COMMENTARY

An example design of large-scale next-generation sequencing study for bone mineral density

Hou-Feng Zheng[1,2]

[1]Departments of Medicine and Human Genetics, McGill University, Montreal, Quebec, Canada. [2]Departments of Epidemiology and Biostatistics, Lady Davis Institute, Jewish General Hospital, Montreal, Quebec, Canada.

Commentary on: Styrkarsdottir U, Thorleifsson G, Sulem P, Gudbjartsson DF, Sigurdsson A, Jonasdottir A, Jonasdottir A, Oddsson A, Helgason A, Magnusson OT, Walters GB, Frigge ML, Helgadottir HT, Johannsdottir H, Bergsteinsdottir K, Ogmundsdottir MH, Center JR, Nguyen TV, Eisman JA, Christiansen C, Steingrimsson E, Jonasson JG, Tryggvadottir L, Eyjolfsson GI, Theodors A, Jonsson T, Ingvarsson T, Olafsson I, Rafnar T, Kong A, Sigurdsson G, Masson G, Thorsteinsdottir U, Stefansson K. Nonsense mutation in the *LGR4* gene is associated with several human diseases and other traits. Nature 2013;**497**:517–520.

Genome-wide association studies (GWAS) have successfully identified more than 60 common loci for bone mineral density (BMD) and osteoporosis,[1] including a recent large-scale meta-analysis of GWAS within the GEnetic Factors for Osteoporosis (GEFOS) Consortium that identified 32 novel BMD loci and 6 fracture loci.[2] Despite the large number of loci identified by GWAS, only 5.8% of the variance in femoral neck BMD has been explained by genome wide-significant single-nucleotide polymorphisms (SNPs) in the largest meta-analysis to date.[2] The fact that small effects from common variants are replicable across large studies, and that no locus contributes a substantial amount to trait variance, suggests that osteoporosis may either have an infinitesimal allelic architecture, wherein many alleles across the allele frequency spectrum have a small effect on risk, or that rare variants of low minor allele frequency (MAF) contribute to the phenotype.[3] Recently, next-generation sequencing technologies in conjunction with bioinformatics has enabled us to sequence and analyze whole-genome sequences from large human cohorts, making it possible to identify rare mutation with large effect for common traits, such as BMD.

A recent publication by Styrkarsdottir *et al.*[4] in *Nature* has begun to illustrate the effect of rare variants on BMD. In this study, 2230 Icelanders were whole-genome sequenced and the haplotypes of sequenced individuals were then imputed into an additional 95 085 Icelandic GWAS samples. To increase power, this study pooled BMD phenotypes of different sites (hip, spine or whole body) together to make a new binary phenotype, low BMD, consisting of samples with standardized BMDs below − 1 s.d. Those samples that had a measured BMD above − 1 s.d. or had not had their BMD measured were combined as a general control group. In total, the low BMD group included 4931 individuals and the general control group comprised 69 034 individuals. Genetic associations with low BMD were tested in 34.2 million sequence variants. Two loci were identified to be associated with low BMD at a genome-wide significance threshold ($P < 5 \times 10^{-8}$). The first locus is with a common

variant located at 13q14 (rs8001611) that has been reported previously,[5,6] and this SNP is located ∼160 kb from gene *TNFSF11* (*RANKL*). The second locus is a novel association, consisting of a group of correlated rare variants at 11p14, including a variant (hg18_chr11:27369242_A, odds ratio (OR) = 4.30, $P = 1.3 \times 10^{-10}$, MAF = 0.174%) that introduces a nonsense codon (c.376C > T) into exon 4 of the *LGR4* gene. Gene expression analysis of carriers and non-carriers of this variant in adipose tissue and published knockout mice data supported this association.

In this first whole-genome sequencing study for BMD, it was actually a GWAS design, which is to compare the frequency of single variant in cases and controls to test the association. Indeed, we could consider this as an extension of traditional GWAS, as both common and rare variants were investigated in the study. A very big sample size was achieved in this study because cases were pooled and a general control group was used. The continuous trait of BMD was converted to a binary trait—low BMD, so that BMD from different sites could be all included in this study (**Figure 1**). The use of a general control group was a coarse yet efficient approach for increasing power; this design also successfully applied in a systemic lupus erythematosus GWAS,[7] in which a series of samples of various diseases, such as psoriasis and vitiligo, were taken as controls. However, because the true status of general controls is unknown, a large sample size is required to dilute any misclassification.

Imputation methods infer untyped SNP genotypes based on linkage disequilibrium (LD) with typed SNPs. Therefore, an attractive study design would be to sequence a small set of samples and impute the rare variants into a larger data set.[8] It has been shown that rare variant imputation is more difficult than that for common variants; however, an increasing number of rare variants can achieve sufficient imputation quality given a large enough number of samples in the reference panel.[9] In the study by Styrkarsdottir *et al.*,[4] 2230 samples were whole
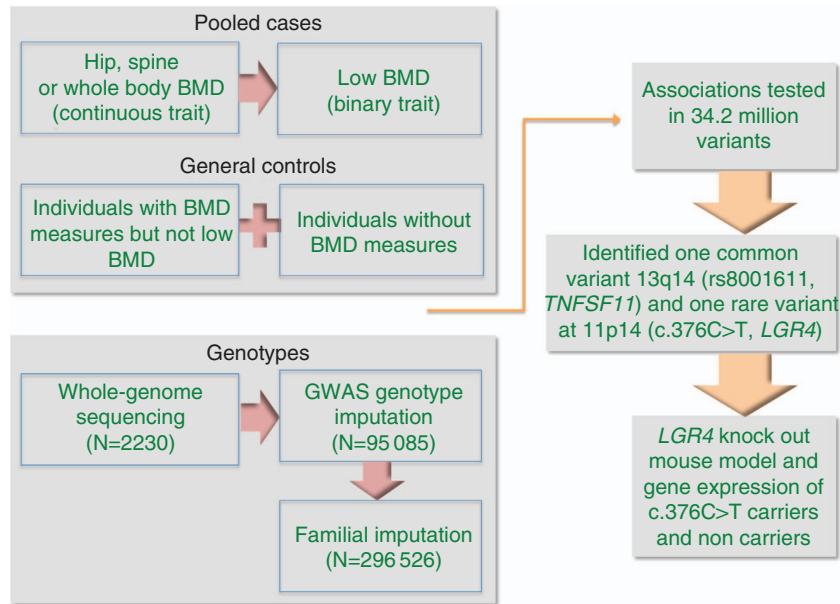
Figure 1 Schematic overview of study design for this large-scale next-generation sequencing study for bone mineral density.

genome sequenced, using these 4460 haplotypes as reference panel, most of the variants with MAF from frequency of singleton to 5% could be imputed,[9] therefore, 95 085 samples who were genome-wide genotyped were imputed and included in the analysis (**Figure 1**). Successful imputation in this manner can increase sample size dramatically, thus improving statistical power. In this study, the imputation of rare variant c.376C > T was validated and improved by direct genotyping, resulting in a slightly stronger association with low BMD.

It has been hypothesized that many GWAS signals may not reflect the causal polymorphism, but instead are synthetic associations arising from rare variants on the same haplotype background.[10] In this study, a common variant (rs10835187) at the same locus as c.376C > T at 11p14 was previously reported to be associated with BMD.[2] However, the effect of the common variant on low BMD was much weaker (OR = 1.06, P = 0.031), and conditional analysis showed that the common and rare variant represent two independent BMD association signals in the 11p14 region. Moreover, fine mapping analysis revealed that rs10835187 acted through gene *LIN7C* or *BDNF* rather than gene *LGR4*. Thus, the rare nonsense variant did not synthesize the common BMD GWAS signal in the 11p14 locus in this study. Styrkarsdottir *et al.*[4] also calculated the LD between these two variants with r^2 measurement ($r^2 < 0.0001$); however, D′ could be a better parameter to assess the LD between common and rare variants, because r^2 takes the MAF into account, two variants with very different MAF will give poor r^2-value, even they are in complete LD.

The variant c.376C > T was not identified in independent replication studies in Danish and Australian samples. However, functional evidence for its role in bone physiology has been shown in knockout mice, where $Lgr4^{-/-}$ mice have delayed osteoblast differentiation and mineralization during embryonic bone formation.[11] Moreover, Styrkarsdottir *et al.*[4] detected reduced levels of mutated *LGR4* messenger RNA isolated from white blood cells and adipose tissue of heterozygous c.376C > T carriers.

As we transition to whole-genome sequencing and rare variant identification for osteoporosis, we should regard the relative advantages of this approach compared to traditional GWAS. First, evidence from current GWAS does not preclude the important contribution of rare genetic variation to osteoporosis. In fact, large-scale sequencing study is an extension of GWAS, in which more variants with lower MAF are included. To study both common variant and rare variant may provide a more comprehensive genetic map for osteoporosis; therefore, one expectation of sequencing study is to find the real causal variants for osteoporosis, or at least to fine map the previous identified GWAS loci. Moreover, sequencing samples could serve as reference haplotypes to infer rare variants from genome-wide genotyping individuals with imputation technique that is described above. This strategy could not only re-use previous GWAS data, but more importantly, also increase the sample size without substantially increasing costs. Therefore, imputation of rare variants into individuals who have undergone genome-wide genotyping from the whole-genome sequenced haplotypes offers a cost-efficient strategy to achieve necessary sample sizes, resulting in increased statistical power.

As the merits of next-generation sequencing start to become evident, another more general question becomes pertinent to the osteoporosis research community: 'Will rare mutation identification for BMD be as successful as genome-wide association studies?' The answer to this question, at least for a short term, is that there will be a limited success for rare variant identification for multiple reasons. First, sample size will remain relatively small for study due to the high cost of next-generation sequencing, thus reducing statistical power to detect rare variants. Second, rare variants may be specific to certain ethnic population, making findings difficult to replicate. As was demonstrated in this study where the mutation c.376C > T was absent in Danish women and Australian samples, however, it should be noted that the sample size of the Danish and Australian populations (~4000 samples) is substantially smaller than the ~10 000 samples from the Icelandic population.

Finally, the analytic methods for rare variant, such as Sequence Kernel Association Test,[12] are developed to be robust for particular genetic models, whereas the true probalistic genetic model underlying complex trait is never known before the analysis.

This study represents an important milestone for the identification of genetic determinants for BMD. Moreover, other large-scale sequencing projects also hold great promise, such as UK10K project (http://www.uk10k.org/), that have whole-genome sequenced thousands of samples with BMD measurements. Further advancements, such as the reduction in sequencing costs, development of new analytical tools and organization of large-scale international collaborative meta-analysis, will position the osteoporosis research community to realize the full potential of mapping all rare and common variation to BMD and osteoporosis.

Conflict of Interest

The author declares no conflict of interest.

Acknowledgements

References

1. Richards JB, Zheng HF, Spector TD. Genetics of osteoporosis from genome-wide association studies: advances and challenges. *Nat Rev Genet* 2012;**13**:576–588.
2. Estrada K, Styrkarsdottir U, Evangelou E, Hsu YH, Duncan EL, Ntzani EE *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**:491–501.
3. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era–concepts and misconceptions. *Nat Rev Genet* 2008;**9**:255–266.
4. Styrkarsdottir U, Thorleifsson G, Sulem P, Gudbjartsson DF, Sigurdsson A, Jonasdottir A *et al.* Nonsense mutation in the *LGR4* gene is associated with several human diseases and other traits. *Nature* 2013;**497**:517–520.
5. Rivadeneira F, Styrkarsdottir U, Estrada K, Halldorsson BV, Hsu YH, Richards JB *et al.* Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. *Nat Genet* 2009;**41**:1199–1206.
6. Styrkarsdottir U, Halldorsson BV, Gretarsdottir S, Gudbjartsson DF, Walters GB, Ingvarsson T *et al.* Multiple genetic loci for bone mineral density and fractures. *N Engl J Med* 2008;**358**: 2355–2365.
7. Han JW, Zheng HF, Cui Y, Sun LD, Ye DQ, Hu Z *et al.* Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat Genet* 2009;**41**:1234–1237.
8. Zeggini E. Next-generation association studies for complex traits. *Nat Genet* 2011;**43**:287–288.
9. Zheng HF, Ladouceur M, Greenwood CM, Richards JB. Effect of genome-wide genotyping and reference panels on rare variants imputation. *J Genet Genomics* 2012;**39**:545–550.
10. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. *PLoS Biol* 2010;**8**:e1000294.
11. Luo J, Zhou W, Zhou X, Li D, Weng J, Yi Z *et al.* Regulation of bone formation and remodeling by G-protein-coupled receptor 48. *Development* 2009;**136**:2747–2756.
12. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* 2011;**89**:82–93.