GSE | Genetics
Selection
Evolution

RESEARCH ARTICLE

CrossMark

Differential contribution of genomic regions to marked genetic variation and prediction of quantitative traits in broiler chickens

Rostam Abdollahi-Arpanahi[1,2*], Gota Morota[3], Bruno D. Valente[1,4], Andreas Kranis[5,6], Guilherme J. M. Rosa[1,7] and Daniel Gianola[1,4,7]

Abstract

Background: Genome-wide association studies in humans have found enrichment of trait-associated single nucleotide polymorphisms (SNPs) in coding regions of the genome and depletion of these in intergenic regions. However, a recent release of the ENCyclopedia of DNA elements showed that ~80 % of the human genome has a biochemical function. Similar studies on the chicken genome are lacking, thus assessing the relative contribution of its genic and non-genic regions to variation is relevant for biological studies and genetic improvement of chicken populations.

Methods: A dataset including 1351 birds that were genotyped with the 600K Affymetrix platform was used. We partitioned SNPs according to genome annotation data into six classes to characterize the relative contribution of genic and non-genic regions to genetic variation as well as their predictive power using all available quality-filtered SNPs. Target traits were body weight, ultrasound measurement of breast muscle and hen house egg production in broiler chickens. Six genomic regions were considered: intergenic regions, introns, missense, synonymous, 5′ and 3′ untranslated regions, and regions that are located 5 kb upstream and downstream of coding genes. Genomic relationship matrices were constructed for each genomic region and fitted in the models, separately or simultaneously. Kernel-based ridge regression was used to estimate variance components and assess predictive ability. Contribution of each class of genomic regions to dominance variance was also considered.

Results: Variance component estimates indicated that all genomic regions contributed to marked additive genetic variation and that the class of synonymous regions tended to have the greatest contribution. The marked dominance genetic variation explained by each class of genomic regions was similar and negligible (~0.05). In terms of prediction mean-square error, the whole-genome approach showed the best predictive ability.

Conclusions: All genic and non-genic regions contributed to phenotypic variation for the three traits studied. Overall, the contribution of additive genetic variance to the total genetic variance was much greater than that of dominance variance. Our results show that all genomic regions are important for the prediction of the targeted traits, and the whole-genome approach was reaffirmed as the best tool for genome-enabled prediction of quantitative traits.

Background

To date, analysis of pathways and post-genome-wide association studies (GWAS) have focused on genic

*Correspondence: r.abdollahi@ut.ac.ir
[1] Department of Animal Sciences, University of Wisconsin, Madison, WI, USA
Full list of author information is available at the end of the article

regions of the genome as evidenced by the emergence of exome sequencing. Exons are functional sequences of the genome which, taken together, represent an important part of the genome that is actually translated into protein. Moreover, genotyping exons is less expensive than whole-genome sequencing. However, a recent release of the ENCyclopedia of DNA Elements (ENCODE) showed

Abdollahi-Arpanahi *et al. Genet Sel Evol* (2016) 48:10

Page 2 of 13

that about 62 % of the genome is transcribed into RNA, which added to the evidence that has accumulated on transcription-factor-binding sites, chromatin structure, DNA methylation, histone modification and other regulatory regions, indicates that about 80 % of a genome has a biochemical function [1]. Nonetheless, DNA sequences in intergenic regions are considered as "dark matter" or "dark matter transcripts" [2] since their role is still ambiguous. Recent research has shown that 43 % of the regions that are detected in GWAS point to intergenic regions (outside of the promoter and transcribed regions), and 45 % to introns [3]. Nevertheless, missense codons and promoter regions are significantly enriched for trait-associated single nucleotide polymorphisms (SNPs), while intergenic regions are significantly underrepresented [3, 4].

On the one hand, most GWAS have used very stringent significance thresholds to avoid false positives due to multiple-testing and, as a result, many variants with small effects have been missed. These also include rare variants that have large effects but explain a small proportion of the variance [5]. On the other hand, in whole-genome prediction, the prediction of genetic merit of individuals is based on the effect of all variants estimated simultaneously. Such an approach does not suffer from multiple-testing, stringent significance thresholds and unrealistic assumptions like linkage equilibrium (LE) between markers, since linkage disequilibrium (LD) is pervasive, especially for agricultural species.

The contribution of genic and non-genic regions of the genome to additive genetic variance has been investigated in humans [6–8], dairy and beef cattle [9] and plants [10]. There is, however, some disagreement between the findings from these studies. For instance, Yang et al. [8] stated that genic regions contributed more additive genetic variation than non-genic regions. Koufariotis et al. [9] also pointed out that the classes of missense and synonymous genomic regions explained most of the additive genetic variation. On the contrary, Gusev et al. [7] reported that DNaseI hypersensitivity sites explained most of the additive genetic variation for 11 common diseases. However, a study by Do et al. [11] on feed intake and its component traits in pigs indicated that the contribution of each SNP to total genomic variance was similar for genic and non-genic regions.

Morota et al. [12] studied the predictive ability of various genomic regions for three chicken broiler traits. They found that the enrichment or depletion of genomic regions in terms of predictive ability was trait-dependent and that the whole-genome approach had the best predictive power regardless of trait. Erbe et al. [13] compared the predictive ability of SNPs in transcribed regions with that of SNPs in intergenic regions and found

that the transcribed part of the genome of dairy cattle performed better, with a 0.03 increase in predictive correlation for Jersey cattle traits. However, these studies fitted genic and non-genic regions using a single kernel (single genomic relationships matrix) approach. An alternative approach would be to use multiple kernels (multiple genomic relationship matrices) that are tailored to each specific genomic region.

Partitioning genetic variation into marked additive and dominance components has been explored [14–18] but knowledge on the contribution of different genomic regions to non-additive genetic variation is lacking. Partitioning the genome into classes of SNPs allows one to target genomic regions of interest.

While Morota et al. [12] used the same dataset to evaluate the predictive performance of different genomic regions, some limitations of their study led us to conduct additional research, i.e.: (1) they used a single-kernel approach, while the multiple-kernel approach is more flexible for prediction and, thus, can be used to improve accuracy of prediction and to decrease learning complexity and training time e.g., [19]; (2) the non-additive contribution of different genomic regions was not considered; (3) variance components estimates were not studied; and (4) overlapping genic regions complicated the interpretation of results, thus partitioning the genome into distinct segments may produce a clearer picture. Therefore, the aim of our study was to investigate the relative contribution of genic and non-genic regions to marked additive and dominance genetic variation for body weight (BW), breast muscle (BM) and hen house egg production (HHP) in broiler chickens. We also evaluated the predictive ability of different genomic regions for yet-to-be observed phenotypes. To our knowledge, this is the first study that addresses the contribution of genomic regions to marked non-additive genetic variation by fitting one class of genomic regions at a time.

Methods
Data
A total of 1351 birds from a commercial broiler chicken line were provided by Aviagen Ltd. This broiler line has undergone several generations of selection using genetic evaluations based on multiple-trait pedigree best linear unbiased prediction (P-BLUP). The following traits were studied: BW, ultrasound of BM at 35 days of age, and HHP defined as the total number of eggs laid between weeks 28 and 54 per bird. Phenotype records for BW and BM were pre-corrected for a combined effect of sex (525 males and 826 females), hatch week, contemporary group of parents and pen in the growing farm, whereas phenotype records for HHP were adjusted for hatch effects. Phenotypic records were merged with SNP

Abdollahi-Arpanahi *et al. Genet Sel Evol* (2016) 48:10

Page 3 of 13

genotype records on individuals that were genotyped using the Affymetrix 600K chip. More details on genotyping and phenotypic data are in Kranis et al. [20] and Abdollahi-Arpanahi et al. [21]. In total, there were 1346, 1331, and 819 individuals scored for BW, BM and HHP, respectively.

A total of 580,954 SNP genotypes were originally available in the dataset. SNPs that departed from Hardy–Weinberg equilibrium ($p < 10^{-6}$) based on a Chi square test, SNPs that had a minor allele frequency (MAF) <0.01, or a missing rate >0.05 were excluded from the analysis. Missing SNPs were imputed using Beagle [22]. After editing, 354,364 SNPs remained for the analysis. Mean MAF was equal to 0.27. Only SNPs on 28 autosomes (GGA1–28, GGA for *Gallus gallus* chromosome) were included, which covered 919 Mb of the *G. gallus* genome. Data editing was done with the PLINK software [23]. The following coding was used for SNP substitution effects in the additive genotype matrix ($\mathbf{X}_A$) i.e. 0 for "aa", 1 for "Aa", and 2 for "AA".

SNP annotation
Chromosome information and physical positions of SNPs were obtained using the annotation file downloaded from the Animal Genome Database (http://www.animalgenome.org/repository/chicken/). We mapped the Aviagen marker coordinates file to Gallus_gallus_4.0 Ensembl VEP tool (release 75) (ftp://ftp.ensembl.org/pub/release-75/variation/VEP/arrays/) and to the animal genome annotated file of chicken data. Each SNP was examined to determine if it was located in genic or non-genic regions. Six classes of disjoint genomic regions were formed, namely, introns, missense (non-synonymous), synonymous, 5′ and 3′ untranslated regions (UTR), regions that are located 5 kb upstream and downstream of gene ("up-down" class), and intergenic regions, which included SNPs that were not annotated as being in genic regions. Missense genomic regions are regions where a point mutation causes a single nucleotide change in a codon that codes for a different amino acid. Conversely, the annotation class of synonymous regions includes SNPs for which the substitution of one base for another in a coding region does not modify the resulting amino acid. Regulatory regions were defined as regions located 5 kb upstream and downstream of genes. In our study, intergenic regions consisted of SNPs without any assignment to the aforementioned annotation categories. Numbers of SNPs and their allelic frequency for each class of genomic regions are in Table 1. After quality control, 43,600 SNPs were not found in the Ensembl database or animal genome annotated database and were discarded from all further analyses. Finally, 310,764 SNPs were retained for the final analysis.

Table 1 Number of SNPs assigned to each genomic region and the corresponding mean and standard deviations (SD) of minor allelic frequencies (MAF)

Classes of genomic regions	Number of SNPs	Mean MAF	SD MAF
Intergenic	139,394	0.27	0.13
Intron	124,734	0.27	0.13
Missense	1658	0.27	0.13
Synonymous	5620	0.27	0.13
UTR	3044	0.28	0.13
Upsteam and downstream	36,314	0.27	0.13
All markers	310,764	0.27	0.13

Statistical models
To explore the variance and predictive ability of each of the six aforementioned classes of genomic regions, 12 additive and dominance parametric kernels were constructed. Kernel methods have been reviewed in Gianola et al. [24], Gianola and van Kaam [25], de los Campos et al. [26] and Morota and Gianola [27]. In order to capture signals from genotypes to phenotypes through the construction of kernel matrices $\mathbf{G}$ (additive genomic relationships) and $\mathbf{D}$ (dominance genomic relationships), three scenarios were considered: (1) fitting additive kernels ($\mathbf{G}$) for each genomic region or fitting one kernel ($\mathbf{G}$ matrix) for all SNPs without distinction of genomic regions; (2) fitting additive and dominance kernels ($\mathbf{G} + \mathbf{D}$) jointly for each genomic region; and (3) fitting six additive kernels jointly ($\mathbf{G}_1 + \mathbf{G}_2 + \cdots + \mathbf{G}_6$), with each kernel linked to one genomic region. All additive kernels are parametric (linear) kernels that were constructed following VanRaden's [28] genomic relationship matrices based on SNP information. While a joint analysis that involves 12 additive and dominance kernels is appealing, it was not considered here because in terms of convergence issue, it is too computationally demanding for at least two reasons: (1) small sample size; and (2) poor mixing since the additive and dominance kernels were not orthogonal to each other. Each of the statistical models is described below.

Scenario (1): Separate analysis with additive kernels (G)
In this analysis, we fitted models by accounting for the similarity relationship matrix (kernel) within each of the six classes of genomic regions. Phenotypes that were pre-corrected for systematic effects were analyzed trait by trait with the following model:

$$y_i = \mu + g(\mathbf{x}_i) + e_i, \tag{1}$$

where y_i is the pre-corrected phenotype on bird i; μ is the intercept; $g(\mathbf{x}_i)$ is a linear function of SNP genotypes $\mathbf{x}_i$, and e_i is the residual of the model for bird i. We assumed

Abdollahi-Arpanahi *et al. Genet Sel Evol* (2016) 48:10

Page 4 of 13

that the genetic signals (*g*) were represented by $\mathbf{G}\boldsymbol{\alpha}$, where $\mathbf{G}$ is an n × n kernel matrix indexed by the observed SNP covariates such that $\mathbf{G} \sim \mathbf{X}\mathbf{X}^{\mathrm{T}}$, where $\mathbf{X}$ is a SNP genotype matrix. $\mathbf{G}$ resulted from a centered and standardized $\mathbf{X}$ matrix, then divided by the number of SNPs, as proposed by VanRaden [28] and Yang et al. [29], and $\boldsymbol{\alpha}$ is the vector of the reproducing kernel Hilbert spaces (RKHS) regression coefficients that is estimated as the solution that minimizes:

$$l(\boldsymbol{\alpha}|\lambda) = (\boldsymbol{y} - \mathbf{G}\boldsymbol{\alpha})'(\boldsymbol{y} - \mathbf{G}\boldsymbol{\alpha}) + \lambda\boldsymbol{\alpha}'\mathbf{G}\boldsymbol{\alpha},$$

where $\boldsymbol{\alpha} \sim N(0, \mathbf{G}^{-1}\sigma_g^2)$, σ_g^2 is the additive variance captured by markers, and λ is a regularization factor. Since $g = \mathbf{G}\boldsymbol{\alpha}$, the hundreds of thousands of SNP predictors are featured into a number of functions that is equal to the number of observations, i.e. 1346, 1331 or 819 individuals for BW, BM, and HHP, respectively.

We can now rewrite Eq. (1) in matrix form as:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{G}\boldsymbol{\alpha} + \mathbf{e},$$

where $\mathbf{1}$ is a vector of ones with the appropriate dimension, and $\mathbf{e}$ is a vector of model residuals with $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$, where σ_e^2 is the residual variance. We assume that the variance–covariance structure of the above model was:

$$\mathbf{V} = \mathbf{G}\sigma_g^2 + \mathbf{I}\sigma_e^2.$$

The prediction of genetic values in a testing set ($\hat{\boldsymbol{g}}_{test}$) is given by:

$$\hat{\boldsymbol{g}}_{test} = \mathbf{G}_{test,strain}\mathbf{G}_{train}^{-1}\hat{\boldsymbol{g}}_{train},$$

where $\mathbf{G}_{test,strain}$ is a rectangular matrix of genomic relationships between training and testing individuals, which represents a subset of the total $\mathbf{G}$ constructed from all individuals in the training and testing sets, $\boldsymbol{G}_{train}$ is the genomic relationship between individuals in the training set, and $\hat{\boldsymbol{g}}_{train}$ is the vector of predicted genetic signals of individuals in the training set.

Scenario (2): Joint analysis of additive and dominance kernels (G + D) for each genomic region

The standard RKHS regression model can be represented in this case as:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{G}\boldsymbol{\alpha}_1 + \mathbf{D}\boldsymbol{\alpha}_2 + \mathrm{e}. \tag{2}$$

where $\boldsymbol{\alpha}_1 \sim N(0, \mathbf{G}^{-1}\sigma_g^2)$ and $\boldsymbol{\alpha}_2 \sim N(0, \mathbf{D}^{-1}\sigma_d^2)$ are unknown regression vectors in RKHS, where σ_d^2 is the marked dominance variance and $\mathbf{D}$ is the dominance relationship matrix. To build $\mathbf{D}$, we created an incidence matrix ($\mathbf{X_D}$) for effects due to dominance $\mathbf{X_D} = (x_{Dij})$. Elements of $\mathbf{X_D}$ are equal to $-2q^2$, $2pq$ and $-2p^2$ for genotypes aa, Aa and AA, respectively, where *p* is the frequency of A and $q = 1 - p$. Then, $\mathbf{X_D}\mathbf{X_D'}$ is standardized

at $4\sum_{i=1}^{m} p_i^2 q_i^2$ as in Vitezica et al. [30], where *m* is the number of SNPs. With this structure of $\mathbf{G}$ and $\mathbf{D}$, the variance–covariance structure was:

$$\mathbf{V} = \mathbf{G}\sigma_g^2 + \mathbf{D}\sigma_d^2 + \mathbf{I}\sigma_e^2.$$

The additive genetic values of individuals in the testing set were predicted with the equation in Scenario (1), and $\hat{\boldsymbol{d}}_{test} = \mathbf{D}_{test,train}\mathbf{D}_{train}^{-1}\hat{\boldsymbol{d}}_{train}$ was used for dominance values.

Scenario (3): Joint analysis of six additive kernels (G₁ + G₂ + ⋯ + G₆)

In the joint analysis of the six classes of genomic regions, we used the following statistical model:

$$\mathbf{y} = \mathbf{1}\mu + \sum_{t=1}^{h} \mathbf{G}_t\boldsymbol{\alpha}_t + \mathbf{e}, \tag{3}$$

where $\mathbf{y}$ is the vector of observations; $\mathbf{1}$ is a vector of ones; μ is an intercept; $h = 6$ is the number of genomic regions; $\boldsymbol{\alpha}_t \sim (0, \mathbf{G}_t^{-1}\sigma_{gt}^2)$ with $t = 1, 2, \ldots, 6$ is a regression vector in RKHS for genomic region t; $\mathbf{G}_t$ is a matrix of additive genomic relationships for the *t*th genomic region; $\sigma_{g_t}^2$ is the variance that is captured by SNPs in the *t*th genomic region; and $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$ is a vector of model residuals, where σ_e^2 is the residual variance, and $\mathbf{I}$ is an identity matrix. Therefore, the joint density of the six random vectors and of the residual term is:

$$p\left(\mathbf{e}, \mathbf{g}_1, \ldots, \mathbf{g}_t | \sigma_e^2, \sigma_{g_1}^2, \ldots, \sigma_{g_t}^2\right)$$
$$= N\left(\mathbf{e}|0, \mathbf{I}\sigma_e^2\right) \prod_{t=1}^{h} N\left(\mathbf{g}_t|0, \mathbf{G}_t\sigma_{g_t}^2\right).$$

The marginal distribution of the data in model (3) has an expected value of $\mathbf{1}\mu$ and the variance–covariance matrix is $\mathbf{V} = \sum_{t=1}^{h} \mathbf{G}_t\sigma_{gt}^2 + \mathbf{I}\sigma_e^2$. It was assumed that there was no covariance between effects of SNPs from different genomic regions. Prediction of genetic value obtained with SNPs in region t for individuals in the testing set was as follows:

$$\hat{\boldsymbol{g}}_{t,test} = \mathbf{G}_{t,test,train}\mathbf{G}_{t,train}^{-1}\hat{\boldsymbol{g}}_{t,train},$$

where notations are as in Scenario (1), except that t indicates the $\mathbf{G}$ matrix of *t*th genomic region.

To illustrate the difference or similarity in information that is captured by the additive genomic relationships in each of the classes of genomic regions, the Euclidean distance (ED) between each pair of genomic relationships was calculated using the formula:

Abdollahi-Arpanahi *et al. Genet Sel Evol* (2016) 48:10

Page 5 of 13

$$ED = \sqrt{\sum_{i=1} (a_i - b_i)^2}, \tag{4}$$

where a and b are the corresponding elements of the genomic relationship matrices for each annotation class. This distance matrix was subsequently fed to an R package for clustering and drawing heat map plots [31].

Implementation of Bayesian analysis

To implement the procedures within a Bayesian framework, a flat prior was assigned to μ, and independent scaled inverse Chi square distributions were assigned to the variance components. The hyper-parameters for each of the inverse Chi square distributions were equal to 5 for degrees of freedom and the scale parameter was calculated according to the default value of the BGLR package [32]. All analyses were conducted using RKHS regression as implemented in the BGLR package.

The Bayesian model was run via Gibbs sampling. For each fitted model, a Markov chain Monte Carlo (MCMC) with 500,000 samples was run and the first 50,000 samples were discarded as burn-in. Subsequently, 450,000 samples were obtained and thinned at a rate of 50, resulting in 9000 mildly correlated samples for posterior inference. Convergence diagnostics and statistical and graphical analysis of Gibbs sampling were checked by visual inspection of trace plots of some parameters (i.e. variance components) and with the Coda [33] package.

Predictive ability

The predictive ability of our RKHS regression models was assessed by cross-validation (CV). Specifically, a fourfold CV scheme was applied by assigning animals randomly to one of four separate subsets. Of these four subsets, three were combined to form a training set and one was used as testing set. Each of the four subsets was used as a testing set only once. Since the CV distribution was dispersed because of the small size of the sample, the above fourfold CV was replicated 15 times, at random, and results were averaged over replications. Predictive abilities were assessed via Pearson's product-moment correlation between pre-adjusted phenotypes and predicted phenotypes (predicted genetic values plus intercept), and via prediction mean-squared error (MSE).

Results

Variance components

Narrow-sense genomic heritabilities were estimated by fitting all SNPs together and were equal to 0.29 ± 0.04 for BW, 0.33 ± 0.04 for BM and 0.24 ± 0.04 for HHP. These estimates agreed with those from our previous study using the same dataset but a different number of SNPs and a restricted maximum likelihood approach to estimate variances [21]. The genomic heritability associated with each class of genomic regions in the separate (Scenario 1) and joint analyses (Scenario 3) is presented in Fig. 1. Estimates obtained by fitting the six classes of genomic regions separately or together differed significantly: i.e. estimates from the separate analyses were much larger than those from the joint analysis for all classes of genomic regions and all traits. In the separate analyses, the variance attributable to each class of genomic regions was overestimated because of the LD between SNPs in different regions, whereas in the joint analysis, all classes of genomic regions acted together and leveraged the polygenic basis of each trait, possibly resulting in more accurate estimates. The mean LD (r^2) 0.32 was observed for SNPs that were separated by <20 kb and it dropped to 0.21 when the distance between SNPs reached 100 kb for the current population. This amount of r^2 was almost the same within and between SNPs in different genomic regions.

Interestingly, in both analyses, all classes of genomic regions made significant contributions to the narrow-sense genomic heritability with the class of synonymous regions having a slightly stronger impact. If the estimates of narrow-sense genomic heritability for each class of genomic regions in the separate analysis are added up, the genomic heritability would also be out of range. However, in the joint analysis, the sum of the six component estimates was similar to the estimate from the whole-genome analysis for each trait. This reinforces the concept that LD causes single SNP regression (GWAS) to capture effects due to other SNPs, and the same is observed with variance component estimates.

In the joint analysis of BW, the estimated genomic heritability ranged from 0.05 to 0.10 for the different classes of genomic regions (Fig. 1). The largest estimate was attributed to the class of synonymous regions. For BM (joint analysis, Fig. 1), the estimates ranged from 0.05 for the class of missense regions to up to 0.09 for that of synonymous regions, and the distribution of the estimates of genomic heritability over the six classes for BM was similar to that for BW. For HHP (joint analysis, Fig. 1), the estimated genomic heritability ranged from 0.06 for the class of missense regions to 0.07 for the UTR class. Differences in the relative contribution of each class of genomic regions were similar in the joint and separate analyses. However, the difference in estimates of genomic heritability between classes of genomic regions was larger in the separate analysis than in the joint analysis. The posterior standard deviations of the genomic heritability in the separate analysis were larger than in the joint analysis.

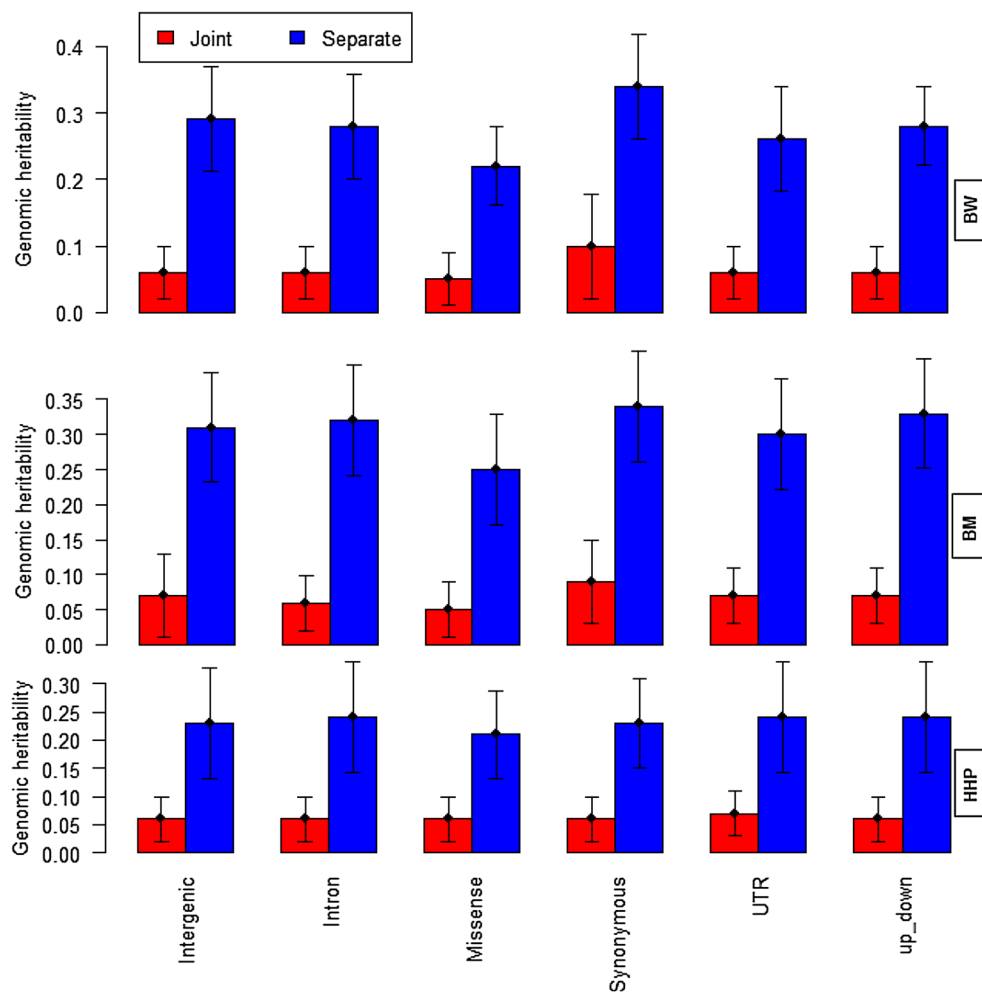Although the number of SNPs within coding DNA regions (synonymous and missense) was smaller than in

Abdollahi-Arpanahi *et al. Genet Sel Evol* (2016) 48:10

Page 6 of 13

Fig. 1 Marked genomic heritability estimates with SNPs partitioned into six classes of genomic regions for body weight (BW), ultra-sound of breast muscle (BM) and hen house egg production (HHP). *Red bars* and *blue bars* show joint and separate analyses, respectively. "*up-down*" indicates regions that are located 5 kb upstream and downstream of the gene. The *whiskers* represent 95 % confidence interval

non-coding regions (introns and intergenic), the magnitude of the genetic variance for these regions was similar and, in some cases, larger than for the non-coding regions. In order to check the influence of number of SNPs within a class of genomic regions, we picked 1500 random SNPs from each class and performed the same analysis to estimate genetic parameters with ten replications. The results (not shown here) produced a similar pattern as those for all the SNPs within each class. Hence, we concluded that our results were neither driven by differences in allelic frequency distributions, which were similar (Table 1), nor by differences between the number of SNPs in genic and intergenic regions.

Estimates of additive and dominance genomic heritability from the separate analysis for each class of genomic regions are in Table 2 (Scenario 2). Results were similar to those obtained with additive kernels only (Fig. 1). These results corroborated that all genomic regions contributed to the marked additive and dominance genetic variation and, for BW and BM, synonymous regions had the greatest contribution to the additive genetic variance among all annotation classes. Relative to the estimates from the separate analysis in Fig. 1, for most genomic regions, there was a decrease in the estimates of the marked additive genomic heritability, which seemed to move to the marked dominance variance. Regardless of the trait under study, the contribution of dominance genomic variance to total genetic variation was negligible (~0.05) and the contributions of each class of genomic regions to dominance variance were almost the same.

The heat map and a hierarchical clustering indicated high similarity between additive genomic relationships

Abdollahi-Arpanahi *et al. Genet Sel Evol* (2016) 48:10

Page 7 of 13

Table 2 Estimates of additive and dominance genomic heritability of SNPs partitioned into six classes of genomic regions for body weight (BW), ultra-sound of breast muscle (BM) and hen house egg production (HHP)

Class of genomic regions	BW		BM		HHP	
	$h^2_{mA} \pm$ SD	$h^2_{mD} \pm$ SD	$h^2_{mA} \pm$ SD	$h^2_{mD} \pm$ SD	$h^2_{mA} \pm$ SD	$h^2_{mD} \pm$ SD
Intergenic	0.22 ± 0.04	0.05 ± 0.02	0.24 ± 0.05	0.07 ± 0.02	0.16 ± 0.04	0.05 ± 0.01
Intron	0.23 ± 0.03	0.05 ± 0.02	0.25 ± 0.05	0.06 ± 0.02	0.17 ± 0.04	0.06 ± 0.02
Missense	0.18 ± 0.03	0.04 ± 0.01	0.20 ± 0.03	0.04 ± 0.01	0.15 ± 0.03	0.05 ± 0.01
Synonymous	*0.28 ± 0.05*	0.04 ± 0.02	*0.28 ± 0.05*	0.05 ± 0.01	0.17 ± 0.04	0.05 ± 0.01
UTR	0.21 ± 0.04	0.05 ± 0.01	0.23 ± 0.04	0.07 ± 0.02	0.17 ± 0.04	0.06 ± 0.01
Up-down stream	0.22 ± 0.04	0.06 ± 0.02	0.26 ± 0.05	0.07 ± 0.02	0.18 ± 0.04	0.05 ± 0.01
All markers	0.23 ± 0.04	0.06 ± 0.02	0.26 ± 0.05	0.07 ± 0.02	0.17 ± 0.04	0.06 ± 0.02

Each genomic region was fitted separately

Numbers in italics indicate the largest estimates among all classes of genomic regions

h^2_{mA} Additive genomic heritability, h^2_{mD} dominance genomic heritability, *SD* posterior standard deviation

matrices that were computed by using SNPs within different genomic regions (Fig. 2). All genomic relationship matrices, except the dominance relationship matrices for intron and intergenic regions, were grouped in the same cluster. There was a large correlation between corresponding elements of **G** from different genomic regions (>0.90), with an average of 0.94. We also found that the correlation between corresponding elements of **D** was >0.83, with an average of 0.90. The correlation between additive and dominance relationships ranged from 0.70 to 0.82 with an average value of 0.76. Thus, these genomic relationship matrices are not orthogonal to each other, and there is some confounding between the estimated parameters. This clearly complicates the separation and interpretation of estimates. Genomic relationship kernels that are "orthogonal" to each other could probably enhance inference or prediction ability but such kernels are not straightforward to construct and further research is needed on this issue.

Predictive ability

Figure 3 represents the predictive correlation (left panel) and MSE (right panel) from the additive kernel (**G**) and additive and dominance kernels (**G** + **D**) for each genomic region [Scenarios (1), (2), respectively]. Here, each genomic region was analyzed separately, thus the LD between SNPs in different genomic regions can affect the results. The gain in predictive ability from the combined use of additive and dominance (**G** + **D**) kernels was negligible. In terms of predictive correlation, minor differences between classes of genomic regions were observed but, according to the MSE metric, all genomic regions performed similarly. For BW, when only additive or joint additive and dominance (**G** + **D**) kernels were fitted, the predictive correlation due to SNPs in synonymous and UTR genomic regions was greater than for other genic regions. The lowest predictive ability was obtained for

the class of missense genomic regions. In general, for BW and BM, the predictive correlation obtained for different classes of genomic regions was similar. Overall, our results indicated that all classes of genomic regions influence the prediction of yet-to-be observed phenotypes.

For HHP, when additive and dominance kernels (**G** + **D**) were fitted jointly, the class of intron regions had the largest predictive correlation, but this superiority was not observed in terms of MSE. In agreement with genomic heritability estimates (Table 2), the MSE for genic regions was slightly smaller than for intron regions (HHP, Fig. 3).

Predictive ability of the six classes of genomic regions for BW, BM and HHP when the statistical model fitted all genomic regions jointly is shown in Fig. 4. In agreement with the estimates of genomic heritability (Fig. 1), SNPs within the class of synonymous regions resulted in a better predictive correlation than other classes of genomic regions irrespective of trait, whereas in terms of MSE the predictive performance of the six classes of genomic regions was almost the same (Fig. 4). In terms of predictive correlation, the predictive abilities of the class of synonymous regions and of all sets of SNPs were the same whereas, in terms of MSE, using all SNPs resulted in a better predictive ability than when using any single class of genomic regions across all traits. The class of missense genomic regions led to the smallest predictive correlation.

Discussion

Which parts of the genome contribute relatively more to the genetic variation of a complex trait is an important question in quantitative genetics. In human and dairy cattle studies, missense SNPs are over-represented in trait-associated variants e.g., [3], which is in agreement with their major role for protein sequence changes [34]. More recently, with the availability of transcriptomic
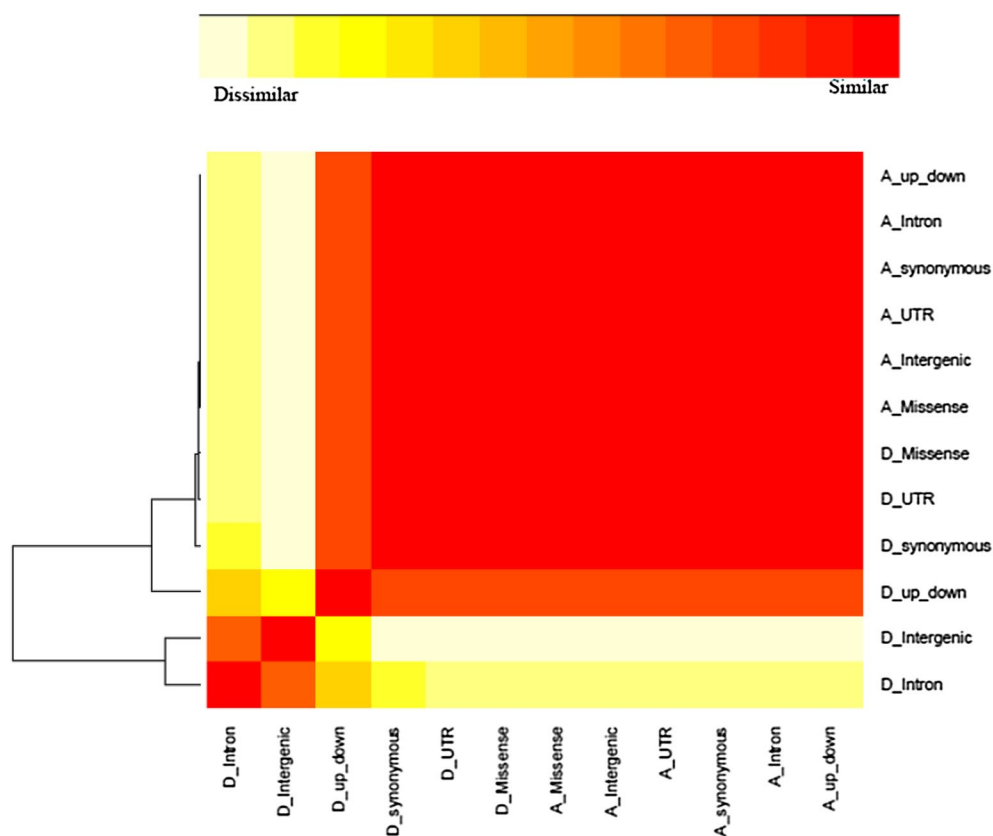
Abdollahi-Arpanahi *et al. Genet Sel Evol* (2016) 48:10

Page 8 of 13

Fig. 2 Heat map visualizing the degree of similarity between additive and dominance genomic relationship matrices for each genomic region. Very similar matrices are indicated in *red* and very dissimilar matrices are in *white*. Variable names that begin with an "A_" denote additive relationships and those with a "D_" denote dominance relationships. "up-down" indicates regions that are located 5 kb upstream and downstream of the gene

data and findings from the ENCODE projects, it has been reported that 80 % of the genome has a biochemical function [1]. The statistical framework that was developed by Fisher [35] assumes that most traits are affected by an infinite number of genes and that each contribute very little to the variance of the trait, and are randomly distributed across the genome. Here, we partitioned the chicken genome into six classes of genic and non-genic regions and investigated their contribution to marked additive and dominance genetic variation, and to predictive performance. We found that all classes of genomic regions contributed to genetic variation and that this contribution was slightly greater for SNPs within synonymous regions. Variance component estimates can be regarded as measures of goodness of fit, but better fit will not necessarily lead to increased predictive accuracy for future samples because of issues such as model overfitting. In terms of MSE, all annotation classes resulted in a similar predictive ability regardless of the trait under study. The MSE is a better and more flexible metric for comparing models than predictive correlations. The predictive correlations are bounded between 0 and 1, while

MSE can move from 0 to infinity. Furthermore, MSE addresses both prediction bias and variability, whereas predictive correlations provide only a measure of association e.g., [36, 37].

Missense regions consistently yielded the lowest performance, whereas synonymous regions produced the highest performance in terms of predictive correlations. However, the predictive ability of genic regions with a few 1000 SNPs was more or less the same as that of non-genic regions with a hundred thousand SNPs. This agrees with the findings of Do et al. [11] who showed that predictive accuracy and prediction bias of genomic regions did not significantly differ from those of randomized SNP groups.

In agreement with the infinitesimal theory and with Morota et al. [12], we found that the whole-genome approach is a better choice for prediction than using genomic regions individually. In general, our results highlight the importance of having SNPs that cover the entire genome, which suggests that many nucleotides play a role in connecting genotypes to phenotypes.
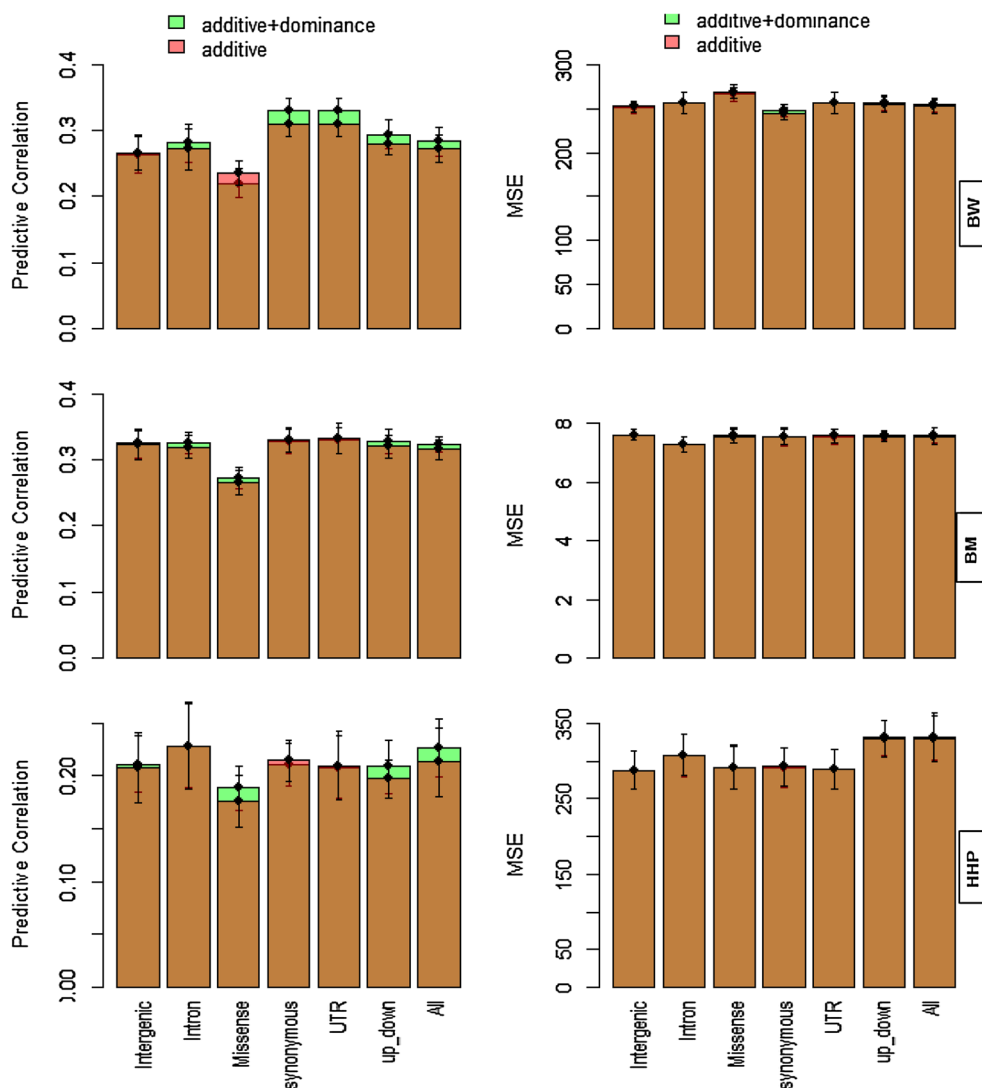
Abdollahi-Arpanahi *et al. Genet Sel Evol* (2016) 48:10

Page 9 of 13

Fig. 3 Predictive correlations and prediction mean-squared errors (MSE) resulting from different classes of genomic regions for body weight (BW), ultra-sound of breast muscle (BM) and hen house egg production (HHP) from the separate analysis of additive (**G**) and additive + dominance (**G** + **D**) kernels for each genomic region. The results were based on a fourfold cross-validation with 15 replications. "*up-down*" indicates regions that are located 5 kb upstream and downstream of the gene. "All" means that all SNPs were used to construct **G**. The *whiskers* represent 95 % confidence intervals and overlapping *bars* are in *bronze color*

Our results for BW and BM agree with the findings of Koufariotis et al. [9] who reported that synonymous parts of the genome explained a larger proportion of the additive variance than other genomic regions in dairy cattle. Importantly, recent studies have demonstrated some functional outcomes of synonymous mutations [10, 38]. In the fields of genetics and pharmacology, there is increasing interest for synonymous codon changes, which do not alter amino acids, [38, 39], since over 50 human diseases have been associated with synonymous mutations [39]. Furthermore, some studies have reported that synonymous codons can affect protein folding and

function of translated proteins and, therefore, may be under selective pressure [40]. Other functions for synonymous mutations, such as the splicing of precursor mRNAs, alteration of the secondary structure of mRNA and effects on mRNA stability, have also been described [38, 39].

However, the association between SNPs in intron regions and complex traits cannot be ruled out since elements within introns have been demonstrated to have regulatory functions [41, 42]. Studies in human genetics have found that the DNaseI hypersensitivity sites are some of the most enriched regions for trait-associated
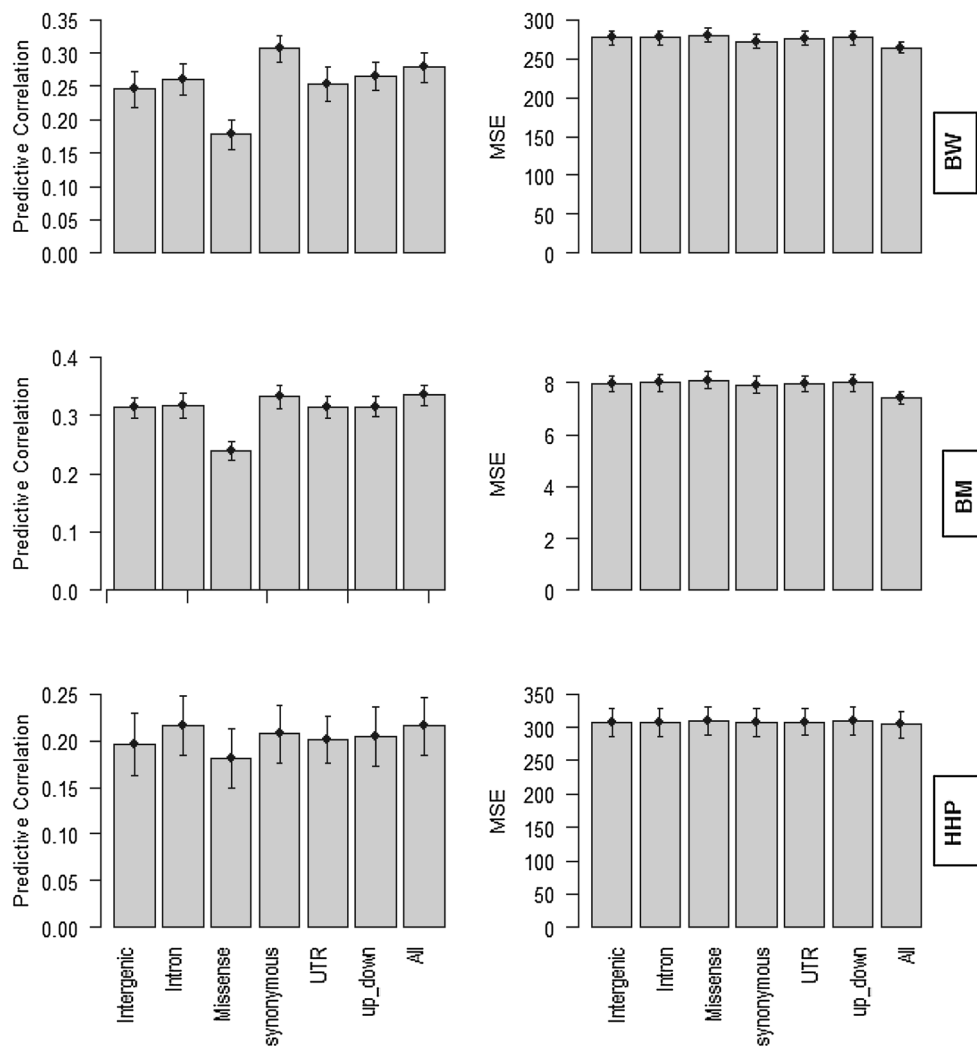
Abdollahi-Arpanahi *et al. Genet Sel Evol* (2016) 48:10

Page 10 of 13

Fig. 4 Predictive correlations and prediction mean-squared error (MSE) for the six classes of genomic regions for body weight (BW), ultra-sound of breast muscle (BM) and hen house egg production (HHP) from the joint analysis of all additive kernels $(\mathbf{G}_1 + \mathbf{G}_2 + \cdots + \mathbf{G}_6)$. The results were based on a fourfold cross-validation with 15 replications. "*up-down*" indicates regions that are located 5 kb upstream and downstream of the gene. "All" means that all SNPs were used to construct **G**. The *whiskers* represent 95 % confidence intervals

SNPs [7]. Regulatory functions such as promoter, enhancer or transcription factor binding sites (TFBS) are located in the genomic regions of the upstream and downstream annotation classes, and Gusev et al. [7] reported that this part of the genome is responsible for part of the genetic variation in 11 human diseases. A slightly better predictive performance of the class of intergenic regions for broiler traits was also suggested in other studies e.g., [12], and hundreds of structural and copy number variants in intronic and intergenic regions were identified in the two Silkie and Taiwanese native chicken breeds [43]. In general, our results suggest that the variants within the non-coding genome are also

important, thus attention should not be limited to variants within the protein-coding regions only.

Our heat map and cluster analysis showed a high level of similarity between genomic relationships based on SNPs in different genomic regions. Hence, if a study aims at partitioning the genetic variation into different classes of genomic regions or into additive and non-additive components, larger sample sizes must be used and non-orthogonality of genomic relationship matrices must be taken into account for proper interpretation. Increasing sample size may improve the power of detecting and distinguishing between minor additive and non-additive effects.

Abdollahi-Arpanahi *et al. Genet Sel Evol* (2016) 48:10

Page 11 of 13

In agreement with our results, Yang et al. [8] partitioned the genome into genic and non-genic regions and found that genic regions explained more genetic variation than intergenic regions for height and body mass index (BMI) in humans. Morota et al. [12] found that, for BW and HHP, non-genic regions performed marginally better than genic regions, while for BM, genic regions resulted in a better predictive performance than non-genic regions. However, there are differences between our study and that of Morota et al. [12]. In the later, coding DNA sequences were clustered together and a tenfold CV design was used, while in our study, annotation classes were distinct and a fourfold CV layout was used. We also used a multiple-kernel approach, while in Morota et al. [12] a single-kernel approach was applied. A multiple-kernel approach helps to eliminate signals that one annotation class has on another annotation class.

We observed that additive genetic variance had a major contribution to total genetic variance. Several studies have reported that the contribution of non-additive variance to total variation is minor [44, 45], while evidence of interactions within and between loci is pervasive [45]. Mäki-Tanila and Hill [46] concluded that multilocus interactions make significant contributions to the additive variance and does not lead to large increases in the non-additive genetic variance, but if the heterozygosity level is high at multiple loci, epistatic loci explain a large part of the non-additive variance. However, Munoz et al. [14] showed that, for tree height, additive and non-additive components of the genetic variance were similar in magnitude. Quantifying non-additive genetic variance precisely requires the setting-up of orthogonal additive and dominance genomic relationship matrices (kernels) and the assumption of LE [47]. LD exacerbates non-orthogonality [18, 48] and construction of orthogonal kernels needs further investigation. Therefore, the decomposition of the variance obtained in this study via the two kernels ($\mathbf{G} + \mathbf{D}$) must be interpreted with caution, because we cannot rule out the possibility that a single kernel captures multiple sources of genetic variation. Morota et al. [18] reported that, for health traits in dairy cattle, the non-additive genetic variance contributed greatly to the genetic variance. However, our results showed that the contribution of dominance variance to total variance was negligible, and that all genic and non-genic regions represented similar contributions to the dominance variation captured by SNPs. Visscher et al. [49], using identity-by-descent (IBD) coefficients, showed that additive and dominance relationship coefficients were highly correlated (theoretical correlation was equal to 0.89), with an empirical correlation between additive and dominance relationships of 0.91. This illustrates that partitioning genetic variance into additive and

dominance variance is indeed very difficult. In terms of predictive ability, our findings agree with a recent study on Fleckvieh cattle, where inclusion of dominance in the model did not increase the accuracy of predicted breeding (and total genetic) values of milk production and conformation traits [50].

There are some limitations in our study, i.e. (1) small sample size; although our sample of 1346 birds was sufficient to estimate the heritability for BW with a reasonable precision, for traits with a smaller heritability such as HHP (or for separating additive variance from dominance variance), larger sample sizes are necessary; (2) the precision of the annotated chicken genomic regions was maybe not sufficiently high, but one can assume that in the near future more precisely annotated data will be available; (3) a genomic annotation study using a ~300K SNP dataset cannot include all genic and non-genic SNPs; the availability of whole-genome sequence data is expected to identify all causal SNPs and, thus, results should be more accurate given the large size of the genotyped samples [51]. However, it has been argued that whole-genome sequencing data would not significantly improve predictive correlations using current genomic relationship-based methods (e.g., GBLUP or single step GBLUP) [52, 53]; (4) due to non-orthogonality between additive and dominance relationships, it was difficult to disentangle the additive variance from the dominance variance; with larger datasets, the issue of the correlation between additive and dominance coefficients should be less important, and additional non-additive effects, such as additive-by-additive or additive-by-dominance interactions could be estimated [49]; and (5) the variants that we investigated represent a set of SNPs that were selected to have a high MAF and to be evenly spaced across the genome (ascertainment bias), rather than a complete set of variants in the population. With the availability of whole-genome sequence data, this limitation should be alleviated.

Conclusions

Our results provide information about the quantitative impact of coding and non-coding DNA regions on complex traits. All genomic regions regardless of the trait under study contribute similarly to the additive and dominance variances. However, the contribution of dominance variance to the total genetic variation was minor compared to that of the additive variance. In agreement with variance component estimates, the predictive ability of all genomic regions was similar except for the class of missense regions, which led to a lower predictive correlation. However, the whole-genome approach provided a better predictive ability than that obtained from classes of genomic regions considered individually.

Abdollahi-Arpanahi *et al. Genet Sel Evol* (2016) 48:10

Page 12 of 13

Author details

[1] Department of Animal Sciences, University of Wisconsin, Madison, WI, USA. [2] Department of Animal and Poultry Science, College of Aburaihan, University of Tehran, Pakdasht, Iran. [3] Department of Animal Science, University of Nebraska, Lincoln, NE, USA. [4] Department of Dairy Science, University of Wisconsin, Madison, WI, USA. [5] Aviagen Ltd, Midlothian, UK. [6] The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian, UK. [7] Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI, USA.

References

1. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489:57–74.
2. van Bakel H, Nislow C, Blencowe BJ, Hughes TR. Most "dark matter" transcripts are associated with known genes. PLoS Biol. 2010;8:e1000371.
3. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci USA. 2009;106:9362–7.
4. Kindt AS, Navarro P, Semple CA, Haley CS. The genomic signature of trait-associated variants. BMC Genomics. 2013;14:108.
5. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461:747–53.
6. Davis LK, Yu D, Keenan CL, Gamazon ER, Konkashbaev AI, Derks EM, et al. Partitioning the heritability of tourette syndrome and obsessive compulsive disorder reveals differences in genetic architecture. PLoS Genet. 2013;9:e1003864.
7. Gusev A, Lee SH, Trynka G, Finucane H, Vilhjálmsson BJ, Xu H, et al. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. Am J Hum Genet. 2014;95:535–52.
8. Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, et al. Genome partitioning of genetic variation for complex traits using common SNPs. Nat Genet. 2011;43:519–25.
9. Koufariotis L, Chen YPP, Bolormaa S, Hayes BJ. Regulatory and coding genome regions are enriched for trait associated variants in dairy and beef cattle. BMC Genomics. 2014;15:436.
10. Li X, Zhu C, Yeh CT, Wu W, Takacs E, Petsch K, et al. Genic and non-genic contributions to natural variation of quantitative traits in maize. Genome Res. 2012;22:2436–44.
11. Do DN, Janss LL, Jensen J, Kadarmideen HN. SNP annotation-based whole genomic prediction and selection: an application to feed efficiency and its component traits in pigs. J Anim Sci. 2015;93:2056–63.
12. Morota G, Abdollahi-Arpanahi R, Kranis A, Gianola D. Genome-enabled prediction of quantitative traits in chickens using genomic annotation. BMC Genomics. 2014;15:109.
13. Erbe M, Hayes B, Matukumalli L, Goswami S, Bowman P, Reich C, et al. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J Dairy Sci. 2012;95:4114–29.
14. Muñoz PR, Resende MF, Gezan SA, Resende MDV, de los Campos G, Kirst M, et al. Unraveling additive from nonadditive effects using genomic relationship matrices. Genetics. 2014;198:1759–68.
15. Nishio M, Satoh M. Including dominance effects in the genomic BLUP method for genomic evaluation. PLoS One. 2014;9:e85792.
16. Su G, Christensen OF, Ostersen T, Henryon M, Lund MS. Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. PLoS One. 2012;7:e45293.
17. Sun C, VanRaden P, O'Connell J, Weigel K, Gianola D. Mating programs including genomic relationships and dominance effects. J Dairy Sci. 2013;96:8014–23.
18. Morota G, Boddhireddy P, Vukasinovic N, Gianola D, DeNise S. Kernel-based variance component estimation and whole-genome prediction of pre-corrected phenotypes and progeny tests for dairy cow health traits. Front Genet. 2014;5:56.
19. Gönen M, Alpaydin E. Multiple kernel learning algorithms. J Mach Learn Res. 2011;12:2211–68.
20. 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. 2013;14:59.
21. Abdollahi-Arpanahi R, Pakdel A, Nejati-Javaremi A, Moradi Shahrbabak M, Morota G, Valente B, et al. Dissection of additive genetic variability for quantitative traits in chickens using SNP markers. J Anim Breed Genet. 2014;131:183–93.
22. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet. 2009;84:210–23.
23. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.
24. Gianola D, Fernando RL, Stella A. Genomic-assisted prediction of genetic value with semiparametric procedures. Genetics. 2006;173:1761–76.
25. Gianola D, van Kaam JB. Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. Genetics. 2008;178:2289–303.
26. de los Campos G, Gianola D, Rosa GJ, Weigel KA, Crossa J. Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. Genet Res. 2010;92:295–308.
27. Morota G, Gianola D. Kernel-based whole-genome prediction of complex traits: a review. Front Genet. 2014;5:363.
28. VanRaden PM. Efficient methods to compute genomic predictions. J Dairy Sci. 2008;91:4414–23.
29. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010;42:565–9.
30. Vitezica ZG, Varona L, Legarra A. On the additive and dominant variance and covariance of individuals within the genomic selection scope. Genetics. 2013;195:1223–30.
31. R Core Team. (2015). R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. http://www.R-project.org/. (Accessed 11 Jan 2015).
32. Pérez P, de los Campos G. Genome-wide regression and prediction with the BGLR statistical package. Genetics. 2014;198:483–95.
33. Raftery AE, Lewis SM. Comment: one long run with diagnostics: implementation strategies for Markov chain Monte Carlo. Stat Sci. 1992;7:493–7.
34. Stenson PD, Mort M, Ball EV, Howells K, Phillips AD, Thomas NS, et al. The human gene mutation database: 2008 update. Genome Med. 2009;1:13.
35. Fisher RA. The correlation between relatives on the supposition of Mendelian inheritance. Trans R Soc Edinb. 1918;52:399–433.
36. Abdollahi-Arpanahi R, Morota G, Valente BD, Kranis A, Rosa GJM, Gianola D. Assessment of bagging GBLUP for whole genome prediction of broiler chicken traits. J Anim Breed Genet. 2015;132:218–28.
37. González-Recio O, Rosa GJ, Gianola D. Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. Livest Sci. 2014;166:217–31.
38. Hunt RC, Simhadri VL, Iandoli M, Sauna ZE, Kimchi-Sarfaty C. Exposing synonymous mutations. Trends Genet. 2014;30:308–21.
39. Sauna ZE, Kimchi-Sarfaty C. Understanding the contribution of synonymous mutations to human disease. Nat Rev Genet. 2011;12:683–91.

Abdollahi-Arpanahi *et al. Genet Sel Evol* (2016) 48:10

Page 13 of 13

40. Parmley JL, Hurst LD. How common are intragene windows with KA > KS owing to purifying selection on synonymous mutations? J Mol Evol. 2007;64:646–55.
41. Van Laere AS, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, et al. A regulatory mutation in *IGF2* causes a major QTL effect on muscle growth in the pig. Nature. 2003;425:832–6.
42. Wu Y, Pi JS, Pan AL, Pu YJ, Du JP, Shen J, et al. An SNP in the *MyoD1* gene intron 2 associated with growth and carcass traits in three duck populations. Biochem Genet. 2012;50:898–907.
43. Fan WL, Ng CS, Chen CF, Lu MYJ, Chen YH, Liu CJ, et al. Genome-wide patterns of genetic variation in two domestic chickens. Genome Biol Evol. 2013;5:1376–92.
44. Hill WG, Goddard ME, Visscher PM. Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet. 2008;4:e1000008.
45. Mackay TF. Epistasis and quantitative traits: using model organisms to study gene-gene interactions. Nat Rev Genet. 2014;15:22–33.
46. Mäki-Tanila A, Hill WG. Influence of gene interaction on complex trait variation with multilocus models. Genetics. 2014;198:355–67.
47. Cockerham CC. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. Genetics. 1954;39:859–82.
48. Gianola D, Hospital F, Verrier E. Contribution of an additive locus to genetic variance when inheritance is multi-factorial with implications on interpretation of GWAS. Theor Appl Genet. 2013;126:1457–72.
49. Visscher PM, Medland SE, Ferreira MA, Morley KI, Zhu G, Cornes BK, et al. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. PLoS Genet. 2006;2:e41.
50. Ertl J, Legarra A, Vitezica ZG, Varona L, Edel C, Emmerling R, et al. Genomic analysis of dominance effects on milk production and conformation traits in Fleckvieh cattle. Genet Sel Evol. 2014;46:40.
51. Daetwyler HD, Capitan A, Pausch H, Stothard P, Van Binsbergen R, Brøndum RF, et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat Genet. 2014;46:858–65.
52. Meuwissen T, Goddard M. Accurate prediction of genetic values for complex traits by whole-genome resequencing. Genetics. 2010;185:623–31.
53. Hayes BJ, MacLeod IM, Daetwyler HD, Bowman PJ, Chamberlian AJ, Vander Jagt CJ, et al. Genomic prediction from whole genome sequence in livestock: the 1000 bull genomes project. In: Proceedings of the 10th world congress of genetics applied to livestock production, 17–22 August; Vancouver; 2014. https://asas.org/docs/default-source/wcgalp-proceedings-oral/183_paper_10441_manuscript_1644_0.pdf?sfvrsn=2.