



# Identification of candidate genes and mutations in QTL regions for chicken growth using bioinformatic analysis of NGS and SNP-chip data

Muhammad Ahsan<sup>1</sup>, Xidan Li<sup>1</sup>, Andreas E. Lundberg<sup>1</sup>, Marcin Kierczak<sup>1</sup>, Paul B. Siegel<sup>2</sup>, Örjan Carlborg<sup>1</sup> and Stefan Marklund<sup>1</sup> \*

<sup>1</sup> Division of Computational Genetics, Department of Clinical Sciences, Swedish University of Agricultural Sciences, Uppsala, Sweden

<sup>2</sup> Department of Animal and Poultry Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

## Edited by:

Ji Qi, Fudan University, China

## Reviewed by:

Gaurav Sablok, Istituto Agrario San

Michele, Italy

Qiangfeng Cliff Zhang, Stanford

University, USA

## \*Correspondence:

Stefan Marklund, Division of  
Computational Genetics, Department  
of Clinical Sciences, Swedish  
University of Agricultural Sciences,  
Box 7078, SE-750 07 Uppsala,  
Sweden  
e-mail: stefan.marklund@slu.se

Mapping of chromosomal regions harboring genetic polymorphisms that regulate complex traits is usually followed by a search for the causative mutations underlying the observed effects. This is often a challenging task even after fine mapping, as millions of base pairs including many genes will typically need to be investigated. Thus to trace the causative mutation(s) there is a great need for efficient bioinformatic strategies. Here, we searched for genes and mutations regulating growth in the Virginia chicken lines – an experimental population comprising two lines that have been divergently selected for body weight at 56 days for more than 50 generations. Several quantitative trait loci (QTL) have been mapped in an F2 intercross between the lines, and the regions have subsequently been replicated and fine mapped using an Advanced Intercross Line. We have further analyzed the QTL regions where the largest genetic divergence between the High-Weight selected (HWS) and Low-Weight selected (LWS) lines was observed. Such regions, covering about 37% of the actual QTL regions, were identified by comparing the allele frequencies of the HWS and LWS lines using both individual 60K SNP chip genotyping of birds and analysis of read proportions from genome resequencing of DNA pools. Based on a combination of criteria including significance of the QTL, allele frequency difference of identified mutations between the selected lines, gene information on relevance for growth, and the predicted functional effects of identified mutations we propose here a subset of candidate mutations of highest priority for further evaluation in functional studies. The candidate mutations were identified within the *GCG*, *IGFBP2*, *GRB14*, *CRIM1*, *FGF16*, *VEGFR-2*, *ALG11*, *EDN1*, *SNX6*, and *BIRC7* genes. We believe that the proposed method of combining different types of genomic information increases the probability that the genes underlying the observed QTL effects are represented among the candidate mutations identified.

**Keywords:** candidate genes, growth, functional prediction, genetic divergence, QTL, SNP, resequencing

## INTRODUCTION

Economically important production traits in domestic animals are generally complex, i.e., determined by factors that may include both genetic and environmental regulators. This is also true for many diseases in humans and animals. Thus, while it is often highly desirable to understand the regulation of specific complex traits, the task can be extremely challenging. For example, regions identified by quantitative trait loci (QTL) analysis will even after fine mapping of the QTL typically indicate regions including millions of base pairs and hundreds of genes that need to be explored to find causative mutation(s).

In this study our aim was to develop a bioinformatics strategy to mine already identified QTL regions to identify candidate genes for growth trait in chicken. The QTLs have been identified for body weight at 56 days of age in the Virginia chicken lines – an experimental population comprising two lines that have been divergently selected for body weight at 56 days for more than 50 generations at Virginia Tech (Dunnington and Siegel, 1996; Marquez et al.,

2010; Dunnington et al., 2013). Both lines started from the same base population, which was produced from crosses of seven partially inbred lines of White Plymouth Rocks and now differ by more than 10-fold in body weight at selection age. Individuals from the 41st generation of these High-Weight selected (HWS) and Low-Weight selected (LWS) lines were used as founders in a QTL mapping pedigree and several QTL regions were mapped in an F2 intercross between the lines (Jacobsson et al., 2005). These regions have subsequently been replicated and fine mapped using an Advanced Intercross Line (Besnier et al., 2011). Candidate genes and mutations were here sought in the regions of the QTLs where the greatest allele frequency differences between HWS and LWS founder lines of the QTL cross were observed by individual SNP-chip genotyping and next generation sequencing (NGS) of DNA pools from the HWS and LWS. Based on a bioinformatic analysis of these regions and the SNPs detected by NGS we present candidate genes and mutations of high priority for further investigations in order to explain the observed QTL effects.

## MATERIALS AND METHODS

Here, we present a bioinformatic strategy that in a structured and objective way helps to prioritize candidate genes for further study in mapped QTL regions by integrating information from multiple sources. First, the region to be evaluated further is narrowed down by, at each SNP-location in the evaluated region, calculating a combined score for the potential that each part of the region harbors a mutation underlying the phenotype. This is done by combining the statistical support from significance of the QTL effect at the particular marker, which is a measurement of the effect of the alternative alleles on the studied phenotype, with two measures of the genetic divergence between the founder lines (i.e., allele-frequency differences) at the particular location, which is an indicator of the direct or indirect selective pressure on the region due to an association with the phenotypes of importance when generating the divergent founder lines. Then, all the polymorphisms in the prioritized region are evaluated in more detail to select the most likely genes affecting the analyzed trait and bioinformatically predict the potential functional effects of each identified polymorphism. The details of the procedure, and its application to our particular chicken dataset, are described with a flowchart in **Figure 1** and in the text below.

### MAPPED QTL REGIONS TO BE EXPLORED

We studied seven fine-mapped QTL on chicken chromosomes 1–5, 7, and 20, with previously observed effects on body-weight at selection age in a QTL mapping pedigree founded with HWS and LWS chickens from generation 41 (Jacobsson et al., 2005; Besnier et al., 2011). The fine-mapping of the QTL was previously reported by Besnier et al. (2011) where the effect of each SNP in the QTL regions was estimated using a Flexible InterCross Analysis model (Rönnegård et al., 2008). The statistical QTL support curve across the regions from the analysis based on this model (Model B in the original paper) was here used for identification and evaluation of candidate regions.

### INDIVIDUAL GENOME-WIDE 60 K SNP CHIP GENOTYPING

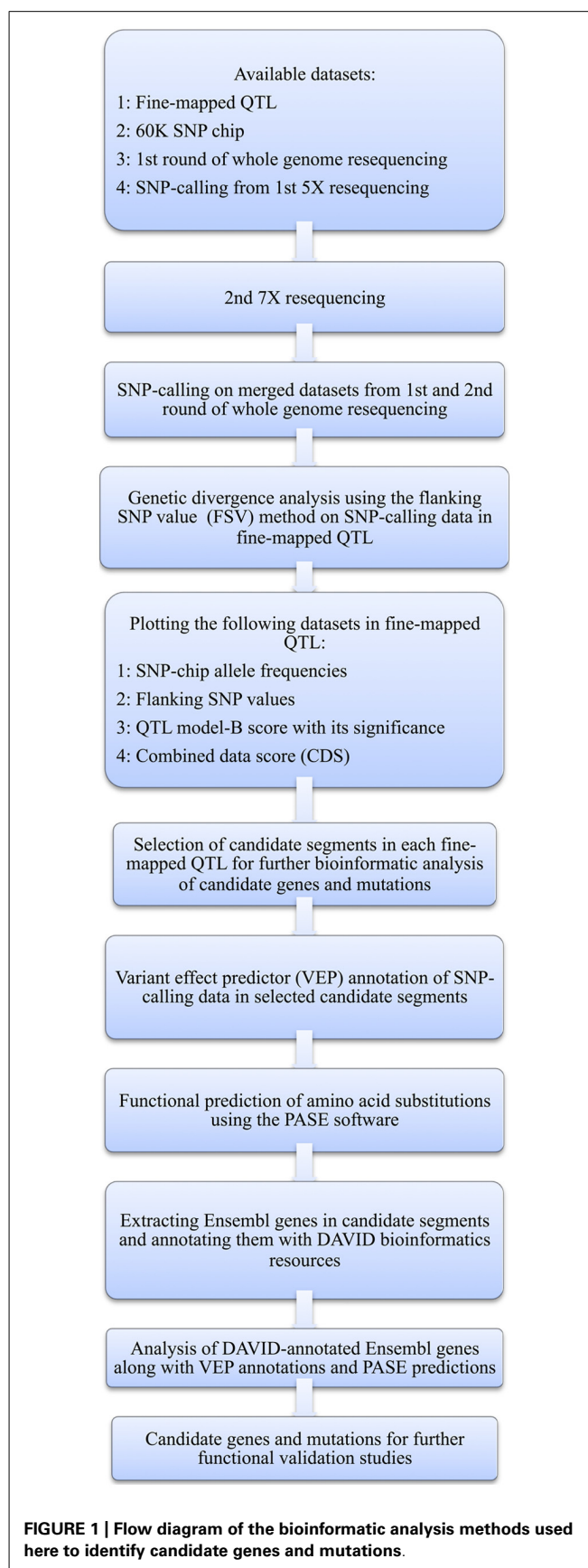
Genome-wide 60K SNP chip genotypes of 20 individuals from each of the HWS and LWS lines, generation 41 (Marklund and Carlborg, 2010) was available. We used these genotype data to estimate the allele-frequency differences between the lines across the QTL regions to be explored.

### GENOME RESEQUENCING OF POOLED POPULATION-SAMPLES AND SNP-CALLING

Genome resequencing was performed in two separate runs using DNA pools from the HWS and LWS lines. The data from the two experiments were combined to maximize the sensitivity in the SNP detection.

For earlier studies DNA from two pools of genomic DNA, one from each of the HWS and LWS lines, were used to generate resequencing data with 5× average depth coverage for each line. The reads were aligned to the Red Jungle Fowl's (RJF) reference genome assembly (WUGSC 2.1/galGal3, May 2006; Marklund and Carlborg, 2010; Rubin et al., 2010).

For the current and future studies the second round of resequencing was performed using two new pools of DNA samples.



**FIGURE 1 | Flow diagram of the bioinformatic analysis methods used here to identify candidate genes and mutations.**

The individuals selected for each pool were guided by data from earlier performed 60K SNP-chip genome-wide genotyping. From each line, the eight individuals with the most non-representative genotype pattern in the QTL regions were selected to increase the possibilities for detection of variation within lines and thereby allow improved precision in the fine mapping of regions with high degree of between-line fixation. The ABI SOLiD resequencing was carried out by the Uppsala Genome Center using mate-pair libraries and 50 bases per read with  $\sim 7\times$  depth coverage in each line. We aligned the reads to the RJF reference genome assembly (WUGSC 2.1/galGal3, May 2006) using the MOSAIK software (Lee et al., 2013). The resequencing datasets from the two rounds of sequencing were combined for SNP calling based on a total of  $\sim 12\times$  depth coverage in each line. However in each line SNP alleles were called at each SNP position as determined using the threshold of three non-RJF reads that we set for SNP detection including the total number of reads from both lines (i.e.,  $\sim 24\times$  depth coverage) to increase the sensitivity. The GigaBayes software, a newer version of PolyBayes (Marth et al., 1999), was used for SNP calling.

#### GENETIC DIVERGENCE ANALYSIS USING THE FLANKING-SNP-VALUE METHOD IN RESEQUENCING DATA

We applied the flanking-SNP-value (FSV) method (Marklund and Carlborg, 2010) to the resequencing data from the HWS and LWS lines across the selected QTL regions. The FSV method computes estimated allele frequency differences between the HWS and LWS lines for each evaluated SNP position based on information from the SNP itself as well as from data of flanking SNPs in both directions within an interval presumed to show a high degree of linkage disequilibrium with the SNP. Thus, the input data for FSV computation are the AB scores at all these positions, which in each line indicate the proportion of resequenced reads that are in agreement with reference sequence of RJF.

#### A COMBINED SCORE FOR CANDIDATE GENE PRIORITIZATION

The allele frequency differences based on the individual SNP genotyping, the genetic divergence estimates (FSV) from the population-pool genome resequencing were plotted across the QTL regions together with the QTL support-curve from the QTL fine-mapping (Besnier et al., 2011). A combined data score (CDS) was also calculated based on these three information sources as:

$$\text{CDS} = \{[(\text{FSVscore} + \text{SNPchip-allele freq.})/2] + (\text{Normalized score of QTL\_ModelB})\}/2$$

The CDS was plotted to provide an objective statistic to prioritize regions for further analysis and evaluations of candidate genes and mutations. In most cases the regions were selected above the QTL significance and with high CDS.

#### IDENTIFICATION OF CANDIDATE GENES AND MUTATIONS IN PRIORITIZED REGIONS

Genes were identified in the prioritized regions within the QTL using the Ensembl database (version 67; Flicek et al., 2012). The general functions and gene annotations for each gene was compiled using information from the Database for Annotation, Visualization and Integrated Discovery (DAVID; Huang et al.,

2009a,b). DAVID integrates annotations for genes from different omics databases including, for instance, gene ontology (GO), KEGG and PANTHER.

All SNPs detected by resequencing in selected candidate regions were analyzed with variant effect predictor (VEP) from Ensembl (McLaren et al., 2010). VEP maps the locations of SNPs, insertions and deletions to different functional parts of Ensembl genes, transcripts and regulatory sequences. It differentiates coding SNPs in exons as synonymous or non-synonymous and shows amino acid substitutions. For some species, however not in chicken, it also predicts the functional consequences of non-synonymous SNPs (nsSNPs) on carrying proteins. We analyzed nsSNPs in protein coding sequences in the prioritized QTL regions using an in-house developed tool for prediction of amino acid substitutions based on their physicochemical properties (PASE) and evolutionary conservation (Li et al., 2013).

The DAVID annotated gene list was then filtered to identify the most likely candidate genes for growth in each QTL region. This was done by selecting the genes that had been associated with any of the following growth-related keywords: growth, development, morphogenesis, formation, proliferation, differentiation, regeneration, mineralization, elongation, biosynthetic, biogenesis, and organization. This set of terms was selected arbitrarily from ontology literature. The whole annotated gene list description was also reviewed to ensure no obvious candidates for growth were omitted.

## RESULTS

In an earlier study, Besnier et al. (2011) fine-mapped a number of QTL affecting body-weight at 8 weeks of age (**Table 1; Figures 2A–E**). The evaluated QTL regions are located on chicken chromosomes 1–5, 7, and 20 and cover in total 121.4 Mbp of the genome.

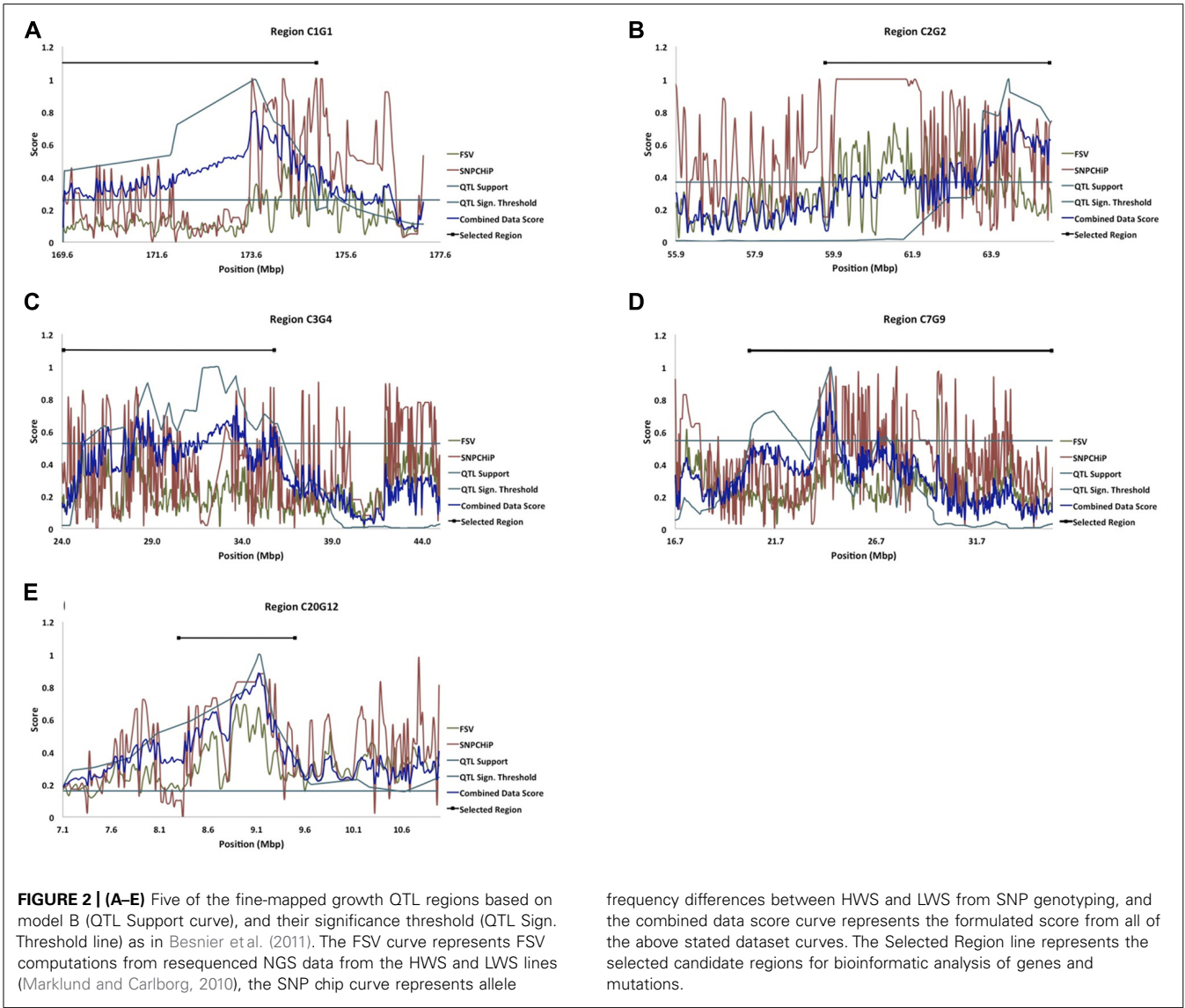
Using the prioritizations strategy described above, 44.7 Mbp of these original QTL were selected using the combined information from the QTL analysis and estimates of differences in allele frequencies between the lines inferred from SNP chip genotyping and FSV computation (**Table 2**).

**Table 1 | Fine-mapped growth QTL regions with significance according to Besnier et al. (2011).**

GGA <sup>1</sup>	QTL <sup>2</sup>	Region name	Start (Mbp <sup>3</sup> )	End (Mbp <sup>3</sup> )	Size (Mbp)
1	Growth1	C1G1	169.6	181.0	11.4
2	Growth2	C2G2	47.9	65.4	17.5
3	Growth4	C3G4	24.0	68.0	43.9
4	Growth6	C4G6	1.3	13.5	12.1
5	Growth8	C5G8	33.6	39.0	5.3
7	Growth9	C7G9	10.9	35.4	24.5
20	Growth12	C20G12	7.1	13.8	6.7
Total					121.4

<sup>1</sup> GGA: *Gallus Gallus* Autosome; <sup>2</sup> QTL names as in Besnier et al. (2011);

<sup>3</sup> Coordinates based on the Chicken (*Gallus gallus*) assembly v 2.1/galGal3



**Table 2 | Candidate regions selected based on QTL data and allele frequency differences between the lines inferred from SNP chip genotyping and FSV computation from resequencing. The selected percentages of the QTL regions significant with model B, are given (Besnier et al., 2011).**

Region name	Start Mbp <sup>1</sup>	End Mbp <sup>1</sup>	Size (Mbp)	QTL support <sup>2</sup>	Ensembl genes <sup>3</sup>
C1G1	169.6	175.0	5.4	5.4	97
C2G2	59.7	65.4	5.7	2.1	52
C3G4	24.1	35.8	11.7	10.3	142
C4G6	10.6	12.9	2.3	0.0	62
C5G8	34.2	36.8	2.6	0.0	20
C5G8	38.2	39.0	0.8	0.0	16
C7G9	20.4	35.4	15.0	4.3	209
C20G12	8.3	9.5	1.2	1.2	38
Total			44.7	23.3	636

<sup>1</sup> Coordinates based on the Chicken (*Gallus gallus*) assembly v 2.1/galGal3; <sup>2</sup>Size of the selected regions significant with QTL model B (Besnier et al., 2011); <sup>3</sup>Number of Ensembl genes in the initial list in the selected regions



**Table 3 | The variant effect predictor summary of SNPs in selected candidate segments of the QTL regions (according to Table 2).**

Location within gene	Region							Total
	C1G1	C2G2	C3G4	C4G6	C5G8	C7G9	C20G12	
3Prime UTR	200	93	200	153	73	348	75	1142
3Prime UTR, Splice site		1			1			2
5Prime UTR	22	9	44	20	3	50	28	176
5Prime UTR, Splice site				1		4		5
Coding unknown			1			1		2
Downstream	6118	2636	5318	2373	1395	7930	1384	27154
Essential splice site	2	3	6	1	1	4	3	20
Non-synonymous coding	215	82	255	92	60	470	80	1254
Non-synonymous coding, Splice site	6	4	8	5	3	17	1	44
splice site, Intronic	78	37	133	33	24	191	31	527
Stop gained	5		7	3	2	10		27
Stop gained, Non-synonymous coding	1							1
Synonymous coding	350	208	543	165	99	1113	159	2637
Synonymous coding, splice site	9	9	12	5	6	20	12	73
Upstream	5506	2626	5755	2570	1200	8312	1479	27488
Within mature miRNA						1		1
Within non-coding gene	4		2	1	3	12	4	26
Within non-coding gene, splice site						1		1
Total	12516	5708	12284	5422	2870	18484	3256	60540

In **Table 3**, we provide a summary of the results obtained using the Ensembl VEP tool. Nearly 61,000 SNPs (excluding intergenic and intronic SNPs) were found to be located within functional elements across the selected candidate segments in this analysis. In **Table 4**, we provide a selection of one or two of the best candidate mutations in each region.

## DISCUSSION

In this study we have developed and applied a bioinformatic strategy to search for candidate mutations affecting body weight at 56 days in several QTL regions that were previously identified and fine-mapped in an intercross between two divergently selected chicken lines. Given the 40 generations of divergent selection for body weight it is reasonable to assume that many of the underlying functional mutations will display a relatively large allele frequency difference, or complete fixation, between the lines. This assumption is supported by earlier work with the lines that many regions across the genome have been driven to fixation for alternative alleles in the lines and that most selection has been on standing genetic variation present in the common base-population at the onset of selection (Johansson et al., 2010). At a smaller number of selected loci mutations might have arisen after the initiation of selection. It is, however, unlikely that the QTL evaluated here are due to such new mutations as they are identified using a statistical analysis that assumes that the crossed lines are fixed for alternative QTL alleles.

To narrow down the target regions and identify the most plausible mutations, we used several independent sources of information. First, measurements of the genetic divergence between the founder lines of the intercross were used as indicators of regions that have been under strongest selection. Both individual SNP chip genotyping and genome resequencing of pools of individuals were used to provide stability and high-resolution in the estimates of the allele frequency difference between the lines.

The potential functional impact of genes and SNPs located within the target regions was bioinformatically evaluated to identify a set of candidate mutations to be further tested and evaluated in functional studies. In regions where there exist several possible candidate genes, our use of a combined and objective selection criteria helped to localize the most promising candidate genes and mutations. The genes and mutations listed in **Table 4** qualified as the strongest candidates underlying the observed QTL. Among these, the glucagon (GCG) gene on chromosome 7 (C7G9 region) is perhaps the most obvious candidate gene due to its well-documented effect on appetite (Suzuki et al., 2010), a trait for which the HWS and LWS lines show an extreme difference. No non-synonymous mutations were found in the glucagon gene, but a mutation was identified in a downstream CpG island with a large (0.87) estimated allele frequency difference between the lines (AFD), and possibly a regulatory effect on glucagon gene expression. The C7G9 region also included mutations in CpG islands with even larger AFD estimates and possibly regulatory roles in genes that in turn can regulate other genes with effects

Table 4 | Candidate mutations identified in the evaluated QTL regions.

Region	SNP (bp) <sup>1</sup>	Gene	SNP location <sup>2</sup>	No of AA <sup>3</sup> reads; depth coverage	Qual <sup>4</sup>	AFD <sup>5</sup>	PC Score <sup>6</sup>	EC Score <sup>7</sup>	PE Score <sup>8</sup>
C1G1	174634021	Asparagine-linked glycosylation 11 homolog ( <i>ALG11</i> )	CpG island, upstream	7; 10	72	0.97	N/A	N/A	N/A
C2G2	63823523	Endothelin 1 ( <i>EDN1</i> )	CpG island, upstream	3; 13	53	0.95	N/A	N/A	N/A
C3G4	33678270	Cysteine rich transmembrane BMP regulator 1 ( <i>CRIM1</i> )	Protein code, NS K/I	10; 19	182	0.97	0.67	0.63	0.42
C4G6	12044024	Similar to receptor tyrosine kinase ( <i>VEGFR-2</i> )	CpG island, upstream	4; 8	82	0.97	N/A	N/A	N/A
C4G6	12902414	Fibroblast growth factor 16 ( <i>FGF16</i> )	CpG island, downstream	8; 16	175	0.95	N/A	N/A	N/A
C5G8	38316301	Sorting nexin 6 ( <i>SNX6</i> )	CpG island, upstream	8; 14	142	0.97	N/A	N/A	N/A
C7G9	21686625	Growth factor receptor-bound protein 14 ( <i>GRB14</i> )	CpG island, downstream	3; 12	52	0.97	N/A	N/A	N/A
C7G9	22711910	Glucagon ( <i>GCG</i> )	CpG island, downstream	3; 9	46	0.87	N/A	N/A	N/A
C7G9	24802616	Insulin-like growth factor binding protein 2 ( <i>IGFBP2</i> )	Protein code, synonymous, CpG island	4; 8	69	0.95	N/A	N/A	N/A
C20G12	8715398	Baculoviral IAP repeat-containing 7 ( <i>BIRC7</i> )	Protein code, NS I/V	5; 8	65	0.97	0.29	0.14	0.04

<sup>1</sup> Coordinates based on the Chicken (*Gallus gallus*) assembly v 2.1/galGal3; <sup>2</sup> Location of the SNP in gene and also amino acid substitution in case of non-synonymous (NS) SNP; <sup>3</sup> Total number of reads in both lines representing the alternate allele (AA) versus the total depth coverage across the SNP position; <sup>4</sup> The Phred scaled probability that a REF/ALT polymorphism exists at this site given sequencing data. Because the Phred scale is  $-10 * \log(1 - p)$ , a value of 10 indicates a 1 in 10 chance of error, while a 100 indicates a 1 in 10<sup>10</sup> chance; <sup>5</sup> Allele frequency difference between the chicken lines as estimated using the GigaBayes software given that a total of 19 individuals from each line were included in the pools; <sup>6</sup> Physico-chemical score of amino acid substitution calculated using PASE (Li et al., 2013); <sup>7</sup> Evolutionary conservation score of amino acid substitution calculated using PASE (Li et al., 2013); <sup>8</sup> Combined score of PC and EC of amino acid substitution calculated using PASE (Li et al., 2013).

on body weight. Such mutations were found in the insulin-like growth factor binding protein 2 (*IGFBP2*) and the growth factor receptor-bound protein 14 (*GRB14*; e.g., Holt and Siddle, 2005) genes. The *IGFBP5* gene is also located in this target region but at this stage we have not found sufficient support for any strong candidate mutation in that gene. The IGF binding proteins can specify the actions of insulin-like growth factors which have key roles in vertebrate growth and development (e.g., Wood et al., 2005). Interestingly, the possibly regulatory *IGFBP2* mutation reported here is located in a coding sequence that is a part of a CpG island. Even though it is a synonymous mutation it may affect the *IGFBP2* expression through mechanisms of codon usage, GC content and/or mRNA stability and folding (reviewed by Shabalina et al., 2013). Overexpression of *IGFBP2* has been shown to reduce postnatal body weight gain in transgenic mice (Hoeftlich et al., 1999). The *GRB14* gene encodes a cellular adapter protein that can bind to receptor tyrosine kinases and intracellular proteins and thereby be involved in various processes. For example, it can bind and modify the signals from the insulin receptor and insulin-like growth factor 1 and its implication in growth regulation has been shown (reviewed by Holt and Siddle, 2005).

Strong candidate genes and mutations were also found in QTL regions on chromosome 3 (C3G4) and 4 (C4G6). In the C3G4 region, the gene encoding the cysteine rich transmembrane BMP regulator 1 (*CRIM1*), showed a non-synonymous mutations with large allele frequency difference between the lines and high PE scores (i.e., combined PC and EC scores; **Table 4**) with the PASE tool. *CRIM1* interactions with growth factors may be important for the development of the central nervous system (CNS) and other organs (Kolle et al., 2000). Perhaps most interesting is the impact the *CRIM1* gene possibly has on the CNS because Ka et al. (2009) reported genes that regulate neuronal plasticity to be differentially expressed between the HWS and LWS lines in the brainstem and hypothalamus. Moreover, electrolytic hypothalamus lesions has been shown to increase appetite in the LWS but not in the HWS line which further supports that CNS is highly involved in the differences between these chicken lines (Burkhart et al., 1983).

In the C4G6 region, candidate CpG island mutations were identified within the fibroblast growth factor 16 (*FGF16*) and vascular endothelial growth factor receptor 2 (*VEGFR-2*) genes. *FGF16* is known to be involved in embryonic development and cell growth (Antoine et al., 2006) whereas the *VEGFR-2* gene has been reported to be of importance for angiogenesis (Patterson et al., 1995).

In the chromosome 1 QTL region (C1G1) we also found a candidate mutation, possibly regulatory, in the asparagine-linked glycosylation 11 homolog (*ALG11*) gene. *ALG11* has been reported to be involved in biosynthetic processes and required for normal growth in yeast (Cipollo et al., 2001).

The chromosome 2 QTL region (C2G2) showed CpG island mutations at the endothelin 1 (*EDN1*) gene with the two chicken lines fixed for opposite alleles. *EDN1* is known for roles in regulation of blood pressure and development (Kurihara et al., 1994).

In the regions on chromosome 5 (C5G8) and 20 (C20G12) the genes found in the analysis were less obvious candidates. However,

such genes may still have key roles in processes with complex and indirect effects on growth-related traits. Keeping this in mind, we consider mutations identified in the sorting nexin 6 (*SNX6*; Caldwell et al., 2005; C5G8 region) and baculoviral IAP repeat-containing 7 (*BIRC7*; Kasof and Gomes, 2001; C20G12 region) genes are of most interest to investigate further.

In conclusion, the described combination of data from QTL mapping, next-generation sequencing, SNP chip genotyping and bioinformatic analysis has provided a list of plausible candidate genes and mutations that will facilitate further verification and experimental evaluation. The support for this list from different types of data and analysis enhances the probability that the selected genes and mutations underlying the QTL effects are an unbiased selection of genes and that the contributing gene(s) are included in the set. Further studies based on this list may therefore reveal mutations which underlie the observed QTL effects and can increase our understanding of growth regulation as well as be more emphasized in animal breeding programs with genomic selection.

## AUTHORS CONTRIBUTIONS

Muhammad Ahsan and Xidan Li carried out the region-targeted computation and analysis using the different sources of data and took part in the planning of the study. Marcin Kierczak and Andreas E. Lundberg performed the assembly of the SOLID resequencing datasets. Stefan Marklund initiated and planned the study. Paul B. Siegel and Örjan Carlborg contributed with comments and advice. Muhammad Ahsan and Stefan Marklund drafted the manuscript and all co-authors contributed to the final version.

## ACKNOWLEDGMENTS

We would like to thank the USDA Chicken GWMAS Consortium, Cobb Vantress, and Hendrix Genetics for access to the developed 60K SNP Illumina iSelect chicken array, DNA landmarks for 60K array genotyping and the Uppsala Genome Center for ABI SOLID sequencing.

This work was financially supported by a EURI award to Örjan Carlborg and a Future research leader grant to Örjan Carlborg from the Swedish Foundation for Strategic Research. The contribution of Muhammad Ahsan was supported by his scholarship from the Higher Education Commission of Pakistan (HEC).

## REFERENCES

- Antoine, M., Wirz, W., Tag, C. G., Gressner, A. M., Wycislo, M., Muller, R., et al. (2006). Fibroblast growth factor 16 and 18 are expressed in human cardiovascular tissues and induce on endothelial cells migration but not proliferation. *Biochem. Biophys. Res. Commun.* 346, 224–233. doi: 10.1016/j.bbrc.2006.05.105
- Besnier, F., Wahlberg, P., Rönnegård, L., Weronica, E. K., Andersson, L., Siegel, P. B., et al. (2011). Fine mapping and replication of QTL in outbred chicken advanced intercross lines. *Genet. Sel. Evol.* 43, 3. doi: 10.1186/1297-9686-43-3
- Burkhart, C. A., Cherry, J. A., Van Krey H. P., and Siegel, P. B. (1983). Genetic selection for growth rate alters hypothalamic satiety mechanisms in chickens. *Behav. Genet.* 13, 295–300. doi: 10.1007/BF01071874
- Caldwell, R. B., Kierzek, A. M., Arakawa, H., Bezzubov, Y., Zaim, J., Fiedler, P., et al. (2005). Full-length cDNAs from chicken bursal lymphocytes to facilitate gene function analysis. *Genome Biol.* 6, R6. doi: 10.1186/gb-2004-6-1-r6
- Cipollo, J. E., Trimble, R. B., Chi, J. H., Yan, Q., and Dean, N. (2001). The yeast *ALG11* gene specifies addition of the terminal alpha 1,2-Man to the Man(5)GlcNAc(2)-PP-dolichol N-glycosylation intermediate formed on the

- cytosolic side of the endoplasmic reticulum. *J. Biol. Chem.* 276, 21828–21840. doi: 10.1074/jbc.M010896200
- Dunnington, E. A., and Siegel, P. B. (1996). Long-term divergent selection for eight-week body weight in White Plymouth Rock chickens. *Poult. Sci.* 75, 1168–1179. doi: 10.3382/ps.0751168
- Dunnington, E. A., Honaker, C. F., McGilliard, M. L., and Siegel, P. B. (2013). Phenotypic responses of chickens to long-term, bidirectional selection for juvenile body weight – historical perspective. *Poult. Sci.* 92, 1724–1734. doi: 10.3382/ps.2013-03069
- Flicek, P., Amode, M. R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., et al. (2012). Ensembl 2012. *Nucleic Acids Res.* 40, D84–D90. doi: 10.1093/nar/gkr991
- Hoeflich, A., Wu, M., Mohan, S., Foll, J., Wanke, R., Froehlich, T., et al. (1999). Overexpression of insulin-like growth factor-binding protein-2 in transgenic mice reduces postnatal body weight gain. *Endocrinology* 140, 5488–5496. doi: 10.1210/en.140.12.5488
- Holt, L. J., and Siddle, K. (2005). Grb10 and Grb14: enigmatic regulators of insulin action – and more? *Biochem. J.* 388, 393–406. doi: 10.1042/BJ20050216
- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13. doi: 10.1093/nar/gkn923
- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211
- Jacobsson, L., Park, H. B., Wahlberg, P., Fredriksson, R., Perez-Enciso, M., Siegel, P. B., 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
- Johansson, A. M., Pettersson, M. E., Siegel, P. B., and Carlborg, Ö. (2010). Genome-Wide Effects of Long-Term Divergent Selection. *PLoS Genet.* 6:e1001188. doi: 10.1371/journal.pgen.1001188
- Ka, S., Lindberg, J., Strömstedt, L., Fitzsimmons, C., Lindqvist, N., Lundberg, J., et al. (2009). Extremely different behaviours in high and low body weight lines of chicken are associated with differential expression of genes involved in neuronal plasticity. *J. Neuroendocrinol.* 21, 208–216. doi: 10.1111/j.1365-2826.2009.01819.x
- Kasof, G. M., and Gomes, B. C. (2001). Livin, a novel inhibitor of apoptosis protein family member. *J. Biol. Chem.* 276, 3238–3246. doi: 10.1074/jbc.M003670200
- Kolle, G., Georgas, K., Holmes, G. P., Little, M. H., and Yamada, T. (2000). CRIM1, a novel gene encoding a cysteine-rich repeat protein, is developmentally regulated and implicated in vertebrate CNS development and organogenesis. *Mech. Dev.* 90, 181–193. doi: 10.1016/S0925-4773(99)00248-8
- Kurihara, Y., Kurihara, H., Suzuki, H., Kodama, T., Maemura, K., Nagai, R., et al. (1994). Elevated blood pressure and craniofacial abnormalities in mice deficient in endothelin-1. *Nature* 368, 703–710. doi: 10.1038/368703a0
- Lee, W. P., Stromberg, M., Ward, A., Stewart, C., Garrison, E., and Marth, G. T. (2013). MOSAIK: a hash-based algorithm for accurate next-generation sequencing read mapping. arXiv preprint arXiv: 1309.1149.
- Li, X., Kierczak, M., Shen, X., Ahsan, M., Carlborg, Ö., and Marklund, S. (2013). PASE: a novel method for functional prediction of amino acid substitutions based on physicochemical properties. *Front. Genet.* 4:21. doi: 10.3389/fgene.2013.00021
- Marklund, S., and Carlborg, Ö. (2010). SNP detection and prediction of variability between chicken lines using genome resequencing of DNA pools. *BMC Genomics* 11, 655. doi: 10.1186/1471-2164-11-665
- Marquez, G. C., Siegel, P. B., and Lewis, R. M. (2010). Genetic diversity and population structure in lines of chickens divergently selected for high and low 8-week body weight. *Poult. Sci.* 89, 2580–2588. doi: 10.3382/ps.2010-01034
- Marth, G. T., Korf, I., Yandell, M. D., Yeh, R. T., Gu, Z. J., Zakeri, H., et al. (1999). A general approach to single-nucleotide polymorphism discovery. *Nat. Genet.* 23, 452–456. doi: 10.1038/70570
- McLaren, W., Pritchard, B., Rios, D., Chen, Y. A., Flicek, P., and Cunningham, F. (2010). Deriving the consequences of genomic variants with the Ensembl API and SNP effect predictor. *Bioinformatics* 26, 2069–2070. doi: 10.1093/bioinformatics/btq330
- Patterson, C., Perrella, M. A., Hsieh, C. M., Yoshizumi, M., Lee, M. E., and Haber, E. (1995). Cloning and functional analysis of the promoter for KDR/flk-1, a receptor for vascular endothelial growth factor. *J. Biol. Chem.* 270, 23111–23118. doi: 10.1074/jbc.270.39.23111
- Rönnegård, L., Besnier, F., and Carlborg, Ö. (2008). An improved method for quantitative trait loci detection and identification of within-line segregation in F-2 intercross designs. *Genetics* 178, 2315–2326. doi: 10.1534/genetics.107.083162
- Rubin, C. J., Zody, M. C., Eriksson, J., Meadows, J. R. S., Sherwood, E., Webster, M. T., et al. (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* 464, 587–591. doi: 10.1038/nature08832
- Shabalina, S. A., Spiridonov, N. A., and Kashina, A. (2013). Sounds of silence: synonymous nucleotides as a key to biological regulation and complexity. *Nucleic Acids Res.* 41, 2073–2094. doi: 10.1093/nar/gks1205
- Suzuki, K., Simpson, K. A., Minnion, J. S., Shillito, J. C., and Bloom, S. R. (2010). The role of gut hormones and the hypothalamus in appetite regulation. *Endocr. J.* 57, 359–372. doi: 10.1507/endocrj.K10E-077
- Wood, A. W., Duan, C., and Bern, H. A. (2005). Insulin-like growth factor signaling in fish. *Int. Rev. Cytol.* 243, 215–285. doi: 10.1016/S0074-7696(05)43004-1

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

Received: 17 July 2013; paper pending published: 15 August 2013; accepted: 17 October 2013; published online: 05 November 2013.

Citation: Ahsan M, Li X, Lundberg AE, Kierczak M, Siegel PB, Carlborg Ö and Marklund S (2013) Identification of candidate genes and mutations in QTL regions for chicken growth using bioinformatic analysis of NGS and SNP-chip data. *Front. Genet.* 4:226. doi: 10.3389/fgene.2013.00226

This article was submitted to *Bioinformatics and Computational Biology*, a section of the journal *Frontiers in Genetics*.

Copyright © 2013 Ahsan, Li, Lundberg, Kierczak, Siegel, Carlborg and Marklund. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.