# Including phenotypic causal networks in genome-wide association studies using mixed effects structural equation models

**Running Head:** Structural equation modeling for association studies

Mehdi Momen[1], Ahmad Ayatollahi Mehrgardi[1*], Mahmoud Amiri Roudbar [1], Andreas Kranis[2], Renan Mercuri Pinto[3,4], Bruno D. Valente[4], Gota Morota[5,] Guilherme J. M. Rosa [4,6], Daniel Gianola[4,6,7]

[1] Department of Animal Science, Faculty of Agriculture, Shahid Bahonar University of Kerman (SBUK), Kerman, Iran
[2] Roslin Institute, University of Edinburgh, Midlothian, UK, EH25 9PS
[3] Department of Exact Sciences, University of São Paulo - ESALQ, Piracicaba-SP, Brazil
[4] Department of Animal Sciences, University of Wisconsin, Madison, WI, USA
[5] Department of Animal Science, University of Nebraska-Lincoln, Lincoln, Nebraska, USA
[6] Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI, USA
[7] Department of Dairy Science, University of Wisconsin, Madison, WI, USA

[*]Corresponding author: mehrgardi@uk.ac.ir

Email addresses:

MM: momenmehdi@yahoo.com

AAM: mehrgardi@uk.ac.ir

MA: mahmoud.amiri225@gmail.com

GM: morota@unl.edu

AK: andreas.kranis@roslin.ed.ac.uk

RMP: rpinto@wisc.edu

BDV: bvalente@wisc.edu

GJMR: grosa@wisc.edu

DG: gianola@ansci.wisc.edu

# Abstract

## Background

Phenotypic networks describing putative causal relationships among multiple phenotypes can be used to infer single-nucleotide polymorphism (SNP) effects in genome-wide association studies (GWAS). In GWAS with multiple phenotypes, reconstructing underlying causal structures among traits and SNPs using a single statistical framework is essential for understanding the entirety of genotype-phenotype maps. A structural equation model (SEM) can be used for such purposes.

## Methods

We applied SEM to GWAS (SEM-GWAS) in chickens, taking into account putative causal relationships among body weight (BW), breast meat (BM), hen-house production (HHP), and SNPs. We assessed the performance of SEM-GWAS by comparing the model results with those obtained from traditional multi-trait association analyses (MTM-GWAS).

## Results

Three different putative causal path diagrams were inferred from highest posterior density (HPD) intervals of 0.75, 0.85, and 0.95 using the inductive causation algorithm. A positive path coefficient was estimated for BM→BW, and negative values were obtained for BM→HHP and BW→HHP in all implemented scenarios. Further, the application of SEM-GWAS enabled the decomposition of SNP effects into direct, indirect, and total effects, identifying whether a SNP effect is acting directly or indirectly on a given trait. In contrast, MTM-GWAS only captured overall genetic effects on traits, which is equivalent to combining the direct and indirect SNP effects from SEM-GWAS.

## Conclusions

Although MTM-GWAS and SEM-GWAS use the same probabilistic models, we provide evidence that SEM-GWAS captures complex relationships and delivers a more comprehensive understanding of SNP effects compared to MTM-GWAS. Our results showed that SEM-GWAS provides important insight regarding the mechanism by which identified SNPs control traits by partitioning them into direct, indirect, and total SNP effects.

**Key words:** Causal structure, GWAS, multiple traits, path analysis, SEM, SNP effect

# Background

Genome-wide association studies (GWAS) have become a standard approach for investigating relationships between common genetic variants in the genome (e.g., single-nucleotide polymorphisms, SNPs) and phenotypes of interest in human, plant, and animal genetics [1-4]. A typical GWAS is based on univariate linear or logistic regression of phenotypes on genotypes for each SNP individually while often adjusting for the presence of nuisance covariates [5]. A statistically significant association indicates that SNPs may be in strong linkage disequilibrium (LD) with quantitative trait loci (QTLs) that contribute to the trait etiology. Alternatively, multi-trait model GWAS (MTM-GWAS) can be used to test for genetic associations among a set of traits [6-8]. It has been established that MTM-GWAS reduces false positives and increases the statistical power of association tests, explaining the recent popularity of this method. MTM-GWAS can be used to study genetic associations of multiple traits; however, it does not identify factors that mediate relationships between the detected effects and dependencies involving complex traits.

Complex traits are the product of various cryptic biological signals that may affect a trait of interest either directly or indirectly through other intermediate traits [9]. A standard regression cannot describe such complex relationships between traits and QTLs properly. For instance, some traits may simultaneously act as both dependent and independent variables. Structural equation modeling (SEM) is an extended version of Wright's path analysis [10, 11] that offers a powerful technique for modeling causal networks. In a complex genotype-phenotype setting involving many traits, a given trait can be influenced not only by genetic and systematic factors but also by other traits (as covariates) as well. Here, QTLs may not affect the target trait directly; instead, the effects may be mediated by upstream traits in a causal network. Indirect effects may therefore constitute a proportion of perceived pleiotropy, and these concepts apply to sets of heritable traits, organized as networks, that are common in biological systems. An example from dairy cattle production systems, described by Gianola and Sorensen [11], is that higher milk yield increases the risk of a particular disease, such as mastitis, while the prevalence of the disease may negatively affect milk yield As another example, Varona, et al. [12] explored a causal link from litter size to average piglet weight in two pig breeds. In humans, obesity is a key factor influencing insulin resistance, which subsequently causes type 2 diabetes. Lists of causal networks across human diseases and candidate genes are described in Kumar and Agrawal [13] and Schadt [14].

Although MTM-GWAS is a valuable approach, it only captures correlations or associations among traits and does not provide information about causal relationships. Knowledge of the causal structures underlying complex traits is essential, as correlation does not imply causation. For example, a correlation between two traits, T1 and T2, could be attributed to a direct effect of T1 on T2 or T2 on T1, or to additional variables that jointly influence both traits [15]. Likewise, if we know a "causal" SNP is linked to a QTL, we can imagine three possible scenarios: 1) causal

($SNP \rightarrow T1 \rightarrow T2$), 2) reactive ($SNP \rightarrow T2 \rightarrow T1$), or 3) independent ($T1 \leftarrow SNP \rightarrow T2$). Scenarios (1) and (2) do not cause pleiotropy but produce association.

A SEM methodology has the ability to handle complex genotype-phenotype maps in GWAS, placing an emphasis on causal networks [16]. Therefore, SEM-based GWAS (SEM-GWAS) may provide a better understanding of biological mechanisms and of relationships among a set of traits than MTM-GWAS. SEM can potentially decompose the total SNP effect on a trait into direct and indirect (i.e., mediated) contributions. However, SEM-derived GWAS has yet not been discussed or applied fully in quantitative genetic studies yet. Our objective was to illustrate the potential utility of SEM-GWAS by using three production traits in broiler chickens genotyped for a battery of SNPs as a case example.

# Methods

## Data set

The analysis included records for 1,351 broiler chickens provided by Aviagen Ltd. (Newbridge, Scotland) for three phenotypic traits: body weight (BW), ultrasound of breast muscle (BM) at 35 days of age, and hen-house egg production (HHP), defined as the total number of eggs laid between weeks 28 and 54 per bird. The sample consisted of 274 full-sib families, 326 sires, and 592 dams. More details regarding population and family structure were provided by Momen, et al. [17]. A pre-correction procedure was performed on the phenotypes to account for systematic effects such as sex, hatch week, pen, and contemporary group for BW, BM, and HHP.

Each bird was genotyped for 580,954 SNP markers with a 600k Affymetrix SNP [18] chip (Affymetrix, Inc., Santa Clara, CA, USA). The Beagle software program [19] was used to impute missing SNP genotypes, and quality control was performed using PLINK version 1.9 [20]. After

118 removing markers that did not fulfill the criteria of minor allele frequencies $< 1\%$, call rate > 95%,
119 and Hardy–Weinberg equilibrium (Chi-square test p-value threshold was $10^{-6}$), 354,364 autosomal
120 SNP markers were included in the analysis.

121 ## Multiple-trait model for GWAS

122 MTM-GWAS is a single-trait GWAS model extended to multi-dimensional responses. When only
123 considering additive effects of SNPs, the phenotype of a quantitative trait using the single-trait
124 model can be described as:

125 $$y_i = \sum_{q=1}^{k} x_{iq}\beta_q + w_{ij}s_j + e_i \qquad (1)$$

126 where $y_i$ is the phenotypic trait of individual *i*, $x_{iq}$ is the incidence value for the *i*th phenotype in
127 the *q*th level of systematic environmental effects, $\beta_q$ is the fixed effect of the *q*th systemic
128 environmental effect on the trait, $w_j = (w_1, \dots, w_p)$ is the number of A alleles (i.e., $w_j \in \{0,1,2\}$)
129 in the genotype of SNP marker *j*, and $s_j$ is the allele substitution effect for SNP marker *j*. Strong
130 LD between markers and QTLs coupled with an adequate marker density increases the chance of
131 detecting marker and phenotype associations. Hypothesis testing is typically used to evaluate the
132 strength of the evidence of a putative association. Typically, a *t*-test is applied to obtain p-values,
133 and the statistic is $T_{ij} = \hat{s}_j / se(\hat{s}_j)$, where $\hat{s}$ is the point estimate of the *j*th SNP effect and $se(\hat{s}_j)$ is
134 its standard error.

135 The single locus model described above is naïve for a complex trait because the data typically
136 contain hidden population structure and individuals have varying degrees of genetic similarity [21,
137 22]. Therefore, accounting for covariance structure induced by genetic similarity is expected to

produce better inferences [23]. Ignoring effects that reveal genetic relatedness inflates the residual terms and compromises the ability to detect association. A random effect $g_i$, including a covariance matrix reflecting pairwise similarities between additive genetic effects of individuals, can be included to control population stratification. The similarity metrics can be derived from pedigree information or from whole-genome marker genotypes. This model, extended for analysis of $t$ traits, is given by:

$$y_{il} = \sum_{q=1}^{k} x_{iq}\beta_{ql} + w_{ij}s_{jl} + g_{il} + e_{il} \qquad (2)$$

for $i = 1,2,\cdots,n$, $l = 1,2,\cdots,t$. In this extension, $y_{il}$ is the phenotypic value of the $l$th trait for the $i$th subject, $\beta_{qj}$ is the systematic effect of the $q$th environmental factor $x_{iq}$ on the $l$th trait, $s_{jl}$ is the additive effect of the $j$th marker on the $l$th trait, $w_{ij}$ is as previously defined, and $g_{il}$ and $e_{il}$ are the random polygenic effect and model residual assigned to individual $i$ for trait $l$, respectively. Random effects within a trait follow the multivariate normal distribution, $\begin{bmatrix}\boldsymbol{g}_l \\ \boldsymbol{e}_l\end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix}\mathbf{0} \\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\boldsymbol{K}\sigma^2_{g_l} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}\sigma^2_{e_l}\end{bmatrix}\right)$, where $\boldsymbol{K}$ is a genetic relationship matrix, $\sigma^2_{g_l}$ is the additive genetic variance of trait $l$, $\boldsymbol{I}$ is an identity matrix, and $\sigma^2_{e_l}$ is the residual variance for trait $l$. The multiple-trait model accounts for the additive genetic ($\rho_{ll'}$) and residual correlation ($\lambda_{ll'}$) between a pair of traits $l$ and $l'$.

The positive definite matrix **K** may be a genomic relationship matrix (**G**) computed from marker data, or a pedigree-based matrix (**A**) computed from genealogical information. The **A** matrix describes the expected additive similarity among individuals, while **G** measures the realized

158 fraction of alleles shared. Genomic relationship matrices can be derived in several ways [24-26].
159 Here, we used the form proposed by VanRaden 2008 [24]:

160 $$\mathbf{G} = \frac{\mathbf{MM}'}{2\sum \mathrm{p_j q_j}} \qquad (3)$$

161 where **M** is an $n \times p$ matrix of centered SNP genotypes and $p_j$ and $q_j = 1 - p_j$ are the allele
162 frequencies at marker locus $j$. We evaluated both **A** and **G** in the present study.

163

164 ## Structural equation model association analysis

165 A SEM consists of two essential parts: a measurement model and a structural model. The
166 measurement model depicts the connections between observable variables and their corresponding
167 latent variables. The measurement model is also known as confirmatory factor analysis. The critical
168 part of a SEM is the structural model, which can have three forms. The first consists of observable
169 exogenous and endogenous variables. This model is a restricted version of a SEM known as path
170 analysis [10]. The second form explains the relationship between exogenous and endogenous
171 variables that are only latent. The third type is a model consisting of both manifest and latent
172 variables.

173 SEM can be applied to GWAS as an alternative to MTM-GWAS to study how different causal
174 paths mediate SNP effects on each trait. The following SEM model was considered:

175 $$y_{il} = \mu_l + \sum_{m \in C} y_m \boldsymbol{\lambda}_{lm} + w_{j(l)} s_{j(l)} + g_{il} + \varepsilon_{il} \qquad (4)$$

176 where $C$ is the set of phenotypic traits that directly affect the trait $l$, $\boldsymbol{\lambda}_{lm}$is a structural coefficient
177 representing the effect of trait $m$ on trait $l$, and $g_l \sim N(0, \boldsymbol{K}\sigma_l^2)$ is the polygenic effect of the $l$th trait.
178 The remaining terms are as presented earlier with one important difference: the SNP effects are not

interpreted as overall effects on trait $l$ but instead represent direct effects on trait $l$. Additional indirect effects from the same SNP may be mediated by phenotypic traits in *C*. Each marker is entered into equation (4) one at a time, and its significance is tested. For a discussion of how SEM represents genetic signals on each trait through multiple causal paths, see Wu et al. [27] and Jamrozik and Schaeffer [28]. Despite the difference in interpretation, the distribution of the vector of polygenic effects is assumed to be the same as in the MTM-GWAS model. The same applies to residual terms within a trait. We also consider trait-specific residuals to be independent within an individual. This restriction is required to render structural coefficients likelihood-identifiable. In addition, the interpretation of inferences as having a causal meaning requires imposing the restriction that the residuals' joint distribution be interpreted as the causal sufficiency assumption [29]. In the present study, all exogenous and endogenous variables were observable, and there was no latent variable. Hence, causal structure was assumed between the endogenous variables BM, BW, and HHP.

We considered the following GWAS models, which their causal structures were recovered by the inductive causation (IC) algorithm [29]: (1) MTM-GWAS with pedigree-based kinship **A** (MTM-A) or marker-based kinship **G** (MTM-G), and (2) SEM-GWAS with **A** (SEM-A) or **G** (SEM-G). Although nuisance covariates such as environmental factors can be omitted in the graph, they may be incorporated into the models as exogenous variables. The SEM representation allowed us to decompose SNP effects into direct, indirect, and total effects.

A direct SNP effect is the path coefficient between a SNP as an exogenous variable and a dependent variable without any causal mediation by any other variable. The indirect effects of a SNP are those mediated by at least one other intervening endogenous variable. Indirect effects are calculated by multiplying path coefficients for each path linking the SNP to an associated variable, and then

summing over all such paths [30]. The overall effect is the sum of all direct and indirect effects. By explicitly accounting for complex relationship structure among traits in such a way, SEM provides a better understanding of a genome-wide SNP analysis by allowing us to decompose effects into direct, indirect, and overall effects within a predefined casual framework [31]. MTM-GWAS and SEM-GWAS were compared with the logarithm of the likelihood function (log L), Akaike's Information Criterion (AIC), and the Bayesian Information Criterion (BIC). The model providing the lowest values for these information criteria is considered to fit the data better [27]. MTM-GWAS and SEM-GWAS were fitted using the SNP Snappy strategy, which is implemented in the Wombat software program [32].

## Searching for a phenotypic causal network in a mixed model

In the SEM-GWAS formulation described earlier, the structure of the underlying causal phenotypic network needs to be known. Because this is not so in practice, we used a causal inference algorithm to infer the structure. Residuals are assumed to be independent in all SEM analyses, so associations between observed traits are viewed as due to causal links between traits and by correlations among genetic values (i.e., $g_1$, $g_2$, and $g_3$). Thus, to eliminate confounding problems when inferring the underlying network among traits, we used the approach of Valente, et al. [32] to search for acyclic causal structures through conditional independencies on the distribution of the phenotypes, given the genetic effects. A causal phenotypic network was inferred in two stages: 1) an MTM model [33] was employed to estimate covariance matrices of additive genetic effects and of residuals, and 2) the causal structure among phenotypes from the covariance matrix between traits, conditionally on additive genetic effects, was inferred by the IC algorithm. The residual (co)variance matrix was inferred using Bayesian Markov-chain Monte Carlo [27, 32], with samples drawn from the posterior distribution. The reason for our use of the residual (co)covariances is that the residual

structure could bear information from the joint distribution of all phenotypic traits conditional on their polygenic effects, such that they correct the confounding issues caused by such effects when the traits are genetically correlated [29]. For each query testing statistical independence between traits $y_l$and $y_{l'}$, the posterior distribution of the residual partial correlation $\rho_{y_l,y_{l'}}|S$ was obtained, where $S$ is a set of variables (traits) that are independent. Three highest posterior density (HPD) intervals of 0.75, 0.85, and 0.95 were used to make statistical decisions for SEM-GWAS. We thus considered SEM-A75 (HPD > 0.75), SEM-A85 (HPD > 0.85), SEM-A95 (HPD > 0.95), and SEM-G75 (HPD > 0.75). An HPD interval that does not contain zero declares $y_l$ and $y_{l'}$ to be conditionally dependent.

# Results

Figure 1 shows phenotypic relationship structures recovered by the IC algorithm for the three different HPD intervals. Edges connecting two traits represent non-null partial correlations as indicated by HPD intervals. We compared the two MTM-GWAS and four SEM-GWAS by using the three chicken traits (BW, BM, and HHP). Only causal structures among the three traits are shown in Figure 1, because other parts were the same across the different SEM models. Fully recursive SEM-A75 and SEM-G75 revealed direct effects of BM on BW and HHP, and those of BW on HHP, as well as an indirect effect of BM on HHP. In addition, SEM-A85 detected a direct effect of BM on BW, the direct effect of BW on HHP, and the indirect effect of BM on HHP mediated by BW. Finally, SEM-A95 only identified a direct effect of BM on BW because of a statistically stringent HPD cutoff imposed.

Given the causal structures inferred from the IC algorithm, the following SEM was fitted:

$$\begin{cases} \boldsymbol{y}_1 = \mu + \boldsymbol{Z}_i\boldsymbol{g}_1 + W_{ij}S_j + \boldsymbol{\varepsilon}_i \\ \boldsymbol{y}_2 = \mu + \lambda_{21}\boldsymbol{y}_1 + \boldsymbol{Z}_i\boldsymbol{g}_2 + W_{ij}S_j + \boldsymbol{\varepsilon}_i \\ \boldsymbol{y}_3 = \mu + \lambda_{31}\boldsymbol{y_1} + \lambda_{32}\boldsymbol{y}_2 + \boldsymbol{Z}_i\boldsymbol{g}_3 + W_{ij}S_j + \boldsymbol{\varepsilon}_i \end{cases} \quad (5)$$

Note that only a small number of the entries in the structural coefficient matrix ($\lambda$ in equation 5) are nonzero due to sparsity. These nonzero entries specify the effect of one phenotype on other phenotypes. The corresponding directed acyclic graph is shown in Figure 2 assuming the causal relationships among the three traits, where $y_1$, $y_2$, and $y_3$ represent BM, BW, and HHP, respectively; $SNP_j$ is the genotype of the *j*th SNP; $S_{jl}$ is the direct SNP effect on trait $l$; and the remaining variables are as presented earlier. This diagram depicts a fully recursive structure in which all recursive relationships among the three phenotypic traits are shown. Arrows represent causal connections, whereas double-headed arrows between polygenic effects are correlations.

<< **Figure 1 about here**>>

<< **Figure 2 about here**>>

We examined the fit of each model implemented to assess how well it describes the data (Table 1). Varona, et al. [12] and recently Valente, et al. [34] showed that re-parametrization and reduction of a SEM mixed model yield the same joint probability distribution of observation as in MTM, suggesting that the expected likelihood of SEM and MTM should be the same. As expected, SEM-GWAS and MTM-GWAS showed very similar results (e.g., SEM-A75 vs. MTM-A and SEM-G75 vs. MTM-G). Among the models considered, those involving **G** exhibited slightly better fits. SEM-A85 and SEM-A95, sharing a subset of the SEM-A75 structure, presented almost identical AIC and BIC values. Since these results imply that the recursive model and standard mixed model for GWAS are statistically equivalent in terms of the fitting criteria, the focus of the remainder of the

266 analysis will be on the modeling of SNP (or QTL) effects in the SEM context as an extension of
267 MTM, which accounts for recursive links among the three measured traits.
268
269 **<<Table 1 about here>>**

270 ## Structural coefficients

271 Table 2 presents the causal structural path coefficients for endogenous variables (BM, BW, and
272 HHP). All models have positive effects for BM→BW, whereas the BM→HHP and BW→HHP
273 relationships have negative path coefficients. The latter confirmed the fact that chicken breeding is
274 divided into broiler and layer sections due to the negative genetic correlation between BW and
275 HHP.

276 **<<Table 2 about here>>**

277 Also shown in Table 2 are the magnitudes of the SEM structural coefficient reflecting the intensity
278 of the causality. The positive coefficient $\lambda_{21}$ quantifies the (direct) causal effect of BM on BW.
279 This suggests that a 1-unit increase in BM results in a $\lambda_{21}$-unit increase in BW. Likewise, the
280 negative causal effects $\lambda_{31}$ and $\lambda_{32}$ offer the same interpretation.

281 ## Decomposition of SNP effect paths using a fully recursive model

282 We can decompose SNP effects into direct and indirect effects using Figure 2. The direct effect of
283 the SNP $j$ on $y_3$ (HHP) is given by $d_{SNP_j \to y_3}$: $\hat{S}_{j(y3)}$, where $d$ denotes the direct effect. Note there
284 are only one direct and many indirect paths. We find three indirect paths from $SNP_j$ to $y_3$ mediated
285 by $y_1$ and $y_2$ (i.e., the nodes formed by other traits). The first indirect effect is $ind_{(1)SNP_j \to y_3}$:
286 $\lambda_{32}(\lambda_{21}\hat{S}_{j(y1)})$ in the path mediated by $y_1$ and $y_2$, where $ind$ denotes the indirect effect. The second

indirect effect $ind_{(2)SNP_j \to y_3}: \lambda_{32}\hat{S}_{j(y2)}$, is mediated by $y_2$. The last indirect effect, is $ind_{(3)SNP_j \to y_3}: \lambda_{31}\hat{S}_{j(y1)}$, mediated by $y_1$. Therefore, the overall effect is given by summing all four paths, $T_{SNP_j \to y_3}: \lambda_{32}(\lambda_{21}\hat{S}_{j(y1)}) + \lambda_{32}\hat{S}_{j(y2)} + \lambda_{31}\hat{S}_{j(y1)} + \hat{S}_{j(y3)}$. The fully recursive model of the overall SNP effect is then:

$$\begin{cases} T_{\hat{S}_{j \to y_1}: \hat{S}_{j(y1)}} \\ T_{\hat{S}_{j \to y_2}: \lambda_{21}(\hat{S}_{j(y1)}) + \hat{S}_{j(y2)}} \\ T_{\hat{S}_{j \to y_3}: \lambda_{32}[\lambda_{21}(\hat{S}_{j(y1)}) + \hat{S}_{j(y2)}] + \lambda_{31}(\hat{S}_{j(y1)}) + \hat{S}_{j(y3)}} \end{cases} \quad (6)$$

For $y_1$ (BM), there is only one effect, so the overall effect is equal to the direct effect. For $y_2$ (BW) and $y_3$ (HHP), direct and indirect SNP effects are involved. There are two paths for $y_2$: one indirect, $ind_{S_j \to y_2}: \hat{S}_{j(y1)} \to y_1 \to y_2$, and one direct, $d_{S_j \to y_2}: \hat{S}_{j(y2)} \to y_2$. Here, the SNP effect is direct and mediated thorough other phenotypes according to causal networks in SEM-GWAS (Figures 1 and 2). For instance, the overall SNP effect for $y_3$ into four direct and indirect paths is $T_{\hat{S}_{j \to y_3}}: \lambda_{32}\lambda_{21}\hat{S}_{j(y1)} + \lambda_{32}\hat{S}_{j(y1)} + \lambda_{31}\hat{S}_{j(y1)} + \hat{S}_{j(y3)}$.

The scatter plots in Figure 3 compare the estimated total effects for HHP ($T_{\hat{S}_{j \to y_3}}$) obtained from SEM-GWAS and those from MTM-GWAS. We observed good agreement between SEM-GWAS and MTM-GWAS. The total SNP signals derived from SEM and MTM are the same but SEM provides biologically relevant additional information.

**<<Figure 3 about here>>**

Supplementary Figures S1–S4 present scatter plots of MTM-GWAS and SEM-GWAS signals (SEM-A75, SEM-G75, SEM-A85, and SEM-A95) for the $BM \to BW$ path, which was a common path across all SEM-GWAS considered. These two traits have a genetic correlation of 0.5 (results not shown). We partitioned the SEM causal link into direct, indirect, and overall effects based on

directed links inferred from the IC algorithm with HPD > 0.85, whereas MTM-GWAS captures an overall SNP effect on BW. Scatter plots of the overall effects from SEM-GWAS and those of the total effects from MTM-GWAS indicated almost perfect agreement (top left plots, Supplementary Figures S1–S4). We also observed concomitance between estimated overall and direct effects (top right plots, Supplementary Figures S1–S4). In contrast, there was less agreement in the magnitude of the SNP effects when comparing overall vs. indirect effects (bottom left plots, Supplementary Figures S1–S4). There was no linear relationship between the indirect and direct SNP effects (bottom right plots, Supplementary Figures S1–S4). In short, genetic signals detected in SEM-GWAS were close to those of MTM-GWAS for overall effects because both models are based on a multivariate approach with the same covariance matrix. In all SEM-GWAS, results showed that direct effects contributed to overall effects more than the indirect effects.

## Manhattan plot of direct, indirect, and overall SNP effects

Figure 4 depicts a Manhattan plot summarizing the magnitude of direct (SEM-75A), indirect (SEM-75A), and overall SNP effects (MTM-75A). We plotted the decomposed SNP effects on BW along chromosomes to visualize estimated marker effects from SEM-GWAS and MTM-GWAS. The indirect and direct effects provide a view of SNP effects from a perspective that is not available for the total effect of MTM-GWAS. For instance, many pleiotropic QTLs have positive direct effects on BW but negative effects on BM. There were two estimated SNP effects on chromosomes 1 and 2 that deserve particular attention. These two SNPs are highlighted with black circles and red ovals. The overall effect of the first SNP consisted of large indirect and small direct effects on BM, whereas the opposite pattern was observed for the second SNP, which showed large direct and small indirect effects. Although the overall effects of these SNPs were similar (top Manhattan plot, Figure 4), use of decomposition allowed us to determine that the trait of interest is

affected in different manners: the second SNP effect acted directly on BW without any mediation by BM, whereas the first SNP reflected a large effect mediated by BM on BW. Collectively, new insight regarding the direction of SNP effects can be obtained using the SEM-GWAS methodology.

The corresponding Manhattan plot based on $-\log_{10}$ (p-values) is shown in Supplementary Figure S5. As with the magnitude of effect sizes, the results showed that $-\log_{10}$ (p-values) of estimated overall effects from SEM-A75 and those from MTM-A75 yielded the same significant peaks. We found that some significant indirect SNP effects reached genome-wide significance after correction for multiple-testing using a 5% FDR threshold level (2.752). The most significant SNPs were on chromosomes 1 and 4 (GGA1 and GGA4).

**<<Figure 4 about here>>**

As an illustration, the six most significant SNPs with the highest $-\log_{10}$ (p-values) for each type of decomposed SNP effect are presented in Table 3. Seven candidate genes were identified near the significant SNPs derived from the SNP effects decomposition, with two on GGA7 (*OLA1* and *ZNF385B*), one on GGA3 (*EPHA7*), three on GGA4 (*LOC422264*, *LOC422265*, and *MAEA*), and one on GGA14 (*GRIN2A*). We found that only genes on GGA4 and GGA1 are linked to significant indirect SNP effects that impact HHP. Some studies reported QTLs for BM on GGA1 and for BW on GGA4, stating that these genomic regions contain QTLs related to abdominal fat and growth traits that were detected across diverse chicken populations [35, 36]. One of the two detected genes on GGA14, i.e., *GRIN2A*, which was linked to the SNP Gga_rs313620413, showed significant direct and overall SNPs effects using SEM as well as MTM. Collectively, Gga_rs15390496, Gga_rs16591372, and Gga_rs313620413 SNPs on GGA3, GGA7, and GGA14, which were linked to *EPHA7*, *OLA1*, and *GRIN2A*, respectively, represent candidate genes identified from overall effects of both SEM and MTM (Table 3).

We noted that the six SNPs selected according to the $-\log_{10}$ (p-values) from the direct effect on HHP ( i.e., $d_{SNP_j \rightarrow y_{(HHP)}}$) had small indirect effects ranging from –0.9018 to 0.2983. These indirect effects were negligible compared with their corresponding direct and total effects. Also, exploring the indirect effect sizes of the six most significant SNPs showed that indirect effects that are transmitted from inferred causal networks have the ability to change the magnitude of overall SNP effects, even changing them to the opposite direction (i.e., from positive to negative or vice versa).

It should also be noted that the estimated additive SNP effects obtained from the four SEM-GWAS can be used for inferring pleiotropy. For instance, a pleiotropic QTL may have a large positive direct effect on BW but may exhibit a negative indirect effect coming from BM, which in turn reduces the total QTL effect on BW. Arguably, the methodology employed here would be most effective when the direct and indirect effects of a QTL are in opposite directions. If the direct and indirect QTL effects are in the same direction, the power of SEM-GWAS may be the same as the overall power of MTM-GWAS. The overall effect ($T_{\hat{s}_{j \rightarrow y_{(HHP)}}}$) of a given SNP consisted of large indirect ($ind_{\hat{s}_{j \rightarrow y_{(HHP)}}}$) and small direct ($d_{\hat{s}_{j \rightarrow y_{(HHP)}}}$) effects on HHP, as observed for the top most significant indirect SNPs localized on GGA4 and GAA1, whereas the opposite pattern was observed for the most significant direct SNPs on GAA3, GGA7, and GGA14, which showed large direct and small indirect effects. Although the overall effects of these SNPs from SEM-GWAS and MTM-GWAS were similar, the use of decomposition allowed us to determine that the trait of interest is affected in different manners. For instance, a given SNP effect may largely act directly on HHP without any mediation by BM and BW, whereas another SNP may be transmitting a large effect through a causal path mediated by BM and BW. Collectively, new insight regarding the direction of SNP effects can be obtained using the SEM-GWAS methodology.

375 **<<Table 3 about here>>**

376

377 # Discussion

378 It is becoming increasingly common to analyze a set of traits simultaneously in GWAS by
379 leveraging genetic correlations between traits [37, 38]. In the present study, we illustrated the
380 potential utility of a SEM-based GWAS approach for causal inference and mediation analysis of
381 SNP effects, which has the potential advantage of embedding a pre-inferred causal structure across
382 phenotypic traits [32]. SEM-GWAS, as an extension of standard MTM, accounts for recursive
383 linking of mediating variables that could be either dependent or independent with restriction on a
384 residual covariance. This is a useful approach when multiple mediators influence the final
385 outcomes via either common or distinct biological pathways [39, 40]. SEM-GWAS is achieved by
386 first inferring the structure of networks between phenotypic traits. For this purpose, we used a
387 modified version of the IC algorithm described by Pearl [29] and modified for implementing in
388 quantitative genetics by Valente, et al. [32]. The IC algorithm was used to explore putative causal
389 links among phenotypes obtained from a residual covariance matrix, in a model that accounted for
390 systematic and genetic confounding factors such as polygenic additive effects. It then produced a
391 posterior distribution of partial residual correlations between any possible pairs of variables. Three
392 different causal path diagrams were inferred from HPD intervals of 0.75, 0.85, and 0.95. We
393 observed that the number of identified paths decreased with an increase in the HPD interval value.
394 Only a path connecting BM and BW was present in all HPD intervals considered. Moreover, we
395 found that the partial residual correlation between BM and HHP was weaker than that between BM

and BW. This may explain why the path between BM and HHP was not detected with HPD intervals larger than 0.75.

Estimated path coefficients reflect the strength of each causal link, quantifying the proportion of direct and indirect effects of a given SNP or genes on the outcome of interest via the mediator phenotypic traits or the predefined causal pathway between a set of mediators and the target outcome. For instance, a positive path coefficient from BM to BW suggests that a unit increase in BM directly results in an increase in BW. Our results showed that MTM-GWAS and SEM-GWAS were similar in terms of the goodness of fit as per the AIC and BIC criteria. This finding is in agreement with theoretical work of Gianola and Sorensen [11] and Varona, et al. [12] showing the equivalence between models. Thus, MTM-GWAS and SEM-GWAS produced the same marginal phenotypic distributions and goodness of fit values. A similar approach has been proposed by Li, et al. [16], Mi, et al. [41], and Wang and van Eeuwijk [42]. The main difference between our approach and theirs is that they used SEM in the context of standard QTL mapping, whereas our SEM-GWAS is developed for GWAS based on a linear mixed model.

The advantage of SEM-GWAS over MTM-GWAS is that the former decomposes SNP effects by tracing inferred causal networks. Our results showed that by partitioning SNP effects into direct, indirect, and total components, an alternative perspective of SNP effects can be obtained. As shown in Table 3 and Figure 4, direct and indirect effects may differ in magnitude and sign, acting in the same direction or in an antagonistic manner. Note that the total SNP effects inferred from SEM-GWAS were the same as the estimated SNP effects from MT-GWAS (Table 3). However, knowledge derived from the decomposition of SNP effects may be critical for animal and plant breeders in breaking unfavorable indirect QTL effects or obtaining better SNP effect estimates than those from MTM-GWAS [e.g., 41].

# Conclusion

SEM offers insights into how phenotypic traits relate to each other. We illustrated potential advantages of SEM-GWAS relative to the commonly used standard MTM-GWAS by using three chicken traits as an example. SNP effects pertaining to SEM-GWAS have a different meaning than those in MTM-GWAS. Our results showed that SEM-GWAS enabled the identification of whether a SNP effect is acting directly or indirectly, i.e. mediated, on given trait. In contrast, MTM-GWAS only captures overall genetic effects on traits, which is equivalent to combining direct and indirect SNP effects from SEM-GWAS together. Thus, SEM-GWAS offers more information and provides an alternative view of putative causal networks, enabling a better understanding of the genetic quiddity of traits at the genomic level.

# Conflict of Interest

The authors do not have any conflict of interest.

# Author's contributions

MM carried out the study and wrote the first draft of the manuscript. GJMR and DG designed the experiment, supervised the study and critically contributed to the final version of manuscript. GM contributed to the interpretation of results, provided critical insights, and revised the manuscript. BDV and AAM participated in discussion and reviewed the manuscript. MA, AK and RMP contributed materials and revised the manuscript. All authors read and approved the final manuscript.

# Acknowledgment

The first author wishes to acknowledge the Ministry of Science, Research and Technology of Iran for financially supporting his visit to the University of Wisconsin-Madison. Work was partially supported by the Wisconsin Agriculture Experiment Station under hatch grant 142-PRJ63CV to DG.

## References

1. Hayes B, Goddard M, Genome-wide association and genomic selection in animal breeding This article is one of a selection of papers from the conference "Exploiting Genome-wide Association in Oilseed Brassicas: a model for genetic improvement of major OECD crops for sustainable farming", Genome. 2010; 53: 876-883.
2. Brachi B, Morris G P, Borevitz J O, Genome-wide association studies in plants: the missing heritability is in the field, Genome Biology. 2011; 12: 232.
3. Wang H, Misztal I, Aguilar I, Legarra A, Muir W, Genome-wide association mapping including phenotypes from relatives without genotypes, Genetics Research. 2012; 94: 73-83.
4. Hayes B, Goddard M, Genome-wide association and genomic selection in animal breedingThis article is one of a selection of papers from the conference "Exploiting Genome-wide Association in Oilseed Brassicas: a model for genetic improvement of major OECD crops for sustainable farming", Genome. 2010; 53: 876-883.
5. Sikorska K, Lesaffre E, Groenen P F, Eilers P H, GWAS on your notebook: fast semi-parallel linear and logistic regression for genome-wide association studies, BMC Bioinformatics. 2013; 14: 1-11.
6. Zhou X, Stephens M, Genome-wide efficient mixed-model analysis for association studies, Nat Genet. 2012; 44: 821-824.
7. Korte A, Vilhjálmsson B J, Segura V, Platt A, Long Q, Nordborg M, A mixed-model approach for genome-wide association studies of correlated traits in structured populations, Nature genetics. 2012; 44: 1066-1071.
8. O'reilly P F, Hoggart C J, Pomyen Y, Calboli F C, Elliott P, Jarvelin M-R, Coin L J, MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS, PloS one. 2012; 7: e34861.
9. Falconer D S, Mackay T F, Introduction to quantitative genetics (4th edn), Trends in Genetics. 1996; 12: 280.
10. Wright S, Correlation and causation, Journal of agricultural research. 1921; 20: 557-585.
11. Gianola D, Sorensen D, Quantitative genetic models for describing simultaneous and recursive relationships between phenotypes, Genetics. 2004; 167: 1407-1424.
12. Varona L, Sorensen D, Thompson R, Analysis of Litter Size and Average Litter Weight in Pigs Using a Recursive Model, Genetics. 2007; 177: 1791.
13. Kumar S, Agrawal S, Disease-oriented Causal Networks, Encyclopedia of Systems Biology, Springer, 2013, pp. 593-594.
14. Schadt E E, Chapter 10 - Reconstructing Causal Network Models of Human Disease A2 - Lehner, Thomas, in: Miller B L, State M W (Eds.), Genomics, Circuits, and Pathways in Clinical Neuropsychiatry, Academic Press, San Diego, 2016, pp. 141-160.
15. Rosa G J, Valente B D, De Los Campos G, Wu X-L, Gianola D, Silva M A, Inferring causal phenotype networks using structural equation models, Genetics Selection Evolution. 2011; 43: 6.
16. Li R, Tsaih S-W, Shockley K, Stylianou I M, Wergedal J, Paigen B, Churchill G A, Structural model analysis of multiple quantitative traits, PLoS genetics. 2006; 2: e114.
17. Momen M, Mehrgardi A A, Sheikhy A, Esmailizadeh A, Fozi M A, Kranis A, Valente B D, Rosa G J M, Gianola D, A predictive assessment of genetic correlations between traits in chickens using markers, Genetics Selection Evolution. 2017; 49: 16.
18. Kranis A, Gheyas A A, Boschiero C, Turner F, Yu L, Smith S, Talbot R, Pirani A, Brew F, Kaiser P, Development of a high density 600K SNP genotyping array for chicken, BMC genomics. 2013; 14: 59.

19. Browning S R, Browning B L, Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering, Am J Hum Genet. 2007; 81.
20. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M A, Bender D, Maller J, Sklar P, De Bakker P I, Daly M J, PLINK: a tool set for whole-genome association and population-based linkage analyses, The American Journal of Human Genetics. 2007; 81: 559-575.
21. Listgarten J, Lippert C, Kadie C M, Davidson R I, Eskin E, Heckerman D, Improved linear mixed models for genome-wide association studies, Nat Meth. 2012; 9: 525-526.
22. Gianola D, Fariello M I, Naya H, Schön C-C, Genome-Wide Association Studies with a Genomic Relationship Matrix: A Case Study with Wheat and Arabidopsis, G3: Genes, Genomes, Genetics. 2016; 6: 3241-3256.
23. Kennedy B W, Quinton M, Van Arendonk J a M, Estimation of effects of single genes on quantitative traits, Journal of Animal Science. 1992; 70: 2000-2012.
24. Vanraden P, Efficient methods to compute genomic predictions, J Dairy Sci. 2008; 91: 4414 - 4423.
25. Yang J, Benyamin B, Mcevoy B, Gordon S, Henders A, Nyholt D, Common SNPs explain a large proportion of the heritability for human height, Nat Genet. 2010; 42: 565 - 569.
26. Forni S, Aguilar I, Misztal I, Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information, Genet Sel Evol. 2011; 43: 1.
27. Wu X L, Heringstad B, Gianola D, Bayesian structural equation models for inferring relationships between phenotypes: a review of methodology, identifiability, and applications, Journal of Animal Breeding and Genetics. 2010; 127: 3-15.
28. Jamrozik J, Schaeffer L, Alternative parameterizations of the multiple-trait random regression model for milk yield and somatic cell score via recursive links between phenotypes, 2011.
29. Pearl J, Causal inference in statistics: An overview, Statistics surveys. 2009; 3: 96-146.
30. Jiang G, Chakraborty A, Wang Z, Boustani M, Liu Y, Skaar T, Li L, New aQTL SNPs for the CYP2D6 identified by a novel mediation analysis of genome-wide SNP arrays, gene expression arrays, and CYP2D6 activity, BioMed research international. 2013; 2013.
31. Nock N, Zhang L, Evaluating aggregate effects of rare and common variants in the 1000 Genomes Project exon sequencing data using latent variable structural equation modeling, in BMC proceedings, 2011, Vol. 5, BioMed Central, pp. S47.
32. Valente B D, Rosa G J, De Los Campos G, Gianola D, Silva M A, Searching for recursive causal structures in multivariate quantitative genetics mixed models, Genetics. 2010; 185: 633-644.
33. Henderson C, Quaas R, Multiple trait evaluation using relatives' records, Journal of Animal Science. 1976; 43: 1188-1197.
34. Valente B D, Rosa G J, Gianola D, Wu X-L, Weigel K, Is structural equation modeling advantageous for the genetic improvement of multiple traits?, Genetics. 2013; 194: 561-572.
35. Van Goor A, Bolek K J, Ashwell C M, Persia M E, Rothschild M F, Schmidt C J, Lamont S J, Identification of quantitative trait loci for body temperature, body weight, breast yield, and digestibility in an advanced intercross line of chickens under heat stress, Genetics Selection Evolution. 2015; 47: 96.
36. Sun Y, Zhao G, Liu R, Zheng M, Hu Y, Wu D, Zhang L, Li P, Wen J, The identification of 14 new genes for meat quality traits in chicken using a genome-wide association study, BMC Genomics. 2013; 14: 458.

534 37. Gao H, Zhang T, Wu Y, Wu Y, Jiang L, Zhan J, Li J, Yang R, Multiple-trait genome-wide
535 association study based on principal component analysis for residual covariance matrix, Heredity.
536 2014; 113: 526-532.
537 38. Wu B, Pankow J S, Genome-wide association test of multiple continuous traits using imputed
538 SNPs, Statistics and its interface. 2017; 10: 379.
539 39. Bellavia A, Valeri L, Decomposition of the total effect in the presence of multiple mediators
540 and interactions, American journal of epidemiology. 2017.
541 40. Barfield R, Shen J, Just A C, Vokonas P S, Schwartz J, Baccarelli A A, Vanderweele T J, Lin
542 X, Testing for the indirect effect under the null for genome-wide mediation analyses, Genetic
543 epidemiology. 2017; 41: 824-833.
544 41. Mi X, Eskridge K, Wang D, Baenziger P S, Campbell B T, Gill K S, Dweikat I, Bovaird J,
545 Regression-based multi-trait QTL mapping using a structural equation model, Statistical
546 applications in genetics and molecular biology. 2010; 9.
547 42. Wang H, Van Eeuwijk F A, A New Method to Infer Causal Phenotype Networks Using QTL
548 and Phenotypic Information, PLOS ONE. 2014; 9: e103997.

549

550

551

552

# Tables

**Table 1. Model comparison criteria: logarithm of the restricted maximum likelihood function (log L), Akaike's information criteria (AIC), Schwarz Bayesian information criteria (BIC) to evaluate model fit for two MTM and four SEM models.**

| Model | Maximum log L | -1/2 AIC | -1/2 BIC |
|---|---|---|---|
| MTM-A | -7093.480 | -7105.48 | -7142.436 |
| SEM-A75 | -7098.370 | -7110.415 | -7147.321 |
| SEM-A85 | -7095.188 | -7107.188 | -7144.143 |
| SEM-A95 | -7097.517 | -7109.517 | -7146.470 |
| MTM-G | -6529.270 | -6541.276 | -6578.232 |
| SEM-G75 | -6537.391 | -6549.391 | -6586.34 |

A: pedigree-based relationship matix, G: VanRaden's mrker-based relationship matrix

564

565

566

**Table 2 Estimates of three causal structural coefficients ($\lambda$) derived from four different structural models. BM: breast meat. BW: body weight. HHP: hen-house production. SEM-75: HPD > 0.75. SEM-G75: HPD > 0.75. SEM-A85: HPD > 0.85. SEM-A95: HPD > 0.95.**

| Path | Structural Models | | | |
|---|---|---|---|---|
| | SEM-A75 | SEM-G75 | SEM-A85 | SEM-A95 |
| $\lambda_{BM\rightarrow BW}(\lambda_{21})$ | 2.13 | 2.19 | 2.14 | 2.14 |
| $\lambda_{BM\rightarrow HHP}(\lambda_{31})$ | -0.17 | -0.28 | *** | *** |
| $\lambda_{BW\rightarrow HHP}(\lambda_{32})$ | -0.27 | -0.096 | -0.31 | *** |

567

**Table 3. Six most significant SNPs selected according** $-\log_{10}$ **(p-values) and their effects, using the full recursive SEM (SEM-A75) and MTM (MTM-A75).** $d_{S_j \to y(HHP)}$, $ind_{S_j \to y(HHp)}$, $T_{S_{j \to y(HHP)}}$ **and** $MTM_{S_{j \to y_{(HHP)}}}$**, represents, direct, indirect and overall from SEM and MTM effects of** j-*th* **SNP on HHP.**

| | CHR | SNP name | candidate genes | $-\log_{10}$ (p-values) | | | | Type of SNP effect | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $d_{S_j \to y(HHP)}$ | $ind_{S_j \to y(HHP)}$ | $T_{S_{j \to y(HHP)}}$ | $MTM_{S_{j \to y(HHP)}}$ | $d_{S_j \to y(HHP)}$ | $ind_{S_j \to y(HHP)}$ | $T_{S_{j \to y(HHP)}}$ | $MTM_{S_{j \to y(HHP)}}$ |
| Top SNPs for direct effects | 14 | Gga_rs313620413 | GRIN2A | **7.4242** | 0.1499 | 9.6599 | 7.4525 | -5.7827 | -0.0498 | -5.8326 | -5.78511 |
| | 7 | Gga_rs16591372 | OLA1 | **7.0868** | 0.2220 | 9.0119 | 6.9783 | -22.5681 | 0.2983 | -22.2698 | -22.3520 |
| | 3 | Gga_rs15390496 | EPHA7 | **7.0209** | 0.2214 | 8.6122 | 7.0297 | -22.4233 | -0.2149 | -22.6382 | -22.4098 |
| | 1 | Gga_rs314001234 | ----- | **7.0147** | 1.1067 | 9.0710 | 7.1653 | -26.6538 | -0.9018 | -27.5556 | -26.9360 |
| | 7 | Gga_rs315626061 | ----- | **6.8300** | 0.3360 | 8.9974 | 6.9529 | 5.1767 | 0.0910 | 5.26783 | 5.22295 |
| | 7 | Gga_rs316509306 | ----- | **6.8241** | 0.3442 | 8.9952 | 6.9485 | 5.1742 | 0.0928 | 5.267116 | 5.22105 |
| Top SNPs for indirect effects | 4 | Gga_rs316082590 | LOC422264 | 0.7137 | **3.6868** | 0.4754 | 0.5696 | -1.2913 | 0.4505 | -0.84073 | -1.07339 |
| | 4 | Gga_rs313358833 | LOC422265 | 0.6449 | **3.2345** | 0.4310 | 0.5202 | -1.2067 | 0.4235 | -0.78322 | -1.01618 |
| | 4 | Gga_rs314615897 | MAEA | 0.1170 | **2.9505** | 0.0474 | 0.0387 | -0.2799 | 0.3853 | 0.105456 | -0.09807 |
| | 1 | Gga_rs15301842 | ----- | 0.0393 | **2.9408** | 0.1436 | 0.0149 | -0.1301 | 0.5053 | 0.375199 | 0.050463 |
| | 1 | Gga_rs314551852 | ----- | 0.0632 | **2.8858** | 0.1100 | 0.0065 | -0.2038 | 0.4994 | 0.295514 | -0.02218 |
| | 1 | Gga_rs317379325 | ----- | 0.1599 | **2.8473** | 0.0070 | 0.0931 | -0.4789 | 0.5000 | 0.021148 | -0.29321 |
| Overall effects | 14 | Gga_rs313620413 | GRIN2A | 7.4242 | 0.1499 | **9.6599** | 7.4525 | -5.7827 | -0.0498 | -5.83262 | -5.7851 |
| | 1 | Gga_rs314001234 | ----- | 7.0147 | 1.1067 | **9.0710** | 7.1653 | -26.653 | -0.9018 | -27.5556 | -26.9360 |
| | 7 | Gga_rs315626061 | ----- | 7.0868 | 0.2220 | **9.0119** | 6.9783 | -22.5681 | 0.2983 | -22.2698 | -22.3520 |
| | 7 | Gga_rs315626061 | ----- | 6.8300 | 0.3360 | **8.9974** | 6.9529 | 5.1767 | 0.0910 | 5.26783 | 5.2229 |
| | 7 | Gga_rs316509306 | ----- | 6.8241 | 0.3442 | **8.9952** | 6.9485 | 5.1742 | 0.0928 | 5.267116 | 5.2210 |
| | 7 | Gga_rs15850017 | ZNF385B | 6.6582 | 0.0499 | **8.6397** | 6.6176 | -20.8591 | -0.0718 | -20.9310 | -20.7681 |
| MTM | 14 | Gga_rs313620413 | GRIN2A | 7.4242 | 0.1499 | 9.6599 | **7.4525** | -5.7827 | -0.0498 | -5.8326 | -5.7851 |
| | 1 | Gga_rs314001234 | ----- | 7.0147 | 1.1067 | 9.0710 | **7.1653** | -26.6538 | -0.9018 | -27.5556 | -26.936 |
| | 3 | Gga_rs15390496 | EPHA7 | 7.0209 | 0.2214 | 8.6122 | **7.0297** | -22.4233 | -0.2149 | -22.6382 | -22.4098 |
| | 7 | Gga_rs16591372 | OLA1 | 7.0868 | 0.2220 | 9.0119 | **6.9783** | -22.5681 | 0.2983 | -22.2698 | -22.352 |
| | 7 | Gga_rs315626061 | ----- | 6.8300 | 0.3360 | 8.9974 | **6.9529** | 5.1767 | 0.0910 | 5.26780 | 5.2229 |
| | 7 | Gga_rs316509306 | ----- | 6.8241 | 0.3442 | 8.9952 | **6.9485** | 5.1742 | 0.0928 | 5.2671 | 5.2210 |

The bold values are $-\log_{10}$ (corrected p-value) for each type of significant SNP effects categories.

568

# Figures

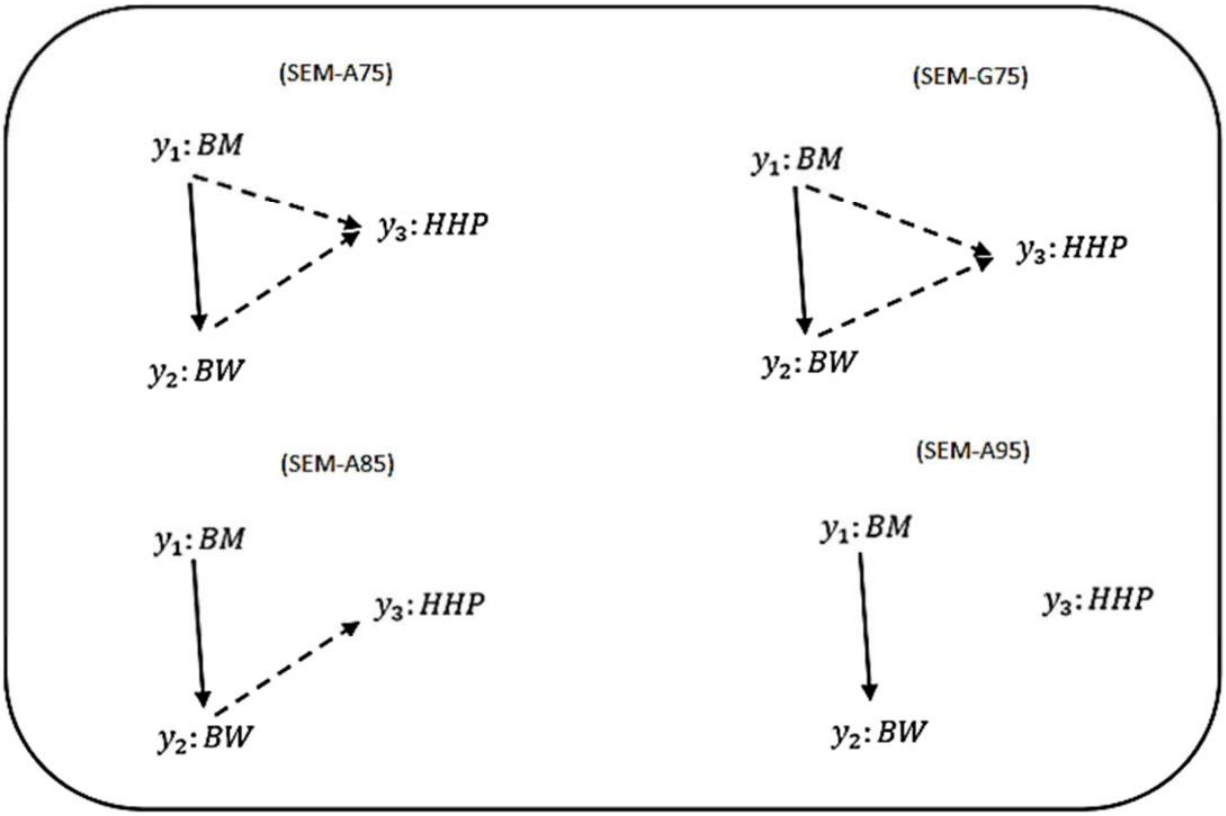



**Figure 1 Causal graphs inferred using the IC algorithm among three traits: breast meat (BM), body weight (BW) and hen-house production (HHP) in the chicken data.** SEM-A75 and SEM-G75 were the inferred fully recursive causal structures with HPD > 0.75 and corrected for genetic confounder using **A** (pedigree-based) and **G** (marker-based) matrices. SEM-A85 and SEM-A95 were obtained with HPD > 0.85 and HPD > 0.95, respectively, corrected with **A**. Arrows indicate direction of causal relationships. Dashed lines indicate negative coefficients, and the continuous arrows indicate positive coefficients.

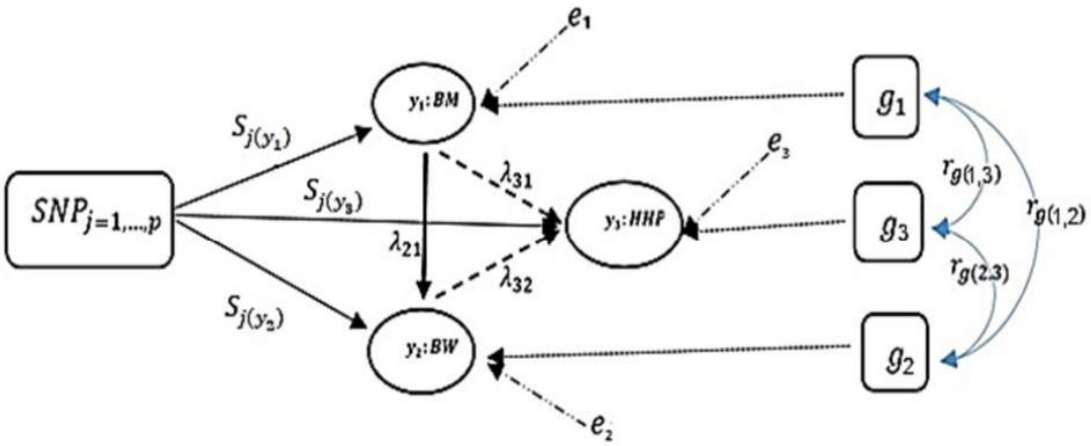



**Figure 2 A diagram for causal path analysis of SNP effects in a fully recursive structural equation model for three traits, $p$ exogenous independent SNP variables, and three correlated polygenic effects.** Arrows indicate the direction of causal effects and dashed lines represent associations among the three phenotypes. Genetic correlation between traits ($r_g$), polygenic effects ($g_l$), environmental effect on trait $l$ ($e_l$), effects of $j$ th SNP on $l$ th trait ($S_{j(y_l)}$), and recursive effect of phenotype $l'$ on phenotype $l$ ($\lambda_{l,l'}$).

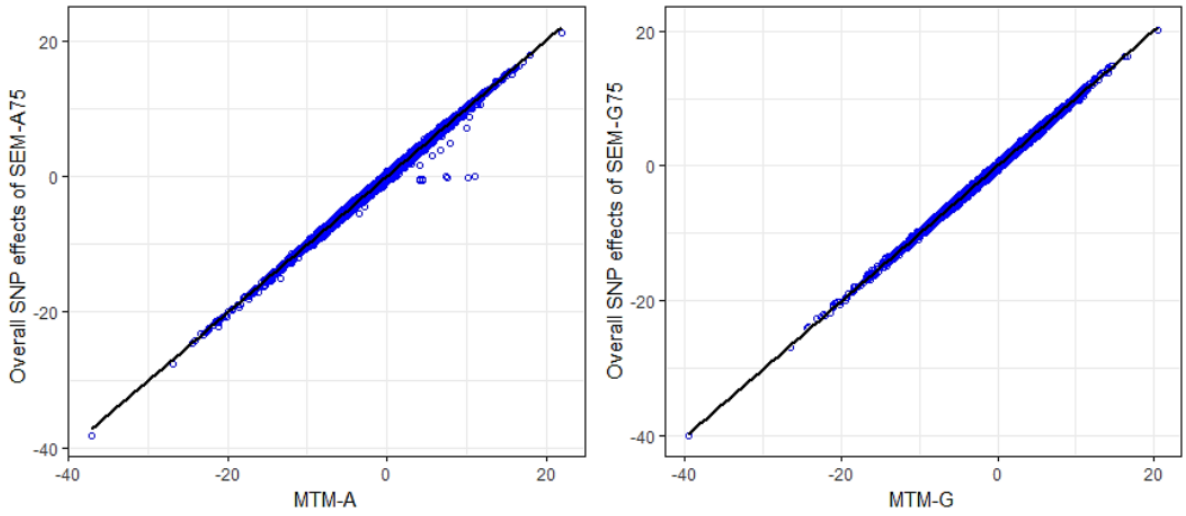
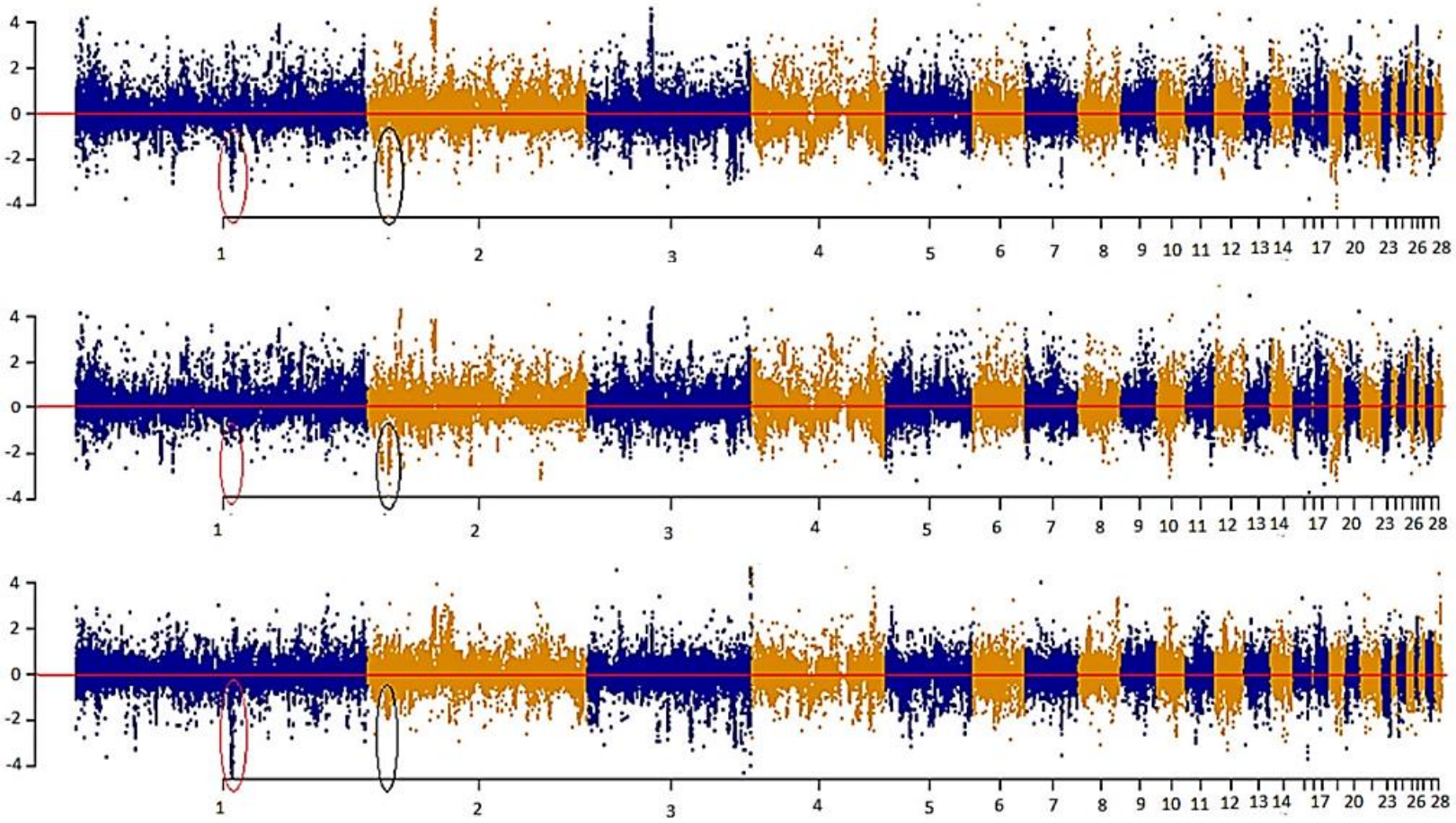



**Figure 3 Comparison of multiple trait (MTM) and fully recursive overall SNP effects obtained with A (pedigree-based) and G (marker-based) from structural equation modeling (SEM)-based GWAS.** Overall effects in SEM are the sum of all direct and indirect effects. HHP: hen-house egg production.

615




616

617 **Figure 4 Manhattan plot showing overall, direct, and indirect SNP effects using a full recursive model based on A matrix for**
618 **body weight (BW).**

619