

Methodology

Open Access

A computational model for sex-specific genetic architecture of complex traits in humans: Implications for mapping pain sensitivity

Chenguang Wang¹, Yun Cheng¹, Tian Liu¹, Qin Li¹, Roger B Fillingim², Margaret R Wallace³, Roland Staud³, Lee Kaplan³ and Rongling Wu*¹

Address: ¹Department of Statistics, University of Florida, Gainesville, FL 32611 USA, ²Department of Community Dentistry and Behavioral Science, University of Florida, Gainesville, FL 32611, USA and ³Department of Molecular Genetics and Microbiology, University of Florida, Gainesville, FL 32611, USA

Email: Chenguang Wang - cgwang@cog.ufl.edu; Yun Cheng - cyuf@ufl.edu; Tian Liu - liut2@gis.a-star.edu.sg; Qin Li - qli@biostat.ufl.edu; Roger B Fillingim - rfilling@ufl.edu; Margaret R Wallace - peggyw@ufl.edu; Roland Staud - Roland.Staud@medicine.ufl.edu; Lee Kaplan - lkaplan@ufl.edu; Rongling Wu* - rwu@ufl.edu

* Corresponding author

Published: 16 April 2008

Received: 29 November 2007

Molecular Pain 2008, 4:13 doi:10.1186/1744-8069-4-13

Accepted: 16 April 2008

This article is available from: <http://www.molecularpain.com/content/4/1/13>

© 2008 Wang et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Understanding differences in the genetic architecture of complex traits between the two sexes has significant implications for evolutionary studies and clinical diagnosis. However, our knowledge about sex-specific genetic architecture is limited largely because of a lack of analytical models that can detect and quantify the effects of sex on the complexity of quantitative genetic variation. Here, we derived a statistical model for mapping DNA sequence variants that contribute to sex-specific differences in allele frequencies, linkage disequilibria, and additive and dominance genetic effects due to haplotype diversity. This model allows a genome-wide search for functional haplotypes and the estimation and test of haplotype by sex interactions and sex-specific heritability. The model, validated by simulation studies, was used to detect sex-specific functional haplotypes that encode a pain sensitivity trait in humans. The model could have important implications for mapping complex trait genes and studying the detailed genetic architecture of sex-specific differences.

Background

Differences in males and females (sexual dimorphism) is ubiquitous in many biological aspects [1-3]. In humans, sexually dimorphic traits include those from morphological shapes and body size to brain development to disease susceptibility [4,5]. Substantial differences are also observed in sensitivities to pain and pain-killing drugs, and susceptibility to developing chronic pain between men and women [6-8]. All these sex-specific differences are due to varying expression of genes on the X/Y chromosome and autosomes, thought to result from differences in cellular and hormonal environments between the two sexes [9]. A growing body of research has been conducted

to elucidate the genetic control of sexual dimorphism in various complex phenotypes by gene mapping approaches [4,5,10,11]. Despite these efforts, however, little is known about the genetic architecture underlying sex-related variation in a quantitative trait.

Since sex is easily determined, the effects of sex on morphological, developmental and pathological traits can be directly observed. However, characterizing the impacts of sex on the genetic architecture of these traits has been challenged by a lack of powerful statistical approaches. The motivation of this article is to develop a statistical and computational model that can systematically search for

sex-specific genes contributing to quantitative variation and formulate testable hypotheses regarding the interplay between sexes and gene expression. Our model is principally different from those used in many previous studies that are aimed to detect sex-specific quantitative trait loci (QTLs) based on linkage or linkage disequilibrium analysis [2,4,5,12,13]. Our model will be founded on the statistical framework constructed by Liu et al. [14] to detect the effects and diversity of haplotypes constructed by single nucleotide polymorphisms (SNPs) that are genotyped at candidate genes or genome-wide [15]. Our model has been generalized to allow the test of sex differences in haplotype frequencies, allele frequencies and linkage disequilibria between different SNPs as well as additive and dominant effects of haplotypes on complex traits. It has power to identify sex-specific DNA sequence variants that encode complex phenotypes in men and women.

Model

Notation

Suppose there is a diversity of haplotypes constructed by two SNPs each with two alleles designated as 1 and 0. Let p and q be the 1-allele frequencies for the first and second SNP, respectively. Thus, the 0-allele frequencies at these two different SNPs will be $1 - p$ and $1 - q$. The two SNPs that are segregating in a natural human population form four haplotypes, [11], [10], [01], and [00], whose frequencies are constructed by allele frequencies and linkage disequilibrium (D) between the two SNPs, i.e.,

$$\begin{aligned} p_{11} &= pq + D \\ p_{10} &= p(1 - q) - D \\ p_{01} &= (1 - p)q - D \\ p_{00} &= (1 - p)(1 - q) + D \end{aligned}$$

The parameters contained in equation (1) can be used to describe some important aspects of the genetic structure and diversity of a natural population. Thus, differences in genetic architecture between the two different sexes can be characterized by these sex-specific parameters. Because it is easy to derive the closed forms for estimating haplotype frequencies [14], we will estimate the linkage disequilibrium from the estimated haplotype frequencies.

Let $\Theta_{M_p} = (p_{M_{11}}, p_{M_{10}}, p_{M_{01}}, p_{M_{00}})$ and $\Theta_{F_p} = (p_{F_{11}}, p_{F_{10}}, p_{F_{01}}, p_{F_{00}})$ be the vectors of haplotype frequencies among males and females, respectively. All the genotypes for the two SNPs are consistent with diplo-

types, except for the double heterozygote, 10/10, that belongs to a diplotype of either [11] [00] or [10] [01] (Table 1). Assuming that the population is at Hardy-Weinberg equilibrium, the frequency of a diplotype is

expressed as the product of the frequencies of the two haplotypes that construct it. Table 1 characterizes the differences in diplotype frequencies between the males and females.

If haplotypes triggers an effect on a quantitative trait, this means that at least one haplotype performs differently from the rest of the haplotypes. Without loss of generality, let haplotype [11] be such a distinct haplotype, called *risk haplotype*, designated as A . All the other non-risk haplotypes, [10], [01] and [00], are collectively expressed as \bar{A} . The risk and non-risk haplotypes form three *composite diplotypes* AA (symbolized as 2), $A\bar{A}$ (symbolized as 1) and $\bar{A}\bar{A}$ (symbolized as 0). The genotypic values of the three composite diplotypes may be different between the two sexes, arrayed in $(\mu_{M_2}, \mu_{M_1}, \mu_{M_0})$ for the males and $(\mu_{F_2}, \mu_{F_1}, \mu_{F_0})$ for the females, respectively. Let (a_M, d_M) and (a_F, d_F) be the additive and dominance genetic effects due to the risk and non-risk haplotypes in males and females, respectively.

Likelihoods

Assume that a total of n subjects (including n_M males and n_F females) sampled from the population are phenotyped for a quantitative trait. In each sex, there are nine possible genotypes for the two SNPs, each genotype with an observed number generally expressed as $n_{r_1 r'_1 / r_2 r'_2}^M$ for the males and $n_{r_1 r'_1 / r_2 r'_2}^F$ for the females ($r_1 \geq r'_1, r_2 \geq r'_2, r_3 \geq r'_3 = 1, 0$).

Many physiological traits scale with body weight (W) according to a power function with a certain allometric exponent. Thus, we implement this allometric scaling law to describe the phenotypic value of a trait for subject i within male or female subpopulations in terms of the haplotypes considered as

$$\begin{aligned} y_{M_i} &= \alpha_M W_{M_i}^{\beta_M} + x_{M_i} a_M + z_{M_i} d_M + e_{M_i}, \\ y_{F_i} &= \alpha_F W_{F_i}^{\beta_F} + x_{F_i} a_F + z_{F_i} d_F + e_{F_i}, \end{aligned}$$

where (α_M, β_M) or (α_F, β_F) are body weight-related allometric coefficients, (x_{M_i}, z_{M_i}) or (x_{F_i}, z_{F_i}) are the indicator variables associated with the additive and dominance effects, respectively, and e_{M_i} or e_{F_i} is the residual error, normally distributed as $N(0, \sigma_M^2)$ or $N(0, \sigma_F^2)$. The geno-

Table 1: Diplotypes and their frequencies for each of nine genotypes at two SNPs, and composite diplotypes for one assumed sex-specific risk haplotype chosen from four possible haplotypes.

Genotype	Diplotype		Risk Haplotype in Males				Risk Haplotype in Females				
	Configuration	Male Freq.	Female Freq.	[11]	[10]	[01]	[00]	[11]	[10]	[01]	[00]
11/11	[11] [11]	$p_{M_{11}}^2$	$p_{F_{11}}^2$	AA	$\bar{A}\bar{A}$	$\bar{A}\bar{A}$	$\bar{A}\bar{A}$	BB	$\bar{B}\bar{B}$	$\bar{B}\bar{B}$	$\bar{B}\bar{B}$
11/10	[11] [10]	$2p_{M_{11}}p_{M_{10}}$	$2p_{F_{11}}p_{F_{10}}$	$A\bar{A}$	$A\bar{A}$	$\bar{A}\bar{A}$	$\bar{A}\bar{A}$	$B\bar{B}$	$B\bar{B}$	$\bar{B}\bar{B}$	$\bar{B}\bar{B}$
11/00	[10] [10]	$p_{M_{10}}^2$	$p_{F_{10}}^2$	$\bar{A}\bar{A}$	AA	$\bar{A}\bar{A}$	$\bar{A}\bar{A}$	$\bar{B}\bar{B}$	BB	$\bar{B}\bar{B}$	$\bar{B}\bar{B}$
10/11	[11] [01]	$2p_{M_{11}}p_{M_{01}}$	$2p_{F_{11}}p_{F_{01}}$	$A\bar{A}$	$\bar{A}\bar{A}$	$A\bar{A}$	$\bar{A}\bar{A}$	$B\bar{B}$	$\bar{B}\bar{B}$	$B\bar{B}$	$\bar{B}\bar{B}$
10/10	$\begin{Bmatrix} [11][00] \\ [10][01] \end{Bmatrix}$	$\begin{Bmatrix} 2p_{M_{11}}p_{M_{00}} \\ 2p_{M_{10}}p_{M_{01}} \end{Bmatrix}$	$\begin{Bmatrix} 2p_{F_{11}}p_{F_{00}} \\ 2p_{F_{10}}p_{F_{01}} \end{Bmatrix}$	$\begin{Bmatrix} A\bar{A} \\ \bar{A}\bar{A} \end{Bmatrix}$	$\begin{Bmatrix} \bar{A}\bar{A} \\ A\bar{A} \end{Bmatrix}$	$\begin{Bmatrix} \bar{A}\bar{A} \\ A\bar{A} \end{Bmatrix}$	$\begin{Bmatrix} A\bar{A} \\ \bar{A}\bar{A} \end{Bmatrix}$	$\begin{Bmatrix} B\bar{B} \\ \bar{B}\bar{B} \end{Bmatrix}$	$\begin{Bmatrix} \bar{B}\bar{B} \\ B\bar{B} \end{Bmatrix}$	$\begin{Bmatrix} \bar{B}\bar{B} \\ B\bar{B} \end{Bmatrix}$	$\begin{Bmatrix} B\bar{B} \\ \bar{B}\bar{B} \end{Bmatrix}$
10/00	[10] [00]	$2p_{M_{10}}p_{M_{00}}$	$2p_{F_{10}}p_{F_{00}}$	$\bar{A}\bar{A}$	$A\bar{A}$	$\bar{A}\bar{A}$	$A\bar{A}$	$\bar{B}\bar{B}$	$B\bar{B}$	$\bar{B}\bar{B}$	$B\bar{B}$
00/11	[01] [01]	$p_{M_{01}}^2$	$p_{F_{01}}^2$	$\bar{A}\bar{A}$	$\bar{A}\bar{A}$	AA	$\bar{A}\bar{A}$	$\bar{B}\bar{B}$	$\bar{B}\bar{B}$	BB	$\bar{B}\bar{B}$
00/10	[01] [00]	$2p_{M_{01}}p_{M_{00}}$	$2p_{F_{01}}p_{F_{00}}$	$\bar{A}\bar{A}$	$\bar{A}\bar{A}$	$A\bar{A}$	$A\bar{A}$	$\bar{B}\bar{B}$	$\bar{B}\bar{B}$	$B\bar{B}$	$B\bar{B}$
00/00	[00] [00]	$p_{M_{00}}^2$	$p_{F_{00}}^2$	$\bar{A}\bar{A}$	$\bar{A}\bar{A}$	$\bar{A}\bar{A}$	AA	$\bar{B}\bar{B}$	$\bar{B}\bar{B}$	$\bar{B}\bar{B}$	BB

Two alleles for each of the two SNPs are denoted as 1 and 0, respectively. Genotypes at different SNPs are separated by a slash. Diplotypes are the combination of two bracketed maternally and paternally derived haplotypes. By assuming different haplotypes as a risk haplotype (denoted as A for males and B for females), composite diplotypes are accordingly defined and their genotypic values are given.

typic values of composite diplotypes and variance are arrayed by a quantitative genetic parameter vector $\Theta_{M_q} = (\alpha_M, \beta_M, a_M, d_M, \sigma_M^2)$ for the males and $\Theta_{F_q} = (\alpha_F, \beta_F, a_F, d_F, \sigma_F^2)$ for the females, respectively.

The log-likelihood of haplotype frequencies, genotypic values of composite diplotypes and residual variances given sex-specific phenotypic (y_M, y_F) and SNP data (S_M, S_F) is factorized into two parts, expressed as

$$\begin{aligned} & \log L(\Theta_{M_p}, \Theta_{M_q}; \Theta_{F_p}, \Theta_{F_q} | y_M, S_M; y_F, S_F) \\ = & \log L(\Theta_{M_p}, \Theta_{F_p} | S_M, S_F) + \log L(\Theta_{M_q}, \Theta_{F_q} | y_M, S_M, \Theta_{M_p}; y_F, S_F, \Theta_{F_p}) \\ = & \log L(\Theta_{M_p} | S_M) + \log L(\Theta_{F_p} | S_F) + \log L(\Theta_{M_q} | y_M, S_M, \Theta_{M_p}) + \log L(\Theta_{F_q} | y_F, S_F, \Theta_{F_p}) \end{aligned}$$

where equation (4) is derived from equation (3) because the males and females are assumed to be independent, and

$$\begin{aligned} \log L(\Theta_{k_p} | S_k) = \text{constant} & & \log L(\Theta_{k_q} | y_k, S_k, \Theta_{k_p}) = \\ +2n_{11/11}^k \log p_{k_{11}} & & \sum_{i=1}^{n_{11/11}^k} \log f_{k_2}(y_{k_i}) \\ +n_{11/10}^k \log(2p_{k_{11}}p_{k_{10}}) & & + \sum_{i=1}^{n_{11/10}^k} \log f_{k_1}(y_{k_i}) \\ +2n_{11/00}^k \log p_{k_{10}} & & + \sum_{i=1}^{n_{11/00}^k} \log f_{k_0}(y_{k_i}) \\ +n_{10/11}^k \log(2p_{k_{11}}p_{k_{01}}) & & + \sum_{i=1}^{n_{10/11}^k} \log f_{k_1}(y_{k_i}) \\ +n_{10/10}^k \log(2p_{k_{11}}p_{k_{00}} + 2p_{k_{10}}p_{k_{01}}) & & + \sum_{i=1}^{n_{10/10}^k} \log[\phi_k f_{k_1}(y_{k_i}) + (1 - \phi_k) f_{k_0}(y_{k_i})] \\ +n_{10/00}^k \log(2p_{k_{10}}p_{k_{00}}) & & + \sum_{i=1}^{n_{10/00}^k} \log f_{k_0}(y_{k_i}) \\ +2n_{00/11}^k \log p_{k_{01}} & & + \sum_{i=1}^{n_{00/11}^k} \log f_{k_0}(y_{k_i}) \\ +n_{00/10}^k \log(2p_{k_{01}}p_{k_{00}}) & & + \sum_{i=1}^{n_{00/10}^k} \log f_{k_0}(y_{k_i}) \\ +2n_{00/00}^k \log p_{k_{00}} & & + \sum_{i=1}^{n_{00/00}^k} \log f_{k_0}(y_{k_i}) \end{aligned}$$

where $f_{k_j}(y_{k_i})$ is a normal distribution density function of composite diplotype j ($j = 2, 1, 0$) for sex k , and

$$\phi_k = \frac{pk_{11}pk_{00}}{pk_{00}pk_{00} + pk_{10}pk_{01}}$$

is the relative proportion of diplotype [11] [00] within the double heterozygote for sex k .

It can be seen from equation (3) or (4) that maximizing $L(\Theta_{M_p}, \Theta_{M_q}; \Theta_{F_p}, \Theta_{F_q} | y_M, S_M, y_F, S_F)$ is equivalent to maximizing $\log(\Theta_{k_p} | S_k)$ and $\log L(\Theta_{k_p} | y_k, S_k, \Theta_{k_p})$ individually in equation (5).

The EM algorithm

A closed-form solution for the EM algorithm [14] has been derived to estimate the unknown parameters that maximize the two sex-specific likelihoods of (5). The estimates of sex-specific haplotype frequencies are based on the log-likelihood function $(\Theta_{k_p} | S_k)$, whereas the estimates of sex-specific genotypic values of composite diploypes and the residual variance are based on the log-likelihood function $L(\Theta_{k_p} | y_k, S_k, \hat{\Theta}_{k_p})$. These two different types of parameters can be estimated using a two-stage hierarchical EM algorithm (see ref. [14] for a detailed implementation).

Model selection

According to equation (5), the summed likelihood across the sexes, $L(\Theta_{M_p}, \Theta_{M_q} | y_M, S_M) + L(\Theta_{F_p}, \Theta_{F_q} | y_F, S_F)$, is formulated by assuming that haplotype [11] is a risk haplotype. However, a real risk haplotype is unknown from raw data (y_k, S_k) . An additional step for choosing the most likely risk haplotype should be implemented. The simplest way to do so is to calculate the likelihood values by assuming that any one of the four haplotypes can be a risk haplotype (Table 1). Thus, we obtain four possible likelihood values as follows:

No.	Haplotype	Risk	Likelihood
1	[11]	$L_1(\hat{\Theta}_{M_p}, \hat{\Theta}_{1M_q} y_M, S_M) + L_1(\hat{\Theta}_{F_p}, \hat{\Theta}_{1F_q} y_F, S_F)$	
2	[10]	$L_2(\hat{\Theta}_{M_p}, \hat{\Theta}_{2M_q} y_M, S_M) + L_2(\hat{\Theta}_{F_p}, \hat{\Theta}_{2F_q} y_F, S_F)$	
3	[01]	$L_3(\hat{\Theta}_{M_p}, \hat{\Theta}_{3M_q} y_M, S_M) + L_3(\hat{\Theta}_{F_p}, \hat{\Theta}_{3F_q} y_F, S_F)$	
4	[00]	$L_4(\hat{\Theta}_{M_p}, \hat{\Theta}_{4M_q} y_M, S_M) + L_4(\hat{\Theta}_{F_p}, \hat{\Theta}_{4F_q} y_F, S_F)$	

The largest likelihood value calculated is thought to correspond to the most likely risk haplotype. Under an optimal

risk haplotype, we estimate sex-specific quantitative genetic parameters $\hat{\Theta}_{M_q}$ and $\hat{\Theta}_{F_q}$.

Hypothesis tests

The genetic architecture of a quantitative trait is characterized by population (including haplotype frequencies, allele frequencies, and linkage disequilibria) and quantitative genetic parameters (including haplotype effects and mode of inheritance for haplotypes). The model proposed provides a meaningful way for estimating the genetic architecture of a trait and further testing sex-specific differences in genetic control.

After haplotype frequencies are estimated, allele frequencies and linkage disequilibrium between the two SNPs with each sex can be calculated as

	Male	Female
Allele frequency for SNP 1	$p_M = p_{M_{11}} + p_{M_{10}}$	$p_F = p_{F_{11}} + p_{F_{10}}$
Allele frequency for SNP 2	$q_M = p_{M_{11}} + p_{M_{01}}$	$q_F = p_{F_{11}} + p_{F_{01}}$
Linkage disequilibrium	$D_M = p_{M_{11}}p_{M_{00}} - p_{M_{10}}p_{M_{01}}$	$D_F = p_{F_{11}}p_{F_{00}} - p_{F_{10}}p_{F_{01}}$

The influence of haplotypes on a quantitative trait is quantified in terms of the additive (a) and dominant genetic effects (d), and the mode of inheritance (ρ), which are estimated for each sex. Each of these population and quantitative genetic parameters can be tested when appropriate hypotheses are formulated.

Overall genetic control

Haplotype effects on the trait, i.e., the existence of functional haplotypes, in both male and female populations can be tested using the following hypotheses expressed as

$$H_0 : \mu_{M_j} \equiv \mu_M \text{ and } \mu_{F_j} \equiv \mu_F \text{ for } j = 2, 1, 0$$

$$H_1 : \text{At least one equality in } H_0 \text{ does not hold}$$

The log-likelihood ratio test statistic (LR) under these two hypotheses can be similarly calculated,

$$LR = -2[\log L_0(\tilde{\mu}_M; \tilde{\mu}_F | y_M; y_F) - \log L_1(\hat{\Theta}_{M_q}; \hat{\Theta}_{F_q} | y_M, S_M, \hat{\Theta}_{M_p}; y_F, S_F, \hat{\Theta}_{F_p})],$$

where the L_0 and L_1 are the plug-in likelihood values under the null and alternative hypotheses of (8), respectively. Although the critical threshold for determining the existence of a functional haplotype can be based on empirical permutation tests, the LR may asymptotically follow a χ^2 distribution with four degrees of freedom, so that the threshold can be obtained from the χ^2 distribution table.

Sex-specific population genetic architecture

The male and female populations may be different in terms of population genetic parameters. Such sex-specific

differences can be tested by formulating the following hypotheses

$$H_0 : p_M = p_F$$

$$H_1 : p_M \neq p_F$$

for allele frequency at SNP 1,

$$H_0 : q_M = q_F$$

$$H_1 : q_M \neq q_F$$

for allele frequency at SNP 2, and

$$H_0 : D_M = D_F$$

$$H_1 : D_M \neq D_F$$

for the linkage disequilibrium between the two SNPs.

For each of the hypotheses (10)–(12), the LR values are calculated, which are each thought to asymptotically follow a χ^2 -distribution with one degree of freedom. Sex-specific differences in overall population genetic architecture can be tested with the null hypothesis $H_0: p_M = p_F, q_M = q_F, \text{ and } D_M = D_F$, with the corresponding LR value to be χ^2 -distributed with three degrees of freedom.

Sex-specific quantitative genetic architecture

Sex-specific differences in overall quantitative genetic architecture can be tested by formulating the hypotheses

$$H_0 : a_M = a_F \text{ and } d_M = d_F$$

$$H_1 : \text{At least one equality in } H_0 \text{ does not hold}$$

The LR value calculated under the null and alternative hypotheses is suggested to follow a χ^2 -distribution with two degrees of freedom. The rejection of the null hypothesis implies that the effects of the same haplotype are different between the two sexes. If there exists a sex-specific difference, the next step is to test whether this difference is due to the additive or dominant genetic effects, or both.

Sex-specific risk haplotypes

In the preceding sections, the same risk haplotype was assumed between the male and female populations. It is possible that the two sexes have different risk haplotypes. Let $\mu_{M_{j_m}}$ ($j_m = 2, 1, 0$) and $\mu_{F_{j_f}}$ ($j_f = 2, 1, 0$) be the genotypic values of composite diplotypes for the males and females constructed by a sex-specific risk haplotype. By reformulating the likelihood $\log L(\Theta_{k_p} | y_k, S_k, \Theta_{k_p})$ of equation (5) based on sex-specific composite diplotypes, these genotypic values can be estimated with the EM algorithm. A best combination of risk haplotypes between the two sexes can be determined from the AIC values.

Multi-locus haplotyping

Three-SNP model

Consider three associated SNPs, $S_1, S_2,$ and S_3 , each with two alleles denoted by 1 and 0. Let p, q and r , and D_{12}, D_{13}, D_{23} and D_{123} be the 1-allele frequencies for the three SNPs, and the linkage disequilibria between SNPs 1 and 2, SNPs 1 and 3, SNPs 2 and 3 and among the three SNPs, respectively. Eight haplotypes, [111], [110], [101], [100], [011], [010], [001] and [000], formed by these three SNPs, have sex-specific frequencies arrayed in $\Theta_{k_p} = (p_{k_{111}}, p_{k_{110}}, p_{k_{101}}, p_{k_{100}}, p_{k_{011}}, p_{k_{010}}, p_{k_{001}}, p_{k_{000}})$ for sex k . Each of these haplotype frequencies is constructed by allele frequencies at different SNPs and their linkage disequilibria of different orders, expressed as

$$\begin{aligned} p_{k_{111}} &= p_k q_k r_k + p_k D_{k_{23}} + q_k D_{k_{13}} + r_k D_{k_{12}} + D_{k_{123}} \\ p_{k_{110}} &= p_k q_k (1 - r_k) - p_k D_{k_{23}} - q_k D_{k_{13}} + (1 - r_k) D_{k_{12}} - D_{k_{123}} \\ p_{k_{101}} &= p_k (1 - q_k) r_k - p_k D_{k_{23}} + (1 - q_k) D_{k_{13}} - r_k D_{k_{12}} - D_{k_{123}} \\ p_{k_{100}} &= p_k (1 - q_k) (1 - r_k) + p_k D_{k_{23}} - (1 - q_k) D_{k_{13}} - (1 - r_k) D_{k_{12}} + D_{k_{123}} \\ p_{k_{011}} &= (1 - p_k) q_k r_k + (1 - p_k) D_{k_{23}} - q_k D_{k_{13}} - r_k D_{k_{12}} - D_{k_{123}} \\ p_{k_{010}} &= (1 - p_k) q_k (1 - r_k) - (1 - p_k) D_{k_{23}} + q_k D_{k_{13}} - (1 - r_k) D_{k_{12}} + D_{k_{123}} \\ p_{k_{001}} &= (1 - p_k) (1 - q_k) r_k - (1 - p_k) D_{k_{23}} - (1 - q_k) D_{k_{13}} - r_k D_{k_{12}} + D_{k_{123}} \\ p_{k_{000}} &= (1 - p_k) (1 - q_k) (1 - r_k) + (1 - p_k) D_{k_{23}} + (1 - q_k) D_{k_{13}} + (1 - r_k) D_{k_{12}} + D_{k_{123}}, \end{aligned}$$

Sex-specific population genetic architecture can be tested by comparing the differences in allele frequencies (p_k, q_k, r_k) and linkage disequilibria of different orders ($D_{k_{12}}, D_{k_{13}}, D_{k_{23}}, D_{k_{123}}$) between the males and females.

In a natural population, there are 27 genotypes for the three SNPs. The frequency of each genotype is expressed in terms of haplotype frequencies. Some genotypes are consistent with diplotypes, whereas the others that are heterozygous at two or more SNPs are not. Each double heterozygote contains two different diplotypes. One triple heterozygote, i.e., 10/10/10, contains four different diplotypes, [111] [000] (in a probability of $2p_{k_{111}}p_{k_{000}}$), [110] [001] (in a probability of $2p_{k_{110}}p_{k_{001}}$), [101] [010] (in a probability of $2p_{k_{101}}p_{k_{010}}$) and [100] [011] (in a probability of $2p_{k_{100}}p_{k_{011}}$). The relative frequencies of different diplotypes for this double or triple heterozygote are a function of haplotype frequencies (Supporting Information Table 1). The integrative EM algorithm can be employed to estimate the MLEs of haplotype frequencies. A general formula for estimating haplotype frequencies can be derived.

By assuming [111] as a risk haplotype (labeled by A) and all the others as non-risk haplotypes (labelled by \bar{A}), the formulation of genotypic values for three composite diplotypes, μ_2 for AA, μ_1 for A \bar{A} and μ_0 for $\bar{A}\bar{A}$ can be derived. Similar procedures described for the two-SNP model can be obtained to estimate and test sex-specific additive and dominance genetic effects when a haplotype contains three SNPs.

L-SNP model

It is possible that the two- and three-SNP models are too simple to characterize genetic variants for quantitative variation. We can develop a model that includes an arbitrary number of SNPs whose sequences are associated with the phenotypic variation. A key issue for the multi-SNP sequencing model is how to distinguish among 2^{-1} different diplotypes for the same genotype heterozygous at loci. The relative frequencies of these diplotypes can be expressed in terms of haplotype frequencies.

Consider a functional haplotype that contains L SNPs among which there exist linkage disequilibria of different orders. The two alleles, 1 and 0, at each of these SNPs are symbolized by r_1, \dots, r_L , respectively. Let $p_{r_1}^k, \dots, p_{r_L}^k$ be the allele frequencies for these different SNPs within sex k . A haplotype frequency, denoted as $p_{r_1 r_2 \dots r_L}$, is decomposed into the following components:

$$\begin{aligned}
 & p_{r_1 r_2 \dots r_L}^k \\
 &= p_{r_1}^k p_{r_2}^k \dots p_{r_L}^k && \text{No LD} \\
 &+ (-1)^{r_1 + r_2} p_{r_1}^k \dots p_{r_{L-2}}^k D_{k(l-1)l} + \dots + (-1)^{r_1 + r_2} p_{r_3}^k \dots p_{r_L}^k D_{k12} && \text{Digenic LD} \\
 &+ (-1)^{r_1 + r_2 + r_3} p_{r_1}^k \dots p_{r_{L-3}}^k D_{k(l-2)(l-1)l} + \dots + (-1)^{r_1 + r_2 + r_3} p_{r_4}^k \dots p_{r_L}^k D_{k123} && \text{Trigenic LD} \\
 &+ \dots \\
 &+ (-1)^L (-1)^{r_1 + \dots + r_L} D_{k1 \dots L} && \text{L-genic LD}
 \end{aligned}$$

where D_k 's are the linkage disequilibria of different orders among particular SNPs for sex k .

Sex-specific difference in terms of allele frequencies and linkage disequilibria between different SNPs as well as haplotype additive and dominance effects can be tested by formulating the corresponding hypotheses.

Results

Pain genetics study

The model proposed was used to detect differences in the genetic architecture of pain sensitivity between men and women. Genetic and phenotypic data were from a pain genetics project in which 237 subjects (including 143 men and 94 women) from five different races were sampled for six SNPs at three candidate genes. As a demonstration of the utilization of the model, we will focus on two SNPs,

OPRDT80G (with two alleles T and G) and OPRDT921C (with two alleles T and C), at the delta opioid receptor. Pain testing procedures followed Fillingim et al. [16]. The phenotypic values of traits were subtracted by the means for each race to remove the effect due to races.

These two SNPs construct four haplotypes, [TC], [TT], [GC], and [GT], which yield 10 diplotypes, [TC] [TC], [TC] [TT], [TT] [TT], [TC] [GC], [TC] [GT], [TT] [GC], [TT] [GT], [GC] [GC], [GC] [GT], and [GT] [GT] and nine genotypes, TT/CC, TT/CT, TT/TT, TG/CC, TG/CT, TG/TT, GG/CC, GG/CT and GG/TT. Based on the observed numbers of each genotype in the male and female populations, we estimated sex-specific haplotype frequencies (Table 1). The pattern of haplotype distribution is consistent between the two sexes, with haplotypes [TC] and [TT] jointly occupying a majority proportion in the populations. Haplotype [GT] is very rare, with the frequency close to zero. SNP OPRDT80G has a low heterozygosity because the frequency of its commoner allele (T) is closer to 0.90, whereas there is a high heterozygosity for SNP OPRDT921C in terms of its averaged allele frequencies. The two SNPs are highly significantly associated at $p = 3.41 \times 10^{-5}$ for males and $p = 9.63 \times 10^{-5}$ for females, with a normalized linkage disequilibrium of $D' = 1.00$, because alleles T from OPRDT80G and T from OPRDT921C as well as alleles G from OPRDT80G and C from OPRDT921C tend to form the same haplotypes more frequently than at random. There is no sex-specific difference in allele frequencies at the two SNPs and their linkage disequilibrium.

By assuming that one of the haplotypes is a risk haplotype, we estimated the effects of each haplotype on a pain sensitivity trait, assessed with a baseline pressure pain threshold measured at the ulna, in the pooled male and female population. The likelihoods of haplotype [TC], [TT] and [GC] as a risk haplotype are -594.8, -594.5, and -596.1, and thus the most likely risk haplotype is [TT]. The genotypic values of composite diplotypes constructed by this risk haplotype and its non-risk haplotype counterpart were estimated and compared between different sexes. In both males and female, the three composite diplotypes do not display significant genetic differences in the pain trait studied, but the directions of the additive and dominance effects are different between the two sexes (Table 2). In males, the non-risk haplotype tends to increase pressure pain thresholds, and it is overdominant to the risk haplotype, leading to increased pressure pain thresholds at a marginal significance level ($p = 0.058$) (Fig. 1). By contrast, in females, the non-risk haplotype tends to reduce pressure pain thresholds, and it also tends to be overdominant to the risk haplotype by reducing pressure pain thresholds. These discrepancies in both effect size and direction (Fig. 2) make the overall quantitative genetic

Table 2: The estimates and tests of population genetic structure for two SNPs, OPRDT80G (with alleles T and G) and OPRDT921C (with two alleles C and T), and quantitative genetic effects of haplotypes constructed by these two SNPs on baseline pressure pain thresholds measured at the ulna in males and females.

Genetic Parameter	Male		Female		Sex-specific LR	Sex-specific p -value
	MLE	p -value	MLE	p -value		
Haplotype Frequency						
\hat{p}_{TC}	0.364		0.396			
\hat{p}_{TT}	0.505		0.463			
\hat{p}_{GC}	0.131		0.142			
\hat{p}_{GT}	0.000		0.000			
Allele Frequency and Linkage Disequilibrium						
\hat{p}_T (OPRDT80G)	0.869		0.858		0.079	0.778
\hat{q}_C (OPRDT921C)	0.495		0.537		0.578	0.447
\hat{D}	-0.066	3.41×10^{-5}	-0.066	9.63×10^{-5}	0.001	0.973
Additive and Dominant Effects and Inheritance Mode for Risk Haplotype [TT]						
\hat{a} and \hat{d}		0.129		0.548	31.44	1.49×10^{-7}
\hat{a}	-0.2	0.431	0.3	0.481	4.72	0.030
\hat{d}	0.7	0.058	-0.5	0.350	24.86	6.16×10^{-7}
$\hat{\rho}$	2.9		1.8		0.63	0.427

architecture of the pain sensitivity trait significantly differ between the two sexes ($p = 1.49 \times 10^{-7}$) (Table 2). Although the additive genetic effect displays a gene by sex interaction at the $p = 0.03$ significance level, a gene by sex interaction for the dominance effect is highly significant at $p = 6.16 \times 10^{-7}$. No significant difference was observed in inheritance mode between males and females.

Monte carlo simulation

Simulation studies were performed to test the statistical properties of the model proposed. Given a certain sample size (n), we simulated two SNPs by assuming different allele frequencies and linkage disequilibria between two sexes. The hypothesized allele frequencies at the two SNPs are $p_M = 0.5$, $q_M = 0.6$ and $D_M = 0.1$ for males and $p_F = 0.8$, $q_F = 0.9$ and $D_F = 0.06$ for females. By postulating one of the four haplotypes constructed by the two SNPs as a risk haplotype, we calculated the genetic variance among three composite diplotypes using the additive effect $a = 0.6$ and dominance effect $d = 0.8$, from which the residual vari-

ance was calculated when the heritability (H^2) of a trait is given. The phenotypic values of the trait were simulated by assuming that they follow a normal distribution under four different simulation designs, (1) $n = 100$ and $H^2 = 0.1$, (2) $n = 400$ and $H^2 = 0.1$, (3) $n = 100$ and $H^2 = 0.4$, and (4) $n = 400$ and $H^2 = 0.4$.

The new model was used to analyze the simulated SNP and phenotypic data with the results tabulated in Table 3. Population genetic parameters including allele frequencies (p and q) and linkage disequilibrium (D) can well be estimated, with increasing precision when the sample size increases from 100 to 400. The power to detect the given sex-specific differences in allele frequencies and linkage disequilibrium is as high as 0.95 even with a modest sample size (100). Although the traditional model that does not implement sex-specific differences can provide precise estimates of these parameters, the estimates are generally between the true values of males and females.

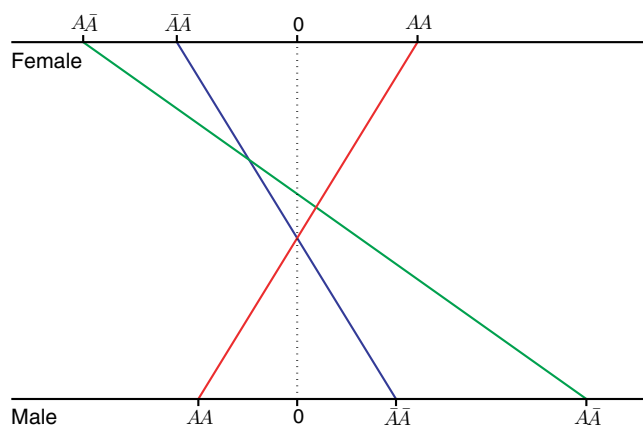


Figure 1
 Different genotypic values of baseline pressure pain thresholds measured at the ulna for composite diplotypes, AA, AĀ, and ĀĀ, constructed by risk haplotype [TT] and non-risk haplotype in males and females. The origin indicates the mean of the genotypic values between the two homozygotes.

The estimation of quantitative genetic parameters including the additive (*a*) and dominance effects (*d*) needs the determination of an optimal risk haplotype. When all possible risk haplotypes were assumed for the simulated data, we found that the true risk haplotype gave the largest likelihood among the four possible cases. In general, quantitative genetic parameters can well be estimated, but the estimation precision increases dramatically with sample size and heritability. Although the additive effect can be obtained with reasonable precision at a modest heritability (0.1) with a modest sample size (100), the precise estimation of the dominance effect relies upon a larger heritability and sample size. Also, the given difference in the additive effect between two sexes can be detected with great power, even when both the sample size and heritability are small. But the same size of sex-specific difference in the dominance effect can be detected with the same power only when the sample size is 400 and heritability is 0.4 (Table 3). The traditional model gave biased estimates of the sex-specific additive and dominance effects regardless of increasing sample size and heritability.

We conducted an additional simulation study, in which the data simulated under the assumption of no sex-specific differences in all genetic parameters were analyzed by the new and traditional model. As expected, both the models provide reasonable estimates of population and quantitative genetic parameters, with estimation precision increasing with increasing sample size and heritability (data not shown). This, in conjunction with the results in

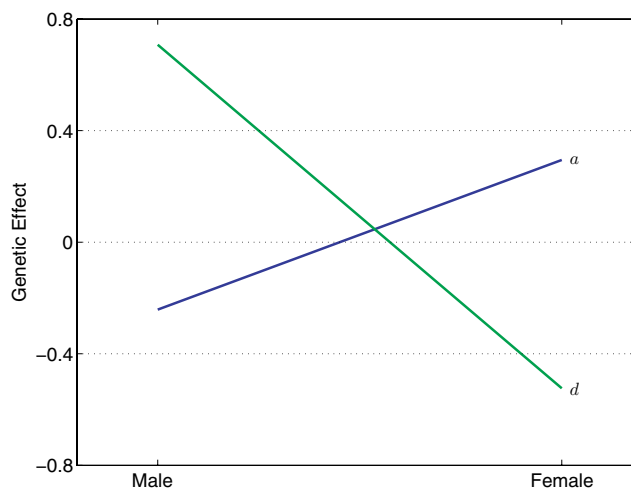


Figure 2
 Different additive (*a*) and dominant genetic effects (*d*) of haplotypes on baseline pressure pain thresholds measured at the ulna in males and females.

Table 3, suggests that the new model provides a general tool for study the genetic architecture of a complex trait, regardless of whether the genetic control of the trait is sex-specific.

Discussion

The genetic architecture of a quantitative trait is complex in terms of interactions between its underlying genetic factors and various environments including sex [1-3]. However, in many current studies, gene by sex interactions are often ignored simply because existing analytical models are not incorporated by environmental factors. While different phenotypic expressions of a trait between the two sexes can be easily measured [4,5], sex-specific discrepancy in the genetic control of the trait can be discerned only when a sophisticated model is used. There is strong evidence for sex-specific genetic influence [17,18] even for traits that display no sexual dimorphism [19]. Although quantitative genetic models are available to estimate sex-specific heritabilities due to aggregative effects of many genes [5,19-22] or map sex-specific QTLs for phenotypic variation [4,5,11-13,23], the model proposed in this article can dissect sex-specific genetic control at the DNA sequence levels.

Our model is founded on the conceptual framework for haplotyping a trait with single nucleotide polymorphisms (SNPs) formulated by Liu et al. [14]. Since haplotypes constructed by physically associated SNPs are thought to affect the expressivity of a complex trait [23,24], it is more precise to characterize such haplotype effects by incorporating gene by gene and gene by sex interactions. Lin and Wu [25] extended Liu et al.'s [14] model to estimate hap-

Table 3: The MLEs of population and quantitative genetic parameters and the standard errors of the estimates obtained by the new model and the power to detect sex-specific differences under different simulation designs. Parameter estimates by a conventional model are also given.

Simulation Design	Genetic Parameter	Sex		Power	Traditional Model
		Male	Female		
1	<i>p</i>	0.5025 ± 0.0360	0.7829 ± 0.0265	0.95	0.6433 ± 0.0224
	<i>q</i>	0.5964 ± 0.0349	0.8732 ± 0.0193	0.95	0.7347 ± 0.0201
	<i>D</i>	0.0961 ± 0.0194	0.0181 ± 0.0099	0.80	0.0779 ± 0.0120
	<i>a</i>	0.6055 ± 0.2008	0.3045 ± 0.1277	0.24	0.4789 ± 0.1192
	<i>d</i>	0.3061 ± 0.3371	0.6093 ± 0.1798	0.19	0.5843 ± 0.1875
	2	<i>p</i>	0.5013 ± 0.0172	0.7980 ± 0.0141	1.00
<i>q</i>		0.6011 ± 0.0177	0.8958 ± 0.0101	1.00	0.7484 ± 0.0100
<i>D</i>		0.0986 ± 0.0098	0.0084 ± 0.0053	1.00	0.0775 ± 0.0061
<i>a</i>		0.6050 ± 0.1114	0.3056 ± 0.0728	0.67	0.4907 ± 0.0616
<i>d</i>		0.2980 ± 0.1604	0.5986 ± 0.0933	0.42	0.5715 ± 0.0966
3		<i>p</i>	0.5008 ± 0.0366	0.7835 ± 0.0269	1.00
	<i>q</i>	0.5945 ± 0.0338	0.8731 ± 0.0195	1.00	0.7338 ± 0.0194
	<i>D</i>	0.0965 ± 0.0178	0.0161 ± 0.0088	0.86	0.0747 ± 0.0115
	<i>a</i>	0.5989 ± 0.0924	0.3012 ± 0.0505	0.85	0.4582 ± 0.0483
	<i>d</i>	0.2934 ± 0.1352	0.6021 ± 0.0652	0.64	0.5355 ± 0.0715
	4	<i>p</i>	0.5014 ± 0.0181	0.7974 ± 0.0136	1.00
<i>q</i>		0.6000 ± 0.0171	0.8968 ± 0.0102	1.00	0.7484 ± 0.0099
<i>D</i>		0.1001 ± 0.0092	0.0077 ± 0.0048	1.00	0.0745 ± 0.0058
<i>a</i>		0.6003 ± 0.0439	0.3010 ± 0.0301	1.00	0.4684 ± 0.0251
<i>d</i>		0.3008 ± 0.0659	0.6006 ± 0.0372	0.99	0.5499 ± 0.0360

lotype-haplotype interactions. The new model reported here can not only estimate sex-specific genetic parameters, but also provide a series of statistical procedures for testing sex-specific differences in the genetic architecture of quantitative variation. In a natural population, the structure and pattern of genetic variation can be studied by population genetic parameters, such as haplotype frequencies, allele frequencies and linkage disequilibria. Thus, the understanding of differences in these parameters between the two sexes help to infer the sex-specific genetic structure of a natural population and its evolutionary processes. As shown through simulation studies, our model is alert to discern sex-specific differences in basic population genetic parameters.

Liu et al. [14] assumed a so-called reference or risk haplotype that triggers an effect on complex traits in a different way from the other haplotypes. Thus, the combinations between the risk and non-risk haplotypes (composite diplotypes) will perform differently, depending on the type of combination, i.e., risk by risk, risk by non-risk and

non-risk by non-risk. Liu et al. [14] proposed the concepts of the additive effect due to the substitution of the non-risk (or risk) haplotype by the risk (or non-risk) haplotype and the dominance effect due to the interaction between the risk and non-risk haplotypes. These concepts have been integrated into the current model that allows the test of additive by sex and dominance by sex interaction effects. Simulation studies suggest that the new model displays adequate power to detect differences in these quantitative genetic parameters between the sexes, although the detection of sex-specific dominance effects needs a much larger sample size and/or heritability level.

Our model was used to analyze a real data set for pain genetics. Our analyses of a pain sensitivity trait-baseline pressure pain threshold measured at the ulna-by estimating genetic parameters and testing their sex-specific differences in a combination of male and female samples revealed that males and females have different population structure at two SNPs genotyped from a candidate gene (delta opioid receptor) for human pain and that haplo-

types exert different genetic effects on the trait between the two sexes. A further test indicates that the risk haplotype [TT] detected by the model exemplifies sex-specific modes of inheritance in affecting the pain trait. While there are no differences among composite diplotypes in males, a significant additive effect was detected in females. Both additive and dominance effects due to the risk haplotype identified are different between the two sexes. Anholt and Mackay [2] described three major mechanisms that explain sex-specific difference in trait control, i.e., sex-specific effects (a gene affects only one sex), sex-biased effects (a gene affects both sexes but to different degrees), and sex-antagonistic effects (a gene affects both sexes but in opposite directions). In our example, the functional haplotype detected affects the pain trait in a sex-antagonistic effect manner, a mechanism thought to help the maintenance of genetic variation in natural populations [26].

In practice, failure to model sex-specific architecture may significantly hamper the ability to detect signals of functional genetic variants in genomewide screens. Although combining male and female data to increase sample size are tempting approaches to increase power, the estimates in this way will be biased from true sex-specific differences. Possible mechanisms that cause sex differences include parent-of-origin effects [27], linkage to or interaction with sex chromosomes, or differences arising from sex-specific hormonal environments. Our gene by sex interaction model that is incorporated by these mechanisms can be modified to consider interactions between genes and any other environments such as life style. Our interaction models should provide a more powerful tool to draw a detailed and precise picture of the genetic architecture of any complex traits that are important to human health.

Authors' contributions

CW, YC, TL and QL wrote the programs and performed data analyses. RF, MW, and RL designed the experiment. LK performed the experiment. RW conceived the idea and wrote the paper. All authors read and approved the final manuscript.

Acknowledgements

The preparation of this manuscript is partially supported by a joint NSF/NIH grant (0540745) to RW and an NIH grant (R01 NS041670-05A1) to RBF.

References

- Mackay TF: **The genetic architecture of quantitative traits: lessons from *Drosophila***. *Curr Opin Genet Dev* 2004, **14**(3):253-257.
- Anholt RR, Mackay TFC: **Quantitative genetic analyses of complex behaviours in *Drosophila***. *Nature Reviews Genetics* 2004, **5**:838-849.
- Foerster K, Coulson T, Sheldon BC, Pemberton JM, Clutton-Brock TH, Kruuk LEB: **Sexually antagonistic genetic variation for fitness in red deer**. *Nature* 2007, **447**:1107-1110.
- Weiss LA, Pan L, Abney M, Ober C: **The sex-specific genetic architecture of quantitative traits in humans**. *Nature Genetics* 2006, **38**:218-222.
- Weiss LA, Abney M, Cook EH, Ober C: **Sex-specific genetic architecture of whole blood serotonin levels**. *American Journal of Human Genetics* 2005, **76**:33-41.
- Fillingim RB: **Sex-related influences on pain: a review of mechanisms and clinical implications**. *Rehabilitation Psychology* 2003, **48**:165-174.
- Fillingim RB: **Sex, gender and pain: The biopsychosocial model in action XX vs. XY**. *The International Journal of Sex Differences in the Study of Health, Disease and Aging* 2003, **1**:98-101.
- Craft RM, Mogil JS, Aloisi AM: **Sex differences in pain and analgesia: the role of gonadal hormones**. *European Journal of Pain* 2004, **8**:397-411.
- Rinn JL, Snyder M: **Sexual dimorphisms in mammalian gene expression**. *Trends in Genetics* 2005, **21**:298-305.
- Mogil JS, Richards SP, O'Toole LA, Helms ML, Mitchell SR, Kest B, Belknap JK: **Identification of a sex-specific quantitative trait locus mediating nonopioid stress-induced analgesia in female mice**. *Journal of Neuroscience* 1997, **17**:7995-8002.
- Zhao W, Ma CX, Cheverud JM, Wu RL: **A unifying statistical model for QTL mapping of genotype-sex interaction for developmental trajectories**. *Physiological Genomics* 2004, **19**:218-227.
- Butterfield RJ, Roper RR, Rhein DM, Melvold RW, Haynes L, Ma RZ, Doerge RW, Teuscher C: **Sex-specific QTL govern susceptibility to Theilers murine encephalomyelitis virus-induced demyelination (TMEVD)**. *Genetics* 2003, **163**:1041-1046.
- Fernandez JR, Vogler GP, Tarantino LM, Vignetti S, Plomin R, McClearn GE: **Sex-exclusive quantitative trait loci influences in alcohol-related phenotypes**. *American Journal of Medical Genetics* 1999, **88**:647-652.
- Liu T, Johnson JA, Casella G, Wu RL: **Sequencing complex diseases with HapMap**. *Genetics* 2004, **168**:503-511.
- The International HapMap Consortium: **The International HapMap Project**. *Nature* 2003, **426**:789-794.
- Fillingim RB, Kaplan L, Staud R, Ness TJ, Glover TL, Campbell CM, Mogil JS, Wallace MR: **The A118G single nucleotide polymorphism of the mu-opioid receptor gene (OPRM1) is associated with pressure pain sensitivity in humans**. *Journal of Pain* 2005, **6**:159-167.
- Jensen H, Sather B-E, Ringsby TH, Tufto J, Griffith SC, Ellegren H: **Sexual variation in heritability and genetic correlations of morphological traits in house sparrow (*Passer domesticus*)**. *Journal of Evolutionary Biology* 2003, **16**:1296-1307.
- Leips J, Mackay TFC: **Quantitative trait loci for life span in *Drosophila melanogaster*: interactions with genetic background and larval density**. *Genetics* 2002, **155**:1773-1788.
- Schousboe K, Visscher PM, Henriksen JE, Hopper JL, Sorensen TIA, Kyvik KO: **Twin study of genetic and environmental influences on glucose tolerance and indices of insulin sensitivity and secretion**. *Diabetologia* 2003, **46**:1276-1283.
- van Beijsterveldt CE, Bartels M, Hudziak JJ, Boomsma DI: **Causes of stability of aggression from early childhood to adolescence: a longitudinal genetic analysis in Dutch twins**. *Behavioral Genetics* 2003, **33**:591-605.
- Keski-Rahkonen A, Neale BM, Bulik CM, Pietil KH, Pietilainen KH, Rose RJ, Kaprio J, Rissanen A: **-bf Intentional weight loss in young adults: Sex-specific genetic and environmental effects**. *Obesity Research* 2005, **13**:745-753.
- Ober C, Pan L, Phillips N, Parry R, Kurina LM: **Sex-specific genetic architecture of asthma-associated quantitative trait loci in a founder population**. *Curr Allergy Asthma Rep* 2006, **6**(3):241-246.
- Bader JS: **The relative power of SNPs and haplotype as genetic markers for association tests**. *Pharmacogenomics* 2001, **2**(1):11-24.
- Judson R, Stephens JC, Windemuth A: **The predictive power of haplotypes in clinical response**. *Pharmacogenomics* 2000, **1**:5-26.
- Lin M, Wu RL: **Detecting sequence-sequence interactions for complex diseases**. *Current Genomics* 2006, **7**:59-72.
- Rice WR: **Sexually antagonistic genes: experimental evidence**. *Science* 1992, **256**:1436-1439.
- Reik W, Walter J: **Genomic imprinting: parental influence on the genome**. *Nature Reviews Genetics* 2001, **2**:21-32.