Current Genomics, 2016, *17*, 439-443

Analytical Models For Genetics of Human Traits Influenced By Sex

Chaeyoung Lee*

Department of Bioinformatics and Life Science, Soongsil University, Seoul, Republic of Korea

C. Lee

Abstract: Analytical models usually assume an additive sex effect by treating it as a covariate to identify genetic associations with sex-influenced traits. Their underlying assumptions are violated by ignoring interactions of sex with genetic factors and heterogeneous genetic effects by sex. Methods to deal with the problems are compared and discussed in this article. Especially, heterogeneity of genetic variance by sex can be assessed employing a mixed model with genetic relationship matrix constructed from genome-wide nucleotide variant information. Estimating genetic architecture of each sex would help understand different prevalence, course, and severity of complex diseases between women and men in the era of personalized medicine.

Keywords: Complex trait, Genetic heterogeneity, Genetic relationship matrix, Genetic variance, Genome-wide association study, Mixed model.

Received on: June 01, 2015 **Revised on: January 18, 2016** **Accepted on: March 14, 2016**

1. INTRODUCTION

Sex is determined by genetic information in human. It extensively controls physical, psychological, and behavioral traits which are also influenced by genetic information. Sex-related genetics may be categorized by how sex affects phenotypic traits, including sex itself as a trait (i.e., sex determination), sex-linked trait, sex-limited trait, and sex-influenced trait. Sex determination depends on the existence of Y chromosome, especially the sex-determining region Y gene on the chromosome [1]. Males with translocation of the gene can have chromosomes of XX. Females with mutation of the gene can have chromosomes of XY. Sex defined as a binary trait has been proven to be more complex as accumulating research reveals its genetics. Different molecular mechanisms operate in sexual development not only by the gene, but also by other autosomal genes that can control hormone balances [2]. Sex might be expressed in terms of an ordinal trait. Sex can be categorized as typical female, female with subtle variation, female with moderate variation, sex with XX and testicular disorder of sex development (DSD), sex with ovotesticular DSD, sex with XY and DSD, male with moderate variation, male with subtle variation, and typical male [3].

Sex-linked traits are determined by genes located on the sex chromosomes. Sex-limited traits are determined by genes located on autosomes and express only in one sex. While these traits are responsible considerably for sexual dimorphism, sex-influenced traits do not show distinctive expression between women and men. Numerous analytical models have treated sex as an essential factor. However, dissection of genetic factors associated with sex has been limitedly explored to examine sex-influenced traits. This is because many of the traits are intricate that they can hardly be explained by a few genes or sex-specific genes (mitochondrial and sex-chromosomal genes). Also, few appropriate analytical methods have been employed to explain sex-influenced traits. Here, I would like to discuss assumption violations in identifying genetic factors for sex-influenced traits and propose appropriate methods to overcome problems.

2. GENETIC ANALYSIS OF SEX-INFLUENCED TRAITS

Most analytical models for sex-influenced traits assume an additive sex effect by treating it as a covariate in models or adjusting it preliminarily (Fig. **1A**). This leads to inference on genetic and environmental effects associated with the term "sex" in the lump [4]. Sex has been interchangeably used with gender contrary to their original meanings of biological and social characteristics of women and men, either of which has connotations of biology and sociology. Sex is more likely to possess biological connotations. Gender is more likely to possess social connotations. A problem is that the analytical models most likely ignore interactions of sex with genetic factors. 'Sex' considerably depends on hormones which control the whole body. Thus, genetic effects related to hormones are likely to interact with other genetic effects. Furthermore, different social environments by 'gender' interact with genetic effects. This is the reason why gene-by-sex interaction might be explained as gene-by-gene interaction and/or gene-by-environment interaction in analytical models. The gene-by-sex interaction has been shown in hypertension, schizophrenia, rheumatoid arthritis, and recombination rate [5, 6]. From an evolutionary point of

*Address correspondence to this author at the Department of Bioinformatics and Life Science, Soongsil University, 369 Sangdo-ro, Dongjak-gu, Seoul 156-743, Republic of Korea; Tel/Fax: +82-2-820-0455, +82-2-824-4383; E-mail: clee@ssu.ac.kr

1875-5488/16 $58.00+.00

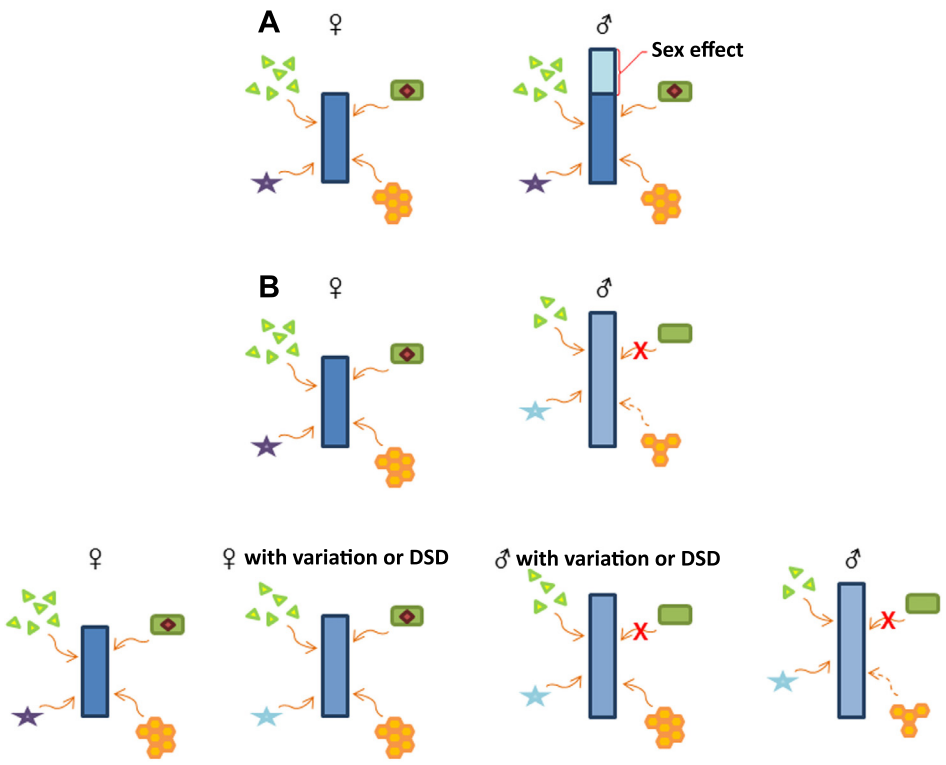

Fig. (1). Schematic concepts for analyzing quantitative trait with sex as a covariate (**A**) and for analyzing quantitative trait with data partitioned by sex (**B**) or by a wider spectrum of sex (**C**). Bar size indicates quantitative phenotype. Circular and polygonal figures indicate genetic products. Environmental effect is not presented to avoid confounding. DSD: disorder of sex development.

view, the gene-by-sex interaction can be produced by sex-specific or sexually antagonistic selection [7]. Nevertheless, the gene-by-sex interaction has been ignored mostly even in studies designed to identify other gene-by-gene or gene-by-environment interaction [8, 9].

Most analytical models assume homogeneous variances by sex. Specifically, homogeneous residual variances are assumed for fixed models. Homogeneous genetic variances are additionally assumed for mixed models. These assumptions can be violated for sex-influenced traits. Sometimes variances are proportional to their means. In this case, heterogeneity can be controlled by using standardized values. However, the homogeneity problem cannot be overcome by this approach for many sex-influenced traits. For example, heritability (a portion of phenotypic variance explained by genetic effects) of men was larger than that of women for body mass index and triglyceride level [10]. This heterogeneous heritability must be attributed to difference in individual genetic effects between women and men. Furthermore, genetic correlation between women and men was far from one for all quantitative traits analyzed using single nucleotide variants selected by significance threshold of 0.01: 0.76 for body mass index, 0.80 for waist-to-hip ratio, 0.73 for pulse pressure, 0.78 for high-density lipoprotein cholesterol, 0.65 for triglyceride, 0.79 for low-density lipoprotein cholesterol, and 0.73 for glucose level. Such heterogeneous genetic effects by sex are diluted by analytical models assuming their homogeneity. More seriously, this can produce false negative genetic associations when genetic effects exist only in one sex.

3. SOLUTION TO VIOLATED ASSUMPTIONS

The problem of ignoring gene-by-sex interaction can be solved by including interaction term in the analytical model (Table **1B**). The problem of heterogeneous variance can be solved by introducing a scaling factor to dispersion parameters. More efficiently, both problems can be simultaneously solved by analyzing data by sex (Fig. **1B**, Table **1A**). The data partitioned by sex might be analyzed separately. Genetic effects can then be assessed by sex. Heterogeneity of genetic variance by sex is assessed employing a Hendersonian mixed model with genetic relationship matrix (Table **1C**). This matrix can be constructed by assessing genetic relationships among individuals using pedigree information or nucleotide variant information. Especially, polygenic relationship matrix can be constructed using a large number of single nucleotide polymorphisms in genome-wide association studies (GWAS) [11-13]. The mixed model methodology with polygenic covariance structure can control for population stratification which often produces spurious genetic associations in GWAS [14]. Simultaneous analysis of both sexes might be also feasible by bivariate mixed model analysis (Table **1D**) [15]. This model treats females as one trait and males as another trait. This enables us to estimate genetic correlation between females and males. A careful analytical model is needed to deal with polygenic effects of sex chromosomes. Polygenic effects of Y chromosome should be independently included in analytical model only for phenotypes of men because of its absence in women and thus absent genetic correlation between women and men. Polygenic effects of X chromosomes might be simultaneously assessed with

Table 1. Analytical models for estimating sex-specific genetic effects on complex traits.

Analytical Model[1]	F/M	F/B	Vg	Ve	Cg	PS
A: $y_m = g_m + e_m \quad \& \quad y_w = g_w + e_w$ $e_m \sim N(0, \sigma^2_{e_m}),\ e_w \sim N(0, \sigma^2_{e_w})$	F	F	N	Y	N	N
B: $y = s + g + (s \times g) + e$ $e \sim N(0, \sigma^2_e)$	F	F	N	N	N	N
C: $\mathbf{y}_m = \mathbf{X}_m\mathbf{B}_m + \mathbf{Z}_m\mathbf{p}_m + \mathbf{e}_m \quad \& \quad \mathbf{y}_w = \mathbf{X}_w\mathbf{B}_w + \mathbf{Z}_w\mathbf{p}_w + \mathbf{e}_w$ $\mathbf{p}_m \sim N(\mathbf{0}, \mathbf{A}_m\sigma^2_{p_m}),\ \mathbf{e}_m \sim N(\mathbf{0}, \mathbf{I}\sigma^2_{e_m})$ $\mathbf{p}_w \sim N(\mathbf{0}, \mathbf{A}_w\sigma^2_{p_w}),\ \mathbf{e}_w \sim N(\mathbf{0}, \mathbf{I}\sigma^2_{e_w})$	M	F	Y	Y	N	Y
D: $\begin{pmatrix}\mathbf{y}_m \\ \mathbf{y}_w\end{pmatrix} = \begin{pmatrix}\mathbf{X}_m & 0 \\ 0 & \mathbf{X}_w\end{pmatrix}\begin{pmatrix}\mathbf{B}_m \\ \mathbf{B}_w\end{pmatrix} + \begin{pmatrix}\mathbf{Z}_m & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_w\end{pmatrix}\begin{pmatrix}\mathbf{p}_m \\ \mathbf{p}_w\end{pmatrix} + \begin{pmatrix}\mathbf{e}_m \\ \mathbf{e}_w\end{pmatrix}$ $\begin{pmatrix}\mathbf{p}_m \\ \mathbf{p}_w\end{pmatrix} \sim N\left(\begin{bmatrix}\mathbf{0} \\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\mathbf{A}\sigma^2_{p_m} & \mathbf{A}\sigma_{p_m,p_w} \\ \mathbf{A}\sigma_{p_m,p_w} & \mathbf{A}\sigma^2_{p_w}\end{bmatrix}\right),\ \begin{pmatrix}\mathbf{e}_m \\ \mathbf{e}_w\end{pmatrix} \sim N\left(\begin{bmatrix}\mathbf{0} \\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\mathbf{I}\sigma^2_{e_m} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\sigma^2_{e_w}\end{bmatrix}\right)$	M	F	Y	Y	Y	Y
$f(\mathbf{B}, \mathbf{p}, \sigma^2_{p_m}, \sigma^2_{p_w}, \sigma_{p_m,p_w}, \sigma^2_{e_m}, \sigma^2_{e_w} \mid \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{A})$ E: $\propto f(\mathbf{y} \mid \mathbf{B}, \mathbf{p}, \sigma^2_{p_m}, \sigma^2_{p_w}, \sigma_{p_m,p_w}, \sigma^2_{e_m}, \sigma^2_{e_w}) \times f(\mathbf{p} \mid \sigma^2_{p_m}, \sigma^2_{p_w}, \sigma_{p_m,p_w})$ $\times f(\sigma^2_{p_m}, \sigma^2_{p_w}, \sigma_{p_m,p_w}) \times f(\sigma^2_{e_m}, \sigma^2_{e_w})$ $f(\mathbf{y} \mid \mathbf{B}, \mathbf{p}, \sigma^2_{p_m}, \sigma^2_{p_w}, \sigma_{p_m,p_w}, \sigma^2_{e_m}, \sigma^2_{e_w}) \sim N,\ f(\mathbf{p} \mid \sigma^2_{p_m}, \sigma^2_{p_w}, \sigma_{p_m,p_w}) \sim N$ $f(\sigma^2_{p_m}, \sigma^2_{p_w}, \sigma_{p_m,p_w}) \sim IW,\ f(\sigma^2_{e_m}, \sigma^2_{e_w}) \sim IG$	M	B	Y	Y	Y	Y

[1]Joint posterior density is presented for Bayesian approach. Fixed models are presented in scalar forms. Mixed models are presented in matrix and vector forms (in bold). y: sex-influenced phenotype, g: SNP effect, e: residual, m(w) in subscript: men(women), s: sex effect, **B** : fixed effects (including SNP effect), p: polygenic effect, **A**: genetic relationship matrix, **I**: identity matrix, **X** and **Z**: design matrices, $\boldsymbol{\sigma}^2_v$: variance component for v, $\boldsymbol{\sigma}_{v1,v2}$: covariance component between v1 and v2, N: Normal distribution, IW: inverse Wishart distribution, IG: inverse Gamma distribution. F/M: fixed model/mixed model, F/B: frequentist/Bayesian, Vg: sex-specific genetic variance (Yes/No), Ve: sex-specific residual variance (Yes/No), Cg: genetic correlation between men and women (Yes/No), PS: control of population stratification (Yes/No).

autosomal polygenic effects or independently assessed, depending on the assumption about two alleles from homologous X chromosomes. This is due to the imbalance in dose between women and men. The sex-stratified bivariate mixed model can be extended to analyses for multiple traits [16]. For example, an analytical model for two diseases of schizophrenia and rheumatoid arthritis has been designed to treat schizophrenia-male, schizophrenia-female, rheumatoid arthritis-male, and rheumatoid arthritis-female as four different traits [6].

A caution with the use of mixed model in estimating fixed and random effects should be attached on its underlying statistical property. Many researchers believe that the best linear unbiased estimation (BLUE) and best linear unbiased prediction (BLUP) can be achieved respectively for fixed and random effects from mixed model analyses. The BLUE and BLUP are assumed to have known variance and covariance components of random effects and residuals. However, variance and covariance components are not known for specific data in reality. Variance and covariance components should be estimated using the same data as used for BLUE and BLUP, typically employing restricted maximum likelihood (REML). That is, the REML estimates are utilized instead of known variance and covariance components. Strictly speaking, fixed and random effects are not estimated as BLUE and BLUP anymore. A Bayesian approach with Gibbs sampling as a Markov chain Monte Carlo could overcome the problem by estimating variance and covariance components and BLUE genetic effects simultaneously (Table **1E**). The Bayesian approach results in exact distributions on parameters and reduced sensitivity to outliers [17]. These merits help draw better inference on genetic parameters which often involve high-degree complexity. Another issue is unrealistic assumption of homogeneous contributions to genetic variance, producing potential bias in genetic variance [18]. Heterogeneous effects can be introduced by a Bayesian method with priors on numbers of ma-

jor SNPs [19] or by penalty based on functions of each SNP effect [20]. Such reasonable approaches should be incorporated in the analysis to identify more accurate sets of genetic variants by sex and their heterogeneous effect sizes. Genetic architecture of a complex trait can be constructed by sex and further extended by more stratified sex as shown in (Fig. **1C**). Furthermore, the analysis would provide sex-dependent genetic potential of each individual for specific complex phenotype. This implies that genetic value inherited from an individual to daughter is different from that inherited from the same individual to son.

4. CLOSING REMARKS

Criticism might be raised for reduced statistical power because sample size decreases in half from partitioning data by sex. Nevertheless, more accurate estimates may be obtained by scrutinizing genetic heterogeneity by sex. False positive and negative associations may be reduced accordingly. Furthermore, the burden of reduced power is dramatically improved as emerging technologies for sequencing DNA are in rapid progress. The cost has been substantially decreasing. It is time to move towards estimating genetic architecture of each sex to understand genetics of sex itself and complex traits related to sex. Genetic effects of each sex should also be estimated for transient traits such as RNA and protein levels. For example, identification of expression quantitative trait loci can be conducted by sex to show sex-dependent gene regulation, which would help understand underlying biological mechanisms of sex-influenced traits.

Research efforts identifying sex-dependent genetic factors of diseases would provide insights on genetic dissection to explain different prevalence, course, and severity of complex diseases between women and men in the era of personalized medicine. This may make it possible to prescribe different health-seeking behavior between women and men. Ultimately, heterogeneous genetic architecture between women and men will contribute healthier life in the future through in-depth understanding of underlying determinants on their health inequalities.

CONFLICT OF INTEREST

The author confirms that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by a grant (Grant No. NRF-2012M3A9D1054705) funded by the National Research Foundation of Korea, the Ministry of Education, Science, and Technology.

REFERENCES

[1] Sinclair, A.H.; Berta, P.; Palmer, M.S.; Hawkins, J.R.; Griffiths, B.L., Smith, M.J.; Foster, J.W.; Frischauf, A.M.; Lovell-Badge, R.; Goodfellow, P.N. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature*, **1990**, *346*, 240-244.

[2] Arbeitman, M.N.; Kopp, A.; Siegal, M.L.; Doren, M.V. The genetics of sex: exploring differences. *Genetics*, **2014**, *197*, 527-529.

[3] Ainsworth, C. Sex redefined. *Nature*, **2015**, *518*, 288-291.

[4] Short, S.E.; Yang, Y.C.; Jenkins, T.M. Sex, gender, genetics, and health. *Am. J. Public Health*, **2013**, *103*, S93-S101.

[5] Ober, C.; Loisel, D.A.; Gilad, Y. Sex-specific genetic architecture of human disease. *Nat. Rev. Genet.,* **2008**, *9*, 911-922.

[6] Lee, S.H.; Byrne, E.M.; Hultman, C.M.; Kähler, A.; Vinkhuyzen, A.A.; Ripke, S.; Andreassen, O.A.; Frisell, T.; Gusev, A.; Hu, X.; Karlsson, R.; Mantzioris, V.X.; McGrath, J.J.; Mehta, D.; Stahl, E.A.; Zhao, Q.; Kendler, K.S.; Sullivan, P.F.; Price, A.L.; O'Donovan, M.; Okada, Y.; Mowry, B.J.; Raychaudhuri, S.; Wray, N.R.; Schizophrenia Working Group of the Psychiatric Genomics Consortium and Rheumatoid Arthritis Consortium International; Schizophrenia Working Group of the Psychiatric Genomics Consortium authors; Byerley, W.; Cahn, W.; Cantor, R.M.; Cichon, S.; Cormican, P.; Curtis, D.; Djurovic, S.; Escott-Price, V.; Gejman, P.V.; Georgieva, L.; Giegling, I.; Hansen, T.F.; Ingason, A.; Kim, Y.; Konte, B.; Lee, P.H.; McIntosh, A.; McQuillin, A.; Morris, D.W.; Nöthen, M.M.; O'Dushlaine, C.; Olincy, A.; Olsen, L.; Pato, C.N.; Pato, M.T.; Pickard, B.S.; Posthuma, D.; Rasmussen, H.B.; Rietschel, M.; Rujescu, D.; Schulze, T.G.; Silverman, J.M.; Thirumalai, S.; Werge, T.; Schizophrenia Working Group of the Psychiatric Genomics Consortium collaborators; Agartz, I.; Amin, F.; Azevedo, M.H.; Bass, N.; Black, D.W.; Blackwood, D.H.; Bruggeman, R.; Buccola, N.G.; Choudhury, K.; Cloninger, R.C.; Corvin, A.; Craddock, N.; Daly, M.J.; Datta, S.; Donohoe, G.J.; Duan, J.; Dudbridge, F.; Fanous, A.; Freedman, R.; Freimer, N.B.; Friedl, M.; Gill, M.; Gurling, H.; De Haan, L.; Hamshere, M.L.; Hartmann, A.M.; Holmans, P.A.; Kahn, R.S.; Keller, M.C.; Kenny, E.; Kirov, G.K.; Krabbendam, L.; Krasucki, R.; Lawrence, J.; Lencz, T.; Levinson, D.F.; Lieberman, J.A.; Lin, D.Y.; Linszen, D.H.; Magnusson, P.K.; Maier, W.; Malhotra, A.K.; Mattheisen, M.; Mattingsdal, M.; McCarroll, S.A.; Medeiros, H.; Melle, I.; Milanova, V.; Myin-Germeys, I.; Neale, B.M.; Ophoff, R.A.; Owen, M.J.; Pimm, J.; Purcell, S.M.; Puri, V.; Quested, D.J.; Rossin, L.; Ruderfer, D.; Sanders, A.R.; Shi, J.; Sklar, P.; St Clair, D.; Scott Stroup, T.; Van Os, J.; Visscher, P.M.; Wiersma, D.; Zammit, S.; Rheumatoid Arthritis Consortium International authors; Louis Bridges, S.; Choi, H.K.; Coenen, M.J.; de Vries, N.; Dieud, P.; Greenberg, J.D.; Huizinga, T.W.; Padyukov, L.; Siminovitch, K.A.; Tak, P.P.; Worthington, J.; Rheumatoid Arthritis Consortium International collaborators; De Jager, P.L.; Denny, J.C.; Gregersen, P.K.; Klareskog, L.; Mariette, X.; Plenge, R.M.; van Laar, M.; van Riel, P. New data and an old puzzle: the negative association between schizophrenia and rheumatoid arthritis. *Int. J. Epidemiol.*, **2015**, *44*(5).

[7] Gilks, W.P.; Abbott, J.K.; Morrow, E.H. Sex differences in disease genetics: evidence, evolution, and detection. *Trends Genet.,* **2014**, *30*, 453-463.

[8] Kong, M.; Kim, Y.; Lee, C. A strong synergistic epistasis between FAM134B and TNFRSF19 on the susceptibility to vascular dementia. *Psychiatr. Genet.,* **2011**, *21*, 37-41.

[9] Garver, W.S. Gene-diet interactions in childhood obesity. *Curr. Genomics*, **2011**, *12*, 180-189.

[10] Lee, D.; Lee, C. Age- and gender-dependent heterogeneous proportion of variation explained by SNPs in quantitative traits reflecting human health. *AGE*, **2015**, *37*, 19.

[11] Kang, H.M.; Sul, J.H.; Service, S.K.; Zaitlen, N.A.; Kong, S.; Freimer, N.B.; Sabatti, C.; Eskin, E. Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.,* **2010**, *42*, 348-354.

[12] Zhang, Z.; Ersoz, E.; Lai, C.Q.; Todhunter, R.J.; Tiwari, H.K.; Gore, M.A.; Bradbury, P.J.; Yu, J.; Arnett, D.K.; Ordovas, J.M.; Buckler, E.S. Mixed linear model approach adapted for genomewide association studies. *Nat. Genet.,* **2010**, *42*, 355-360.

[13] Yang, J.; Lee, S.H.; Goddard, M.E.; Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.,* **2011**, *88*, 76-82.

[14] Shin, J.; Lee, C. A mixed model reduces spurious genetic associations produced by population stratification in genome-wide association studies. *Genomics*, **2015**, *105*, 191-196.

[15] Lee, S.H.; Yang, J.; Goddard, M.E.; Visscher, P.M.; Wray, N.R. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics*, **2012**, *28*, 2540-2542.

[16] Maier, R.; Moser, G.; Chen, G.B.; Ripke, S.; Cross-Disorder Working Group of the Psychiatric Genomics Consortium; Coryell, W.; Potash, J.B.; Scheftner, W.A.; Shi, J.; Weissman, M.M.; Hultman, C.M.; Landén, M.; Levinson, D.F.; Kendler, K.S.; Smoller, J.W.; Wray, N.R.; Lee, S.H. Joint analysis of psychiatric disorders in-

creases accuracy of risk prediction for schizophrenia, bipolar disorder, and major depressive disorder. *Am. J. Hum. Genet.*, **2015**, *96*, 283-294.

[17] Bayarri, M.J.; Berger, J.O. The interplay of Bayesian and frequentist analysis. *Stat. Sci.*, **2004**, *19*, 58-80.

[18] Ryoo, H.; Lee, C. Underestimation of heritability using a mixed model with a polygenic covariance structure in a genome-wide association study for complex traits. *Eur. J. Hum. Genet.,* **2014**, *22*, 851-854.

[19] Lee, S.H.; van der Werf, J.H.; Hayes, B.J. Goddard, M.E., Visscher, P.M. Predicting unobserved phenotypes for complex traits from whole-genome SNP data. *PLoS Genet.,* **2008**, *4*, e1000231.

[20] Yi, N.J.; Xu, S.H. Bayesian LASSO for quantitative trait loci mapping. *Genetics,* **2008**, *179*, 1045-1055.