Edinburgh Research Explorer

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

OPEN ACCESS

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

Eirini Tarsani*, **Georgios Theodorou, Irida Palamidi, Antonios Kominakis**

Department of Animal Science, Agricultural University of Athens, Athens, Greece

Email address:
etarsani@aua.gr (E. Tarsani)
*Corresponding author

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. 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 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.

2. Material and Methods

2.1. 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).

2.2. 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).

2.3. 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 *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), β is the 55 x 1 vector of corresponding coefficients of fixed effects, *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), α 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, *u* is the vector of random polygenic effects, and *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.

2.4. 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].

2.5. Proportion of Variance Explained (PVE) 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.

2.6. Estimation of Allelic Effects Based on PVE

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

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

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

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

2.7. 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).

2.8. 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 $\geq$ 2.0% were considered putative evidence of large genetic effects [24] while those with 2.0<PVE $\geq$ 1.0% were considered evidence for moderate genetic effects [24, 25].

2.9. 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.

3. Results

3.1. 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

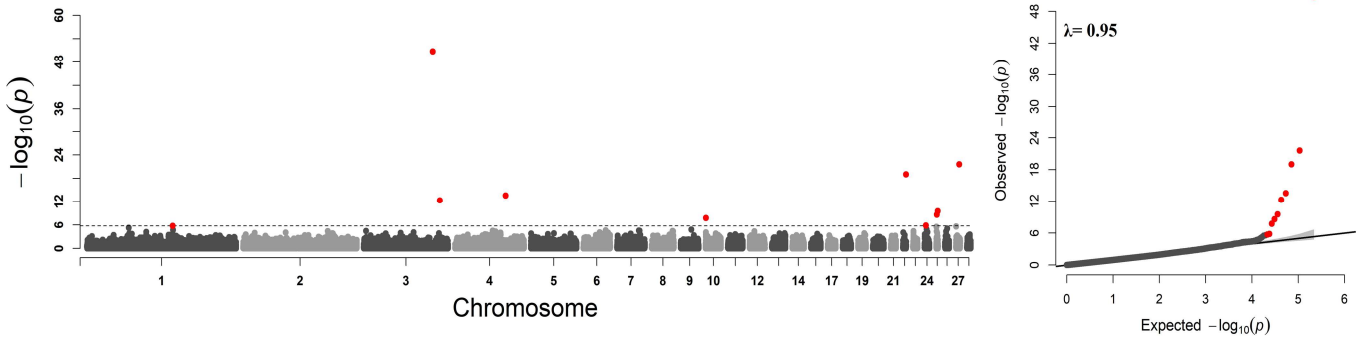with highest p-values on the log$_{10}$ scale (Table 1).

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. Red 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/).*

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	-log$_{10}$ (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

3.2. Detection Efficacy Across the Various Sampling Scenarios and Estimated PVE
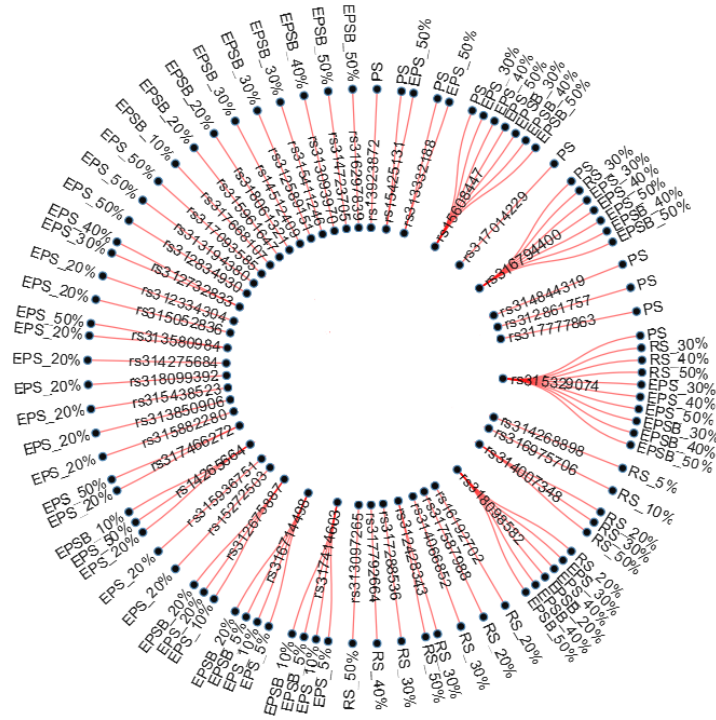

Figure 2. *Radial network of the genome-wide significant SNPs detected across the various sampling scenarios (PS: whole population sampling, RS: random sampling, EPS: extreme phenotype sampling and EPSB: extreme phenotypes sampling binary case). Figure was constructed using the data. tree and networkD3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

Figure 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).

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	-$\log_{10}$ (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
	rs314007348	4	66,225,823	2.64E-08	7.5777	0.0019	2.1494	0.447
20	*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
	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
30	*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
	rs313097265	2	93,140,828	7.79E-42	41.1081	1.68E-36	2.7875	0.101
50	*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

EPS								
Sample proportion (%)	SNP ID	GGA	Position (bp)[a]	p-value	-log$_{10}$ (p-value)	FDR p-value	PVE (%)	MAF
	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
40	*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
	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
50	*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
	rs14265664	2	138,095,717	4.08E-08	7.3887	0.00294	7.1316	0.070
10	*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
	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
20	*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
	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
30	*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
	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
40	*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
	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
50	*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
	rs314007348	4	66,225,823	234,093
30	*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
	rs15608447	4	66,459,916	0
30	*rs316794400*	22	5,149,585	0
	rs315329074	27	6,920,352	0
	rs15608447	4	66,459,916	0
40	*rs316794400*	22	5,149,585	0
	rs315329074	27	6,920,352	0
	rs15425131	3	90,795,168	0
	rs313332188	3	99,991,484	0
	rs15608447	4	66,459,916	0
50	*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

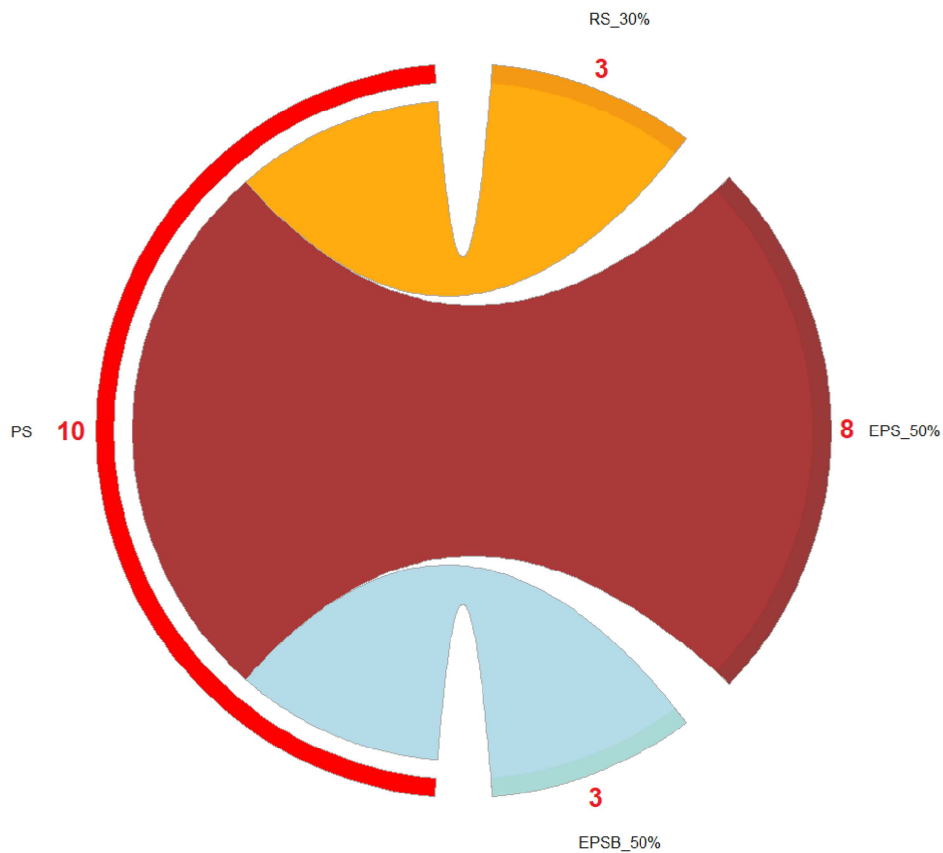| Extreme phenotypes binary case				
Sample proportion (%)	**SNP ID**	**GGA**	**Position (bp)**	**Distance from PS SNP (bp)**
30	*rs15608447*	4	66,459,916	0
	rs315329074	27	6,920,352	0
40	*rs15608447*	4	66,459,916	0
	rs316794400	22	5,149,585	0
	rs315329074	27	6,920,352	0
50	*rs15608447*	4	66,459,916	0
	rs316794400	22	5,149,585	0
	rs315329074	27	6,920,352	0

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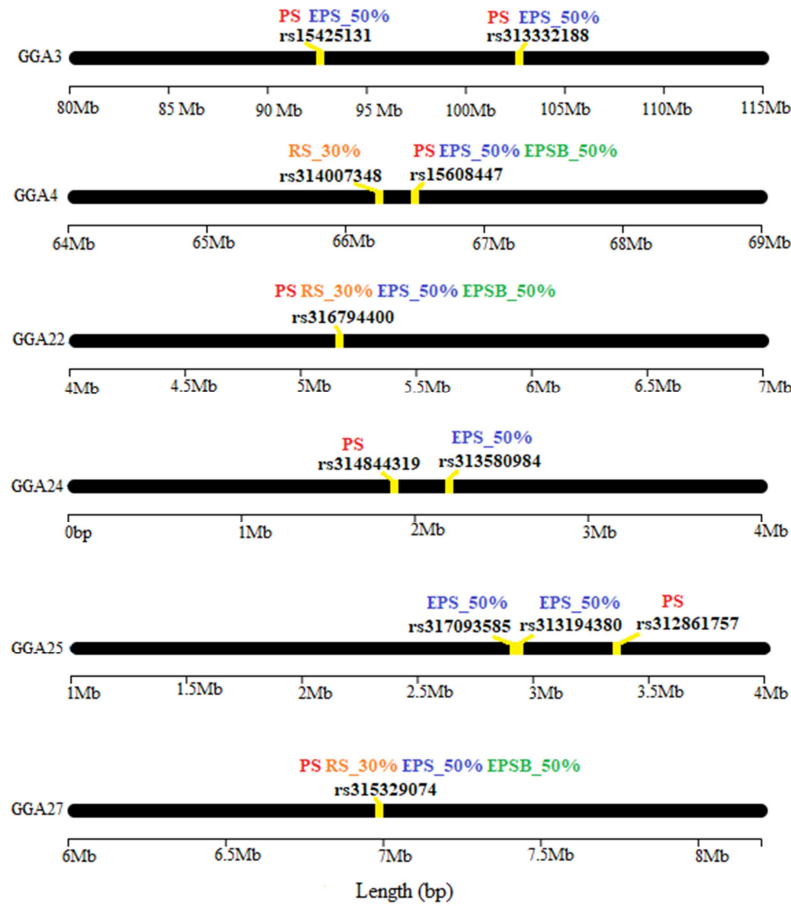
Figure 4. *Positions of SNPs detected across sampling scenarios (RS: random sampling, EPS: extreme phenotype sampling and EPSB: extreme phenotypes sampling binary case) in relation to position of SNPs in whole population sampling (PS) on the same autosomes. SNP positions denoted by yellow color. Figure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

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.

3.3. 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 ≥ 2.0% (Table 4) while three markers (*rs14265664*, *rs316794400 and rs15608447*) had PVE<2.0% and PVE ≥ 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).

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* *LOC112531900*	Wattles weight (127117, Animal QTLdb)
rs316794400	22	5,149,585	A	Upstream gene variant Downstream gene variant Intron variant	*SLC20A1* *ANXA4* *SLAIN2*	Breast muscle percentage (95429, Animal QTLdb), Abdominal fat weight (96666, Animal QTLdb)
rs15608447	4	66,459,916	G	Downstream gene variant, intron variant, Non coding transcript variant, Upstream gene variant	*LOC107053243* *LOC112532289*	-

4. Discussion

4.1. 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.

4.2. 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 $\beta=2.9$ g was consistently detected across almost all sampling cases, while *rs15425131* (*MYOM2* gene) with lower MAF=0.091 and highest effect size ($\beta=3.9$ 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].

5. 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.

Conflict of Interest

The authors declared no conflict of interest.

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/173513a8eb8a0adfb2cd59e90 248333ce4c9.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.