INRA and Région Bretagne. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: olivier.demeure@rennes.inra.fr

⁹ These authors contributed equally to this work.

Introduction

Over the past decade, a lot of studies that aimed at dissecting the genetics of complex traits have been carried out, focused on identifying causal genes and polymorphisms for disease phenotypes or traits of economic interest, on humans, animal models and livestock species [1,2].

For such studies, genome-wide association study (GWAS), using high densities of genetic markers, based on linkage disequilibrium (LD) analyses, and permitting the identification of short length QTL regions is now commonly used. However, as technologies (high density SNP arrays) allowing LD approaches were only recently available, most of the published studies on complex traits in livestock species were based on linkage analysis (LA) approaches (as described in QTLdb [1]), and therefore described larger QTL regions.

Such large regions highlighted as impacting a complex trait using LA contain dozens of genes. Therefore, in general, only genes already known for having a link with the traits of interest are studied, while most of the genes are not even considered, as

they have no functional characterization. Even doing so, for many traits the number of potential candidate genes is high and studying them one by one is time consuming. This probably explains why, while thousands of QTL were detected, only very few causal polymorphisms were identified [3–6]. For a few years, with the advent of next-generation sequencing (NGS), and the highly decreased whole-genome sequencing costs associated, it is now possible to sequence the whole genome of a few individuals and to access without *a priori* to all polymorphisms from key individuals, which is critical to identify causal polymorphism underlying QTL regions [7,8]. The aim of this study was to combine QTL and NGS information to characterize regions affecting adiposity in chicken. This led to the identification of 216 missense SNPs, 5 nonsense SNPs and 3 coding indels occurring in 77 genes that underlay two QTLs. Using conservation- and functionality-based filters aiming at prioritizing polymorphisms, this number was reduced to 76 functional polymorphisms in 41 genes including 21 functional polymorphisms in 10 genes related to energetic metabolism.

Methods

Experimental design

A F_2 design of 561 offspring in 5 F_1 sire families [9] was created by inter-crossing two experimental meat-type chicken lines, the lean line and the fat line, that were divergently selected on abdominal fatness [10]. 801 backcross animals in 6 sire families derived from the F_2 design were also used.

Broilers were fed *ad libitum* using conventional starter diet from 0 to 3 week and grower diet from 4 to 9 week. At nine weeks of age, blood was collected from all animals of the F_2 and BC designs before slaughter. Body weight and abdominal fat weight were measured for each F_2 and BC animal. The experimental unit where birds were kept is registered by the French Ministry of Agriculture with the license number B-37-175-1 for animal experimentation. Except blood collection, no manipulation was performed before slaughtering. Slaughtering and blood collection were performed in accordance with guideline of ethics committee in Animal Experimentation of Val de Loire that approved this study.

Genetic markers

The F_1 sires were genotyped for a set of 9126 SNPs covering the available genome (assembly 2.1 WASHUC2). A subset of 1536 SNPs was selected using MarkerSet [11], based on marker location and heterozygosity in the F_1 population to maximize both genome coverage and marker informativity. The average density was one SNP each 0.66 cM, *i.e.* one SNP for 3 Mb. The 1362 offspring were then genotyped for those 1536 SNPs, at the National Genotyping Center (CNG, Evry, France) using Illumina GoldenGate technology (Illumina, San Diego, CA, USA). MendelSoft [12] was used to correct data for Mendelian inconsistencies. Out of the 1536 markers, 191 were eliminated due to technical or inconsistency issues (call rate lower than 85% and/or Mendelian errors higher than 5%). The chicken linkage consensus map build by Groenen *et al.* [13] was used to determine the genetic location of markers. Location of markers unavailable in the consensus map was extrapolated based on flanking markers.

QTL mapping

QTL interval mapping was performed using QTLMap software [14]. A mixture of half and full-sib families was considered as pedigree structure, and only sire meioses were studied. For abdominal fat (AF) QTL interval mapping, sex (2 levels) and hatch group (5 levels) were used as fixed effects, while body weight at nine weeks (BW9) was used as co-variable to adjust data. A likelihood ratio test (LRT) was performed at each cM to compare the fit of two models (*i.e.* the model with a QTL at the location considered *vs.* the model without fitting any QTL effect). Chromosome-wide significance thresholds were evaluated through empirical calculations obtained by simulations under the null hypothesis. A total of 10,000 simulations was performed for each trait \times chromosome combination and maximum LRT quantiles were calculated according to Harrel and Davis method [15]. Confidence intervals on QTL positions were estimated by the drop-off method [16]. Similarly to the reduction of one logarithm of odds (LOD) when using LOD scores, the maximum LRT value was reduced by 3.84 (a χ^2 distribution with one degree of freedom for $p < 0.05$) to determine a threshold. Region boundaries were then defined by the LRT locations crossing this threshold upstream and downstream of the LRT peak. The substitution effect of QTL was estimated in each sire family at the position of the LRT maximum and the significance was evaluated using a Student *t*-test.

Whole genome re-sequencing

DNA-seq libraries from 8 sire samples were prepared using the TruSeq DNA Sample Preparation Kit (Illumina, San Diego, CA) according to the manufacturer's instructions. Briefly, paired-end libraries with a 250-bp insert size were generated using the Illumina TruSeq DNA Sample Prep Kit. The libraries were quantified using QPCR Library Quantification Kit (Agilent), controlled on a High Sensitivity DNA Chip (Agilent) and sequenced in paired-end 2×100 bp on Illumina HiSeq 2000 with TruSeq v3 Kit. Sequencing produced an average of 92% of uniquely mapped reads, *i.e.* 20.4 Gb, which stands for a sequencing depth of 19.7 X.

DNA-seq data preprocessing, variant and genotype callings

The read sets obtained by sequencing whole genome were aligned against the *Gallus gallus* WASHUC2.1 reference genome from Ensembl 58 using BWA v0.7.0 [17]. All alignment bam files have been indexed and filtered. PCR duplicates were removed using SAMtools rmdup. Only reads with a unique mapping hit and a phred mapping quality score greater than 30 were kept. All these steps were performed using SAMtools v0.1.19 [18]. Genome Analysis ToolKit v2.4.9 (GATK) [19] was then used for base quality score re-calibration, indel re-alignment, and variant calling with UnifiedGenotyper by using default parameters, as suggested in GATK Best Practices recommendations [19,20]. Standard hard filtering parameters were finally used on SNPs and indels sets, according to GATK Best Practices recommendations [19,20]. Briefly, we filtered out SNPs characterized by: $QD < 2.0$ (quality by depth for non reference samples), $MQ < 40$ (mapping quality across all samples), $FS > 60.0$ (phred-scaled *p*-value of Fisher's exact test for strand bias), $MQRankSum < -12.5$ (mapping quality rank sum test) and $ReadPosRankSum < -8$ (read position rank sum test for the distance from the end of the read for reads with alternate allele); and we filtered out indels characterized by: $QD < 2.0$, $ReadPosRankSum < -20.0$, $InbreedingCoeff < -0.8$ and $FS < 200.0$. Finally, we removed genotypes for which the global depth DP (depth) was under 3 or higher than 60 (mean depth + 6σ), using VCFtools [21].

Polymorphism and gene annotations

To identify positional candidate genes in highlighted QTL regions, AnnotQTL software [22] was used, providing the location of genes in a specific region using NCBI and Ensembl databases, and filtering them onto ontology annotation criterion. Thus, genes belonging to "lipid metabolic process" Gene Ontology (GO) class, and to "diabetes mellitus" and "obesity" Online Mendelian Inheritance in Men (OMIM) classes were considered as functional candidate genes.

Variant annotation was performed using a two steps procedure. First, we used Ensembl Variant Effect Predictor (VEP) [23] for a global annotation on WASHUC2.1, allowing us to focus on SNPs associated to "missense", "stop gained" or "stop lost" coding consequences, and on indels associated to "inframe insertion", "inframe deletion" or "frameshift" coding consequences. After a validation of these annotations with VEP on the latest reference genome Galgal4, these variants were finely analyzed using the NGS-SNP tool [24] allowing to add meta-information about conservation between orthologous sequences. For this second step, we considered conservation information between *Gallus gallus*, *Canis lupus familiaris*, *Bos taurus*, *Mus musculus*, *Rattus norvegicus*, *Sus scrofa* and *Homo sapiens*. We considered two types of conservation score provided by NGS-SNP. The first one

defined by Grant *et al.* [24], termed “alignment score change” or a score, and based on the log-odds scoring matrix BLOSUM62, allows to range the similarity between the variant and the reference amino acid resulting from a coding variant, and amino acids in orthologous sequences. Considering the absolute value for this score, largest scores indicate that highly conserved amino acid residues are impacted and that substitution involved are less likely according to BLOSUM62 matrix. After studying how alignment score change was distributed for SNPs located on the two QTL regions, we considered as functional SNPs those having a $|a|$ score higher than the first quartile value of the distribution. The second score we used was provided by the Sorting Tolerant From Intolerant (SIFT) algorithm [25], based on the principle of protein evolution and using sequence homology based approach to classify amino acid substitutions. This latter approach is based on the hypothesis that highly conserved positions tend to be intolerant to substitutions, while those with a low degree of conservation tolerate most substitutions. We considered all SNPs classified as “deleterious” by SIFT as functional (excepted for “stop gained” SNPs, for which SIFT annotation is not available). Regarding indels, as they impact sequence in a much more complex pattern than SNPs, those annotations are not available.

Polymorphism validations by Sanger sequencing

Polymorphism validation was performed by Sanger sequencing for 8 coding SNPs and 3 coding indels. Targeted sequences were first PCR-amplified using 50 ng DNA with a Taq Uptitherm kit (Interchim). Amplicons were then purified and sent to GATC-Biotech (Konstanz, Germany) for Sanger sequencing using primers described in Table S3. Sanger traces related to indels were analyzed using Mixed Sequence Reader [26].

Results and Discussion

QTL analysis revealed two regions involved in the regulation of abdominal fatness

Whole genome QTL analysis for the abdominal fatness trait on 1362 offspring in 11 sire families using 1536 markers led to the identification of two QTLs mapped on GGA7 ($p < 0.05$) and GGA19 ($p < 0.01$) (Table 1). QTL effects were estimated at 0.54 and 0.45 phenotypic standard deviations and confidence interval at 10.2 cM (i.e. 6.01 Mb) and 5.7 cM (i.e. 2.91 Mb) for QTL on GGA7 and GGA19, respectively. These two regions were previously described as affecting AF ([27–29] for GGA7, [30] for GGA19, Figure 1), which reinforce the interest of focusing on them.

NGS data allowed the identification of functional polymorphisms in coding regions within QTL regions

As in all QTL fine-mapping studies, we first characterized genes located in each QTL region. Using AnnotQTL [22], based on both NCBI and Ensembl databases, all genes located in both QTL regions were listed. This led to the identification of 77 genes and 84 positional candidate genes located on GGA7 and GGA19 QTL regions, respectively.

Using whole genome DNA sequencing data, we then characterized polymorphisms in both regions and revealed 39,781 and 19,755 SNPs and 4,613 and 1,829 indels (Table 2). As it is difficult to annotate and determine the impact of polymorphisms on regulatory regions – an issue that is still important in model species due to the lack of annotation in non-coding regions [31–33] –, only polymorphisms having an impact on coding sequences were further considered.

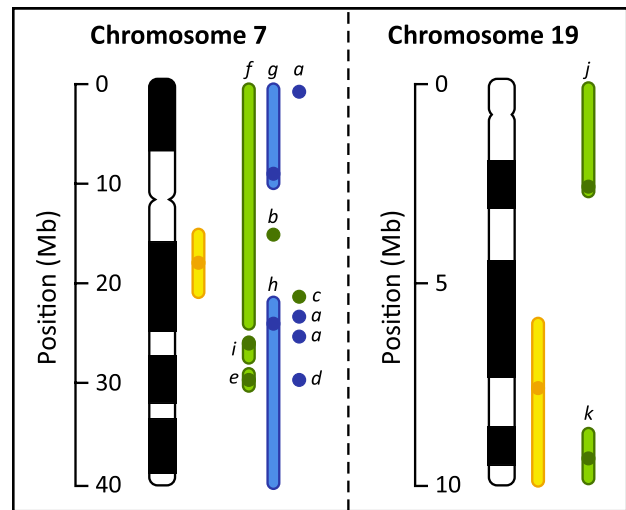


Figure 1. Chromosomal location of present and previously published QTLs related to abdominal fat weight. Empty boxes encompass the confidence interval of the QTL, when available. Plain boxes point out the QTL peak location, when available. QTLs colored in red are genome-wide significant ($p < 0.05$), while those colored in blue are suggestive QTLs ($p < 0.2$). QTLs described in the present study are colored in orange. ^a Ankra-Badu *et al.* [30], ^b Zhou *et al.* [54], ^c McElroy *et al.* [55], ^d Jennen *et al.* [56], ^e Tatsuda *et al.* [57], ^f Ikeobi *et al.* [28], ^g Lagarrigue *et al.* [27], ^h Park *et al.* [58], ⁱ Wang *et al.* [59], ^j Nadaf *et al.* [60], ^k Demeure *et al.* [9], ^l Tian *et al.* [61]. doi:10.1371/journal.pone.0111299.g001

With this aim, we used the Variant Effect Predictor (VEP) [23] tool from Ensembl to annotate pre-selected polymorphisms. We therefore focused on “missense”, “stop lost” or “stop gained” annotated coding SNPs, and on “frameshift”, “inframe insertion” and “inframe deletion” annotated coding indels. After a validation procedure of those annotations using VEP on the latest version of the chicken genome, we finally highlighted 120 (including three “stop gained” SNPs and one “frameshift” indel) and 104 (including two “stop gained” SNP, one “frameshift” indels, and one “inframe insertion” indels) candidate polymorphisms occurring in 35 and 42 genes in GGA7 and GGA19 QTL regions, respectively (Table 2).

Considering that important positions in protein and nucleotides sequence have been conserved throughout evolution, we then applied a filter taking conservation into account using NGS-SNP [24]. Indeed, high conservation rate through evolution among different vertebrates may reveal a high selective pressure, and therefore a major impact of substitutions on final protein function. Therefore, selecting SNPs impacting highly conserved regions on orthologous genes may help to focus on potential causal polymorphism underlying QTLs. In this study, we considered two criteria to estimate conservation at a given locus: the first one BLOSUM62-based [24] (SNPs with a $|a|$ score > 0.27 impact conserved amino-acid residues), and the other one based on SIFT prediction [25] (SNPs annotated as deleterious, see Methods). Considering SNPs that were respectful to either conservation criteria, or to both, the number of functional candidate SNPs was 38 SNPs on GGA7 and to 30 SNPs on GGA19 distributed on 17 and 21 genes, respectively (Table 2, Table S1).

Even if for nonsense SNPs and for indels those criteria cannot be robustly evaluated (see Methods), they were further considered as functional, due to the high impact they may have on final protein products (*i.e.* loss of a part of the protein sequence). Indeed, three indels were identified, located in *PRRI1* and two

Table 1. QTL analysis results.

| Chromosome | GGA7 | GGA19 |
|---------------------------------|------------|------------|
| QTL location (cM) | 58 | 52 |
| Size (cM) | 10.2 | 5.7 |
| Size (Mb) | 6.01 | 2.91 |
| LRT | 31.9 | 31.5 |
| Significance level ¹ | * | ** |
| QTL effect ² | 0.54 | 0.45 |
| Flanking marker – | rs15853071 | rs15850508 |
| Flanking marker + | rs14615490 | rs13576125 |

1: * 5%; ** 1%; Chromosome wide significance.

2: Substitution effect, expressed in phenotypic standard deviation.

doi:10.1371/journal.pone.0111299.t001

unknown genes (ENSGALG00000021856 and ENSGALG00000005578), with 60%, 12% and 97% of protein loss consequence, respectively. Similarly, three nonsense SNPs out of five were likely to have a drastic effects on the protein structure; they impact *SCN1A*, ENSGALG00000021856 and *OPN1LW* genes, and cause 69%, 41%, and 92% of the proteins loss, respectively (Table 3). None of them have been related to lipid metabolism and therefore represent strong positional but not functional candidate genes for the two QTL regions for adiposity.

To sum up, we finally considered 43 and 36 polymorphisms as strong functional candidates, impacting 18 and 23 genes on GGA7 and GGA19 QTL, respectively (Table 2, Table S1 and Table S2). Among those latter one, we selected 3 indels and 8 SNPs to perform Sanger sequencing. All SNP were validated as being polymorphic in our experimental designs. Concerning the 3 indels, this analysis led to the validation of an existing polymorphism in each case, and confirmed the open read frame shift for two of them, while a fine analysis of the one occurring on the chromosome 7 revealed it was not impacting the coding frame as first predicted by VEP and NGS-SNP, but was instead leading to the elongation of the encoded protein. As variants annotation rely on the use of a reference genome, miss-assemblies in such reference could lead to erroneous prediction. Moreover, short indels are usually constituted with homopolymers or tandem repeats, which tends to show higher error rates in re-sequencing data, and negatively impact the mappability of reads supporting them [34]. It appears therefore mandatory to consider carefully indels deeply impacting protein sequences highlighted with high-throughput sequencing approaches and to perform fine validation for further consideration.

Gene functional information allowed prioritization on genes related to adiposity and associated candidate polymorphisms

Those polymorphisms, selected for being both coding and being functional (*i.e.* having a high propensity for impacting final protein product) were considered as strong candidates underlying QTLs. But, even if the causal polymorphisms may be located in genes that have not already been described as involved in the lipid metabolism, genes known as being related to the trait of interest stand for the first strong candidate genes to be considered. Using gene functional annotations could therefore allow prioritizing polymorphisms among pre-selected candidates. Considering the phenotype targeted in our study, such functional genes were selected using GO and OMIM databases (and genes related to lipid metabolic process, diabetes mellitus and obesity, see Methods). We identified, 6 and 4 functional candidate genes among the 18 and 23 genes with at least one functional polymorphism previously selected on GGA7 and GGA19 QTL regions, respectively. All these genes and associated polymorphisms (17 and 4 for the two regions) are listed in Table S1 and S2. Location and impact of the mutation on the protein and conservation between species are presented in Figure S1.

Among those functional candidates, *SLC25A12* that encodes the Ca²⁺-regulated mitochondrial aspartate-glutamate carrier, is involved in carbohydrates and glucose metabolism and is known as a major autism spectrum disorder susceptibility gene and support oxidative phosphorylation and energy production [35]. *GRB14* and *DOC2B* are involved in regulation of insulin secretion, through insulin receptor binding abilities for the first one [36–39] and through insulin vesicle-mediated secretion and uptake capacity for the second one [39–42]. *LRP2* encodes for an

Table 2. Selection of candidate polymorphisms.

| | | In QTL region | And affecting protein sequence | And potentially functional |
|-------|------------------|---------------|--------------------------------|----------------------------|
| GGA7 | Number of SNPs | 39781 | 119 | 41 |
| | Number of indels | 4613 | 1 | 1 |
| | Number of genes | 77 | 35 | 18 |
| GGA19 | Number of SNPs | 19755 | 102 | 32 |
| | Number of indels | 1829 | 2 | 2 |
| | Number of genes | 84 | 42 | 24 |

doi:10.1371/journal.pone.0111299.t002

Table 3. Distribution of functional polymorphisms.

| Chromosome | Ensembl gene ID | HGNC | Functional missense SNPs ¹ | Nonsense SNPs ² | Coding indels ³ |
|---------------------------|---------------------------|---------------------------|---------------------------------------|----------------------------|----------------------------|
| GGA7 | ENSGALG00000011149 | PLA2R1 | 7 (12) | - | - |
| | ENSGALG00000011153 | LY75 | 3 (5) | - | - |
| | ENSGALG00000010858 | LRP2 | 2 (6) | 1 (255, 5%) | - |
| | ENSGALG00000020703 | GRB14 | 2 (3) | - | - |
| | ENSGALG00000010891 | ABCB11 | 1 (2) | - | - |
| | ENSGALG00000009545 | SLC25A12 | 1 (1) | - | - |
| | ENSGALG00000010943 | SCN1A | 1 (1) | 1 (816, 69%) | - |
| | ENSGALG00000021856 | - | - | 1 (41, 41%) | 1 (12, 12%) |
| | ENSGALG00000010933 | XIRP2 | 5 (15) | - | - |
| | ENSGALG00000011068 | COBLL1 | 4 (14) | - | - |
| | ENSGALG00000011052 | SLC38A11 | 3 (5) | - | - |
| | ENSGALG00000010956 | TTC21B | 2 (9) | - | - |
| | ENSGALG00000014209 | GPR155 | 2 (3) | - | - |
| | ENSGALG00000013235 | PDK1 | 1 (1) | - | - |
| | ENSGALG00000009583 | GORASP2 | 1 (2) | - | - |
| | ENSGALG00000020737 | KLHL23 | 1 (3) | - | - |
| | ENSGALG00000011110 | DPP4 | 1 (2) | - | - |
| | ENSGALG00000011172 | LOC429030 | 1 (7) | - | - |
| | GGA19 | ENSGALG00000023554 | PIGW | 1 (8) | - |
| ENSGALG00000005420 | | AATF | 1 (4) | - | - |
| ENSGALG00000005084 | | TRIM37 | 1 (2) | - | - |
| ENSGALG00000004917 | | DOC2B | 1 (1) | - | - |
| ENSGALG00000005037 | | TEX14 | 4 (13) | 1 (131, 9%) | - |
| ENSGALG00000004924 | | OPN1LW | - | 1 (324, 9%) | - |
| ENSGALG00000021526 | | PRR11 | 2 (6) | - | 1 (176, 60%) |
| ENSGALG00000005578 | | - | - | - | 1 (133, 97%) |
| ENSGALG00000005061 | | PPM1E | 3 (4) | - | - |
| ENSGALG00000005279 | | BRIP1 | 2 (2) | - | - |
| ENSGALG00000005230 | | MED13 | 2 (2) | - | - |
| ENSGALG00000005295 | | BCAS3 | 2 (2) | - | - |
| ENSGALG00000005468 | | SYNRG | 1 (6) | - | - |
| ENSGALG00000005350 | | USP32 | 1 (5) | - | - |
| ENSGALG00000005489 | | DDX52 | 1 (4) | - | - |
| ENSGALG00000005516 | | HEATR6 | 1 (4) | - | - |
| ENSGALG00000005173 | | TUBD1 | 1 (4) | - | - |
| ENSGALG00000005594 | | OMG | 1 (3) | - | - |
| ENSGALG00000005269 | | INTS2 | 1 (2) | - | - |
| ENSGALG00000005285 | | TBX4 | 1 (2) | - | - |
| ENSGALG00000005362 | | - | 1 (2) | - | - |
| ENSGALG00000005868 | | RAP1GAP2 | 1 (2) | - | - |
| ENSGALG00000011040 | | SCN2A | 1 (2) | - | - |
| ENSGALG00000005126 | | DHX40 | 1 (1) | - | - |

1: Number of SNPs having a potential impact on protein function; Number of total SNPs affecting protein sequence is given in brackets.

2: Number of SNPs having a nonsense impact; Number of amino acids and percentage of protein sequence that are lost are given in brackets.

3: Number of coding indels; Number of amino acids and percentage of protein sequence that are lost are given in brackets.

doi:10.1371/journal.pone.0111299.t003

endocytic receptor known as megalin, which internalizes a variety of ligands such as nutrients, signalling molecules, hormones and lipoproteins and regulates hepatic lipid flow-through [43,44]. *ABCB11* operate the release of bile salt on the canalicular

membrane of hepatocytes and is associated to intrahepatic cholestasis [45,46]. *PLA2R1* is involved in the modulation of eicosanoid production [47,48]. *LY75* is involved in antigen presentation and endocytosis, and has been identify as a

susceptibility locus for type 2 diabetes mellitus [49,50]. *TRIM37*, which encodes a peroxysomal protein with E3 ubiquitin ligase known to cause mulibrey nanism when mutated, is also related to severe insulin resistance syndrome [51]. Finally, *PIGW* is involved in the glycosylphosphatidylinositol synthesis [52] and *AATF* in apoptosis inhibition [53].

Conclusions

Using NGS data allowed the identification of 120 (GGA7) and 104 (GGA19) polymorphisms having an impact on protein sequence. While all of them are good candidates, conservation- and functionality-based filters applied on polymorphisms or genes functional annotation allows us to prioritize a list, with 45 (GGA7) and 36 (GGA19) polymorphisms that might have a strong effect on the protein function (including 7 coding indels or nonsense SNPs), 17 (GGA7) and 4 (GGA19) of them being in functional candidate genes. Even if we cannot exclude that the causal polymorphisms may be located in regulatory regions, this strategy gives a complete overview of the candidate polymorphisms in coding regions, prioritize them and open the way to further validation, by genetic approaches using other populations phenotyped for the traits of interest, or by molecular and cellular functional approaches.

References

- Hu Z-L, Park CA, Wu X-L, Reecy JM (2013) Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Res* 41: D871–D879. doi:10.1093/nar/gks1150.
- Laulederkind SJF, Hayman GT, Wang S-J, Smith JR, Lowry TF, et al. (2013) The Rat Genome Database 2013 – data, tools and users. *Brief Bioinform* 14: 520–526. doi:10.1093/bib/bbt007.
- Georges M (2007) Mapping, Fine Mapping, and Molecular Dissection of Quantitative Trait Loci in Domestic Animals. *Annu Rev Genomics Hum Genet* 8: 131–162. doi:10.1146/annurev.genom.8.080706.092408.
- Grisart B, Famir F, Karim L, Cambisano N, Kim JJ, et al. (2004) Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. Available: <http://orbi.ulg.ac.be/handle/2268/101941>. Accessed 18 March 2014.
- Clop A, Marcq F, Takeda H, Pirotin D, Tordoir X, et al. (2006) A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nat Genet* 38: 813–818. doi:10.1038/ng1810.
- Le Bihan-Duval E, Nadaf J, Berri C, Pitel F, Graulet B, et al. (2011) Detection of a Cis eQTL Controlling BMCO1 Gene Expression Leads to the Identification of a QTG for Chicken Breast Meat Color. *PLoS ONE* 6: e14825. doi:10.1371/journal.pone.0014825.
- Rat Genome Sequencing and Mapping Consortium, Baud A, Hermesen R, Guryev V, Stridh P, et al. (2013) Combined sequence-based and genetic mapping analysis of complex traits in outbred rats. *Nat Genet* 45: 767–775. doi:10.1038/ng.2644.
- Edwards SL, Beesley J, French JD, Dunning AM (2013) Beyond GWAS: Illuminating the Dark Road from Association to Function. *Am J Hum Genet* 93: 779–797. doi:10.1016/j.ajhg.2013.10.012.
- Demeure O, Duclos MJ, Bacciu N, Mignon GL, Filangi O, et al. (2013) Genome-wide interval mapping using SNPs identifies new QTL for growth, body composition and several physiological variables in an F2 intercross between fat and lean chicken lines. *Genet Sel Evol* 45: 36. doi:10.1186/1297-9686-45-36.
- Leclercq B, Blum JC, Boyer JP (1980) Selecting broilers for low or high abdominal fat: Initial observations. *Br Poult Sci* 21: 107–113. doi:10.1080/00071668008416644.
- Demeure O, Lecercq F (2008) MarkerSet: a marker selection tool based on markers location and informativity in experimental designs. *BMC Res Notes* 1: 9. doi:10.1186/1756-0500-1-9.
- Givry S de, Bouchez M, Chabrier P, Milan D, Schiex T (2005) Carh ta Gene: multipopulation integrated genetic and radiation hybrid mapping. *Bioinformatics* 21: 1703–1704. doi:10.1093/bioinformatics/bti222.
- Groenen MAM, Wahlberg P, Foglio M, Cheng HH, Megens H-J, et al. (n.d.) A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. *Genome Res* 19: 510–519.
- Filangi O, Mangin B, Goffinet B, Boichard D, Le Roy P (2012) QTLMap, a software for QTL detection in outbred populations.
- Harrell FE, Davis CE (1982) A new distribution-free quantile estimator. *Biometrika* 69: 635–640. doi:10.1093/biomet/69.3.635.
- Ott J (1999) *Analysis of Human Genetic Linkage*. JHU Press. 418 p.
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25: 1754–1760. doi:10.1093/bioinformatics/btp324.

Supporting Information

Figure S1 Multi-species protein alignments for a sub-selection of indels and SNPs impacting peptidique sequence.

(PDF)

Table S1 Description of functional candidate SNPs in the two QTL regions.

(XLSX)

Table S2 Description of functional candidate indels in the two QTL regions.

(XLSX)

Table S3 List of primers used for Sanger sequencing validation.

(XLSX)

Author Contributions

Conceived and designed the experiments: PFR SL OD. Performed the experiments: PFR MB CD DE. Analyzed the data: PFR AD CK. Wrote the paper: PFR SL OD.

- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079. doi:10.1093/bioinformatics/btp352.
- Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, et al. (2013) From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics*. John Wiley & Sons, Inc. Available: <http://onlinelibrary.wiley.com/doi/10.1002/0471250953.bi1110s43.abstract>. Accessed 18 December 2013.
- DePristo MA, Banks E, Poplin RE, Garimella KV, Maguire JR, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43: 491–498. doi:10.1038/ng.806.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158. doi:10.1093/bioinformatics/btr330.
- Lecerf F, Bretaudeau A, Sallou O, Desert C, Blum Y, et al. (2011) AnnotQTL: a new tool to gather functional and comparative information on a genomic region. *Nucleic Acids Res* 39: W328–W333. doi:10.1093/nar/gkr361.
- McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, et al. (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* 26: 2069–2070. doi:10.1093/bioinformatics/btq330.
- Grant JR, Arantes AS, Liao X, Stothard P (2011) In-depth annotation of SNPs arising from resequencing projects using NGS-SNP. *Bioinformatics* 27: 2300–2301. doi:10.1093/bioinformatics/btr372.
- Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 4: 1073–1081. doi:10.1038/nprot.2009.86.
- Chang C-T, Tsai C-N, Tang CY, Chen C-H, Lian J-H, et al. (2012) Mixed sequence reader: a program for analyzing DNA sequences with heterozygous base calling. *ScientificWorldJournal* 2012: 365104. doi:10.1100/2012/365104.
- Lagarigue S, Pitel F, Carré W, Abasht B, Roy PL, et al. (2006) Mapping quantitative trait loci affecting fatness and breast muscle weight in meat-type chicken lines divergently selected on abdominal fatness. *Genet Sel Evol* 38: 85. doi:10.1186/1297-9686-38-1-85.
- Ikeobi CON, Woolliams JA, Morrice DR, Law A, Windsor D, et al. (2002) Quantitative trait loci affecting fatness in the chicken. *Anim Genet* 33: 428–435.
- Jacobsson L, Park H-B, Wahlberg P, Fredriksson R, Perez-Enciso M, et al. (2005) Many QTLs with minor additive effects are associated with a large difference in growth between two selection lines in chickens. *Genet Res* 86: 115–125. doi:10.1017/S0016672305007767.
- Ankra-Badu GA, Shriner D, Bihan-Duval EL, Mignon-Grasteau S, Pitel F, et al. (2010) Mapping main, epistatic and sex-specific QTL for body composition in a chicken population divergently selected for low or high growth rate. *BMC Genomics* 11: 107. doi:10.1186/1471-2164-11-107.
- Paul DS, Soranzo N, Beck S (2014) Functional interpretation of non-coding sequence variation: concepts and challenges. *BioEssays News Rev Mol Cell Dev Biol* 36: 191–199. doi:10.1002/bies.201300126.
- Alexander RP, Fang G, Rozowsky J, Snyder M, Gerstein MB (2010) Annotating non-coding regions of the genome. *Nat Rev Genet* 11: 559–571. doi:10.1038/nrg2814.

33. Ritchie GRS, Dunham I, Zeggini E, Flicek P (2014) Functional annotation of noncoding sequence variants. *Nat Methods* 11: 294–296. doi:10.1038/nmeth.2832.
34. Albers CA, Lunter G, MacArthur DG, McVean G, Ouwehand WH, et al. (2011) Dindel: Accurate indel calls from short-read data. *Genome Res* 21: 961–973. doi:10.1101/gr.112326.110.
35. Napolioni V, Persico AM, Porcelli V, Palmieri L (2011) The Mitochondrial Aspartate/Glutamate Carrier AGC1 and Calcium Homeostasis: Physiological Links and Abnormalities in Autism. *Mol Neurobiol* 44: 83–92. doi:10.1007/s12035-011-8192-2.
36. Kasus-Jacobi A, Perdereau D, Auzan C, Clauser E, Van Obberghen E, et al. (1998) Identification of the rat adapter Grb14 as an inhibitor of insulin actions. *J Biol Chem* 273: 26026–26035.
37. Cooney GJ, Lyons RJ, Crew AJ, Jensen TE, Molero JC, et al. (2004) Improved glucose homeostasis and enhanced insulin signalling in Grb14-deficient mice. *EMBO J* 23: 582–593. doi:10.1038/sj.emboj.7600082.
38. Holt IJ, Siddle K (2005) Grb10 and Grb14: enigmatic regulators of insulin action – and more? *Biochem J* 388: 393–406. doi:10.1042/BJ20050216.
39. Goenaga D, Hampe C, Carré N, Cailliau K, Browaeys-Poly E, et al. (2009) Molecular determinants of Grb14-mediated inhibition of insulin signaling. *Mol Endocrinol* 23: 1043–1051. doi:10.1210/me.2008-0360.
40. Ramalingam L, Oh E, Yoder SM, Brozinick JT, Kalwat MA, et al. (2012) Doc2b is a key effector of insulin secretion and skeletal muscle insulin sensitivity. *Diabetes* 61: 2424–2432. doi:10.2337/db11-1525.
41. Ramalingam L, Oh E, Thurmond DC (2014) Doc2b enrichment enhances glucose homeostasis in mice via potentiation of insulin secretion and peripheral insulin sensitivity. *Diabetologia*. doi:10.1007/s00125-014-3227-7.
42. Li J, Cantley J, Burchfield JG, Meoli CC, Stöckli J, et al. (2014) DOC2 isoforms play dual roles in insulin secretion and insulin-stimulated glucose uptake. *Diabetologia*. doi:10.1007/s00125-014-3312-y.
43. Pieper-Fürst U, Lammert F (2013) Low-density lipoprotein receptors in liver: old acquaintances and a newcomer. *Biochim Biophys Acta* 1831: 1191–1198.
44. Gotthardt M, Trommsdorff M, Nevitt MF, Shelton J, Richardson JA, et al. (2000) Interactions of the low density lipoprotein receptor gene family with cytosolic adaptor and scaffold proteins suggest diverse biological functions in cellular communication and signal transduction. *J Biol Chem* 275: 25616–25624. doi:10.1074/jbc.M000955200.
45. Strautmieks SS, Bull LN, Knisely AS, Kocoshis SA, Dahl N, et al. (1998) A gene encoding a liver-specific ABC transporter is mutated in progressive familial intrahepatic cholestasis. *Nat Genet* 20: 233–238. doi:10.1038/3034.
46. Kubitz R, Dröge C, Stündt J, Weissenberger K, Häussinger D (2012) The bile salt export pump (BSEP) in health and disease. *Clin Res Hepatol Gastroenterol* 36: 536–553. doi:10.1016/j.clinre.2012.06.006.
47. Hanasaki K, Arita H (1999) Biological and pathological functions of phospholipase A(2) receptor. *Arch Biochem Biophys* 372: 215–223. doi:10.1006/abbi.1999.1511.
48. Tamaru S, Mishina H, Watanabe Y, Watanabe K, Fujioka D, et al. (2013) Deficiency of phospholipase A2 receptor exacerbates ovalbumin-induced lung inflammation. *J Immunol* 191: 1021–1028. doi:10.4049/jimmunol.1300738.
49. Greenawalt DM, Sieberts SK, Cornelis MC, Girman CJ, Zhong H, et al. (2012) Integrating Genetic Association, Genetics of Gene Expression, and Single Nucleotide Polymorphism Set Analysis to Identify Susceptibility Loci for Type 2 Diabetes Mellitus. *Am J Epidemiol* 176: 423–430. doi:10.1093/aje/kws123.
50. Staines K, Young JR, Butter C (2013) Expression of Chicken DEC205 Reflects the Unique Structure and Function of the Avian Immune System. *PLoS ONE* 8: e51799. doi:10.1371/journal.pone.0051799.
51. Karlberg N, Jalanko H, Kallijärvi J, Lehesjoki A-E, Lipsanen-Nyman M (2005) Insulin resistance syndrome in subjects with mutated RING finger protein TRIM37. *Diabetes* 54: 3577–3581.
52. Murakami Y, Siripanyapinyo U, Hong Y, Kang JY, Ishihara S, et al. (2003) PIG-W is critical for inositol acylation but not for flipping of glycosylphosphatidylinositol-anchor. *Mol Biol Cell* 14: 4285–4295. doi:10.1091/mbc.E03-03-0193.
53. Page G, Lödige I, Kögel D, Scheidtmann KH (1999) AATF, a novel transcription factor that interacts with Dlk/ZIP kinase and interferes with apoptosis. *FEBS Lett* 462: 187–191. doi:10.1016/S0014-5793(99)01529-X.
54. Zhou H, Deeb N, Evock-Clover CM, Ashwell CM, Lamont SJ (2006) Genome-Wide Linkage Analysis to Identify Chromosomal Regions Affecting Phenotypic Traits in the Chicken. II. Body Composition. *Poult Sci* 85: 1712–1721. doi:10.1093/ps/85.10.1712.
55. McElroy JP, Kim JJ, Harry DE, Brown SR, Dekkers JCM, et al. (2006) Identification of trait loci affecting white meat percentage and other growth and carcass traits in commercial broiler chickens. *Poult Sci* 85: 593–605.
56. Jennen DGJ, Vereijken ALJ, Bovenhuis H, Crooijmans RPMA, Veenendaal A, et al. (2004) Detection and localization of quantitative trait loci affecting fatness in broilers. *Poult Sci* 83: 295–301.
57. Tatsuda K, Fujinaka K (2001) Genetic Mapping of the QTL Affecting Abdominal Fat Deposition in Chickens. *J Poult Sci* 38: 266–274. doi:10.2141/jpsa.38.266.
58. Park H-B, Jacobsson L, Wahlberg P, Siegel PB, Andersson L (2006) QTL analysis of body composition and metabolic traits in an intercross between chicken lines divergently selected for growth. *Physiol Genomics* 25: 216–223. doi:10.1152/physiolgenomics.00113.2005.
59. Wang SZ, Hu XX, Wang ZP, Li XC, Wang QG, et al. (2012) Quantitative trait loci associated with body weight and abdominal fat traits on chicken chromosomes 3, 5 and 7. *Genet Mol Res GMR* 11: 956–965. doi:10.4238/2012.April.19.1.
60. Nadaf J, Pitel F, Gilbert H, Duclos MJ, Vignoles F, et al. (2009) QTL for several metabolic traits map to loci controlling growth and body composition in an F2 intercross between high- and low-growth chicken lines. *Physiol Genomics* 38: 241–249. doi:10.1152/physiolgenomics.90384.2008.
61. Tian J, Wang S, Wang Q, Leng L, Hu X, et al. (2009) A Single Nucleotide Polymorphism of Chicken Acetyl-CoA Carboxylase A Gene Associated with Fatness Traits. *Anim Biotechnol* 21: 42–50. doi:10.1080/10495390903347009.