

# Genome-Wide Association Study Identifies *ALDH7A1* as a Novel Susceptibility Gene for Osteoporosis

Yan Guo<sup>1,9</sup>, Li-Jun Tan<sup>2,9</sup>, Shu-Feng Lei<sup>3</sup>, Tie-Lin Yang<sup>1</sup>, Xiang-Ding Chen<sup>2</sup>, Feng Zhang<sup>1</sup>, Yuan Chen<sup>1</sup>, Feng Pan<sup>1</sup>, Han Yan<sup>1</sup>, Xiaogang Liu<sup>1</sup>, Qing Tian<sup>3</sup>, Zhi-Xin Zhang<sup>1</sup>, Qi Zhou<sup>1</sup>, Chuan Qiu<sup>1</sup>, Shan-Shan Dong<sup>1</sup>, Xiang-Hong Xu<sup>1</sup>, Yan-Fang Guo<sup>1</sup>, Xue-Zhen Zhu<sup>1</sup>, Shan-Lin Liu<sup>2</sup>, Xiang-Li Wang<sup>2</sup>, Xi Li<sup>2</sup>, Yi Luo<sup>2</sup>, Li-Shu Zhang<sup>3</sup>, Meng Li<sup>4</sup>, Jin-Tang Wang<sup>4</sup>, Ting Wen<sup>5</sup>, Betty Drees<sup>3</sup>, James Hamilton<sup>3</sup>, Christopher J. Papasian<sup>3</sup>, Robert R. Recker<sup>6</sup>, Xiao-Ping Song<sup>1</sup>, Jing Cheng<sup>7</sup>, Hong-Wen Deng<sup>1,2,3,8\*</sup>

**1** Key Laboratory of Biomedical Information Engineering of Ministry of Education, and Institute of Molecular Genetics, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, People's Republic of China, **2** Molecular and Statistical Genetics Lab, College of Life Sciences, Hunan Normal University, Changsha, People's Republic of China, **3** School of Medicine, University of Missouri Kansas City, Kansas City, Missouri, United States of America, **4** Department of Orthopaedics, the First Affiliated Hospital, School of Medicine, Xi'an Jiaotong University, Xi'an, People's Republic of China, **5** Department of Orthopaedics, Xiangya Hospital of Central South University, Changsha, Hunan, People's Republic of China, **6** Osteoporosis Research Center, Creighton University, Omaha, Nebraska, United States of America, **7** National Engineering Research Center for Beijing Biochip Technology, Beijing, People's Republic of China, **8** Center of Systematic Biomedical Research, Shanghai University of Science and Technology, Shanghai, People's Republic of China

## Abstract

Osteoporosis is a major public health problem. It is mainly characterized by low bone mineral density (BMD) and/or low-trauma osteoporotic fractures (OF), both of which have strong genetic determination. The specific genes influencing these phenotypic traits, however, are largely unknown. Using the Affymetrix 500K array set, we performed a case-control genome-wide association study (GWAS) in 700 elderly Chinese Han subjects (350 with hip OF and 350 healthy matched controls). A follow-up replication study was conducted to validate our major GWAS findings in an independent Chinese sample containing 390 cases with hip OF and 516 controls. We found that a SNP, rs13182402 within the *ALDH7A1* gene on chromosome 5q31, was strongly associated with OF with evidence combined GWAS and replication studies ( $P = 2.08 \times 10^{-9}$ , odds ratio = 2.25). In order to explore the target risk factors and potential mechanism underlying hip OF risk, we further examined this candidate SNP's relevance to hip BMD both in Chinese and Caucasian populations involving 9,962 additional subjects. This SNP was confirmed as consistently associated with hip BMD even across ethnic boundaries, in both Chinese and Caucasians (combined  $P = 6.39 \times 10^{-6}$ ), further attesting to its potential effect on osteoporosis. *ALDH7A1* degrades and detoxifies acetaldehyde, which inhibits osteoblast proliferation and results in decreased bone formation. Our findings may provide new insights into the pathogenesis of osteoporosis.

**Citation:** Guo Y, Tan L-J, Lei S-F, Yang T-L, Chen X-D, et al. (2010) Genome-Wide Association Study Identifies *ALDH7A1* as a Novel Susceptibility Gene for Osteoporosis. PLoS Genet 6(1): e1000806. doi:10.1371/journal.pgen.1000806

**Editor:** Michel Georges, University of Liège, Belgium

**Received:** July 29, 2009; **Accepted:** December 9, 2009; **Published:** January 8, 2010

**Copyright:** © 2010 Guo et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by the National Institute of Health (R01 AR050496, R21 AG027110, R01 AG026564, P50 AR055081, and R21 AA015973). The study was also funded partially by from the National Natural Science Foundation of China (30570875), Xian Jiaotong University, and the Ministry of Education of China; Huo Ying Dong Education Foundation, HuNan Province, and Hunan Normal University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: dengh@umkc.edu

These authors contributed equally to this work.

## Introduction

Osteoporosis, characterized primarily by low bone mineral density (BMD), is a major public health problem because it increases susceptibility to low-trauma osteoporotic fractures (OF). Hip fractures, which are the most common and severe form of OF, are associated with high morbidity and mortality, as well as tremendous health care expenditures [1]. Due to an aging population, the annual incidence of hip fractures worldwide is predicted to be ~6.27 million by the year 2050, with an estimated cost of ~\$131.5 billion [1]. The ultimate goal of osteoporosis research is to reduce the incidence and prevalence of OF.

Genetic factors play an important role in susceptibility to osteoporosis. Both BMD and OF have high genetic determinations

[2,3,4,5]. BMD has been identified as the major risk factor for susceptibility to OF and is currently the predominant study phenotype for osteoporosis. Variations in BMD account for ~50–70% of the variation in total bone strength [6] and risk of OF [7]. Additional risk factors, including those not readily quantifiable (e.g. bone microstructure [8] and cartilage organization [9]), also contribute to the risk of OF. Most of genetic studies of osteoporosis have focused primarily on the surrogate phenotype BMD, whereas little effort has been expended on the study of OF *per se* as a focal phenotype or on the relevance of genes associated with BMD on OF [3]. The major obstacle to this approach has been assembling a homogeneous sample with a homogeneously defined OF type. Genetic factors associated with variations in BMD and risk of OF overlap, to some extent, but are not all

## Author Summary

Osteoporosis is a major health concern worldwide. It is a highly heritable disease characterized mainly by low bone mineral density (BMD) and/or osteoporotic fractures. However, the specific genetic variants determining risk for low BMD or OF are largely unknown. Here, taking advantage of recent technological advances in human genetics, we performed a genome-wide association study and follow-up validation studies to identify genetic variants for osteoporosis. By examining a total of 11,568 individuals from Chinese and Caucasian populations, we discovered a susceptibility gene, *ALDH7A1*, which is associated with hip osteoporotic fracture and BMD. *ALDH7A1* might inhibit osteoblast proliferation and decrease bone formation. Our finding opens a new avenue for exploring the pathophysiology of osteoporosis.

identical [2]. The ultimate goal of osteoporosis research is to reduce the incidence and prevalence of OF. Therefore, it is useful to conduct genetic studies of OF *per se*, in conjunction with other intermediate phenotypes (e.g. BMD) that influence the risk of OF. This approach can be used to identify quantifiable measures for early prevention and intervention before the adverse clinical outcome, OF, actually occurs.

So far, several specific genes contributing to osteoporosis (i.e. those impacting BMD or risk of OF) have been identified, such as *ESR1* with OF risk, *COL1A1* and *VDR* with BMD and vertebral fracture risk, *OPG* and *LRP5* with BMD [10,11,12,13,14,15]. However, the majority of genetic variants that influence osteoporosis remain unknown. With current high throughput SNP genotyping platforms and our knowledge about the distribution and correlation of SNPs in the human genome (e.g., haplotype structure), genome-wide association study (GWAS) has proven itself to be a feasible, powerful and effective approach for identifying novel genes associated with complex phenotypes. Four recent GWAS's [12,13,16,17] have identified several specific genes for osteoporosis. In the current investigation, based on significant heritability of ~50% for OF [2,4], we utilized a GWAS to identify genetic variants underlying susceptibility to osteoporosis that are directly relevant to the risk of OF. Using the Affymetrix 500K array set, we successfully genotyped a study population of 700 elderly Chinese Han subjects consisting of 350 cases with homogeneous hip OF and 350 healthy matched controls. A follow-up replication study was performed in an independent Chinese sample consisting of 390 cases with hip OF and 516 controls. For SNPs that were identified for OF, we further

examined their relationships with hip BMD in two ethnic groups (Chinese and Caucasians), involving additional 9,962 subjects, in order to determine whether the genetic basis for their contribution to the risk of OF might also be, at least partially, attributable to their effects on variation in BMD.

## Results

### GWAS Discovery Study

The study design included an initial exploratory stage in a Chinese Han sample of moderate size and follow-up replication and validation studies with much larger sample sizes in independent Chinese Han and Caucasian samples. Table 1 details the basic characteristics of the respective samples. In the GWAS discovery stage, a total of 281,533 SNPs passed our quality control criteria for GWAS analyses. A quantile-quantile (QQ) plot is presented in Figure 1. The  $\chi^2$  distributions for the association tests across the SNPs tested showed little evidence of overall systematic bias (genomic inflation factor  $\lambda = 1.02$ ). The highest  $\chi^2$  was consistent with the presence of true association. We further performed the principal component analysis implemented in EIGENSTRAT to guard against possible population stratification. The first two principal components were not associated with case status ( $P$  values > 0.05), further indicating that it is very unlikely that positive associations in this study would be attributable to confounding due to population structure. The association analyses by EIGENSTRAT confirmed, qualitatively, our main results and consequently, the results of the EIGENSTRAT analyses are not detailed here.

Table 2 lists the most promising results from GWAS analyses. We identified five SNPs with  $P$  values <  $5 \times 10^{-6}$  by allelic association analyses. After applying the Bonferroni correction for multiple testing, a single SNP, rs13182402, reached a genome-wide significance level ( $P < 1.78 \times 10^{-7}$ ). SNP rs13182402 achieved a  $P$  value of  $8.53 \times 10^{-9}$  in the allelic test (Bonferroni corrected  $P = 2.40 \times 10^{-3}$ ). The odds ratio (OR) was 2.94 (95% confidence interval (CI): 2.02–4.30) for minor allele G. The frequency of the G allele was 0.162 in cases, and 0.061 in controls. When all covariates were considered simultaneously in a multivariate logistic regression model, this SNP remained a significant predictor of OF risk, independent of age, sex, height, and weight ( $P = 2.21 \times 10^{-8}$ ).

### Assessment of Genome-Wide Findings

**Replication in Chinese.** Replication analyses were performed in an independent Chinese sample containing 390 cases with hip OF and 516 controls. Of the ten genotyped SNPs in the OF replication sample, two SNPs (rs13182402 and rs16894980)

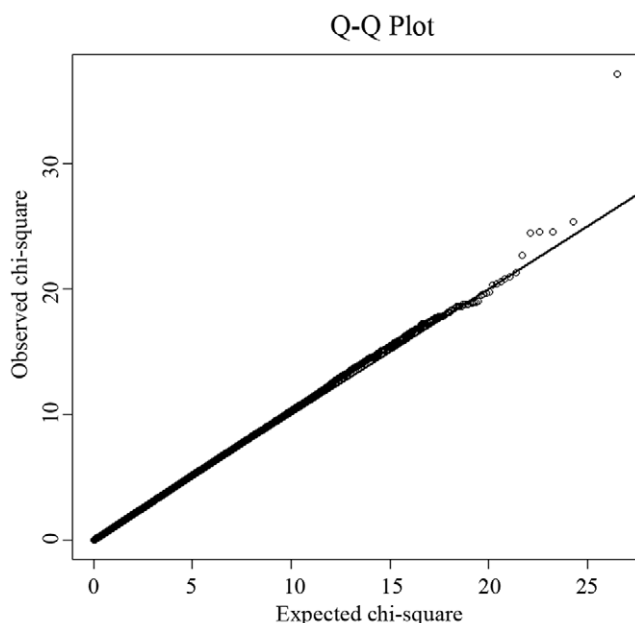
**Table 1.** Characteristics of the study subjects.

|                 | GWAS Sample |             | Chinese replication sample |             | Chinese BMD sample | US-MidWest samples |              | US-Framingham sample |
|-----------------|-------------|-------------|----------------------------|-------------|--------------------|--------------------|--------------|----------------------|
|                 | Case        | Control     | Case                       | Control     |                    | Unrelated          | Related      |                      |
| Number          | 350         | 350         | 390                        | 516         | 2,955              | 1,725              | 2,329        | 2,953                |
| Sex ratio (M/F) | 124/226     | 173/177     | 112/278                    | 188/328     | 1,437/1,518        | 868/857            | 843/1,486    | 1,274/1,679          |
| Age (years)     | 69.4 (7.4)  | 69.5 (6.1)  | 68.1 (12.5)                | 68.2 (6.7)  | 33.1 (14.5)        | 55.0 (16.8)        | 49.0 (15.7)  | 63.1 (12.7)          |
| Weight (kg)     | 59.2 (12.1) | 59.6 (10.8) | 56.8 (9.1)                 | 62.6 (9.8)  | 58.6 (10.2)        | 80.9 (18.0)        | 77.4 (18.3)  | 76.5 (17.1)          |
| Height (cm)     | 162.8 (8.3) | 159.4 (9.2) | 160.9 (7.4)                | 159.6 (8.5) | 163.7 (7.9)        | 170.3 (9.8)        | 167.8 (10.2) | 166.0 (9.9)          |

Abbreviations: M: male; F: female.

Data are shown as mean (standard deviation, SD).

doi:10.1371/journal.pgen.1000806.t001



**Figure 1. Quantile-quantile (QQ) plot of genome-wide allelic association results.** Under the null hypothesis of no association at any locus, the points would be expected to follow the slope line. Deviations from the slope line correspond to loci that deviate from the null hypothesis.  
doi:10.1371/journal.pgen.1000806.g001

were nominally significant (Table 2). However, considering the effect direction, only SNP rs13182402 was successfully replicated ( $P = 1.10 \times 10^{-2}$ ) and the effect was in the same direction as in the initial GWAS sample (OR: 1.66 for allele G, 95% CI: 1.12~2.43). Combining the GWAS discovery and replication samples by meta-analyses, rs13182402 achieved a  $P$  value of  $2.08 \times 10^{-9}$  with an estimated OR of 2.25 (95% CI: 1.72~2.94).

**BMD validation in Chinese and Caucasians.** To further explore the relationship between rs13182402 and osteoporosis risk, we performed association analyses with hip BMD in Chinese and Caucasian samples (Table 3). In the Chinese BMD sample, rs13182402 was associated with reduced hip BMD values ( $P = 2.35 \times 10^{-2}$ ) and the effect size ( $\beta$ ) was estimated to be  $\sim 0.04$  for each copy of the minor allele. This was consistent with its association with an increased risk of hip OF. The contribution of rs13182402 to BMD variation was estimated to be  $\sim 0.68\%$ .

Statistical significance of rs13182402 was consistently achieved in the US-MidWest Caucasian samples (unrelated sample:  $P = 1.92 \times 10^{-2}$ ; related sample:  $P = 1.60 \times 10^{-3}$ ), and the effect is in the same direction as in the Chinese BMD sample. The  $\beta$  was estimated to be 0.043 for each copy of the minor allele in the unrelated sample. The contribution of this SNP to BMD variation in the unrelated sample was estimated to be  $\sim 0.75\%$ .

We further examined the association signal in the US-Framingham Caucasian sample. SNP rs13182402 was consistently significantly associated with hip BMD in the US-Framingham sample ( $P = 3.38 \times 10^{-2}$ ).

We also examined the associations between rs13182402 and spine BMD in the Chinese and Caucasian BMD samples. However, no significant results were found (data not shown), which might be due to the heterogeneity of BMD across different skeletal sites [18].

Finally, using meta-analysis, we combined all of the BMD validation results (one Chinese sample and three Caucasian

**Table 2. Summary of the GWAS and replication studies.**

| dbSNP      | Chr | Position  | Nearest Gene | Allele <sup>a</sup> |              | Replication |              |              |                         | Combined         |                         |                  |
|------------|-----|-----------|--------------|---------------------|--------------|-------------|--------------|--------------|-------------------------|------------------|-------------------------|------------------|
|            |     |           |              | MAF cases           | MAF controls | MAF cases   | MAF controls | MAF controls | P value <sup>b</sup>    | OR (95% CI)      | P value                 | OR (95% CI)      |
| rs13182402 | 5   | 125946047 | ALDH7A1      | G/A                 | 0.162        | 0.061       | 0.085        | 0.053        | 1.10 × 10 <sup>-2</sup> | 1.66 (1.12~2.43) | 2.08 × 10 <sup>-9</sup> | 2.25 (1.72~2.94) |
| rs6039266  | 20  | 8714071   | PLCB1        | A/G                 | 0.078        | 0.023       | -            | -            | -                       | -                | -                       | -                |
| rs6711417  | 2   | 170666928 | UBR3         | G/T                 | 0.304        | 0.427       | 0.356        | 0.370        | 0.521                   | 0.94 (0.77~1.52) | 5.99 × 10 <sup>-4</sup> | 0.77 (0.66~0.89) |
| rs16894980 | 8   | 122333958 | SNTB1        | A/G                 | 0.123        | 0.055       | 0.081        | 0.131        | 9.76 × 10 <sup>-3</sup> | 0.59 (0.39~0.88) | 0.094                   | 1.27 (0.96~1.67) |
| rs3212217  | 5   | 158687708 | IL12B        | C/G                 | 0.384        | 0.415       | 0.461        | 0.479        | 0.450                   | 0.93 (0.77~1.12) | 0.212                   | 0.91 (0.79~1.05) |
| rs12100867 | 14  | 100582840 | DIO3         | G/A                 | 0.251        | 0.359       | 0.347        | 0.296        | 0.100                   | 1.26 (0.96~1.66) | 0.023                   | 0.81 (0.68~0.97) |
| rs12590815 | 14  | 100573762 | DIO3         | C/T                 | 0.196        | 0.292       | 0.235        | 0.243        | 0.808                   | 0.96 (0.71~1.30) | 8.29 × 10 <sup>-4</sup> | 0.72 (0.60~0.87) |
| rs746219   | 20  | 57832814  | PHACTR3      | T/G                 | 0.256        | 0.171       | 0.231        | 0.226        | 0.804                   | 1.03 (0.82~1.29) | 0.831                   | 0.97 (0.82~1.16) |
| rs6064822  | 20  | 57648532  | PHACTR3      | A/G                 | 0.077        | 0.131       | 0.086        | 0.079        | 0.612                   | 1.09 (0.78~1.52) | 0.056                   | 0.78 (0.62~1.01) |
| rs1555364  | 20  | 57644265  | PHACTR3      | C/T                 | 0.066        | 0.119       | 0.096        | 0.071        | 0.178                   | 1.38 (0.86~2.23) | 0.110                   | 0.78 (0.58~1.03) |

Abbreviations: Chr: chromosome; MAF, minor allele frequency; OR, odds ratio; CI, confidence interval.

<sup>a</sup>The former allele represents the minor allele.

<sup>b</sup>All  $P$  values listed in Table 2 are two-sided.

doi:10.1371/journal.pgen.1000806.t002

**Table 3.** BMD validation association results for rs13182402.

| Sample                      | Number | Alleles <sup>a</sup> | MAF   | P value <sup>b</sup>  | $\beta^c$ |
|-----------------------------|--------|----------------------|-------|-----------------------|-----------|
| Chinese BMD sample          | 2,955  | G/A                  | 0.062 | $2.35 \times 10^{-2}$ | -0.040    |
| US-MidWest-unrelated sample | 1,725  | G/A                  | 0.098 | $1.92 \times 10^{-2}$ | -0.043    |
| US-MidWest-related sample   | 2,329  | G/A                  | 0.110 | $1.60 \times 10^{-3}$ | -         |
| US-Framingham sample        | 2,953  | G/A                  | 0.084 | $3.38 \times 10^{-2}$ | -         |
| combined samples            | 9,962  |                      |       | $6.39 \times 10^{-6}$ | -         |

Abbreviations: MAF, minor allele frequency.

<sup>a</sup>The former allele represents the minor allele.

<sup>b</sup>P values were one tailed.

<sup>c</sup>The estimation for SNP effect size was performed under additive model.

<sup>d</sup>The P value from each sample set was combined based on the Stouffer method to quantify the overall association significance.

doi:10.1371/journal.pgen.1000806.t003

samples) to yield a commonly used probability measure. The statistical significance for rs13182402 was significantly improved ( $P = 6.39 \times 10^{-6}$ ). These findings, combined with the results of our GWAS studies, lend strong support for the conclusion that rs13182402 is associated with low hip BMD and increased risk for OF.

### Fine Mapping for Gene Identification

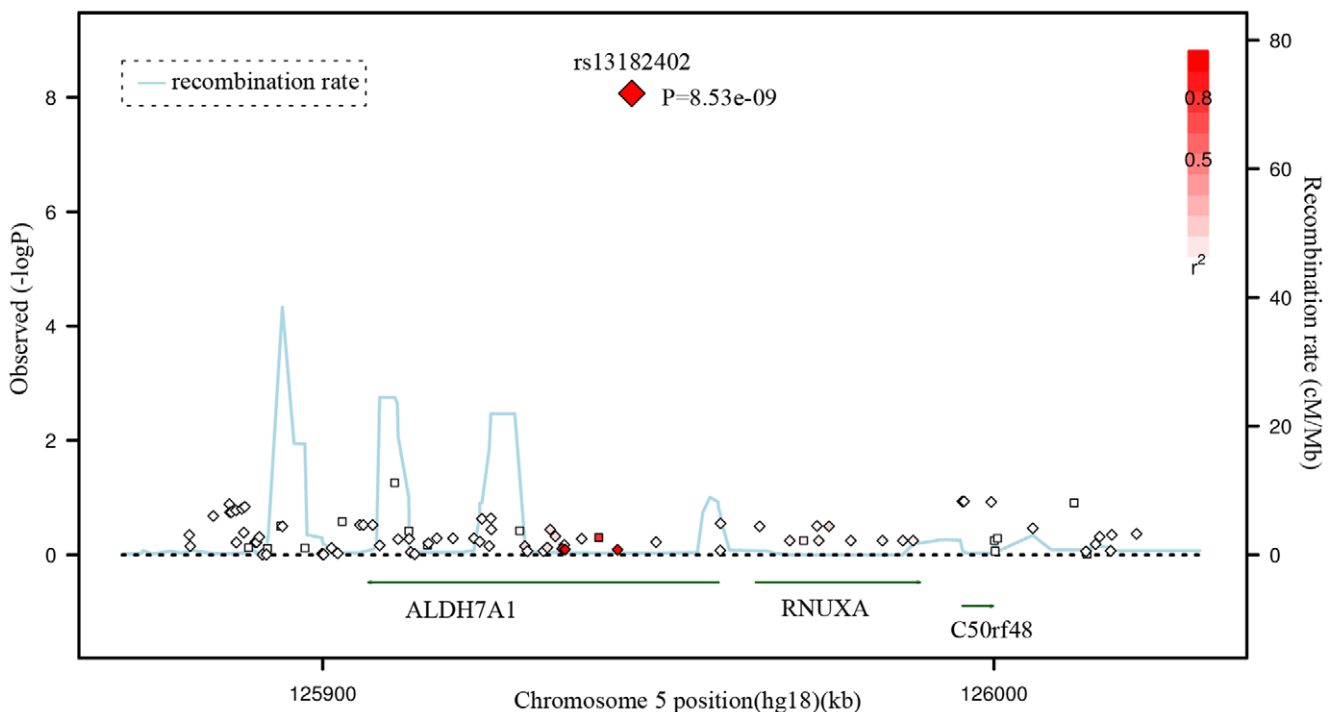
Given the significant evidence for rs13182402, we imputed the genotypes of SNPs located surrounding this SNP based on our GWAS data and Asian HapMap data, and presented a regional association plot in Figure 2. The most significantly associated SNP, rs13182402 (GWAS:  $P = 8.53 \times 10^{-9}$ ), is located 394 bp downstream from exon 5 of the *ALDH7A1* gene (aldehyde dehydrogenase 7 family, member A1) on chromosome 5q31. According to

the FASTSNP program (<http://fastsnp.ibms.sinica.edu.tw>), a change of “A→G” at rs13182402 may lead to removal of binding sites for transcription factors *RORα* and *CdxA*.

### Discussion

In this study, we first performed a GWAS and follow-up replication on OF and identified a novel susceptibility gene (*ALDH7A1*) that significantly impacts the risk for OF *per se*. Next, we examined this gene’s relationship with hip BMD both in Chinese and Caucasian populations, and this gene was consistently associated with hip BMD even across ethnic boundaries. The effect size on BMD was modest and lower than the effect size on OF risk. One interpretation of this differential effect would be that BMD is not the only risk factor for OF; other risk factors also contribute to the risk of OF. It is consistent with and supports our statement in the introduction. It might also be caused by the differences in power between the relatively small hip OF samples compared to the large BMD samples. In addition, because we didn’t have BMD measurements for the hip OF cases, we couldn’t adjust the OR for BMD to see if the risk would be attenuated by the adjustment. However, regardless of this differential effect, the significant association results we identified both for BMD and OF risk strongly support the potential contribution of *ALDH7A1* to the pathogenesis of osteoporosis.

The *ALDH7A1* gene encodes an enzyme of the acetaldehyde dehydrogenase superfamily, which degrades and detoxifies acetaldehyde generated by alcohol metabolism. Acetaldehyde has been shown to inhibit osteoblast proliferation and to decrease bone formation [19]. In addition, previous studies have identified that polymorphisms of the *ALDH2* gene, another member of the acetaldehyde dehydrogenase family, are significantly associated with osteoporosis [20]. Our findings, combined with the above



**Figure 2. Regional Association Plot for rs13182402 on chromosome 5 in the GWAS stage.** The color scheme of a white-to-red gradient reflects lower to higher LD values ( $r^2$ ). The scatter graph indicates the negative logarithm of P-value for each SNP. doi:10.1371/journal.pgen.1000806.g002

lines of evidence, suggest that *ALDH7A1* might be a novel and potential candidate gene contributing to the risk of osteoporosis.

Using the genotyped and imputed genotypes in our GWAS sample of 700 Chinese, we examined the associations between hip OF and the key SNPs identified in previous GWAS on osteoporosis [12,13,16]. Table 4 summarizes the major results. Only two SNPs in *RANKL* were confirmed to be associated with hip OF in our sample, including rs9594759 ( $P=0.020$ ) and rs9594738 ( $P=0.045$ ). The data provided may serve as a reference for other investigators searching for replication for their GWAS results.

An apparent advantage of this study is that our GWAS sample came from a homogenous population with well defined homogeneous phenotype. The genomic control factor was quite close to 1.0 ( $\lambda=1.02$ ) (expected under no population stratification) and, analyses by EIGENSTRAT showed qualitatively supportive results. Thus, our association results are unlikely to be plagued by spurious associations due to population stratification. In particular, since the significant associations with BMD are shown in both Caucasian and Chinese samples, the results are even less likely to be due to population stratification/admixture.

A potential limitation of our study is the relatively small size of the GWAS sample and the replication sample, which might lead to over estimation of the effect size for the significant SNPs identified. However, hip fractures are the most severe OFs followed by high mortality rates, making subjects recruitment difficult. It took us several years to accumulate such a homogeneous hip OF sample. This study represents the best we can do under current conditions to identify genes for OF. Meanwhile, we are keeping the recruitment of hip OF subjects. As a future direction, a new GWAS needs to be implemented on a larger sample to identify more comprehensively novel genes for OF. In addition, since genetic and environmental backgrounds vary for different populations, replication across a wide range of populations is necessary to determine the generality of our findings to the

broader population, or to specific ethnic groups or populations [21].

In summary, using data from over 11,500 individuals, we have identified and validated *ALDH7A1* as a novel susceptibility candidate gene for osteoporosis. Further studies are warranted to explore the generality of our findings for *ALDH7A1* identified by GWAS to other populations, and to determine the mechanisms by which this gene and its products contribute to the pathogenesis of osteoporosis.

## Materials and Methods

### Ethics Statement

The study was approved by the local institutional review boards or the office of research administration of all participating institutions. After signing an informed consent, all subjects received assistance completing a structured questionnaire including anthropometric variables, lifestyles, and medical history.

### Study Samples

The study was initially performed with a GWAS discovery stage for SNPs of potential significance for OF in a Chinese Han, case-control sample. The significance of the SNPs identified in the discovery stage was subsequently confirmed through replication study in another independent Chinese case-control sample. For SNPs identified for OF, we further examined their relationships with hip BMD within/across ethnic groups in a Chinese unrelated BMD sample and three independent Caucasian samples. Table 1 details the basic characteristics of the respective samples, with additional descriptions below.

### GWAS Discovery Sample

The sample for the initial GWAS consisted of 350 patients with osteoporotic (low trauma) hip fractures (including 124 males and 226 females) and 350 elderly controls (including 173 males and

**Table 4.** Comparison of the previous GWAS for BMD and the current GWAS.

| SNP       | Associated gene | Cytoband | Current GWAS $P$ value | Published GWAS $P$ value <sup>a</sup>       | Reference |
|-----------|-----------------|----------|------------------------|---------------------------------------------|-----------|
| rs9479055 | <i>ESR1</i>     | 6q25     | 0.936                  | $7.0 \times 10^{-4}$ (hipBMD <sup>b</sup> ) | [13]      |
| rs1038304 | <i>ESR1</i>     | 6q25     | 0.634                  | $1.8 \times 10^{-5}$ (hipBMD)               | [13]      |
| rs6929137 | <i>ESR1</i>     | 6q25     | 0.970                  | $1.4 \times 10^{-5}$ (hipBMD)               | [13]      |
| rs1999805 | <i>ESR1</i>     | 6q25     | 0.849                  | 0.002(hipBMD)                               | [13]      |
| rs2504063 | <i>ESR1</i>     | 6q25     | 0.141                  | $5.7 \times 10^{-8}$ (SPBMD <sup>c</sup> )  | [13]      |
| rs851982  | <i>ESR1</i>     | 6q25     | 0.192                  | $1.6 \times 10^{-5}$ (hipBMD)               | [13]      |
| rs4870044 | <i>ESR1</i>     | 6q25     | 0.751                  | $9.9 \times 10^{-5}$ (hipBMD)               | [13]      |
| rs3736228 | <i>LRP5</i>     | 11q13    | 0.684                  | $1.9 \times 10^{-5}$ (SPBMD)                | [12]      |
| rs2010281 | <i>MARK3</i>    | 14q32    | 0.331                  | $7.4 \times 10^{-5}$ (hipBMD)               | [16]      |
| rs4355801 | <i>OPG</i>      | 8q24     | 0.481                  | $7.9 \times 10^{-4}$ (SPBMD)                | [12]      |
| rs9594759 | <i>RANKL</i>    | 13q14    | 0.020                  | $4.5 \times 10^{-4}$ (hipBMD)               | [13]      |
| rs9594738 | <i>RANKL</i>    | 13q14    | 0.045                  | $2.7 \times 10^{-4}$ (hipBMD)               | [13]      |
| rs3018362 | <i>RANK</i>     | 18q21    | 0.185                  | $3.5 \times 10^{-5}$ (hipBMD)               | [16]      |
| rs1513670 | <i>SOST</i>     | 17q21    | 0.385                  | $6.0 \times 10^{-5}$ (hipBMD)               | [16]      |
| rs7220711 | <i>SOST</i>     | 17q21    | 0.930                  | $1.4 \times 10^{-4}$ (hipBMD)               | [16]      |
| rs1107748 | <i>SOST</i>     | 17q21    | 0.930                  | $7.2 \times 10^{-5}$ (hipBMD)               | [16]      |

<sup>a</sup> $P$  value reported here was the original  $P$  value in the discovery sample in each GWAS.

<sup>b</sup>hipBMD is the combined BMD at the femoral neck, trochanter and intertrochanter region.

<sup>c</sup>SPBMD: Spine BMD.

doi:10.1371/journal.pgen.1000806.t004



177 females). Since fractures at different skeletal sites may have different underlying pathological mechanisms, we focused exclusively on hip fractures in order to minimize potential clinical and genetic heterogeneity of the study phenotype. All the subjects were unrelated northern Chinese Han adults living in the city of Xi'an and its neighboring areas. Affected individuals with low trauma hip fractures were recruited from the affiliated hospitals and their associated clinics of Xi'an Jiaotong University. Inclusion criteria for cases were (i) onset age of hip OF > 55 years, to make sure all female subjects were postmenopausal, and the onset of OF was largely due to decreased BMD; (ii) age < 80 years to minimize the effect due to age, since previous studies showed that approximately half of females aged 80 years or older have fractures [22]; (iii) fractures occurred with minimal or no trauma, usually due to falls from standing height or less; (iv) the fracture sites were at the femoral neck or inter-trochanter regions; (v) hip fracture was identified/confirmed through diagnosis of orthopedic surgeons/radiologists according to radiological reports and x-rays. Patients with pathological fractures and high-impact fractures (such as due to motor vehicle accidents) were excluded. Patients with chronic diseases before the onset of HF were also excluded.

Healthy control subjects were enrolled by use of local advertisements. They were geography- and age-matched to the cases. Inclusion/exclusion criteria for controls were: (i) age at exam must be > 55 years, without any fracture histories (all female controls were postmenopausal); (ii) subjects with chronic diseases and conditions that might potentially affect bone mass, structure, or metabolism were excluded. Diseases/conditions resulting in exclusion included chronic disorders involving vital organs (heart, lung, liver, kidney, brain), serious metabolic diseases (diabetes, hypo- and hyper-parathyroidism, hyperthyroidism, etc.), other skeletal diseases (Paget's disease, osteogenesis imperfecta, rheumatoid arthritis, etc.), chronic use of drugs affecting bone metabolism (e.g., hormone replacement therapy, corticosteroid therapy, anti-convulsant drugs), and malnutrition conditions (such as chronic diarrhea, chronic ulcerative colitis); (iii) subjects taking anti-bone-resorptive or bone anabolic agents/drugs, such as bisphosphonates were excluded.

### Chinese Replication Sample

For replication of our GWAS findings for hip OF, we used an independent Chinese sample containing 906 unrelated Han subjects (390 cases with hip OF and 516 controls). All subjects were drawn from the same geographic area as the above GWAS discovery sample, and the sample inclusion and exclusion criteria for cases and controls were the same as those adopted in the recruitment of the above GWAS sample.

For SNPs that were identified for OF, we further performed validation analyses to evaluate their relevance with hip BMD (with targeted experimental genotyping of candidate SNPs discovered in the initial GWAS) in two ethnic groups, including a Chinese sample and two US Mid-West Caucasian samples. We finally performed *in silico* validation to compare the association signals of our most promising GWAS results with those achieved in the Framingham Heart Study (FHS) [23].

### Chinese BMD Sample

The Chinese BMD sample contained 2,955 unrelated ethnic Han adults. This sample came from Changsha, Hunan province, which is more than 1,000 km from Xi'an where the sample for the GWAS was recruited. The subjects were randomly selected from an established and expanding database with BMD measurements. The exclusion criteria were the same as those adopted in the

recruitment of healthy control subjects in the GWAS sample, and have been detailed in our earlier publication [2].

### US-MidWest Caucasian BMD Samples

The US-MidWest BMD samples with a total of 4,054 subjects consisted of two independent sample sets, including one sample of unrelated subjects and the other sample of nuclear families, which were all US Caucasians of Northern European origin living in Omaha, Nebraska, and its surrounding regions in Midwestern USA. They were normal healthy subjects defined by the same exclusion criteria as above in Chinese samples. The unrelated sample contained 1,725 subjects. The related sample contained 2,329 subjects from 593 nuclear families.

All hip BMD measurements for the above BMD samples were obtained with dual-energy X-ray absorptiometry using the same type of machine (Hologic 4500) under the same protocol defined by the manufacturer (Hologic Inc., Bedford, MA, USA). The machines were calibrated daily. The coefficients of variation (CV) of the hip BMD measurements were 1.33% for Chinese and 1.40% for Caucasians, respectively.

### US-Framingham BMD Sample

The US-Framingham BMD sample is from the Framingham Osteoporosis Study, an ancillary study of the Framingham SNP Health Association Resource (SHARe) data sets [23]. Details and descriptions about the Framingham Osteoporosis Study have been previously reported [24]. Both genotype and phenotype data were downloaded from dbGaP database (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap>). Data download and usage was authorized by SHARe data access committee (phs000007.v3.p2, phs000078.v3.p2). We have the data on 2,953 phenotyped Caucasian subjects, 448 from the Original cohort (160 men and 288 women) and 2,505 from the Offspring cohort (1,114 men and 1,391 women). The Original Cohort participants had BMD measures by dual x-ray absorptiometry machine (Lunar DPX-L) at the hip performed at exam 24. The Offspring Cohort participants were scanned with the same machine at exam 6/7. As reported before [24], the CV was 1.7% for femoral neck.

### Genotyping and Quality Control

Genomic DNA was extracted from peripheral blood leukocytes using standard protocols. The genome-wide scan was performed using the Affymetrix Human Mapping 500K array set (Affymetrix, Santa Clara, CA, USA) according to the Affymetrix protocol. Data management and analyses were conducted using the Affymetrix GeneChip Operating System. Genotyping calls were determined from the fluorescent intensities using the DM algorithm with a 0.33 *P*-value setting [25] as well as the BRLMM algorithm [26].

Quality control procedures were as follows. First, only samples with a minimum call rate of 95% were included. Due to efforts of repeat experiments, all samples ( $n = 700$ ) met this criteria and the final mean BRLMM call rate reached a high level of 99.02%. Second, out of the initial full-set of 500,568 SNPs, we discarded: 1) SNPs with a call rate < 90% in the total sample ( $n = 54,845$ ); 2) those deviating from Hardy-Weinberg equilibrium (HWE) in controls ( $P < 0.001$ ,  $n = 22,002$ ); 3) those having a minor allele frequency (MAF) < 0.05 in the total sample ( $n = 142,188$ ). Therefore, 281,533 SNPs were available for subsequent analyses.

Based on the initial GWAS results, we selected the 10 most promising SNPs for subsequent genotyping in the Chinese hip OF replication sample based on the following inclusion criteria: (i) *P* values  $\leq 5 \times 10^{-6}$  in the GWAS allelic association analyses (5 SNPs); (ii) *P* values between  $5 \times 10^{-6}$  and  $5 \times 10^{-5}$ , with

neighboring SNPs having  $P$  values  $\leq 10^{-4}$  showing a consistent trend of association (5 SNPs). Genotypes were obtained using MALDI-TOF mass spectrometry on a Sequenom system (Sequenom, Inc., San Diego, CA) with iPLEX assay [27]. Primers were designed using MassARRAY Assay Design 3.1 software. Genotyping quality control procedures leading to SNP exclusion were call rate  $< 90\%$ ,  $MAF < 0.05$  in the total sample and  $P < 0.001$  for deviations from HWE in controls. The average call rate was 98.7% for the Sequenom system and the corresponding consistency of genotyping (replication or concordance rates), as obtained by duplication samples, was 99.8%. Nine of the ten genotyped SNPs were qualified for subsequent association analyses.

The Chinese BMD sample was also genotyped using the Sequenom system, which was the same as that used for the OF replication sample. Genotyping of the two US-MidWest samples was performed by a service company KBioscience (Herts, UK) using a technology of competitive allele specific PCR (KASPar), which is detailed at the website ([www.kbioscience.co.uk](http://www.kbioscience.co.uk)). The US-Framingham sample was genotyped using approximately 550,000 SNPs (Affymetrix 500K mapping array plus Affymetrix 50K supplemental array).

### Statistical Analyses

For the GWAS analyses for OF risk, single-marker allelic association analysis was performed by comparing SNP allele counts among cases and controls with a  $\chi^2$  test. ORs with the corresponding 95% CIs were also computed. For the interesting SNPs identified by allelic tests, we also used a multivariate logistic regression model to examine associations with OF risk, taking into account potential covariates such as age, sex, height, and weight. For the sex chromosome analyses, the affymetrix platform does not assay the Y chromosome. The X chromosome needs to be treated differently from the autosomes since males have only one copy of the X chromosome. As most loci on the X chromosome are subject to X chromosome inactivation, it is reasonable to treat males as if they were homozygous females, and then the assumption was the same as tests for autosomes. All association statistical analyses were carried out using HelixTree 5.3.1 software (Golden Helix, Bozeman, MT, USA). We adjusted for multiple testing by adopting the conservative Bonferroni correction. The genome-wide significance threshold was set at a  $P$  value of less than  $1.78 \times 10^{-7}$  ( $0.05/281,533$  SNPs that passed our quality control check).

To correct for potential population stratification that may lead to spurious association results, we estimated the inflation factor ( $\lambda$ ) for the GWAS sample using a method of genomic control [28].  $\lambda$  was calculated as the median of the observed  $\chi^2$  statistics divided by the median of the expected  $\chi^2$  statistics for the genome-wide SNP set. This led to an  $\lambda$  of 1.02. Results presented in this study were based on adjusting  $\chi^2$  statistics by dividing each of them by 1.02. The data were also analyzed by the principal component analyses implemented in EIGENSTRAT [29] for cross-checking the association results while controlling for admixture.

For OF replication analyses, the same allelic association analysis was performed by  $\chi^2$  tests. For BMD validation analyses,

significant parameters ( $P < 0.05$ ) such as age, sex, height and weight were used as covariates to adjust for the raw BMD values. For the unrelated samples, ANOVA was conducted to achieve the association tests. ANOVA is a model free test and more robust than assuming any genetic models. The independent variable was the genotype, which was divided into three levels corresponding to the three genotypes observed for each SNP (1,1; 1,2; 2,2). Since ANOVA can't give the effect size, we estimated the effect size of significant SNPs using the linear regression assuming the additive model in SAS (SAS Institute Inc., Cary, NC). For the related samples, we used the BMD residuals after adjustment to conduct family-based association tests under an additive model using FBAT program [30]. FBAT is a powerful approach to handle family sample, which tests for differences in probability of transmission of a genotype from parents to offspring based on phenotype. FBAT examines association within families, which is not affected by population stratification bias. To quantify the overall evidence of associations, we performed meta-analyses by using the Mantel-Haenszel method to calculate the  $P$  values and OR for the combined OF samples. For the BMD samples, we used a weighted Z-score method to calculate the combined  $P$  values. The individual Z-score (a standard normal deviate, the statistic associated with a  $P$  value) was weighted by the square root of the sample size of each sample. We added the individual weighted Z-score together and divided by the square root of the total sample size to obtain a combined Z-score and an associated combined  $P$  value (Stouffer method) [31]. If an individual result was nonsignificant and gave no other useful data for calculation of a Z-score, we set it as 0 to calculate the combined probability.

For the interested genomic regions, IMPUTE program [32] was utilized to impute the genotypes of all SNPs located in the regions based on Asian HapMap data. SNPTEST [32] was used to test for associations between the imputed SNPs and OF. SNAP was used to depict the regional association plot [33].

### Acknowledgments

We thank the Framingham Heart Study and the Framingham SHARE project, which are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with Boston University. The Framingham SHARE data used for the analyses described in this manuscript were obtained through dbGaP (phs000007.v3.p2, phs000078.v3.p2). This manuscript was not prepared in collaboration with investigators of the Framingham Heart Study and does not necessarily reflect the opinions or views of the Framingham Heart Study, Boston University, or the NHLBI.

### Author Contributions

Conceived and designed the experiments: HWD. Performed the experiments: YG IJT SFL TLY XDC FZ YC FP HY ZZZ QZ CQ SSD XHX SLL XLW XL YL. Analyzed the data: TLY XL YFG. Contributed reagents/materials/analysis tools: QT XZZ LSZ ML JTW TW RRR XPS JC. Wrote the paper: YG. Manuscript writing revision: BD, CJP, JH.

### References

- Cooper C, Campion G, Melton LJ 3rd (1992) Hip fractures in the elderly: a world-wide projection. *Osteoporos Int* 2: 285–289.
- Deng HW, Mahaney MC, Williams JT, Li J, Conway T, et al. (2002) Relevance of the genes for bone mass variation to susceptibility to osteoporotic fractures and its implications to gene search for complex human diseases. *Genet Epidemiol* 22: 12–25.
- Liu YJ, Shen H, Xiao P, Xiong DH, Li LH, et al. (2006) Molecular genetic studies of gene identification for osteoporosis: a 2004 update. *J Bone Miner Res* 21: 1511–1535.
- Michaelsson K, Melhus H, Ferm H, Ahlborn A, Pedersen NL (2005) Genetic liability to fractures in the elderly. *Arch Intern Med* 165: 1825–1830.
- Yang TL, Chen XD, Guo Y, Lei SF, Wang JT, et al. (2008) Genome-wide copy-number-variation study identified a susceptibility gene, UGT2B17, for osteoporosis. *Am J Hum Genet* 83: 663–674.
- Ammann P, Rizzoli R (2003) Bone strength and its determinants. *Osteoporos Int* 14 Suppl 3: S13–18.
- Marshall D, Johnell O, Wedel H (1996) Meta-analysis of how well measures of bone mineral density predict occurrence of osteoporotic fractures. *BMJ* 312: 1254–1259.
- Hazenberg JG, Taylor D, Lee TC (2007) The role of osteocytes and bone microstructure in preventing osteoporotic fractures. *Osteoporos Int* 18: 1–8.

9. Carrington JL (2005) Aging bone and cartilage: cross-cutting issues. *Biochem Biophys Res Commun* 328: 700–708.
10. Ioannidis JP, Ralston SH, Bennett ST, Brandi ML, Grinberg D, et al. (2004) Differential genetic effects of ESR1 gene polymorphisms on osteoporosis outcomes. *JAMA* 292: 2105–2114.
11. Ralston SH, Uitterlinden AG, Brandi ML, Balcells S, Langdahl BL, et al. (2006) Large-scale evidence for the effect of the COL1A1 Sp1 polymorphism on osteoporosis outcomes: the GENOMOS study. *PLoS Med* 3: e90. doi:10.1371/journal.pmed.0030090.
12. Richards JB, Rivadeneira F, Inouye M, Pastinen TM, Soranzo N, et al. (2008) Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. *Lancet* 371: 1505–1512.
13. Styrkarsdottir U, Halldorsson BV, Gretarsdottir S, Gudbjartsson DF, Walters GB, et al. (2008) Multiple genetic loci for bone mineral density and fractures. *N Engl J Med* 358: 2355–2365.
14. Uitterlinden AG, Ralston SH, Brandi ML, Carey AH, Grinberg D, et al. (2006) The association between common vitamin D receptor gene variations and osteoporosis: a participant-level meta-analysis. *Ann Intern Med* 145: 255–264.
15. van Meurs JB, Trikalinos TA, Ralston SH, Balcells S, Brandi ML, et al. (2008) Large-scale analysis of association between LRP5 and LRP6 variants and osteoporosis. *JAMA* 299: 1277–1290.
16. Styrkarsdottir U, Halldorsson BV, Gretarsdottir S, Gudbjartsson DF, Walters GB, et al. (2009) New sequence variants associated with bone mineral density. *Nat Genet* 41: 15–17.
17. Xiong DH, Liu XG, Guo YF, Tan LJ, Wang L, et al. (2009) Genome-wide association and follow-up replication studies identified ADAMTS18 and TGFB3 as bone mass candidate genes in different ethnic groups. *Am J Hum Genet* 84: 388–398.
18. Deng HW, Li JL, Li J, Davies KM, Recker RR (1998) Heterogeneity of bone mineral density across skeletal sites and its clinical implications. *J Clin Densitom* 1: 339–353.
19. Giuliani N, Girasole G, Vescovi PP, Passeri G, Pedrazzoni M (1999) Ethanol and acetaldehyde inhibit the formation of early osteoblast progenitors in murine and human bone marrow cultures. *Alcohol Clin Exp Res* 23: 381–385.
20. Yamaguchi J, Hasegawa Y, Kawasaki M, Masui T, Kanoh T, et al. (2006) ALDH2 polymorphisms and bone mineral density in an elderly Japanese population. *Osteoporos Int* 17: 908–913.
21. Cooper RS, Tayo B, Zhu X (2008) Genome-wide association studies: implications for multiethnic samples. *Hum Mol Genet* 17: R151–155.
22. Ross PD (1998) Risk factors for osteoporotic fracture. *Endocrinol Metab Clin North Am* 27: 289–301.
23. Cupples LA, Arruda HT, Benjamin EJ, D'Agostino RB Sr, Demissie S, et al. (2007) The Framingham Heart Study 100K SNP genome-wide association study resource: overview of 17 phenotype working group reports. *BMC Med Genet* 8 Suppl 1: S1.
24. Hannan MT, Felson DT, Anderson JJ (1992) Bone mineral density in elderly men and women: results from the Framingham osteoporosis study. *J Bone Miner Res* 7: 547–553.
25. Di X, Matsuzaki H, Webster TA, Hubbell E, Liu G, et al. (2005) Dynamic model based algorithms for screening and genotyping over 100 K SNPs on oligonucleotide microarrays. *Bioinformatics* 21: 1958–1963.
26. Rabbee N, Speed TP (2006) A genotype calling algorithm for affymetrix SNP arrays. *Bioinformatics* 22: 7–12.
27. Braun A, Roth R, McGinniss MJ (2003) Technology challenges in screening single gene disorders. *Eur J Pediatr* 162 Suppl 1: S13–16.
28. Devlin B, Roeder K (1999) Genomic control for association studies. *Biometrics* 55: 997–1004.
29. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38: 904–909.
30. Lange C, DeMeo DL, Laird NM (2002) Power and design considerations for a general class of family-based association tests: quantitative traits. *Am J Hum Genet* 71: 1330–1341.
31. Rosenthal R (1991) *Meta-analytic procedures for social research*. Calif: Sage.
32. Marchini J, Howie B, Myers S, McVean G, Donnelly P (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 39: 906–913.
33. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, et al. (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 24: 2938–2939.