ARTICLE

Bivariate association analysis in selected samples: application to a GWAS of two bone mineral density phenotypes in males with high or low BMD

Aude Saint-Pierre[*,1], Jean-Marc Kaufman[2], Agnes Ostertag[3], Martine Cohen-Solal[3], Anne Boland[4], Kaatje Toye[2], Diana Zelenika[4], Mark Lathrop[4], Marie-Christine de Vernejoul[3] and Maria Martinez[1]

Our specific aims were to evaluate the power of bivariate analysis and to compare its performance with traditional univariate analysis in samples of unrelated subjects under varying sampling selection designs. Bivariate association analysis was based on the seemingly unrelated regression (SUR) model that allows different genetic models for different traits. We conducted extensive simulations for the case of two correlated quantitative phenotypes, with the quantitative trait locus making equal or unequal contributions to each phenotype. Our simulation results confirmed that the power of bivariate analysis is affected by the size, direction and source of the phenotypic correlations between traits. They also showed that the optimal sampling scheme depends on the size and direction of the induced genetic correlation. In addition, we demonstrated the efficacy of SUR-based bivariate test by applying it to a real Genome-Wide Association Study (GWAS) of Bone Mineral Density (BMD) values measured at the lumbar spine (LS) and at the femoral neck (FN) in a sample of unrelated males with low BMD (LS Z-scores ≤ -2) and with high BMD (LS and FN Z-scores > 0.5). A substantial amount of top hits in bivariate analysis did not reach nominal significance in any of the two single-trait analyses. Altogether, our studies suggest that bivariate analysis is of practical significance for GWAS of correlated phenotypes.
European Journal of Human Genetics (2011) **19**, 710–716; doi:10.1038/ejhg.2011.22; published online 23 March 2011

INTRODUCTION

With the availability of high-density maps of single-nucleotide polymorphisms (SNPs), association studies have become popular tools for identifying genes underlying complex human traits and diseases. For most current population-based genome-wide association studies (GWAS), statistical power is often limited because of the complex interplay among factors that influence the etiology of diseases.[1] Increasing sample size and multilocus or multivariate statistical analyses can improve the power for detecting association. Sample size is often restricted because of genotyping costs and limited sample resources. Several studies have demonstrated that analyzing samples selected with extreme values can be more powerful than analyzing samples randomly selected from the population.[2–4] In addition to using selected samples, another approach to increasing association test power is to perform joint analysis of multiple correlated phenotypes. For many common multifactorial traits, several correlated phenotypes are usually recorded for each individual during sample collection, but most often, the phenotypes are analyzed separately in a univariate framework. Joint analysis of correlated phenotypes can theoretically provide greater power than that provided by analysis of individual phenotypes.[3,5–7] Multivariate analysis can also alleviate the multiple testing problem, caused by testing different traits separately, and thereby improve the ability to detect genetic variants whose effects are too small to be detected in univariate analysis.[8] Several multivariate approaches have

been applied to linkage studies of correlated complex phenotypes, such as osteoporosis and bone-related phenotypes.[9–12] Similarly, various methods, often based on generalized estimating equations (GEEs), have been proposed for performing multivariate association tests on population- or family-based data.[13–20] Of the two studies that have investigated the power of bivariate association test in population-based data, one applied the restricted bivariate association test that assumes same quantitative trait locus (QTL) effects on each trait.[16,18] Such constraints in the model may have overestimated or underestimated the relative performance of bivariate over univariate analysis. Finally, GWAS studies using multivariate analysis are rare, especially in samples of subjects selected through their phenotype values, and further investigations using this approach are warranted.[4]

To this aim, we evaluated the statistical properties of joint association analysis of two correlated quantitative traits in samples of unrelated subjects through simulation studies using the seemingly unrelated regression (SUR) bivariate model that allows for different QTL effects on traits. The evaluation was conducted under different situations according to the sample selection design, genetic effects and residual correlation between the traits. We demonstrate the efficacy of SUR-based bivariate test by applying it to simultaneous GWAS analysis of two correlated bone phenotypes, bone mineral density (BMD) at the lumbar spine (LS) and at the femoral neck (FN), which are major risk factors of osteoporosis.

MATERIALS AND METHODS

SUR-based bivariate model

The SUR model[21] is a generalization of a classical linear regression model that consists of several regression equations with potentially different sets of explanatory variables. It thus allows for a differential effect of explanatory variables on phenotypes as well as the possibility that some variables might be associated with only one trait. Let N be the total number of unrelated subjects ($i=1, \ldots, N$), each having observations on two phenotypes y_{ji} ($j=1, 2$). Consider a system of two equations, where the jth equation is of the form: $y_j = X_j \times \beta_j + e_j$; y_j is a $N \times 1$ vector of the phenotypic values, X_j is a $(K_j+1) \times N$ matrix of explanatory variables with K_j representing the number of explanatory variables in the model for phenotype j excluding the intercept; $\beta_j = (\beta_0^j, \beta_1^j, \ldots, \beta_{K_j}^j)^t$ is the $(K_j+1) \times 1$ vector of coefficients and e_j is a $N \times 1$ vector of the residuals errors. The system of SUR can be written as:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \quad (1)$$

The SUR model allows for cross-equation correlation of the residual terms. The covariance matrix of all the residuals is assumed to be normally distributed with mean 0 and covariance matrix $E(ee^t) = \Omega = \Sigma \otimes I_N$ where I_N is a $N \times N$ unit matrix and Σ a 2×2 matrix with the following form:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & rE \times \sigma_1 \sigma_2 \\ rE \times \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix} \quad (2)$$

σ_1^2 and σ_2^2 are the residual variances of Y_1 and Y_2, respectively, and rE is the residual correlation between Y_1 and Y_2.

The SUR model is estimated using the generalized least square method where the covariance matrix Ω is first estimated using ordinary least square regression in system (1). Linear restrictions on coefficients can be tested by an F test. The F statistic for systems of equations is: $F = [(C\beta')^t (C\text{cov}(\beta')C^t)^{-1} (C\beta')]/2$ where C is the matrix of restrictions on coefficients. Under the null hypothesis, the F statistic has a central Fisher distribution with 2 and $2 \times N - K$ degrees of freedom, where K is the total number of estimated coefficients ($K = K_1 + K_2 + 2$). The goodness of fit of the whole system can be measured by the McElroy's r-square (R^2). R^2 is the proportion of covariance because of X taking into account the residual matrix covariance Ω.[22]

Here, we applied the SUR model to test association between two continuous phenotypes in unrelated subjects genotyped at one SNP marker, and X_j is the $N \times 1$ vector of genotypes at the SNP. Under an additive model, the genotype for each individual i, noted g_i, is coded as a function of the number of minor alleles, that is, 0, 1 or 2. We computed the SUR model free of constraints on the regression coefficients, that is, $\beta 1$ and $\beta 2$ were freely estimated. Under the null hypothesis of no association to either one or both phenotypes, the F statistic has a central Fisher distribution with 2 and $2 \times (N-2)$ degrees of freedom. Separate association analyses of Y_1 and Y_2 can be conducted using traditional univariate linear regression model: $y_j = g \times \beta_j + e_j$, where y_j, g and β_j are as described above but now, e_j is assumed to follow a normal distribution $N(0, \sigma_j^2)$. The null hypothesis of no association ($\beta_j = 0$) can be tested against the alternative ($\beta_j \neq 0$) with a Student's statistic (t-test) with $N-2$ degrees of freedom.

Simulation study

We considered genetic models of complex traits and specifically tried to generate correlated data, mimicking as much as possible our real BMD GWAS data (see below). As a strong (~ 0.5) and positive phenotypic covariation exists for BMD values at the LS and at the FN,[23] we generated data for two positively correlated quantitative phenotypes. Further, in real data sets, as causal loci usually contribute a small proportion to the total phenotypic correlation, residual correlation approximates phenotypic correlation between traits. It is also more realistic to assume that the investigator has a priori knowledge on the magnitude and sign of the covariation of the studied phenotypes than on the magnitude and sign of the QTL effect on each phenotype. Therefore, in all our scenarios, the sign of the residual correlation (rE) was positive, but the sign of the induced QTL correlation (rG) was either positive or negative. Also, our BMD GWAS study used a sampling design, with extreme truncate selection of unrelated males, aiming to improve power. Therefore, we also generated samples of subjects drawn from the extremes of the phenotype(s) population distribution.

The main scenarios and parameter settings are shown in Table 1. The different settings allowed us to generate data for a QTL having same or different effect on the two positively correlated phenotypes, and the two sources of covariation (QTL and residual) have same or opposite sign. Briefly, we assumed a biallelic QTL having additive effects (a_j) on Y_j ($j=1, 2$), with minor and major allele frequency q and p, respectively. The QTL contribution to Y_j is the trait-specific QTL heritability, h_j^2. Here, we focussed our power investigation to QTLs explaining a relatively small part of the trait variance, that is, from 0.5 to 3% that, for complex traits, seemed to us more realistic. The genotypic means (m_{jk}) of Y_j are equal to $2q \times a_j$, $(q-p) \times a_j$ and $-2p \times a_j$ when k, the number of minor alleles, is equal to 0, 1 and 2, respectively, and with $a_j = \sqrt{[h_j^2/2pq]}$. We varied the sign of a_j; both were of same or opposite sign and the QTL correlation (rG) was, thus, equal to $+1$ or -1, respectively. We first generated samples of subjects unselected for their traits values (denoted as Su). Second, we generated subjects selected from the 2.5% (ie, trait value ≤ -2) and 30% (ie, trait value >0.5) left and right tail of the population distribution of Y_1 (denoted as S1), respectively. Third, we included Y_2 in the selection design, that is, we selected subjects from the 2.5 and 30% left and right tail of the population distribution of Y_1 and Y_2, respectively, (denoted as S2). These truncate selection criteria (trait value ≤ -2 or >0.5) are the values that we have used in our real BMD GWAS. Under S1 and S2, we generated samples with equal number of subjects drawn from the left ($N/2$) and the right ($N/2$) side of the phenotypes distributions.

Traits values of N (300, 1000)-unrelated subjects were generated as follows. For a given combination of parameter values (rE, h_1^2, h_2^2, rG), we first draw QTL alleles from a binomial distribution with parameter q, and built genotypes under Hardy–Weinberg equilibrium. Then, conditionally on the generated genotype, g_k ($k=0, 1, 2$), we jointly drew the values of Y_1 and Y_2 via a bivariate normal distribution with mean $(m_{1k}, m_{2k})^t$ and variance matrix Ω, given in equation (2). Third, under sampling S1 or S2, we applied the corresponding truncate selection, that is, individuals not fulfilling the selection criteria were withdrawn from the sample. Steps 1–3 were repeated until reaching the required left and right truncated sample sizes of ($N/2$) subjects.

Table 1 Outline of the main scenarios and varying parameter values in the bivariate data simulations

QTL model	Main scenarios Heritability (h_j^2)	Effect size (a_j)[a]	rG	q	h_j^2	Parameter values rE	N	Sampling design
I. Null	$h_1^2 = h_2^2 = 0$	$a_1 = a_2 = 0$	0	0.1/0.4	$h_1^2 = 0$ and $h_2^2 = 0$	0.2/0.6	1000/300	Su/S1/S2
II. No pleiotropic effect	$h_1^2 > 0$; $h_2^2 = 0$	$a_1 > 0$; $a_2 = 0$	0	0.1/0.4	$h_1^2 = (0.5\%/1\%/3\%)$ and $h_2^2 = 0$	0.2/0.6	1000/300	Su/S1/S2
III. Pleiotropic effect								
a_j: Same direction	$h_1^2 = h_2^2$	$a_1 = a_2$	+	0.1/0.4	$h_1^2 = h_2^2 = (0.5\%/1\%/3\%)$	0.2/0.6	1000/300	Su/S1/S2
	$h_1^2 \neq h_2^2$	$a_1 \neq a_2$	+	0.1/0.4	$h_1^2 = 0.5\%$ and $h_2^2 = (1\%/3\%)$; $h_1^2 = 1\%$ and $h_2^2 = 3\%$	0.2/0.6	1000/300	Su/S1/S2
a_j: Opposite direction	$h_1^2 = h_2^2$	$a_1 = -a_2$	−	0.1/0.4	$h_1^2 = h_2^2 = (0.5\%/1\%/3\%)$	0.2/0.6	1000/300	Su/S1/S2
	$h_1^2 \neq h_2^2$	$a_1 \neq -a_2$	−	0.1/0.4	$h_1^2 = 0.5\%$ and $h_2^2 = (1\%/3\%)$; $h_1^2 = 1\%$ and $h_2^2 = 3\%$	0.2/0.6	1000/300	Su/S1/S2

[a] $a_j = \sqrt{[h_j^2/(2pq)]}$, where q is the minor allele frequency.

Each replicate was analyzed with SUR-based bivariate and with two separate univariate analyses using the systemfit package of R software (http://www.r-project.org/) using the genotypes at the QTL, that is, the SNP is the causal variant. The mean and standard deviations of each association statistic (F test and t_1, t_2-tests) were derived from K replicates. Power and type I error rates of each association test were calculated as the proportion of replicates with a test statistic exceeding a given theoretical threshold ($R\alpha$) value, at nominal significance levels, α=5, 1, 0.1 and 10^{-3}%. Type 1 errors were estimated in the settings were $h^2_1=h^2_2=0$ with K=20 000 replicates. Power rates were derived with K=1000 replicates. To compare the performance of bivariate and that of univariate association analysis, we computed the proportion of replicates where t_1 and t_2 were both lower than $R\alpha$. One minus this proportion estimated the probability to detect association to either one of the two phenotypes. To adjust for the two univariate association tests, we applied the Bonferroni correction, that is, we used the theoretical thresholds $R\alpha/2$.

RESULTS
Simulation study
Tables 2 and 3 present the mean (and SD) association statistic of the SUR-based bivariate (F test) and of the traditional univariate tests (*t*-test), respectively, when N=1000 for 66 scenarios under the alternative hypothesis and when q=0.4. For a given QTL heritability value, the results did not vary, as expected, with q.

Bivariate association statistics. In randomly selected samples, the results in Table 2 show several well-established power figures. First, mean F statistics of bivariate association analysis increase with the size of the trait-specific QTL heritability (h^2_1 and/or h^2_2) irrespective of rG and rE. Second, the power is highest in presence (rG$\neq$0) than in

absence (rG=0) of pleiotropic effects: the highest power is achieved when rG=−1, that is, when the correlation induced by the QTL effect and the residual correlation are opposite in sign. Third, the results also confirm that the power of bivariate association test varies with the size of the residual correlation: when rG=0 or rG=−1, the power increases with rE; conversely, when rG=+1, it decreases with rE. These general trends are observed irrespective of the sampling selection designs. Applying extreme truncate selection increases the power of bivariate association analysis, but the optimal selection design depends on the true genetic model. When rG=0 or rG=−1, extreme selection on one trait (S1) is more efficient than extreme selection on both traits (S2). Conversely, when rG=+1, S2 is more efficient than S1. Overall, under Su or S1, the highest mean F statistics are obtained when rG=−1, irrespective of rE. Under S2, the highest power is achieved when rG=+1 or when rG=−1, depending on the size of rE. Interestingly, when the traits are moderately (rE=0.20) correlated, mean F statistics have greater values when rG=+1 than when rG=−1.

Univariate association statistics. Table 3 shows again several well-established power figures. In randomly selected samples, the power of univariate analysis increases with the QTL heritability (h^2_1/h^2_2) and varies little with the size of the residual correlation, rE. For phenotype Y_1, under a given QTL heritability (h^2_1) value, the mean statistic values of all models are similar in the randomly selected samples. Applying extreme truncate selection increases the power of univariate association analysis of Y_1. Under S1, the power remains similar whatever may be the rG value. Under S2, the power is the highest and the lowest for the pleiotropic models rG=+1 and rG=−1, respectively. When rG=−1 or rG=0, the power of univariate association analysis is greater under S1 than under S2. The reverse trend is obtained when rG=+1. For phenotype Y_2, the power of univariate analysis depends on rG and rE. Further, applying extreme selection does not always lead to a gain in power. Indeed, when rG=−1, the power of univariate analysis is the greatest in the unselected samples (Su). When rG=0, the mean *t*-statistic values in the selected samples are biased and inflated. The magnitude of the bias is greater under S2 than under S1. Under S1, the bias increases with rE.

Overall, applying selection criteria on one or both traits is an optimal sampling design when rG=+1: the power of each separate univariate analysis is improved over that in randomly selected samples. When rG=−1, applying extreme truncate selection leads to both a substantial gain and decrease in power for Y_1 and Y_2, respectively. For the situations in which the QTL does not exert pleiotropic effects (rG=0), the highest power of univariate analysis of Y_1 is obtained in the selected samples. However, the mean *t*-statistic values for Y_2, the trait not associated to the QTL, are also increased. Type I error rates of separate univariate analyses may thus be inflated, especially in selected samples and when the residual correlation is high.

Type I error rates. When the QTL/SNP has no effect on Y_1 and Y_2, the values of the mean and standard deviation of both bivariate and univariate association tests are close to the theoretical values, regardless of the residual correlation, minor allele frequency of the studied SNP and of the selection sampling design (Supplementary Table 1A). Indeed, SUR-based bivariate and each separate univariate association tests have correct type I error rates (Supplementary Table 1B). However, the false positive rates of univariate association analyses for detecting association to either or both the two traits are, as expected, inflated: the estimated rates are roughly two times higher than the theoretical rates. Applying a Bonferroni correction (denoted

Table 2 Mean and SD (μF (SD)) of the SUR-based bivariate association statistic (F test) in samples of N=1000 subjects for various parameter settings: QTL heritability (h^2_1/h^2_2), sign of the induced genetic correlation (rG), residual correlation (rE) and sampling selection design

| | | | | Sampling | |
rE	rG	h^2_1/h^2_2	Su μF (SD)	S1 μF (SD)	S2 μF (SD)
0.2	0	0.005/0	3.69 (2.61)	10.10 (4.34)	9.23 (4.18)
		0.01/0	6.17 (3.32)	18.94 (5.86)	17.97 (5.74)
		0.03/0	17.02 (6.08)	59.02 (10.56)	55.3 (10.32)
	+1	0.005/0.005	5.08 (2.99)	11.42 (4.86)	15.17 (5.41)
		0.005/0.01	7.45 (3.66)	14.07 (5.04)	19.08 (6.58)
		0.01/0.01	9.50 (4.42)	22.89 (6.86)	29.84 (7.89)
		0.03/0.03	26.90 (7.34)	72.04 (12.89)	92.88 (14.41)
	−1	0.005/0.005	7.26 (3.92)	13.91 (5.30)	9.57 (4.48)
		0.005/0.01	10.57 (4.36)	17.22 (5.65)	11.05 (4.88)
		0.01/0.01	13.69 (5.42)	27.72 (7.40)	19.29 (6.56)
		0.03/0.03	39.95 (9.40)	89.83 (13.86)	68.63 (13.23)
0.6	0	0.005/0	4.88 (3.04)	11.26 (4.51)	9.96 (4.47)
		0.01/0	8.79 (4.04)	22.54 (6.75)	19.78 (6.16)
		0.03/0	25.04 (7.34)	69.67 (11.85)	62.92 (11.30)
	+1	0.005/0.005	4.09 (2.69)	10.41 (4.49)	12.22 (4.78)
		0.005/0.01	6.36 (3.60)	12.67 (5.06)	15.60 (5.52)
		0.01/0.01	7.33 (3.85)	20.35 (6.34)	23.81 (6.76)
		0.03/0.03	20.42 (6.53)	63.56 (11.19)	73.11 (11.61)
	−1	0.005/0.005	13.70 (5.32)	20.94 (6.59)	16.02 (5.84)
		0.005/0.01	19.71 (6.52)	27.35 (7.55)	21.20 (6.92)
		0.01/0.01	26.06 (7.40)	42.83 (9.58)	34.26 (8.87)
		0.03/0.03	78.65 (14.17)	143.61 (19.15)	124.18 (17.77)

Abbreviations: Su, unselected sample; S1, sample selected on Y_1 distribution; S2, sample selected on Y_1 and Y_2 distributions.

Table 3 Mean and SD (μt (SD)) of the traditional univariate association statistic (*t*-test) in samples of $N=1000$ subjects for various parameter settings: QTL heritability (h^2_1/h^2_2), sign of the induced genetic correlation (rG), residual correlation (rE) and sampling selection design

			Sampling					
			Su		*S1*		*S2*	
rE	rG	h^2_1/h^2_2	$Y_1 \mu t$ *(SD)*	$Y_2 \mu t$ *(SD)*	$Y_1 \mu t$ *(SD)*	$Y_2 \mu t$ *(SD)*	$Y_1 \mu t$ *(SD)*	$Y_2 \mu t$ *(SD)*
0.2	0	0.005/0	2.26 (1.01)	−0.03 (1.00)	4.23 (0.99)	0.98 (1.02)	4.02 (1.00)	2.59 (1.05)
		0.01/0	3.15 (1.01)	−0.02 (0.97)	5.95 (0.98)	1.41 (1.00)	5.78 (0.96)	3.60 (1.00)
		0.03/0	5.53 (1.06)	0.00 (1.00)	10.69 (0.96)	2.33 (0.96)	10.34 (0.98)	6.34 (1.01)
	+1	0.005/0.005	2.23 (1.00)	2.20 (0.98)	4.15 (0.97)	3.21 (1.07)	4.99 (0.97)	4.83 (1.03)
		0.005/0.01	2.19 (0.97)	3.23 (1.02)	4.19 (0.97)	4.21 (0.99)	5.36 (1.02)	5.69 (1.07)
		0.01/0.01	3.18 (1.00)	3.19 (1.04)	5.96 (1.01)	4.72 (1.04)	7.06 (0.96)	6.93 (1.09)
		0.03/0.03	5.57 (1.01)	5.57 (0.98)	10.64 (1.01)	8.60 (1.07)	12.28 (0.96)	12.57 (1.13)
	−1	0.005/0.005	2.20 (1.01)	−2.26 (1.01)	4.18 (1.00)	−1.26 (0.99)	3.15 (0.99)	0.36 (1.02)
		0.005/0.01	2.26 (0.99)	−3.22 (0.97)	4.21 (0.92)	−2.16 (0.99)	2.69 (1.03)	−0.60 (0.99)
		0.01/0.01	3.18 (1.04)	−3.17 (1.04)	5.95 (0.98)	−1.95 (1.01)	4.56 (0.96)	0.44 (0.97)
		0.03/0.03	5.57 (1.01)	−5.58 (1.04)	10.68 (0.97)	−3.91 (1.01)	8.51 (0.96)	0.29 (0.95)
0.6	0	0.005/0	2.23 (1.00)	0.00 (0.96)	4.19 (0.97)	2.35 (0.98)	3.89 (0.99)	2.69 (0.97)
		0.01/0	3.13 (0.99)	−0.05 (0.98)	6.01 (1.00)	3.26 (0.97)	5.64 (0.97)	3.87 (0.98)
		0.03/0	5.55 (1.02)	0.01 (0.98)	10.62 (0.98)	5.44 (0.96)	10.08 (0.96)	6.63 (0.96)
	+1	0.005/0.005	2.22 (1.00)	2.24 (0.99)	4.17 (1.00)	4.06 (1.02)	4.58 (0.99)	4.59 (1.01)
		0.005/0.01	2.25 (1.01)	3.24 (1.03)	4.17 (1.00)	4.79 (1.03)	4.90 (0.94)	5.40 (0.99)
		0.01/0.01	3.17 (1.03)	3.18 (1.00)	5.97 (1.01)	5.83 (1.01)	6.52 (0.96)	6.58 (1.00)
		0.03/0.03	5.55 (1.01)	5.60 (1.03)	10.70 (0.98)	10.47 (0.99)	11.48 (0.93)	11.75 (0.99)
	−1	0.005/0.005	2.20 (1.02)	−2.30 (1.02)	4.13 (0.97)	0.48 (0.95)	3.23 (0.97)	0.81 (0.92)
		0.005/0.01	2.23 (1.00)	−3.21 (1.00)	4.15 (1.00)	−0.20 (1.00)	2.97 (0.99)	0.03 (0.97)
		0.01/0.01	3.10 (1.02)	−3.22 (1.01)	5.99 (0.96)	0.70 (0.97)	4.75 (0.97)	1.11 (0.95)
		0.03/0.03	5.52 (1.04)	−5.60 (1.05)	10.68 (0.95)	0.57 (0.93)	8.89 (0.92)	1.69 (0.88)

Abbreviations: Su, unselected sample; S1, sample selected on Y_1 distribution; S2, sample selected on Y_1 and Y_2 distributions.

as U_b) leads to slightly conservative significance levels, especially when the residual correlation between the traits is strong.

Power comparisons. The power to detect association to either or both of the two traits using SUR-based bivariate analysis was compared with the power of separate univariate analysis of Y_1 and Y_2 adjusted for multiple testing by the Bonferroni correction (denoted as U_b). Figure 1a shows the power curves (at significance of 10^{-5}) against the QTL heritability (h^2_1, h^2_2) when $N=1000$ for moderately (rE=0.2) or strongly (rE=0.6) correlated traits. Power curves under S1 and S2 are shown in Figure 1b, when $h^2_1=h^2_2=0.005$, $N=1000$ and rE=0.2 or 0.6.

In randomly selected samples (Figure 1a), the relative advantage of SUR-based bivariate over univariate association analysis is more obvious not only when rG=−1 and/or the traits are strongly correlated (rE=0.6) but also when rG=+1 and the traits are moderately correlated (rE=0.2). Under S1 (Figure 1b), SUR-based bivariate is slightly less powerful than univariate analysis when rG=+1 and rE=0.6 or when rG=0 and rE=0.2. For strongly correlated traits, the power rates are equal to 94.5% (SUR) *versus* 29.3% (U_b) when rG=−1; 44.0% (SUR) *versus* 32.3% (U_b) when rG=0; and 36.8% (SUR) *versus* 39.9% (U_b) when rG=+1. For moderately correlated traits, the power rates are equal to 64.6% (SUR) *versus* 31.7% (U_b) when rG=−1; 32.9% (SUR) *versus* 34.9% (U_b) when rG=0; and 43.7% (SUR) *versus* 32.6% (U_b) when rG=+1. Under S2 (Figure 1b), SUR-based bivariate shows same or slightly lower power than univariate analysis, except when rG=−1 or when rG=0 and rE=0.6 where it outperforms univariate test. As already noted above, selecting on Y_1 (S1) is the most efficient sampling design when rG=−1 or when rG=0 and the traits are strongly correlated (rE=0.6). Selecting on both traits (S2) is the most efficient design when rG=+1. Overall,

when rE=0.6, the power of SUR is the greatest (94.5%) when rG=−1 and under S1, whereas the power of univariate analysis is the greatest (56.8%) when rG=+1 and under S2. When rE=0.2, the power of SUR and univariate analysis are both the greatest (72.5 and 72.9%) when rG=+1 and under S2. As shown in Supplementary Table 2, all these trends are confirmed under various parameter settings.

Analyses of empirical BMD genome-wide association data
BMD GWAS data. Subjects were recruited from the Network in Europe on Male Osteoporosis Study.[24,25] Subjects selected from this cohort were unrelated males >18 and <68 years of age. In addition, the subjects were selected by bone densitometry (measured at the LS and FN) criteria, having either low BMD (LS Z-scores ≤−2, n=175) or high BMD (both LS and FN Z-scores >0.50, n=155). Further details of the study sample are provided in Supplementary Table 3. Genotyping was carried out at the Centre National de Génotypage (Evry, France) using the Illumina 370K platform (Illumina, San Diego, CA, USA). SNPs and DNA data were subjected to standard quality control analyses with PLINK[26] (details are provided in Supplementary Methods).

Association analysis. Our primary analysis was the joint association analysis of LS Z-scores and FN Z-scores by means of SUR-based bivariate test. For comparison purpose, we also applied separate univariate association analyses of LS and FN Z-scores. We used single-marker analysis assuming additive genetic effects. The mean F statistic of our SUR-based genome-wide association (GWA) analysis was equal to 1.018 (SD=1.022, median=0.70). The mean *t*-statistic of LS and FN were −0.0167 (SD=1.011, median=−0.0165) and −0.0129 (SD=1.006, median=0.0104), respectively. These results indicated that there was no meaningful inflation of univariate as well as bivariate association analyses.
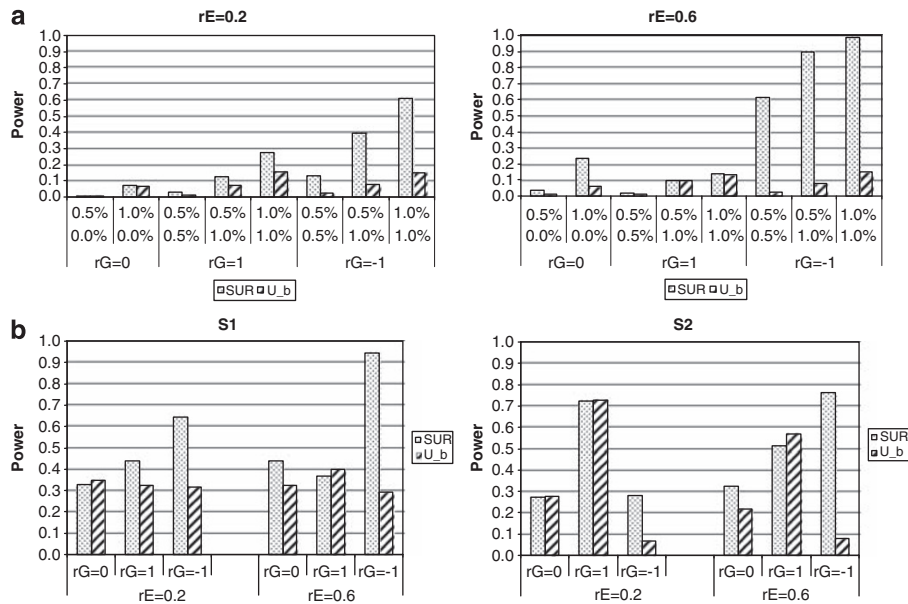
Figure 1 Power rates at $\alpha=10^{-5}$ of SUR-based bivariate analysis and univariate analysis, adjusted for multiple testing by Bonferroni correction (U_b), in samples of $N=1000$ subjects and under various parameters settings: QTL heritability (h_1^2/h_2^2), sign of the induced genetic correlation (rG) and residual correlation (rE). (**a**) Power estimates against QTL heritability for moderately (rE=0.2) or strongly (rE=0.6) correlated traits in randomly selected samples (Su). (**b**) Power estimates under extreme selection (S1 or S2) for moderately (rE=0.2) or strongly (rE=0.6) correlated traits and QTL heritability ($h_1^2=h_2^2=0.005$).

Results. SUR-based bivariate analyses identified a substantial number (35) of SNPs with strong evidence of association (*P*-value $<10^{-4}$). Interestingly, several of the identified SNPs failed to reach nominal (*P*-value $<5\%$) significance under separate univariate analyses for either one or the two BMD phenotypes. Genome-wide bivariate and univariate association results were compared in terms of statistical significance and ranks of the SNPs identified in either one of the two approaches. For each SNP, we kept the lowest *P*-value (denoted as Best_U) of LS or FN univariate association analysis. Univariate *P*-values were not corrected for multiple testing. We ranked the Best_U *P*-values from the lowest to the highest. We similarly ranked the *P*-values from SUR-based bivariate analysis of LS and FN. Figure 2 plots the significance levels in each procedure for the top 100 most associated SNPs identified from SUR-based (Figure 2a) or from univariate (Figure 2b) analyses. We found that a majority (52) of the top SNPs in SUR-based bivariate analysis also show strong ($P<3\times10^{-4}$) association signal in univariate analyses. For a substantial number (16) of the remaining SNPs, univariate analyses fail to reach nominal ($P<5\%$) significance (Figure 2a) On the other hand, all of the top 100 SNPs in univariate analyses (Figure 2b) are also highly significant ($P<8\times10^{-4}$) in bivariate analysis. Table 4 shows details of the association results for the top 10 SNPs in SUR-based and in each separate univariate analysis. The table also shows *P*-values and ranks found in each of the two other procedures. The genetic contributions (R^2 values) of the 10 top SNPs are not great, as expected for any relatively common polymorphic locus. In all, 3 of the top 10 SNPs from bivariate analysis also rank well (ie, are in the set of top 300 SNPs) in univariate analyses of LS and/or FN. They are located on 6q25: rank=2, $P=1.3\times10^{-5}$ (LS) and rank=1, $P=1.2\times10^{-5}$ (FN); on 15q14-q15: rank=2635, $P=8.4\times10^{-3}$ (LS) and rank=3, $P=1.7\times10^{-5}$ (FN); and on 22q13: rank=1, $P=3.5\times10^{-6}$ (LS) and rank=8, $P=3\times10^{-5}$ (FN). All the remaining seven SNPs show a much stronger association signal in bivariate than in univariate analyses, including two of the three best SUR-based association signals. For the most
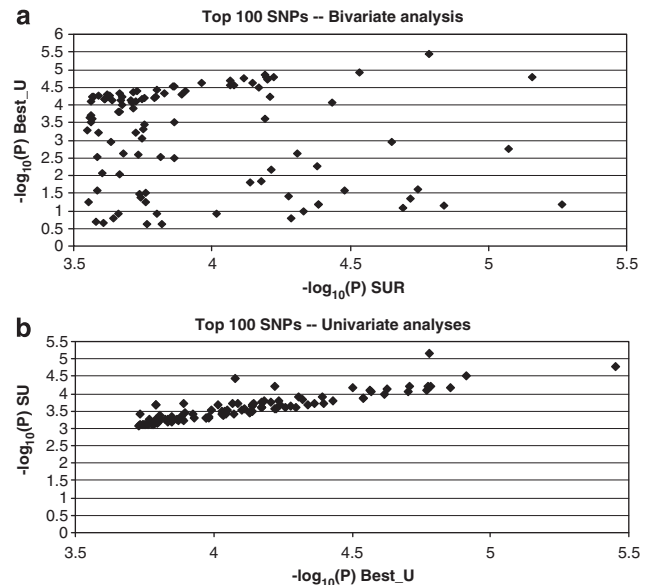

Figure 2 Overlap in significance of results from bivariate and univariate (Best_U) association analysis. (**a**) Top 100 hits in SUR-based bivariate association test: $-\log_{10}$ *P*-values of univariate analysis against $-\log_{10}$ *P*-values of SUR-based bivariate analysis. (**b**) Top 100 hits in univariate association test: $-\log_{10}$ *P*-values of SUR-based bivariate analysis against $-\log_{10}$ *P*-values of univariate analysis.

significant result, on 22q11.2 ($P=5.44\times10^{-6}$), the QTL explains 3.85% of the joint (co)variance of LS and FN. This value likely overestimates the contribution in unselected populations. Nonetheless, univariate analyses failed to detect association ($P>0.07$) with this SNP. Conversely, all the top 20 SNPs identified from univariate analysis of either LS or FN belong to the set of top 42 SNPs from SUR-based bivariate analysis. Overall, our analyses showed that

Table 4 Association results: top 10 most associated SNPs from SUR-based bivariate or from separate univariate analysis of LS and FN BMD

Chr. (locus)	Closest gene	Pos (bp)	SNP	Min.[a]	MAF[b]	SUR P-value	Rank[c]	R (%)[d]	Univ. LS P-value[e]	Rank[c]	R² (%)[e]	Univ. FN P-value[e]	Rank[c]	R² (%)[e]
Top from SUR analysis														
2q37.1	SP100	231 037 761	rs1649866	A	0.33	2.03E−05	8	3.41	0.539	162 064	0.12	0.086	25 794	0.95
		231 042 007	rs1678160	G	0.33	1.45E−05	4	3.52	0.574	172 327	0.10	0.072	21 500	1.04
3q25	LEKR1	158 200 574	rs6799034	C	0.43	1.81E−05	6	3.45	0.970	289 752	0.00	2.41E−02	7166	1.63
6q22.1	LOC643884	113 858 994	rs2049924	A	0.29	8.42E−06	3	3.69	1.80E−03	607	3.09	0.363	109 135	0.27
6q25	TIAM2	155 533 083	rs998318	G	0.31	2.94E−05	10	3.30	1.30E−05	2	5.94	1.22E−05	1	5.98
12p13-p12	PZP – A2MP	9 254 198	rs1017301	C	0.34	2.24E−05	9	3.38	0.223	67 841	0.48	1.13E−03	348	3.36
15q14-15	RYR3	31 680 776	rs2437143	C	0.38	6.97E−06	2	3.75	8.41E−03	2635	2.21	1.65E−05	3	5.80
19p13.11	FAM125A	17 392 450	rs2303680	G	0.41	1.92E−05	7	3.43	0.743	222 673	0.03	4.66E−02	13 886	1.27
22q11.2	SLC2A11	22 534 158	rs2275979	A	0.18	5.44E−06	1	3.85	0.523	157 609	0.13	0.067	20 101	1.08
22q13	LL22NC03	43 026 421	rs3935378	T	0.49	1.64E−05	5	3.48	3.54E−06	1	6.69	2.95E−05	8	5.47
Top from Univ. of LS														
1q32-41	CAMK1G	207 787 465	rs996146	T	0.11	6.39E−05	24	3.06	1.65E−05	3	5.80	5.19E−05	10	5.14
1q41	ESRRG	214 946 534	rs2813711	A	0.13	7.13E−05	29	3.02	3.68E−05	10	5.34	2.36E−05	5	5.59
2q11.2	AFF3	99 943 239	rs11887597	C	0.48	1.08E−04	36	2.90	2.42E−05	7	5.60	0.0005639	156	3.77
3q22.3	PIK3CB	139 883 989	rs531577	C	0.49	1.37E−04	40	2.85	2.89E−05	9	5.53	0.0004341	113	3.95
6q25.2	TIAM2	155 533 083	rs998318	G	0.31	2.94E−05	10	3.30	1.30E−05	2	5.94	1.22E−05	1	5.98
7q11.23	CCDC146	76 658 736	rs10252204	A	0.40	6.29E−05	23	3.07	1.97E−05	5	5.72	0.001236	382	3.32
10q21.2	LOC729184	62 158 387	rs1904418	G	0.32	7.68E−05	31	3.06	1.71E−05	4	5.89	0.0004438	115	3.97
16p13	NPM1P3	5 551 122	rs1969139	C	0.31	1.38E−04	42	2.82	2.88E−05	8	5.48	0.0003619	93	4.01
16q21	SLC38A7	57 262 362	rs9806843	G	0.43	8.55E−05	33	2.97	1.99E−05	6	5.69	0.0005906	169	3.73
22q13.31	LL22NC03	43 026 421	rs3935378	T	0.49	1.64E−05	5	3.48	3.54E−06	1	6.69	2.95E−05	8	5.47
Top from Univ. of FN														
1p21.1	LOC126987	106 644 698	rs1330226	C	0.30	8.56E−05	34	2.97	4.64E−05	18	5.20	2.70E−05	6	5.52
1q32-41	CAMK1G	207 787 465	rs996146	T	0.11	6.39E−05	24	3.06	1.65E−05	3	5.80	5.19E−05	10	5.14
1q41	ESRRG	214 946 534	rs2813711	A	0.13	7.13E−05	29	3.02	3.68E−05	10	5.34	2.36E−05	5	5.59
2p21	C2 or f34	44 846 991	rs11679997	T	0.15	6.73E−05	28	3.05	2.99E−03	960	2.81	3.14E−05	9	5.45
6q25.2	TIAM2	155 533 083	rs998318	G	0.31	2.94E−05	10	3.30	1.30E−05	2	5.94	1.22E−05	1	5.98
7p22.2	SDK1	3 411 837	rs6952184	C	0.09	5.97E−05	20	3.08	4.09E−05	14	5.28	1.68E−05	4	5.79
12q21.31	TSPAN19	83 921 735	rs1581563	G	0.16	8.36E−05	32	2.97	4.31E−05	15	5.25	2.73E−05	7	5.51
15q14-15	RYR3	31 680 776	rs2437143	C	0.38	6.97E−06	2	3.75	8.41E−03	2635	2.21	1.65E−05	3	5.80
20p12-p11.2	NXT1	23 219 822	rs4815192	T	0.46	6.45E−05	26	3.06	9.81E−05	41	4.78	1.39E−05	2	5.92
22q13.31	LL22NC03	43 026 421	rs3935378	T	0.49	1.64E−05	5	3.48	3.54E−06	1	6.69	2.95E−05	8	5.47

[a]Minor allele.
[b]Minor allele frequency.
[c]rank of the identified SNP.
[d]R^2 Of the whole system taking into account the residual (co)variance matrix.
[e]R^2 From linear regression; unadjusted univariate *P*-values.

univariate analysis did not identify new strongly associated SNPs as compared with those detected in bivariate analysis. Conversely, SUR-based analysis identified strongly associated SNPs that were not detected in univariate analysis.

Our study used a design, with extreme truncate selection of unrelated males, aiming to improve power. The approach of studying samples drawn from the extremes of the population distribution of BMD has been used in several linkage studies of BMD variation,[25,27] but rarely in association studies,[28] and to our knowledge, never in samples drawn from the population of males. Owing to our relatively small GWA sample size, no SNP showed evidence of association to either one or both BMD phenotypes at genome-wide significance threshold of 1.7×10^{-7} (0.05/298 783 SNPs). However, we used an extreme truncate selection design that, as shown by our simulation studies, has increased power over unselected samples. Our SUR-based bivariate association analyses identified strong association ($P < 8.4 \times 10^{-6}$) with three genomic regions (6q22.1, 15q14 and 22q11). These SNPs have not yet been reported to be associated with bone density in previous GWAS.[29–31] Two of them, on 15q14-15 and 22q11, are located in genes that are known to be expressed in skeletal muscle:[32,33] *GLUT 11* encoded by *SLC2A11* on 22q11 and *RYR3* on 15q14-15. Because muscle contraction has a major impact on bone density, this might represent an indirect role of these genes on bone density. These genetic variants, whether they are site specific or possibly shared (pleiotropic), may warrant further follow-up genetic studies on BMD and other bone-related phenotypes.

DISCUSSION

We have evaluated the performance of bivariate association analysis based on the SUR model, which allows different genetic models for different traits. To our knowledge, this is the first study to specifically derive the power and the relative performance of bivariate association analysis in selected samples of unrelated subjects. Our main results coincide with well-known power figures,[6–8] and confirmed that bivariate association analysis outperforms univariate analysis when the QTL exerts pleiotropic effects and the relative increase in power is the greatest

when correlation of the QTL is opposite in sign to the residual correlation. The most powerful sampling selection design varied with the genetic model, specifically with the size and the direction of the induced QTL correlation. Applying truncate selection on one trait was found the most efficient sampling design when the genetic and the residual correlations are opposite in signs. The same most efficient design was found when the QTL does not exert pleiotropic effects: the power of the SUR-based bivariate association test was found as good as or better than that of univariate association test, depending on the size of the residual correlation. Finally, when the QTL exerts pleiotropic effects and both sources (QTL and residual) of covariation are of the same sign, applying selection criteria on both traits was found to be the optimal sampling selection design. Under this sampling design, the performance of SUR-based bivariate test relatively to univariate analysis decreases with the size of the residual correlation.

So far, two studies have investigated the power of bivariate association in unselected population-based data, and they both applied bivariate association test based on GEEs.[16,18] The former applied a general GEE-based model that allows, as the SUR model, for different QTL effects on the two traits. The second study used a GEE-based bivariate model that assumed same QTL effects on the phenotypes. Our results are congruent with those reported by the first study. The restricted bivariate test estimates, as the univariate test, a single parameter (ie, the SNP regression coefficients on each trait are all set as equal). Under the restricted bivariate model, the gain in power of bivariate analysis is enhanced and reduced when the QTL has similar effect and when it affects one trait only, respectively. Clearly, rarely, knowledge of this magnitude about a complex trait is known a priori. Thus, we do not recommend using restricted bivariate models even in unselected data.

Our bivariate GWA analysis of LS and FN BMD values, conducted in a sample of unrelated males with low BMD (LS Z-scores ≤ -2) and high BMD (LS and FN Z-scores > 0.5), consistently demonstrated the advantage of the SUR-based bivariate test over separate univariate analysis. All the top hits in univariate analysis also showed strong evidence of association in bivariate analysis. Conversely, additional SNP associations were detected with the bivariate method that did not reach nominal significance in single-trait analyses: this was achieved without adjusting significance of univariate analyses for multiple testing.

In conclusion, our results showed that SUR-based models are useful to detect association for correlated phenotypes. However, our results also showed that similar power levels can be achieved whether the QTL exerts or not pleiotropic effects. Thus, disentangling pure pleiotropic from residual covariation remains a challenge even in bivariate association analysis.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

1 Hirschhorn JN, Daly MJ: Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 2005; **6**: 95–108.

2 Allison DB: Transmission-disequilibrium tests for quantitative traits. *Am J Hum Genet* 1997; **60**: 676–690.

3 Allison DB, Thiel B, St Jean P, Elston RC, Infante MC, Schork NJ: Multiple phenotype modeling in gene-mapping studies of quantitative traits: power advantages. *Am J Hum Genet* 1998; **63**: 1190–1201.

4 Abecasis GR, Cookson WO, Cardon LR: The power to detect linkage disequilibrium with quantitative traits in selected samples. *Am J Hum Genet* 2001; **68**: 1463–1474.

5 Amos C, de Andrade M, Zhu D: Comparison of multivariate tests for genetic linkage. *Hum Hered* 2001; **51**: 133–144.

6 Almasy L, Dyer TD, Blangero J: Bivariate quantitative trait linkage analysis: pleiotropy versus co-incident linkages. *Genet Epidemiol* 1997; **14**: 953–958.

7 Amos CI, Laing AE: A comparison of univariate and multivariate tests for genetic linkage. *Genet Epidemiol* 1993; **10**: 671–676.

8 Jiang C, Zeng ZB: Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics* 1995; **140**: 1111–1127.

9 Wang L, Liu YJ, Xiao P et al: Chromosome 2q32 may harbor a QTL affecting BMD variation at different skeletal sites. *J Bone Miner Res* 2007; **22**: 1672–1678.

10 Pan F, Xiao P, Guo Y et al: Chromosomal regions 22q13 and 3p25 may harbor quantitative trait loci influencing both age at menarche and bone mineral density. *Hum Genet* 2008; **123**: 419–427.

11 Wang XL, Deng FY, Tan LJ et al: Bivariate whole genome linkage analyses for total body lean mass and BMD. *J Bone Miner Res* 2008; **23**: 447–452.

12 Liu XG, Liu YJ, Liu J et al: A bivariate whole genome linkage study identified genomic regions influencing both BMD and bone structure. *J Bone Miner Res* 2008; **23**: 1806–1814.

13 Lange C, Silverman EK, Xu X, Weiss ST, Laird NM: A multivariate family-based association test using generalized estimating equations: FBAT-GEE. *Biostatistics* 2003; **4**: 195–206.

14 Lange C, van Steen K, Andrew T et al: A family-based association test for repeatedly measured quantitative traits adjusting for unknown environmental and/or polygenic effects. *Stat Appl Genet Mol Biol* 2004; **3**: Article17.

15 Jung J, Zhong M, Liu L, Fan R: Bivariate combined linkage and association mapping of quantitative trait loci. *Genet Epidemiol* 2008; **32**: 396–412.

16 Liu J, Pei Y, Papasian CJ, Deng HW: Bivariate association analyses for the mixture of continuous and binary traits with the use of extended generalized estimating equations. *Genet Epidemiol* 2009; **33**: 217–227.

17 Pei YF, Zhang L, Liu J, Deng HW: Multivariate association test using haplotype trend regression. *Ann Hum Genet* 2009; **73**: 456–464.

18 Yang F, Tang Z, Deng H: Bivariate association analysis for quantitative traits using generalized estimation equation. *J Genet Genomics* 2009; **36**: 733–743.

19 Zhang L, Bonham AJ, Li J et al: Family-based bivariate association tests for quantitative traits. *PLoS One* 2009; **4**: e8133.

20 Zhang L, Pei YF, Li J, Papasian CJ, Deng HW: Univariate/multivariate genome-wide association scans using data from families and unrelated samples. *PLoS One* 2009; **4**: e6502.

21 Zellner A: An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *J Am Stat Assoc* 1962; **57**: 348–368.

22 McElroy MB: Goodness of Fit for Seemingly Unrelated Regressions. *J Econometrics* 1977; **6**: 381–387.

23 Livshits G, Deng HW, Nguyen TV, Yakovenko K, Recker RR, Eisman JA: Genetics of bone mineral density: evidence for a major pleiotropic effect from an intercontinental study. *J Bone Miner Res* 2004; **19**: 914–923.

24 Pelat C, Van Pottelbergh I, Cohen-Solal M et al: Complex segregation analysis accounting for GxE of bone mineral density in European pedigrees selected through a male proband with low BMD. *Ann Hum Genet* 2007; **71**: 29–42.

25 Kaufman JM, Ostertag A, Saint-Pierre A et al: Genome-wide linkage screen of bone mineral density (BMD) in European pedigrees ascertained through a male relative with low BMD values: evidence for quantitative trait loci on 17q21-23, 11q12-13, 13q12-14, and 22q11. *J Clin Endocrinol Metab* 2008; **93**: 3755–3762.

26 Purcell S, Neale B, Todd-Brown K et al: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.

27 Sims AM, Shephard N, Carter K et al: Genetic analyses in a sample of individuals with high or low BMD shows association with multiple Wnt pathway genes. *J Bone Miner Res* 2008; **23**: 499–506.

28 Kung AW, Xiao SM, Cherny S et al: Association of JAG1 with bone mineral density and osteoporotic fractures: a genome-wide association study and follow-up replication studies. *Am J Hum Genet* 2010; **86**: 229–239.

29 Richards JB, Kavvoura FK, Rivadeneira F et al: Collaborative meta-analysis: associations of 150 candidate genes with osteoporosis and osteoporotic fracture. *Ann Intern Med* 2009; **151**: 528–537.

30 Rivadeneira F, Styrkarsdottir U, Estrada K et al: Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. *Nat Genet* 2009; **41**: 1199–1206.

31 Styrkarsdottir U, Halldorsson BV, Gretarsdottir S et al: Multiple genetic loci for bone mineral density and fractures. *N Engl J Med* 2008; **358**: 2355–2365.

32 Doege H, Bocianski A, Scheepers A et al: Characterization of human glucose transporter (GLUT) 11 (encoded by SLC2A11), a novel sugar-transport facilitator specifically expressed in heart and skeletal muscle. *Biochem J* 2001; **359**: 443–449.

33 Bertocchini F, Ovitt CE, Conti A et al: Requirement for the ryanodine receptor type 3 for efficient contraction in neonatal skeletal muscles. *EMBO J* 1997; **16**: 6956–6963.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)