ORIGINAL ARTICLE

JBMR®

Meta-Analysis of Genome-Wide Association Studies Identifies Three Loci Associated With Stiffness Index of the Calcaneus

Hsing-Fang Lu,[1,2] Kuo-Sheng Hung,[3,4,5] Hou-Wei Chu,[6] Henry Sung-Ching Wong,[1,3] Jihye Kim,[7,8] Mi Kyung Kim,[7,8] Bo Youl Choi,[7,8] Yu-Ting Tai,[9] Shiro Ikegawa,[2] Er-Chieh Cho,[1,3,10] and Wei-Chiao Chang[1,3,11,12,13]

[1]School of Pharmacy, Taipei Medical University, Taipei, Taiwan
[2]Laboratory of Bone and Joint Diseases, RIKEN Center for Integrative Medical Sciences, Tokyo, Japan
[3]Master Program for Clinical Pharmacogenomics and Pharmacoproteomics, School of Pharmacy, Taipei Medical University, Taipei, Taiwan
[4]Department of Neurosurgery, Taipei Medical University-Wan Fang Hospital, Taipei, Taiwan
[5]Graduate Institute of Injury, Prevention and Control, College of Public Health, Taipei Medical University, Taipei, Taiwan
[6]Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan
[7]Department of Preventive Medicine, College of Medicine, Hanyang University, Seoul, South Korea
[8]Institute for Health and Society, Hanyang University, Seoul, South Korea
[9]Department of Anesthesiology, Taipei Medical University, Taipei, Taiwan
[10]Department of Clinical Pharmacy, School of Pharmacy, College of Pharmacy, Taipei Medical University, Taipei, Taiwan
[11]Department of Pharmacy, Taipei Medical University-Wanfang Hospital, Taipei, Taiwan
[12]Center for Biomarkers and Biotech Drugs, Kaohsiung Medical University, Kaohsiung, Taiwan
[13]Department of Medicine Research, Taipei Medical University-Shuang Ho Hospital, New Taipei City, Taiwan

ABSTRACT

The stiffness index (SI) from quantitative ultrasound measurements is a good indicator of BMD and may be used to predict the risk of osteoporotic fracture. We conducted a genomewide association study (GWAS) for SI using 7742 individuals from the Taiwan Biobank, followed by a replication study in a Korean population ($n = 2955$). Approximately 6.1 million SNPs were subjected to association analysis, and SI-associated variants were identified. We further conducted a meta-analysis of Taiwan Biobank significant SNPs with a Korean population-based cohort. Candidate genes were prioritized according to epigenetic annotations, gene ontology, protein–protein interaction, GWAS catalog, and expression quantitative trait loci analyses. Our results revealed seven significant single-nucleotide polymorphisms (SNPs) within three loci: 7q31.31, 17p13.3, and 11q14.2. Conditional analysis showed that three SNPs, rs2536195 (*CPED1/WNT16*), rs1231207 (*SMG6*), and rs4944661 (*LOC10050636/TMEM135*), were the most important signals within these regions. The associations for the three SNPs were confirmed in a UK Biobank estimated BMD GWAS; these three cytobands were replicated successfully after a meta-analysis with a Korean population cohort as well. However, two SNPs were not replicated. After prioritization, we identified two novel genes, *RAB15* and *FNTB,* as strong candidates for association with SI. Our study identified three SI-associated SNPs and two novel SI-related genes. Overall, these results provide further insight into the genetic architecture of osteoporosis. Further studies in larger East Asian populations are needed. © 2019 American Society for Bone and Mineral Research.

KEY WORDS: STIFFNESS INDEX; GENOME-WIDE ASSOCIATION STUDY; TAIWAN BIOBANK; OSTEOPOROSIS; HEEL BONE MINERAL DENSITY

Introduction

Osteoporosis is a major medical problem worldwide that is commonly diagnosed by measuring BMD.[1] Low BMD may contribute to higher risk of fractures, leading to serious burdens on healthcare providers and patients.[2] As a means to assess heel BMD, quantitative ultrasound (QUS) is currently the most common method, owing to its ease of use and low cost. QUS is a reliable technique for evaluating the bone architecture of the heel.[3–5] This method has been used to predict the risk of fractures based on two parameters, the speed of sound (SOS) and the broadband ultrasound attenuation (BUA),[6] which are positively correlated with BMD measurements from DXA.[7] SOS and BUA may also be mathematically combined to calculate

Received in original form August 24, 2018; revised form January 27, 2019; accepted February 12, 2019. Accepted manuscript online March 19, 2019.
Address correspondence to: Wei-Chiao Chang, School of Pharmacy, Taipei Medical University, No. 250, Wuxing St., Xinyi Dist., Taipei City 11031, Taiwan. E-mail: wcc@tmu.edu.tw (WCC)
Additional Supporting Information may be found in the online version of this article.
Journal of Bone and Mineral Research, Vol. 34, No. 7, July 2019, pp 1275–1283.
DOI: 10.1002/jbmr.3703

the stiffness index (SI), according to the following formula: $SI = (0.67 \times BUA) + (0.28 \times SOS) - 420$.[8] Previous studies indicate that SI can provide a more accurate fracture risk prediction than BUA or SOS alone.[9,10]

Twin- and family-based studies have demonstrated that BMD, as measured by QUS, is highly heritable (82%).[11] This high heritability indicates the importance of identifying specific genetic influences on heel-bone properties.[12] Previous genome-wide association (GWA) meta-analyses have revealed several loci that are associated with heel-bone properties.[13,14] However, these studies focused on populations that were mainly comprised of individuals with European ancestry. Recently, one study identified a novel SOS-associated locus in the Korean population[15]; however, a comprehensive survey of SI-associated variants in East Asian populations is still lacking. Therefore, we decided to assess the genetic influences on heel-bone SI measurements by conducting a genomewide association study (GWAS) to identify candidate risk variants in a Taiwanese population. After identifying candidates in the discovery population, we compared our findings with data from a UK Biobank (UKBB) GWAS, and then performed a replication study using a Korean cohort. Finally, we leveraged the risk variants to identify and prioritize additional candidate SI-associated genes using data from publicly accessible databases.

Materials and Methods

Study population

The Taiwan Biobank is a prospective cohort with individual genotype data and detailed clinical information.[16] The Taiwan Biobank included 109,411 individuals without cancer from the community and hospitals. Among them, 20,117 individuals contained GWAS data. The DNA samples of the Taiwan Biobank were stored at −80°C. Genetic data were obtained by using an Affymetrix array (Axiom genomewide TWB plate; Affymetrix, Santa Clara, CA, USA). A GeneTitan Multi-Channel instrument (Thermo Fisher Scientific, Waltham, MA, USA) was used to conduct automated genotyping. Axiom Analysis Suite version 3.1 (Thermo Fisher Scientific, Waltham, MA, USA) was applied to generate genotyping output data by following the Axiom best-practice workflow. When the SNP did not fit the quality control criteria of the Axiom data analysis, the genotype of the SNP was recorded as a missing value. The subjects who were older than 50 years of age with complete age, sex, BMI, and GWAS information were included. Following GWAPower Detection V1,[17] we decided to assess 8000 individuals at our study design stage. When the total SNPs number equaled 6.5 million (TWB chip), the effect size of the candidate SNP was 0.1,[18] and the linkage disequilibrium (LD) between the causal SNP and the test SNP was estimated as $r^2 = 0.8$; hence, this study needed 8000 subjects to achieve >80% statistical power. These 8000 individuals were randomly selected by the staff of the Taiwan Biobank from all qualified samples ($n = 8848$). The SI of the individuals was measured using a QUS system and calculated by a GE Lunar Achilles InSight bone densitometer (GE Healthcare, Waukesha, WI, USA) automatically. The repeat measurement precision (% coefficient of variation [CV]) of GE Achilles InSight for SI is 2.4%.[19] All Taiwan Biobank participants provided informed consent. This study was conducted in accordance with guidelines approved by the institutional review boards (IRB) of Taipei Medical University (IRB no. 201210039) and the Taiwan Biobank, Academia Sinica (TWBR10505-05 and TWBR10602-02).

Data quality control and SNP imputation

Quality control procedures were conducted using PLINK,[20] which removed SNPs with a minor allelic frequency (MAF) <5%, and a Hardy–Weinberg equilibrium (HWE) p value $< 1 \times 10^{-6}$. Individuals with >5% missing genotypes and a genotype heterozygosity rate out of mean ± 3 SD ($n = 88$), as well as second-degree relatives based on an identity-by-descent threshold larger than 0.1875 ($n = 170$) were excluded. We eliminated 258 individuals because of a quality-control processing issue of the genotyping data. To increase the genome coverage of assessed variants, we imputed untyped SNPs via the Michigan imputation server using the 1000 Genomes Project phase 3 reference panel (https://imputationserver.sph.umich.edu/.) (East Asian ancestry [EAS], with genome coordinates of build GRCh37). After imputation, we filtered variants with an MAF of <5%, an R^2 of <0.3, and an imputation quality score of <0.9.[21]

Genomewide association analysis

The GWAS was conducted using SNPTEST (https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html). The statistical approach of SNPTEST is based on linear regression. The original distribution of SI was skewed; therefore, we normalized the SI to a zero mean and unit variance before analysis to fit it to the underlying statistic model of SNPTEST.[22] Genotype dosages were used to conduct an analysis under the additive genetic model. A principal component analysis (PCA) was conducted with samples from HapMap to inspect the substructure of the population. Four principal components (PC1 to PC4), age, sex, and BMI were included as covariates to regress out their effects in the linear regression model. The R package, qqman, was used to visualize the results of associations as Manhattan and quantile–quantile plots.

Genomic heritability

To estimate the phenotypic variance explained by autosomal variations, we used a restricted maximum likelihood analysis provided by a Genome-wide Complex Trait Analysis (GCTA).[23] In addition, we used an LD score regression to estimate the phenotypic variance proportion explained by the GWAS summary statistics.

Linkage disequilibrium-based clumping

To evaluate LD independent signals in each locus, an LD-based clumping procedure from PLINK was applied to the GWAS results. This approach is based on a greedy algorithm to select the most significant variants within 250 kb, an r^2 of ≥0.8, and a p value of $< 1 \times 10^{-5}$, thereby generating a list of LD-independent SNPs.

Conditional analysis

To identify the presence of independent association signals in each locus, we performed a conditional analysis within a 1 Mb genomic region of each of the LD-independent SNPs. We applied GCTA to this analysis by conditioning on the most significant variant signal. Imputed data of all individuals were used as an LD reference panel. Variants which in highly LD ($r^2 > 0.9$) with the lead SNPs were excluded from the analysis. We applied LocusZoom[24] to generate regional plots.

Replication in a Korean population

The KoGES Yangpyeong is a community cohort in the Korean Multi-Rural Communities Cohort Study (MRCohort) as a part of the Korean Genome Epidemiology Study (KoGES). The MRCohort was initiated to identify risk factors for cardiovascular diseases. Yangpyeong is located 45 km east of Seoul, the capital of South Korea; most of the participants were farmers or housewives. Details of the study design and procedures are described in a previous report.[25] From August 2005 to August 2011, 3564 individuals, age >40 years, were recruited in the KoGES Yangpyeong cohort. In the current study, participants who reported a physician-diagnosed cancer ($n = 76$) and who did not have information regarding their BMI ($n = 6$), SOS/BUA ($n = 67$), and genotype ($n = 420$) were excluded. In addition, we excluded participants with the lowest and highest 0.5% of SOS/BUA to minimize potential impact of outliers ($n = 40$). Ultimately, 2955 participants were included in the data analysis.[15] We measured SI at the calcaneal bone using the Sahara clinical sonometer (Hologic, Bedford, MA, USA), whose repeat measurement precision (% CV) is 4.55 for BUA and 0.52 for SOS, respectively.[26] The genotype data for the KoGES Yangpyeong replication set were generated by the Korean Chip (K-CHIP), which was designed by the Center for Genome Science, Korea National Institute of Health (KNIH) based on the UKBB Axiom Array. IMPUTE2[27] was used to conduct imputation using phase 1 of the 1000 Genomes Project as a reference panel. Genotype data quality-control procedures were conducted using PLINK. The exclusion criteria for SNPs included a p value of HWE $<1 \times 10^{-6}$, info score <0.8, and MAF < 0.01. Individuals with more than 5% missing genotypes, and pairwise identity-by-state values >.905 were excluded. To analyze the GWAS of SI from the KoGES Yangpyeong GWAS cohort, a multiple linear regression model was conducted; age, sex, and BMI were included as covariates. We further conducted a meta-analysis of summary statistics from GWAS significant ($p < 5 \times 10^{-8}$) SNPs of the Taiwan Biobank and KoGES Yangpyeong GWAS cohorts. METAL (https://genome.sph.umich.edu/wiki/METAL_Documentation) was adopted to perform the meta-analysis under fixed effect model.

Functional annotation

The rs (reference SNP cluster ID) number of each variant was annotated using two Bioconductor packages: GenomicRanges and SNPlocs.Hsapiens.dbSNP144.GRCh37. We annotated the locations and corresponding genes of SNPs by ANNOVAR[28] based on the RefSeq hg19 reference genome according to the National Center for Biotechnology Information. After annotating by ANNOVAR, the output generated the genes that the queried variants were located within or nearby. These genes were applied to further analysis directly. To evaluate relationships between SNPs and gene-expression profiles, we queried tissue-specific *cis*-expression quantitative trait locus (*cis*-eQTL) evidence from Haploreg.[29] When the SNP contained any *cis*-eQTL signals, we considered it as a point of our scoring system. We did not limit the signal in a specific window size of the SNP. Therefore, the reported eQTL signals may stand for genes different from our candidate genes. Because there are no bone tissue data in the publicly available eQTL database, we included all currently available data.

The ChromHMM (chromatin hidden Markov model) integrates multiple chromatin datasets of a variety of histone modifications from ENCODE and ROADMAP, including H3K4me1, H3K4me3, H3K27me3, H3K9me3, H3K36me3, and H3K27ac.[30] This model is based on a multivariate HMM, which can classify a genome into different functional regions. We downloaded the ChrommHMM-25-model profiles of 33 tissues and cells from the ENCODE and ROADMAP databases (Supplementary Table 1). By matching the position in each tissue, we annotated 25 ChromHMM states (Supplementary Table 2) to candidate SNPs. Each SNP had 33 corresponding chromatin states. The minimum state in all tissue and cell types was selected as the state of the SNP because the smaller state represents higher DNA accessibility (ie, one, the active transcription starting site has the highest DNA accessibility). To understand the chromatin states within candidate genes, we assigned the state of each SNP to within or nearby genes. Because one gene might contain several different chromatin states from multiple SNPs mapped to the gene, we summed the proportion of each state.

Gene ontology analysis

The corresponding genes of SI-associated variants were characterized using the topGO package. We used the "weight01" method, which is a mixture of the "elim" and "weighted" algorithms as described by Alexa and colleagues[31] and produced more-conservative results by taking the gene ontology (GO) hierarchy into account. We used Fisher's exact test to identify the enrichment of our gene list in biological processes. *P* values were adjusted for multiple comparisons using the Benjamini–Hochberg method.

Knockout mice phenotype enrichment analysis

We downloaded KO mice phenotype data from the Mouse Genome Informatics (MGI) database[32] on June 15, 2017, which comprised MGI phenotypes annotated to genes in a human orthologue. We then assessed the enrichment of SI-associated genes using a gene set overrepresentation analysis (ORA). To perform ORA, we used a one-tailed Fisher's exact test to assess the overrepresentation of genes in specific gene sets. We adopted a hypergeometric test to identify enriched MGI phenotypes, and further adjusted for multiple comparisons to reduce the likelihood of false-positives.

Empirical evidence regarding previous GWASs

To seek supportive evidence of identified SI-associated genes, we compared them with previously reported GWAS results using the gwascat package, which allowed us to explore the data from National Human Genome Research Institute–European Bioinformatics Institute (NHGRI–EBI) GWAS catalog. We mapped the SNPs to the corresponding genes provided by the GWAS catalog and generated a gene list for each phenotype. We then used these gene lists to conduct an ORA and analyze the enrichment of SI-associated genes in all GWAS phenotypes. Notably, we performed this enrichment analysis in a gene-based manner under the consideration that previously reported variants were not exactly the same as the identified SNPs in this study, but their information may be captured by allocating them onto the same gene when they were located in nearby loci to the identified SI-associated SNPs.

Protein–protein interaction analysis

Protein–protein interaction (PPI) evidence from in vivo, in vitro, and yeast two-hybrid experiments were queried from the Human Protein Reference Database (HPRD) Release 9.[33] We

linked genes with directed PPIs with SI-related genes, further generated SI-related genes with PPI annotations, and finally visualized the network using a Fruchterman–Reingold algorithm implemented in the igraph package.

Gene prioritization

We scored each gene using the following criteria: (1) contained nonsynonymous variant(s) associated with the SI; (2) contained eQTL(s); (3) belonged to an enriched GO biological process; (4) belonged to an enriched GWAS catalog phenotype; (5) had PPI annotations; and (6) contained SNP(s) that had potential within DNA-accessible region supported by the ChromHMM 25 model. Each gene that fulfilled one of the criteria was assigned a score of 1, resulting in a final score ranging from 0 to 6 (summation). We defined genes with a score ≥ 3 as biological SI-associated genes.

Results

GWAS revealed 78 genomewide significant SNPs associated with SI

There were 8000 individuals included in the association analysis. After imputation and quality control, approximately 6.1 million common variants were analyzed in 7742 individuals (mean age 58.5 ± 6.2 years). Proportions of the two genders were similar (Supplementary Table 3), and PCA revealed that there was no detectable substructure in the population (Supplementary Fig. 1). A quantile–quantile plot showed that there was no severe genomic inflation of the association results ($\lambda_{GC} = 1.07$; Supplementary Fig. 2). A Manhattan plot of the SI GWAS is shown in Fig. 1. There were 78 SNPs that reached GWA significance ($p < 5 \times 10^{-8}$; Supplementary Table 4) as well as 620 SNPs that achieved the suggestive significance threshold ($p < 1 \times 10^{-5}$). Among the suggestive significant SNPs, 10 were previously reported in the NHGRI–EBI GWAS catalog, including those annotated for bone-related and obesity-related traits (Supplementary Table 5). To remove redundant SNPs with high LD ($r^2 \geq 0.8$) within each locus, we applied an LD-based clumping method to the SNPs. After LD clumping, seven SNPs reached genomewide significance and 73 SNPs achieved suggestive significance (Supplementary Table 6). According to a genomewide complex trait analysis, the combination of all GWAS variants accounted for $23.1 \pm 4\%$ of SI heritability. This proportion was comparable to that calculated by LD score regression ($25.2 \pm 7\%$).

Independent variant loci revealed by conditional analysis

From the association analysis, we identified three cytogenetic regions (7q31.31, 17p13.3, and 11q14.2) that were significantly associated with SI (Fig. 1). We then conducted a conditional analysis to determine the number of independent signals in these regions. After conditioning on rs2536195 in the 7q31.31 locus, no SNPs reached genomewide significance (Supplementary Fig. 3). Therefore, we considered rs2536195 (*CPED/WNT16*) to be the most prominent signal within this locus with a p value of 1.41×10^{-11} (Table 1). Similarly, in the other two significant loci, we found no additional independent signals after conditioning on the top SNPs, rs1231207 ($p = 5.13 \times 10^{-9}$, *SMG6*, 17p13.3) and rs4944661 ($p = 3.92 \times 10^{-8}$, *LOC100506363/TMEM135*, 11q14.2).

Different ethnic populations share common risk variants

To test whether risk loci are shared across different populations, we compared the effect sizes and values of significant SNPs from previous GWAS reports for heel-bone properties[13–15] with the SNPs identified in our analysis of a Taiwanese population (Supplementary Table 7). The findings in the Taiwanese population were similar to those from previous GWAS reports (Supplementary Fig. 4). Recently, Kemp and colleagues conducted a study of 142,487 Caucasian individuals from the UKBB and identified 153 loci associated with heel BMD.[34] We investigated the correlation between the effect size of 620 suggestive significant SNPs of SI and estimated BMD (eBMD). After applying Pearson's correlation test, the correlation coefficient was 0.8 ($p < 2.2 \times 10^{-16}$). All of our genomewide significant SNPs were successfully confirmed as significant signals in this UKBB study (Supplementary Table 4). When we queried our LD-independent significant SNPs in the UKBB summary statistics (UKBB eBMD GWAS), 72 SNPs were found to be common between the studies. Notably, the effect directions of each SNP were also similar between the two studies (Supplementary Fig. 5). Thirty SNPs were genomewide

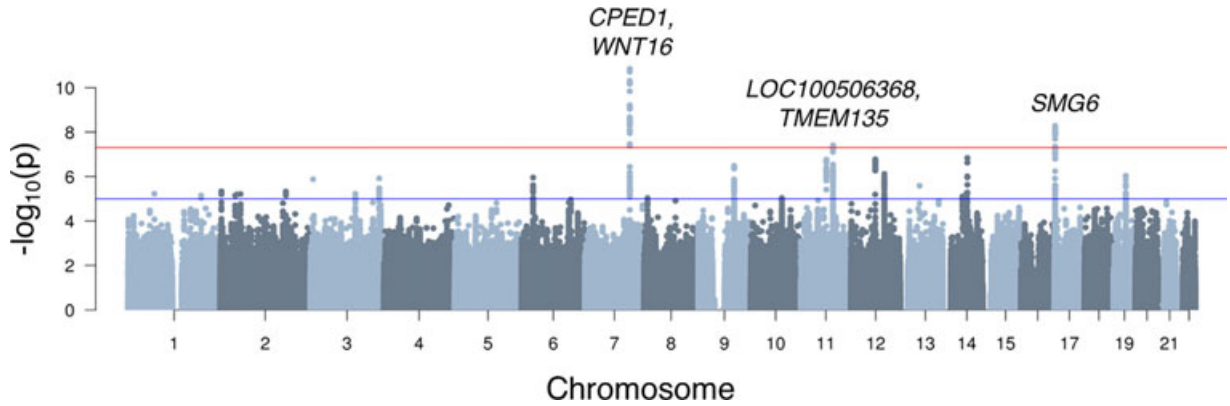

Fig. 1. Genome-wide association of stiffness index in a Taiwanese population. Red and blue lines correspond to $p = 5 \times 10^{-8}$ and $p = 1 \times 10^{-5}$, respectively.

Table 1. Three Conditional Independent Genomewide Significant SNPs of the Stiffness Index

SNP	Cytoband	Chr	Pos	A1	A2	Taiwan Biobank (SI)				UK Biobank (eBMD)				Position in gene	Gene
						Freq	Beta	SE	*p* value	Freq	Beta	SE	*p* value		
rs2536195	7q31.31	7	120959155	A	G	0.544	−0.108	0.016	1.41×10^{-11}	0.603	−0.168	0.003	4.94×10^{-324}	Intergenic	*CPED1, WNT16*
rs1231207	17p13.3	17	2125450	A	C	0.565	−0.089	0.015	5.13×10^{-09}	0.275	−0.052	0.004	9.10×10^{-45}	Intronic	*SMG6*
rs4944661	11q14.2	11	86738825	T	C	0.625	−0.087	0.016	3.88×10^{-08}	0.345	−0.032	0.004	2.20×10^{-24}	Intergenic	*LOC100506368, TMEM135*

SI = stiffness index; eBMD = estimated bone mineral density; Chr = chromosome; Pos = nucleotide position on the chromosome; A1 = allele 1; A2 = allele 2; Freq = frequency of allele 1; Beta = linear regression association coefficient.

significant while five SNPs showed suggestive significance (Supplementary Table 6). All three conditional independent variants passed the genomewide significance threshold (6.6×10^{-9}) in the original study (Table 1).[34]

Replication of 78 SNP associations in a Korean population

We further tested the 78 genomewide significant SNPs in a Korean replication cohort. The baseline characteristics of the Korean population are shown in Supplementary Table 8. Because of quality controls, 59 SNPs remained subsequently (Supplementary Table 4). Among them, 48 SNP reached GWAS significance (5×10^{-8}) after meta-analysis. Because the quality-control process excluded two of the three conditional independent SNPs, only one independent SNP (rs1231207, 17p13.3) was replicated ($p_{meta} = 1.88 \times 10^{-9}$). Nevertheless, signals from the other two cytobands (7q31.31 and 11q14.2) were successfully replicated as well.

SI-associated SNPs were mostly located in functional regulatory regions

Beginning with the 73 identified SNPs that showed suggestive significance ($p < 1 \times 10^{-5}$), we first excluded those SNPs located on the HLA region or without ANNOVAR annotation, leaving 54 GWAS signals. We extended the variant list by including SNPs with $r^2 \geq 0.6$ and a p value <0.05. As a result, we obtained 1656 SNPs mapped to 59 genes. Most of the SNPs on the variant list were found to be located in intronic regions ($n = 1149$, 69.4%) or intergenic regions ($n = 399$, 24.1%; Fig. 2*A*); 12 were found to be located in an exon with eight being nonsynonymous. Because the vast majority of SI-associated SNPs were located within intronic or intergenic regions, we examined the epigenetic profiles of these loci. After annotating the ChromHMM state of each variant, we found that 649 SNPs were located in at least one weakly accessible DNA region per tissue type (Fig. 2*B*). In addition, 39 genes (66.1%) contained more than one weakly transcribed variant, and 18 genes (30.5%) were at least weakly transcribed in 50% of tissues (Fig. 2*C*).

Biological roles of the SI-associated loci

There were 59 candidate genes subjected to ORAs of GWAS catalog phenotypes, GO terms, and KO mouse phenotypes. Our results showed significant enrichment of three highly relevant GWAS catalog phenotypes, including bone QUS (false discovery rate [FDR] $= 1.59 \times 10^{-4}$), SOS (FDR $= 3.72 \times 10^{-4}$), and bone density (FDR = 0.05). These phenotypes passed stringent FDR controls (Supplementary Table 9). Specific genes that were enriched in these three phenotypes included *GPATCH1*, *SPTBN1*, and *WNT16*. Furthermore, according to our GO term ORA, these candidate genes were found to be enriched in 10 biological processes (Supplementary Table 10); however, none of the GO terms remained significant after FDR adjustment. In addition, no significantly enriched KO mouse phenotypes were observed (Supplementary Table 11). We also identified 20 SI-related genes that had PPI annotations, according to the HPRD (Supplementary Fig. 6).

Gene prioritization identified two novel SI-related genes

To prioritize candidate genes, we incorporated information detailing nonsynonymous variants, *cis*-eQTL, minimum

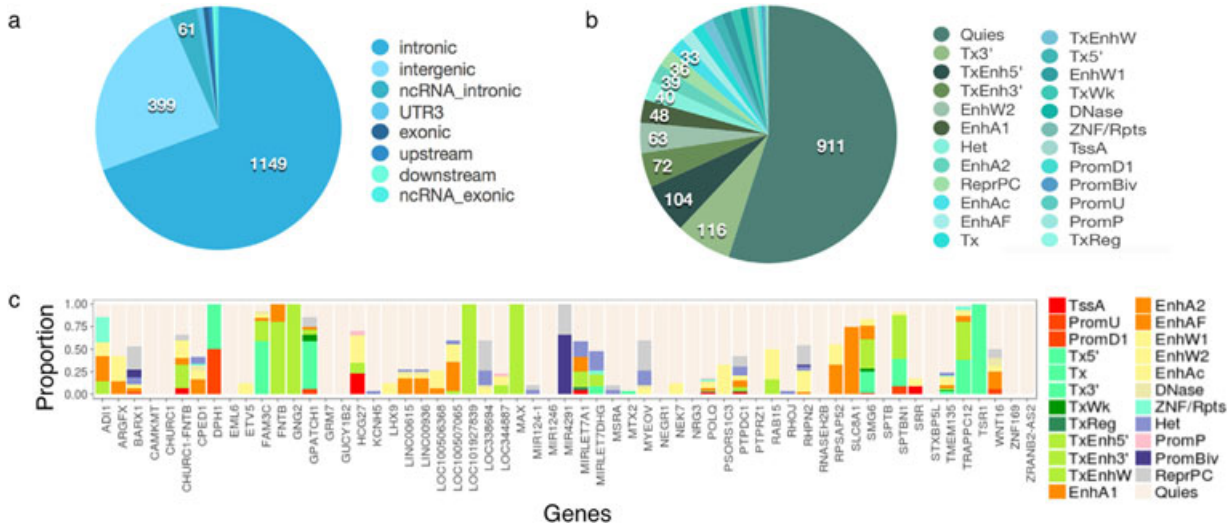

Fig. 2. Functional annotations of 1656 single-nucleotide polymorphisms and 59 genes associated with the stiffness index in the Taiwanese Biobank genomewide association study. (*A*) Functional annotations from ANNOVAR. (*B*) The minimum ChromHMM state across 33 tissue types. The full names of each annotation are listed in Supplementary Table 2. (*C*) The proportion of the minimum ChromHMM state of each gene.

ChromHMM state, GWAS catalog enrichment, GO biological processes, and PPI data (Fig. 3 and Supplementary Table 12). Among the 59 genes, 5 (8.5%) had nonsynonymous variants, 24 (40.7%) contained eQTL variants in multiple tissues, 9 (15.3%) were involved in enriched biological processes, and 7 (11.9%) were significantly enriched in GWAS catalog phenotypes. Based on these parameters, we scored and prioritized the genes. There were 14 genes (23.7%) that were scored for at least three criteria (Supplementary Fig. 7). After prioritization, we considered these 14 genes to be potential biologically relevant SI-related candidate genes (Fig. 3). We further extracted SNPs within ± 100 kb of the candidate gene region, and compared our results with UKBB eBMD GWAS. Two genes, *RAB15* (Fig. 4*A*) and *FNTB* (Fig. 4*B*), did not contain any

Chr	Marker SNP	Gene	Priority score	Annotation		Biology		Pathways	
				Nonsynonymous	ChromHMM	cis-eQTL	GWAS catalog	GO-BP	PPI
17	rs1231207	*SMG6*	6	■	■	■	■	■	■
2	rs2971879	*SPTBN1*	5		■	■	■	■	■
17	rs59904243	*TSR1*	4	■	■	■		■	
19	rs577799800	*RHPN2*	4	■	■	■			■
7	rs2707466	*WNT16*	4	■	■	■	■		
11	rs7395163	*TMEM135*	3		■	■	■		
14	rs11540872	*RAB15*	3		■	■			■
14	rs4902352	*FNTB*	3		■			■	■
17	rs11078780	*DPH1*	3		■	■			■
19	rs553114464	*GPATCH1*	3		■	■	■		
2	rs2041759	*SLC8A1*	3		■	■			■
3	rs500103	*POLQ*	3	■	■	■			
7	rs798901	*CPED1*	3		■	■	■		
7	rs2952565	*FAM3C*	3		■	■	■		

Fig. 3. Candidate genes for the stiffness index identified by in silico analyses. Fourteen genes are considered good candidates after prioritization based on statistical and biological evidence represented as colored cells. We classified our scoring criteria into three main categories: annotation, biology, and pathway, respectively. The color of the cells is according to the category. GO = gene ontology; BP = biological process; PPI = protein–protein interaction.

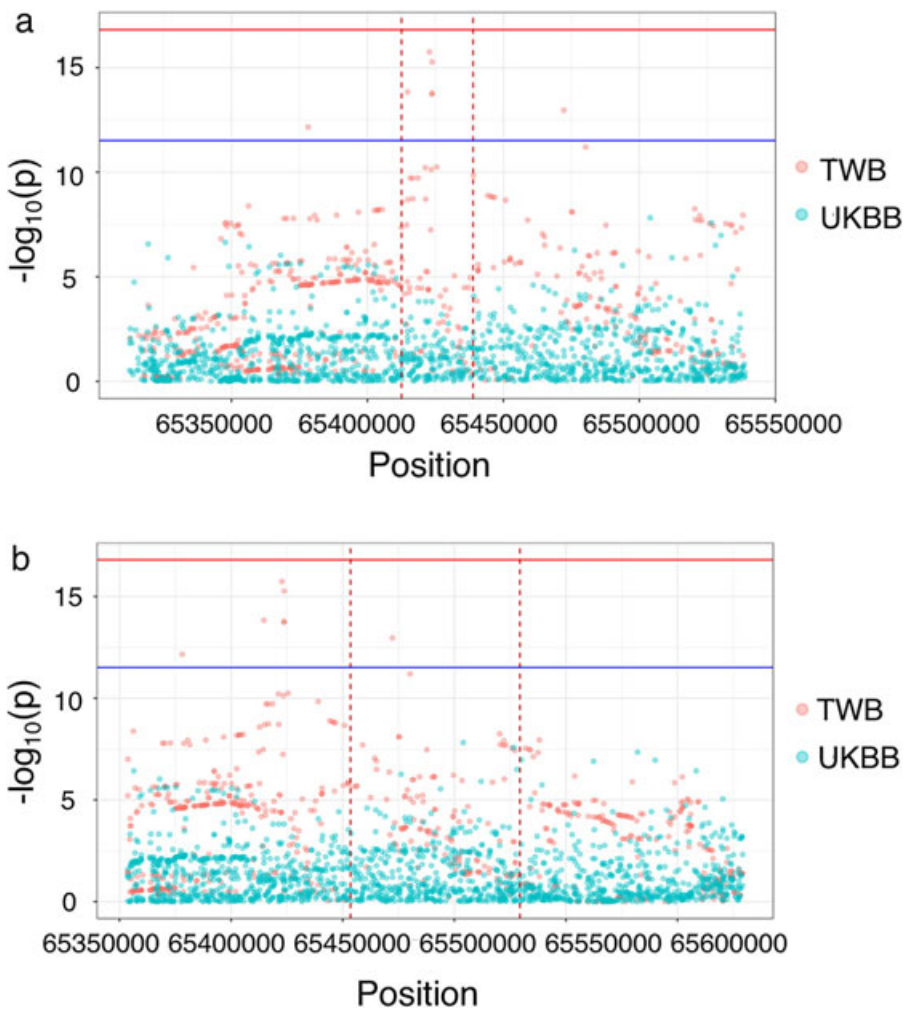

Fig. 4. Novel stiffness index-related genes identified by comparison of the Taiwan Biobank and the UK Biobank genomewide association studies. The *X* axis represents the chromosome position, and the *Y* axis is the negative log10 *p* value. Orange plots are *p* values of variants from the Taiwan Biobank, and green plots from the UK Biobank. Red dashed lines represent the location of the genes. (*A*) *RAB15*. (*B*) *FNTB*.

significant signals in the UKBB data. Therefore, we consider these two genes to be novel SI-related genes.

Discussion

In this study, we used imputed genotypes of 7742 individuals to identify SI-associated candidate loci in a Taiwanese population. Through this analysis, we discovered that rs2536195 (7q31.31), rs1231207 (17p13.3), and rs4944661 (11q14.2) were SI-associated loci with conditionally independent genomewide significance. These three SNPs were significant in the UKBB eBMD GWAS as well. By integrating annotations from the ChromHMM, we added further information regarding the DNA accessibility at each candidate locus. Moreover, several ORAs, based on different databases, were also applied to the candidate genes. After gene prioritization, we identified 14 potential biologically relevant SI-related candidate genes.

According to LD-based clumping, we uncovered five significant genomewide LD-independent variants within the 7q31.31 region. Data were conditioned on rs2536195 (the most significant signal), revealing this SNP as the independent signal in the region. Interestingly, 7q31.31 was previously reported to be associated with heel-bone properties by multiple groups,[13,14] and Kemp and colleagues reported rs2536195 as an independent significant SNP within the locus.[34] Hence, we consider rs2536195 to be a crucial variant for heel-bone BMD in both East Asian and Caucasian populations. The other two loci, 17p13.3 and 11q14.2, also significantly associated with eBMD in the GWAS from the UKBB; however, the independent signals in the previous report were different from what we discovered in the current study. We attribute this to ethnicity-specific factors caused by the differences in MAFs and LD structure. Indeed, the frequency of the A allele at rs1231207 is 0.57 in the current study, but it is 0.26 in the UKBB; the frequency in Asians is 0.62 and in Europeans is 0.26 based on the data from 1000 Genomes Project (Supplementary Table 13). Consistent with this, we observed similar results for rs4944661; the frequency of the T allele is 0.63 in the current study and 0.35 in the UKBB.

In general, the directions of the SNP effects from our Taiwanese population were comparable to genomewide variants reported in previous studies of other ethnicities

(Supplementary Figs. 4, 5). Furthermore, we attempted to replicate our significant SNPs in a Korean population. The different measurement of heel SI may be one of the potential explanations for the difference in the beta value between the two cohorts. For example, the beta of rs1231207 was −0.089 in the results of the Taiwan Biobank, whereas it was −1.343 in the Korean population. Despite having only one independent significant SNP replicated, the other two loci were found to contain alternative significant signals. This may because of the different LD structure within the two populations and/or the limited power of the replication cohort ($n = 2{,}955$). Because two of the three SNPs were not replicated successfully, further studies in larger East Asian populations are needed.

By comparing our SNPs with suggestive significance to catalogued GWAS reports, we found that rs73039434, which is found within an intronic region of *RHPN2*, also showed a significant GWA with childhood obesity ($p = 5.96 \times 10^{-6}$).[35] Thus, this locus may exert pleiotropic effects, impacting both heel bone and obesity-related traits.

Among the candidate genes we identified, *WNT16* (7q31.31),[13,14,18,36,37] *SPTBN1* (2p16.2),[13,14,38–40] *GPATCH1* (19q13.11),[13,14,18] *SLC8A1* (2p22.1),[34,41] and *CPED1* (7q31.31)[13,14,18,42–44] have all been reported to be significantly associated with both BMD and US measurements.[13,14] In addition, *SMG6* (17p13.3)[18] and *FAM3C* (7q31.31)[43] have also been identified as BMD-associated loci previously. Therefore, these seven genes may potentially be involved in regulating BMD without site specificity. Aside from BMD, many of our candidate genes are associated with other phenotypes as well. For example, *DPH1* (17p13.3) is a susceptibility gene for short stature and craniofacial anomalies.[45] *TSR1* (17p13.3) and *POLQ* (3q13.33) have been reported to be candidates for influencing aortic root size[46] and multiple myeloma,[47] respectively. Importantly, *FTNB* and *RAB15*, were identified as novel heel bone property-related genes in our study. *FNTB* (14q23.3) was associated with a blood-related trait,[48] whereas *RAB15* (14q23.3) has not been reported as a disease-associated gene so far.

Annotation of SNPs with the ChromHMM showed that about half of the SI-associated intronic SNPs may participate in gene regulation. This finding emphasized the importance of incorporating epigenomic data into the analysis to estimate the influence of intronic SNPs. In bone tissue, however, epigenetics data are currently limited; therefore, our annotations may not be comprehensive. Additional epigenomic or functional studies of bone cells may uncover further influences of our candidate SNPs.

Our study had some limitations. First, we focused on analyzing common variants (MAF $\geq$ 5%). Because most effect sizes of common variants on phenotypes are modest, increasing the number of study participants would be crucially important to identify rare variants and structural variations of osteoporosis.[49] This limitation may possibly explain our low heritability estimation (23%) compared to twin and family studies (82%). Further studies on large populations should be conducted to uncover the impact of rare variants on SI. Second, we only used SI as the phenotype. Data on BUA and SOS were unavailable. Third, the present study only considered age, sex, and BMI as covariates, whereas a number of other factors may be associated with SI. For example, lifestyle factors (eg, smoking, exercise, alcohol use) and certain medications could influence bone density.[50] Because these risk factors may also have interactions with genes,[51] our current study did not estimate the gene–environment interaction effect. Further studies should consider the impact of environmental exposures.

In conclusion, our study identified 7q31.31, 17p13.3, and 11q14.2 as SI-related loci in ethnically diverse populations. By using several biological annotations for each candidate gene, we uncovered 14 candidate SI-associated genes. These findings enhance our understanding of bone genetics and provide novel insight into the genetic influences on bone-related properties.

Disclosures

Hsing-Fang Lu, Kuo-Sheng Hung, Henry Sung-Ching Wong, Hou-Wei Chu, Bo Youl Choi, Mi Kyung Kim, Jihye Kim, Yu-Ting Tai, Shiro Ikegawa, Er-Chieh Cho, and Wei-Chiao Chang declare that they have no conflict of interest.

Acknowledgments

This work was supported by grants from the Ministry of Science and Technology, Taiwan (MOST 105-2628-B-038-001-MY4; MOST 106-2923-B-038-002), the Higher Education Sprout Project by the Ministry of Education in Taiwan (MOE DP2-107-21121-A-05), the Japan Osteoporosis Foundation, and the research program funded by the Korea Centers for Disease Control and Prevention (2017N-E71001-00).

References

1. Kanis JA, Melton LJ, 3rd, Christiansen C, Johnston CC, Khaltaev N. The diagnosis of osteoporosis. J Bone Miner Res. 1994;9(8):1137–41.
2. Osteoporosis prevention, diagnosis, and therapy. JAMA. 2001; 285(6):785–95.
3. Hans D, Fuerst T, Lang T, et al. How can we measure bone quality? Baillieres Clin Rheumatol. 1997;11(3):495–515.
4. Moayyeri A, Kaptoge S, Dalzell N, et al. Is QUS or DXA better for predicting the 10-year absolute risk of fracture? J Bone Miner Res. 2009;24(7):1319–25.
5. Moayyeri A, Adams JE, Adler RA, et al. Quantitative ultrasound of the heel and fracture risk assessment: an updated meta-analysis. Osteoporos Int. 2012;23(1):143–53.
6. Gonnelli S, Cepollaro C, Agnusdei D, Palmieri R, Rossi S, Gennari C. Diagnostic value of ultrasound analysis and bone densitometry as predictors of vertebral deformity in postmenopausal women. Osteoporos Int. 1995;5(6):413–8.
7. Trimpou P, Bosaeus I, Bengtsson BA, Landin-Wilhelmsen K. High correlation between quantitative ultrasound and DXA during 7 years of follow-up. Euro J Radiol. 2010;73(2):360–4.
8. Hadji P, Hars O, Schüler M, et al. Assessment by quantitative ultrasonometry of the effects of hormone replacement therapy on bone mass. Am J Obst Gyn. 2000;182(3):529–34.
9. Hadji P, Hars O, Wuster C, et al. Stiffness index identifies patients with osteoporotic fractures better than ultrasound velocity or attenuation alone. Maturitas. 1999;31(3):221–6.
10. Gonnelli S, Cepollaro C, Gennari L, et al. Quantitative ultrasound and dual-energy X-ray absorptiometry in the prediction of fragility fracture in men. Osteoporos Int. 2005;16(8):963–8.
11. Howard GM, Nguyen TV, Harris M, Kelly PJ, Eisman JA. Genetic and environmental contributions to the association between quantitative ultrasound and bone mineral density measurements: a twin study. J Bone Miner Res. 1998;13(8):1318–27.
12. Hendrickx G, Boudin E, Van Hul W. A look behind the scenes: the risk and pathogenesis of primary osteoporosis. Nat Rev Rheumatol. Review 2015;11(8):462–74.
13. Moayyeri A, Hsu Y-H, Karasik D, et al. Genetic determinants of heel bone properties: genome-wide association meta-analysis and replication in the GEFOS/GENOMOS consortium. Hum Mol Genet. 2014;23(11):3054–68.

14. Mullin BH, Hua Zhao J, Brown SJ, et al. Genome-wide association study meta-analysis for quantitative ultrasound parameters of bone identifies five novel loci for broadband ultrasound attenuation. Hum Mol Genet. 2017;26(14):2791–802.
15. Hwang J-Y, Kim YJ, Choi BY, Kim B-J, Han B-G. Meta analysis identifies a novel susceptibility locus associated with heel bone strength in the Korean population. Bone. 2016;84:47–51.
16. Fan CT, Lin JC, Lee CH. Taiwan Biobank: a project aiming to aid Taiwan's transition into a biomedical island. Pharmacogenomics. 2008;9(2):235–46.
17. Hong EP, Park JW. Sample size and statistical power calculation in genetic association studies. Genomics Inform. 2012;10(2):117–22.
18. Estrada K, Styrkarsdottir U, Evangelou E, et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. Nat Genet. 2012;44(5):491–501.
19. Greenspan SL, Bouxsein ML, Melton ME, et al. Precision and discriminatory ability of calcaneal bone assessment technologies. J Bone Miner Res. 1997;12(8):1303–13.
20. 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(3):559–75.
21. Lin P, Hartz SM, Zhang Z, et al. A new statistic to evaluate imputation reliability. PLoS One. 2010;5(3):e9697.
22. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet. 2007;39(7):906–13.
23. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88(1):76–82.
24. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010; 26(18):2336–7.
25. Kim Y, Han BG. Cohort profile: The Korean Genome and Epidemiology Study (KoGES) Consortium. Int J Epidemiol. 2017;46(2):e20.
26. Njeh CF, Hans D, Li J, et al. Comparison of six calcaneal quantitative ultrasound devices: precision and hip fracture discrimination. Osteoporos Int. 2000;11(12):1051–62.
27. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009;5(6):e1000529.
28. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
29. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012;40(D1):D930–4.
30. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518(7539):317–30.
31. Alexa A, Rahnenfuhrer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics. 2006;22(13):1600–7.
32. Blake JA, Eppig JT, Kadin JA, Richardson JE, Smith CL, Bult CJ. Mouse Genome Database (MGD)-2017: community knowledge resource for the laboratory mouse. Nucleic Acids Res. 2017;45(D1):D723–9.
33. Keshava Prasad TS, Goel R, Kandasamy K, et al. Human Protein Reference Database—2009 update. Nucleic Acids Res. 2009;37: D767–72.
34. Kemp JP, Morris JA, Medina-Gomez C, et al. Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. Nat Genet. 2017;49:1468.
35. Comuzzie AG, Cole SA, Laston SL, et al. Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. PLoS One. 2012;7(12):e51954.
36. Zheng HF, Tobias JH, Duncan E, et al. WNT16 influences bone mineral density, cortical bone thickness, bone strength, and osteoporotic fracture risk. PLoS Genet. 2012;8(7):e1002745.
37. Zhang L, Choi HJ, Estrada K, et al. Multistage genome-wide association meta-analyses identified two new loci for bone mineral density. Hum Mol Genet. 2014;23(7):1923–33.
38. Styrkarsdottir U, Halldorsson BV, Gretarsdottir S, et al. Multiple genetic loci for bone mineral density and fractures. N Engl J Med. 2008;358(22):2355–65.
39. Styrkarsdottir U, Halldorsson BV, Gretarsdottir S, et al. New sequence variants associated with bone mineral density. Nat Genet. 2009; 41(1):15–7.
40. 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(11):1199–206.
41. Medina-Gomez C, Kemp JP, Trajanoska K, et al. Life-course genome-wide association study meta-analysis of total body BMD and assessment of age-specific effects. Am J Hum Genet. 102(1):88–102.
42. Chesi A, Mitchell JA, Kalkwarf HJ, et al. A trans-ethnic genome-wide association study identifies gender-specific loci influencing pediatric aBMD and BMC at the distal radius. Hum Mol Genet. 2015; 24(17):5053–9.
43. Kemp JP, Medina-Gomez C, Estrada K, et al. Phenotypic dissection of bone mineral density reveals skeletal site specificity and facilitates the identification of novel loci in the genetic regulation of bone mass attainment. PLoS Genet. 2014;10(6):e1004423.
44. Chesi A, Mitchell JA, Kalkwarf HJ, et al. A genomewide association study identifies two sex-specific loci, at SPTB and IZUMO3, Influencing pediatric bone mineral density at multiple skeletal sites. J Bone Miner Res. 2017;32(6):1274–81.
45. Loucks CM, Parboosingh JS, Shaheen R, et al. Matching two independent cohorts validates DPH1 as a gene responsible for autosomal recessive intellectual disability with short stature, craniofacial, and ectodermal anomalies. Human Mutation. 2015;36(10):1015–9.
46. Vasan RS, Glazer NL, Felix JF, et al. Genetic variants associated with cardiac structure and function: a meta-analysis and replication of genome-wide association data. JAMA. 2009;302(2):168–78.
47. Weinhold N, Johnson DC, Chubb D, et al. The CCND1 c.870G>A polymorphism is a risk factor for t(11;14)(q13;q32) multiple myeloma. Nat Genet. 2013;45(5):522–5.
48. van der Harst P, Zhang W, Mateo Leach I, et al. Seventy-five genetic loci influencing the human red blood cell. Nature. 2012; 492(7429):369–75.
49. Park J-H, Gail MH, Weinberg CR, et al. Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants. Proc Natl Acad Sci U S A. 2011; 108(44):18026–31.
50. Sambrook P, Cooper C. Osteoporosis. Lancet. 2006;367(9527): 2010–8.
51. Hunter DJ. Gene-environment interactions in human diseases. Nat Rev Genet. 2005;6(4):287–98.