Title: Increased burden of common risk alleles in children with a significant fracture history.

Authors: Despoina Manousaki[1,2], MD, Anders Kämpe[3,4], MD, Vince Forgetta[1], PhD, Riikka E Makitie[5,6,7], MD, Ghalib Bardai[8], MD, Alexandre Belisle[9], BSc, Rui Li[9], PhD, Sture Andersson[10], MD, Outi Makitie[3,5,6,10], MD, Frank Rauch[8], MD, J Brent Richards[1,2,11,12], MD

[1]Lady Davis Institute for Medical Research, Jewish General Hospital, McGill University, Montreal, Canada

[2]Department of Human Genetics, McGill University, Montreal, Canada

[3]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden

[4]Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden

[5]Research Program for Clinical and Molecular Metabolism, University of Helsinki, Helsinki, Finland

[6]Folkhälsan Institute of Genetics, Helsinki, Finland

[7]Molecular Endocrinology Laboratory, Department of Medicine, Hammersmith Campus, Imperial College London, London, United Kingdom

[8]McGill University, Ingram School of Nursing, and Shriners Hospitals for Children, Montreal, Canada

[9]McGill Genome Center, McGill University, Montreal, Canada

[10]Children's Hospital, Pediatric Research Center, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

[11]Department of Medicine, Epidemiology, Biostatistics & Occupational Health, McGill University, Montreal, Canada

[12]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK

*Corresponding author:

Brent Richards, MD, MSc, Professor of Medicine

Pavillon H-413, Jewish General Hospital

3755 Cote Ste Catherine

Montreal, QC, Canada, H3T 1E2

T: +1 514 340 8222 ext. 24362

F: +1 514 340 7529

brent.richards@mcgill.ca

Short title: Polygenicity of childhood significant fracture history

This article contains Supplemental Material

Disclosure page

The authors have nothing to disclose.

Abstract

Extreme presentations of common disease in children are often presumed to be of Mendelian etiology, but their polygenic basis has not been fully explored. We tested whether children with significant fracture history and no osteogenesis imperfecta (OI) are at increased polygenic risk for fracture. A childhood significant fracture history was defined as the presence of low-trauma vertebral fractures or multiple long bone fractures. We generated a polygenic score of heel ultrasound derived speed of sound, termed "gSOS", which predicts risk of osteoporotic fracture. We tested if individuals from three cohorts with significant childhood fracture history had lower gSOS. A Canadian cohort included 94 children with suspected Mendelian osteoporosis, of which 68 had negative OI gene panel. Two Finnish cohorts included 59 children with significant fracture history and 22 with suspected Mendelian osteoporosis, among which 18 had no OI. After excluding individuals with OI and ancestral outliers, we generated gSOS estimates and compared their mean to that of a UK Biobank subset, representing the general population. The average gSOS across all three cohorts (n=131) was -0.47 SD lower than that in UK Biobank (n=80,027, P= 1.1 x 10^{-5}). The gSOS of 78 individuals with suspected Mendelian osteoporosis was even lower (-0.76 SD, P =5.3 x10^{-10}). Among the 131 individuals with a significant fracture history, we observed 8 individuals with gSOS below minus 2SD from the mean; their mean lumbar spine DXA-derived BMD z-score was -1.7 (SD:0.8). In summary, children with significant fracture history but no OI have an increased burden of common risk alleles. This suggests that a polygenic contribution to disease should be considered in children with extreme presentations of fracture.

Introduction

Children with phenotypes at the extremes of common disease spectrum are sometimes presumed to have a monogenic Mendelian disease. Typical examples include extreme presentations of obesity(1), short or tall stature(2), and autism spectrum disorders(3). However, in most cases, children with such phenotypic presentations investigated with targeted gene panels are found to be negative for mutations in known disease-related genes(4). While such cases could still be caused by Mendelian mutations in unknown genes, it is also possible that such patients have an increased burden of common risk alleles, which individually confer a small risk of disease, but taken together can have large effects on disease risk(5). The extent to which common risk alleles of small effect on disease risk contribute to presumed Mendelian presentations of a disease is not well known.

We sought to explore the contribution of common risk alleles to a bone fragility phenotype in children. The rationale for choosing this phenotype as an example of extreme presentation of common disease in childhood is twofold. First, fractures are indeed common in childhood, affecting as much as half of otherwise healthy children and adolescents(6). However, it is rare for children to have repeated long-bone fractures or low-trauma vertebral compression fractures, and such cases are often considered to be due to an underlying monogenic disorder(7). The most common fracture-related monogenic disorder is osteogenesis imperfecta (OI), a connective tissue disease with typical extraskeletal findings, such as blue sclerae and dentinogenesis imperfecta(8). OI is a Mendelian disorder caused by mutations in more than 20 different genes, with an estimated prevalence of 1 in 20,000(9). These mutations are present in 97 to 99% of cases with typical extraskeletal findings of OI (10), but when these findings are absent, a molecular diagnosis of OI explains less than a third of cases of bone fragility(10). However, if the cause of these fractures

were polygenic, this information could be important for treatment and identification of other family members at risk. The second reason for studying significant fracture history as an example of extreme presentation of common disease, is that bone density is a highly heritable trait, for which hundreds of common risk alleles have been identified through genome-wide association studies (11-16). An emerging method which enables identification of individuals at increased genetic risk for fracture and other common diseases, is through the use of polygenic risk scores, which use genotypes to assess risk of heritable diseases and clinically-relevant traits(17-19). We have recently demonstrated that 24% of the variance, as measured by r^2, in heel ultrasound-derived speed of sound (SOS), a skeletal measure of fracture risk, can be predicted from genotypes alone—approximately five fold more than can be predicted from conventional clinical risk factors(17).

Therefore, in the present study, we computed the genetically predicted heel SOS (gSOS) using a polygenic risk score in three separate cohorts of individuals with a significant fracture history in childhood, but no typical extra-skeletal symptoms, skeletal deformities, or family history of OI. Next, we compared their gSOS to that of the general population, using as a control population, a large sample of 80,027 individuals from UK Biobank, a population-based cohort. Finally, we assessed within these cohorts the number of individuals with extremely low gSOS (defined as below 2 standard deviations [SD] from the mean), and their average lumbar spine BMD.

Materials and Methods

Patients

The population of this study comprised young individuals from three separate cohorts, with a significant fracture history but no typical OI phenotype (**Figure 1**), and a control population from UK Biobank.

The Canadian cohort included 94 subjects under age 21 years with suspected Mendelian cause of their fractures and no family history of OI. Of them, 26 had a positive and 68 had a negative gene panel comprising 21 OI-related genes. 67 of these 68 children with a negative gene panel were genome-wide genotyped (described below). In this cohort significant fracture history was defined as the presence of one or more of the following: one or more long-bone fractures of the lower extremities; two or more long-bone fractures of the upper extremities; or one or more vertebral compression fracture. All individuals were evaluated either at the Shriners Hospital for Children in Montreal or at the Children's Hospital of Eastern Ontario in Ottawa. A detailed description of the cohort of 94 children who underwent OI gene panel sequencing can found in a previous publication(10). This study was approved by the Institutional Review Board of McGill University and the Research Ethics Board at the Children's Hospital of Eastern Ontario. Consent to perform next generation sequencing for the present study was obtained from all the participants and their parents.

The two additional cohorts comprised subjects from Finland with a significant fracture history, as defined by the International Society for Clinical Densitometry (ISCD) criteria. Specifically, participants had sustained at least one vertebral fracture; two long-bone fractures before age 10 years; or three long-bone fractures before age 19 years.

The first Finnish cohort consisted of 59 genome-wide genotyped “fracture-prone” children, aged 4 to 16 years. These children had been previously treated for a radiologically confirmed fracture and assessed for previous fracture history at the Children's Hospital in Helsinki during a 12-month recruitment period. All of them met the criteria for a clinically significant fracture history according to the above ISCD criteria, but only 19 of them met the criteria for childhood osteoporosis, which requires presence of one or more vertebral compression fractures, or in addition to a significant long-bone fracture history a lumbar spine BMD z-score < -2.0. Also, none of these children had secondary osteoporosis or a clinical diagnosis of OI. Since concordance between clinical and molecular diagnosis of OI is high(20), sequencing for *COL1A1* and *COL1A2*, the most common OI genes, was not performed, because it was unlikely that these mutations would explain their bone fragility. However, these children were assessed for mutations in selected OI-related genes (*LRP5*, *PLS3, WNT1*)(21, 22) and copy-number variants (CNVs)(23), and in all cases pathogenic mutations were excluded. A detailed description of this sample can be found in a previous publication (24).

The second Finnish cohort consisted of 22 genome-wide genotyped subjects with a clinical diagnosis of childhood-onset primary osteoporosis and clinical suspicion of a Mendelian disorder. In all individuals OI was clinically excluded, 9 individuals had been assessed with *COL1A1* and *COL1A2* sequencing as part of their clinical assessment and were regarded as having early-onset primary osteoporosis of unknown cause. In the remaining 13 subjects whole-genome sequencing was performed for research purposes. In 4 of these 13 individuals a known monogenic cause was subsequently identified. After exclusion of these 4 subjects, 18 individuals from this cohort (9 with negative *COL1A1* and *COL1A2* sequencing and 9 with negative whole-genome sequencing) were

included in the present study. The studies recruiting the two Finnish cohorts were conducted in accordance with the Declaration of Helsinki, and both studies were approved by the Ethics Review Board at Helsinki University Hospital. All patients and/or their guardians gave a written informed consent before participation.

To evaluate the polygenic risk for fracture in the three cohorts separately, and after combining them to a single "juvenile fracture" cohort, we generated gSOS (described below) and compared the resultant values to those of a large control group representative of the general population, specifically a sample of 80,027 White British individuals from UK Biobank. UK Biobank is a large-scale population-based health study which consists of 502,543 volunteer adult participants (mean age 57 years) from the United Kingdom(25). 426,811 UK Biobank participants have been genome-wide genotyped using Affymetrix arrays, and their genotypes were then imputed to the Haplotype Reference Consortium(26, 27). UK Biobank has ethical approval from the Northwest Multi-centre Research Ethics Committee, and informed consent was obtained from all participants.

Genotypes do not change through the lifetime of an individual, and therefore we would not expect the distribution of gSOS to differ between a healthy adult population (such as UK Biobank) and a healthy pediatric population. Nevertheless, since there could be some response bias in UK Biobank, as a population of mainly elderly people, we assessed the average gSOS of a healthy pediatric cohort, and compared it to that of UK Biobank. To do so, we used available genotype data from the VIDI study, originally designed to test the effects of an interventional dose of vitamin D in a cohort of Finnish infants(28, 29). The VIDI study included 975 healthy new-borns, recruited between January 2013 to June 2014 at Kätilöopisto Helsinki Maternity Hospital, Helsinki, Finland.

All children included in the study had normal birth weight and were born at term from healthy mothers of Northern European descent (30). Thus, we can assume that this healthy Finnish infant population is representative of a healthy pediatric population of European ancestry. Among the 975 VIDI participants, 928 individuals underwent genome-wide genotyping. The VIDI study was approved by the Research Ethics Committee of the Hospital District of Helsinki and Uusimaa. All guardians of the VIDI participants gave a written informed consent before participation.

Genome-wide genotyping, quality control, imputation, and control for population stratification

Genotyping for the Canadian cohort was conducted using the UK Biobank Axiom array at the McGill University and Genome Quebec Innovation Centre (MUGQIC). This array contains ~820,000 SNPs. Quality control conducted at Affymetrix for UK Biobank revealed that a total of 794,409 SNPs were reliable, therefore, this later count of SNPs was used for further quality control, which included removing individuals with genotype missingness > 5% and heterozygosity > 3 SD. Genotype imputation was performed using the Sanger Imputation Service. Genotype data was pre-phased by the Sanger Imputation Service using EAGLE6, (31) following by imputation using PBWT 7(32) to the Haplotype Reference Consortium (r1.1)(31). The imputation process resulted in 39,131,578 autosomal SNPs. Population stratification was assessed using EIGENSTRAT version 6.0.14. To assess for ancestry outlier removal in the principal component analysis (PCA), we combined the genotype data from the Canadian cohort with data from 1000 Genomes phase 3 (N=2,504). Analysis using the PLINK 2 "--pca " command was performed to obtain the top 10 principal components. We used the two top principal components to define outliers.

Genotyping for two Finnish fracture cohorts and of the VIDI study was conducted using the Illumina Infinium Global Screening Array v1.0 at the Human Genomics Facility (HuGe-F) at Erasmus MC, Netherlands. A total of 686,085 markers across the genome where genotyped, which included 682,209 biallelic SNPs. The Will Rayner's HRC prep tool and bcftools was used to fix strand issues and harmonize the data with the Haplotype Reference Consortium (r1.1) reference before imputation(31). For imputation, we used the same pipeline as the one applied to the Canadian cohort. The imputation resulted in a total of 40,359,612 SNPs. PCA controlling for population stratification was conducted using the TRACE software (LASER suite) (33, 34). For this analysis, a 4-dimension reference space was created using 503 individuals from the European subset of the 1000 genomes data phase 3, of which 99 individuals are of Finnish descent. A 20-dimension genetic map was created by applying PCA in each individual, after merging it with the reference population. All individuals were then projected from their 20-dimensional map in to the 4-dimensional reference space. To define outliers (individuals not clustering with the Finnish reference), a 4 standard deviation cut-off was chosen graphically.

gSOS calculation

A detailed description of the algorithm used to calculate heel ultrasound gSOS has been described previously.(17) Briefly, 341,449 individuals from UK Biobank with SOS measures were used to develop polygenic risk score models using LASSO regression. The optimal prediction model was determined in 5,335 separate individuals and termed "gSOS". Its utility in fracture risk screening was tested in 5 validation adult cohorts totaling 10,522 participants, and its performance characteristics for predicting fracture risk were similar to measured SOS and femoral neck BMD. gSOS was calculated in individuals of European descent only, defined by a population ancestry

cluster analysis as UK Biobank individuals being within the same cluster as the GBR (British in England and Scotland) 1000 Genomes population.(17) Computation of gSOS in the three juvenile fracture cohorts and in the VIDI study was performed by multiplying the allele count for each SNP allele included in gSOS with the respective beta estimate, derived from the LASSO regression, and by adding the counts across all the alleles for each individual. Thus, a single score was allocated to each participant. To facilitate interpretation and comparisons, gSOS measures of the juvenile fracture cohort and the control cohort from UK Biobank were merged and then standardized to a mean of 0 and a standard deviation of 1.

Clinical analyses

In all three fracture cohorts, DXA was performed in the antero–posterior direction at the lumbar spine (L1–L4) using a Hologic QDR Discovery device (Hologic Inc., Waltham, MA, USA). Lumbar spine areal BMD results were transformed to age- and gender-specific z-scores (zBMD) using published reference data. A detailed description of the assessment of fractures (vertebral and non-vertebral) in the Canadian and the "fracture-prone" Finnish cohort can be found in previous publications(10, 24).

Statistical Analysis

Descriptive statistics of the Canadian and the two Finnish cohorts, appear as means and standard deviations or percentages. The comparisons of the means of the gSOS measurements in the three fracture cohorts and the combined juvenile fracture cohort with the mean gSOS of UK Biobank, as well as of the means of gSOS between UK Biobank and the VIDI cohort, were performed using a student's t-test.

As a sensitivity analysis, we computed gSOS estimates omitting SNPs with low info-score (<0.4) and compared the newly computed mean gSOS to the mean gSOS of UK Biobank. We also computed the mean gSOS separately in the Canadian and the two Finnish cohorts, and compared them to the control population from UK Biobank. Additionally, we conducted a sensitivity analysis on a "suspicion of monogenic disease" cohort, which was the result of the merging of the Canadian to the "primary osteoporosis" Finnish cohort, since all these individuals had been referred for investigation with OI gene panel. Finally, we also computed the mean gSOS of the subset of individuals among the juvenile fracture cohort who satisfied criteria for childhood osteoporosis.

Results

Cohorts

The clinical characteristics of the three fracture cohorts (excluding individuals who failed genotyping or genotype quality control, and who were outliers in the PCA) appear in **Table 1**. The characteristics of the adult control cohort from UK Biobank are provided in **Supplemental Table 1.**

Genotyping and assessment of population stratification

Among the 67 individuals genotyped in the Canadian cohort, two were removed due to extreme heterozygosity rates. Results of the PCA are depicted in **Supplemental Figure 1**, and demonstrate 5 outliers in EIGENSTRAT, which are outside the 1000 Genomes European super-population, whereas the inliers largely overlap or are in close proximity to the European super-population. These outliers were removed from further analyses. The average SNP info score for SNPs included in the calculation of gSOS was 0.89.

All 77 genotyped samples from the two Finnish fracture cohorts (59 from the "fracture-prone" and 18 from the "primary osteoporosis" cohort) passed quality control. Among the 928 genotyped individuals from the VIDI cohort, 7 samples were excluded due to increased genotype missingness (>5%), and 1 individual was excluded due to a discordant genotype-inferred and reported sex. The average info score for SNPs used to calculate gSOS was 0.92. Results of the PCA for the two Finnish fracture cohorts are shown in **Supplemental Figure 2.** Six individuals from the "fracture-prone" cohort were excluded since they did not cluster with the Finnish reference in the 4-dimensional reference map. Twenty-two ancestral outliers were excluded in the VIDI cohort.

Main analysis

The average gSOS in the combined juvenile fracture cohort (n= 131) was -0.47 SD lower than that in UK Biobank (n=80,027, average gSOS: 0.00, p-value for comparison: 1.1×10^{-5}) **(Table 2)**. **Figure 2** shows the distribution of gSOS values of the juvenile fracture cohort, projected on the distribution of gSOS measurements in UK Biobank. The average gSOS in the Canadian cohort (n= 60) was -0.82 SD and was lower than that in the control population from UK Biobank (p-value for difference: 3.7×10^{-9}). The calculated gSOS in the "primary osteoporosis" Finnish cohort (n=18), was -0.54 SD lower than that of the UK Biobank general population (p-value for comparison: 0.04). In contrast, the 53 individuals from the "fracture-prone" Finnish cohort, had an average gSOS of -0.04 SD, which was comparable to the mean in UK Biobank (p-value for comparison: 0.77). The average gSOS of 898 individuals from the VIDI cohort was -0.01 SD, and it was similar to the average gSOS of UK Biobank (p-value for comparison: 0.69). When comparing the average gSOS of the 3 fracture cohorts and of the entire juvenile fracture cohort to that in the VIDI cohort, we obtained similar results with the respective comparisons with UK Biobank (data not shown). This provides evidence that, as expected, UK Biobank, an adult cohort, and the VIDI cohort, a pediatric cohort, can be used interchangeably as control populations in our study, representing both the general European population.

Sensitivity analyses

The "suspicion of monogenic disease" subset consisted of 78 individuals who had undergone gene panel sequencing or otherwise extensively examined genetically (Canadian cohort and Finnish "primary osteoporosis" cohort). Their average gSOS was -0.76 SD lower than in UK Biobank (p-

value for comparison: 5.3 $x10^{-10}$). The gSOS of the subset of individuals (n=57) who satisfied the definition of "childhood osteoporosis" from all 3 cohorts (n=27 from the Canadian, n=15 from the Finnish "primary osteoporosis" and n=15 from the Finnish "fracture-prone" cohort) was -0.67 SD lower than UK Biobank (p-value for comparison: 1.9 $x10^{-5}$) (**Table 2**).

Among the 20,796 SNPs used to compute gSOS in the Canadian cohort, 1,467 SNPs had low info score (<0.4), while among the 21,716 SNPs used to compute gSOS in the two Finnish cohort, 1,071 SNPs had info score below 0.4. We undertook a sensitivity analysis by recomputing gSOS excluding these SNPs in all three cohorts, and compared the new mean gSOS to that of UK Biobank. We found that gSOS was still lower in the juvenile fracture cohort (average gSOS -0.40 SD) compared to UK Biobank (p-value: 7.8 $x10^{-5}$). The info-score filtered gSOS estimates separately in the three cohorts appear in **Table 2.**

Participants with gSOS below minus 2 SD

We next assessed the number of individuals with fractures with gSOS below 2 SD from the mean of the general population, since this could be an informative and clinically relevant cut-off for fracture risk stratification. Of the 131 individuals in the combined juvenile fracture cohort, 8 individuals (6%) had standardized gSOS below minus 2, compared to 3 individuals (2.5%) which would be expected by chance alone. Among these 8 individuals, there were 5 individuals from the Canadian cohort, 2 individuals from the fracture-prone Finnish cohort, and 1 individual from the primary osteoporosis Finnish cohort. The average zBMD of these 8 individuals was -1.7 (SD: 0.8). Notably, the average heel ultrasound-derived SOS z-score in 1,878 adult participants from the control UK Biobank subset with standardized gSOS below minus 2 was -1.0 (SD:0.69).

Discussion

In this study, we found that young individuals presenting with increased bone fragility and presumed to have a monogenic etiology of their condition tended to have an increased burden of common risk alleles. Specifically, we found that gSOS in this population was, on average, half a standard deviation below that of the general population. These findings imply that an important proportion of patients with apparent monogenic - Mendelian disease and a significant fracture history in childhood and suspicion of OI, could have an increased burden of common risk alleles predisposing them to this clinical presentation.

These results may have clinical implications, since they imply that polygenic predisposition to disease may be present in an important proportion of individuals who are presumed to have Mendelian forms of disease. As our ability to quantify polygenic risk improves through rapidly evolving polygenic risk scores, the ability to estimate this risk may inform clinical care not only of fracture, but of other traits and diseases, such as extreme presentations of autism(3), obesity(1) and short or tall stature(2), which appear to have both monogenic and polygenic forms of the disease. The accuracy of polygenic risk scores(35) depend in large part on the sample size from which they are calculated, suggesting that further increases in sample sizes of genome-wide genotyped populations will help to improve their clinical performance and allow the establishment of clinically relevant cut-offs for risk stratification. In the absence of such cut-offs, polygenic risk scores, although not deterministic, can still be informative and complement other approaches, such as disease-specific gene panel sequencing, whole exome or genome sequencing. In settings where clear cut-offs are set, knowing the polygenic predisposition to disease can directly influence

clinical management, and may also decrease its cost, since it is based upon genome-wide genotyping, which costs below $50 USD, and is considerably less expensive than most sequencing-based tests, whose cost remains over $1,000 USD (genome.gov/sequencingcosts).

Our study has important limitations. Our three fracture cohorts are heterogeneous, with the individuals in the Finnish "fracture-prone" cohort presenting a milder phenotype compared to the two other cohorts. Despite this, when taken together, participants in all three cohorts had a definition of significant fracture history in childhood and no clinical or molecular diagnosis of OI. We chose a genetic predictor of SOS and not of BMD as a measure of fracture risk, since the available SOS GWAS(15) is larger than existing BMD GWASes, which allowed the development of a more performant polygenic risk score for SOS than previous existing polygenic scores for BMD(36). Also, SOS has a gradient of risk for osteoporotic fractures similar to BMD and a recent meta-analysis described its effect on fracture risk, allowing for the development of a SOS-based screening model(37). The predictive performance of gSOS on childhood heel ultrasound SOS has not been tested, yet we assumed that it can be used as a predictor of pediatric BMD, since there is a large overlap between findings of pediatric and adult BMD GWAS (38). Although there may be a variation in the effect sizes of the BMD-related SNPs according to age, with the effects being generally attenuated in the pediatric studies, the actual loci remain largely similar, as indicated by a recent life-course GWAS on total body BMD(38). Also, since there is no consensus on a clinically-relevant cut-off value for gSOS, below which an individual is considered at increased risk for fracture, we were not able to test the performance of gSOS as a screening tool in our juvenile fracture cohort. Replication of our findings in larger cohorts of children with a significant fracture history is warranted, and might allow testing the accuracy of various cut-off values of

gSOS in identifying individuals at risk. Further, the polygenic risk score deployed here was derived from a European-ancestry population and was tested in individuals of European ancestry. The performance of such a score in other ancestral populations remains to be tested. Finally, as mentioned above, it would have been interesting to assess how gSOS affects risk of fracture within individuals with an identified Mendelian cause of bone fragility. This could be the scope of a future study on a larger cohort of children diagnosed with OI.

In summary, we provide evidence that a portion of children with bone fragility and no extraskeletal signs of OI have an increased burden of common gene variants which predispose to osteoporotic fracture in adulthood.

Acknowledgments

We acknowledge all participants in the three cohorts and their families.

The Richards lab is supported by the Canadian Institutes of Health Research (CIHR), the Canadian Foundation for Innovation and the Fonds de Recherche Santé Québec (FRQS). Dr. Richards is supported by a FRQS Clinical Research Scholarship. Dr. Manousaki is supported by the Juvenile Diabetes Research Foundation (JDRF).

Author's roles

BR conceptualized and designed the study. FR, RM, SA and OM collected data. DM and AK collected data and undertook the bioinformatics analyses. AB and RL were involved in the analysis of the genotypes of the Canadian Cohort. DM wrote the first draft of the manuscript. All authors revised critically the article for important intellectual content and approved the final version of the manuscript. BR takes responsibility for the integrity of the data analysis.

References

1. Pettersson M, Viljakainen H, Loid P, Mustila T, Pekkinen M, Armenio M, et al. Copy Number Variants Are Enriched in Individuals With Early-Onset Obesity and Highlight Novel Pathogenic Pathways. J Clin Endocrinol Metab. 2017;102(8):3029-39.
2. Katsanis N. The continuum of causality in human genetic disorders. Genome Biol. 2016;17(1):233.
3. Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, et al. Most genetic risk for autism resides with common variation. Nat Genet. 2014;46(8):881-5.
4. Kalsner L, Twachtman-Bassett J, Tokarski K, Stanley C, Dumont-Mathieu T, Cotney J, et al. Genetic testing including targeted gene panel in a diverse clinical population of children with autism spectrum disorder: Findings and implications. Mol Genet Genomic Med. 2018;6(2):171-85.
5. Timpson NJ, Greenwood CMT, Soranzo N, Lawson DJ, Richards JB. Genetic architecture: the shape of the genetic contribution to human traits and disease. Nat Rev Genet. 2018;19(2):110-24.
6. Jones IE, Williams SM, Dow N, Goulding A. How many children remain fracture-free during growth? a longitudinal study of children and adolescents participating in the Dunedin Multidisciplinary Health and Development Study. Osteoporos Int. 2002;13(12):990-5.
7. Ward LM, Konji VN, Ma J. The management of osteoporosis in children. Osteoporos Int. 2016;27(7):2147-79.
8. Forlino A, Marini JC. Osteogenesis imperfecta. Lancet. 2016;387(10028):1657-71.
9. Marini JC. Osteogenesis imperfecta: comprehensive management. Adv Pediatr. 1988;35:391-426.
10. Bardai G, Ward LM, Trejo P, Moffatt P, Glorieux FH, Rauch F. Molecular diagnosis in children with fractures but no extraskeletal signs of osteogenesis imperfecta. Osteoporos Int. 2017;28(7):2095-101.
11. Richards JB, Rivadeneira F, Inouye M, Pastinen TM, Soranzo N, Wilson SG, et al. Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. Lancet. 2008;371(9623):1505-12.
12. 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(5):491-501.
13. Zheng HF, Forgetta V, Hsu YH, Estrada K, Rosello-Diez A, Leo PJ, et al. Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. Nature. 2015;526(7571):112-7.
14. Kemp JP, Morris JA, Medina-Gomez C, Forgetta V, Warrington NM, Youlten SE, et al. Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. Nat Genet. 2017;49(10):1468-75.
15. Morris JA, Kemp JP, Youlten SE, Laurent L, Logan JG, Chai RC, et al. An atlas of genetic influences on osteoporosis in humans and mice. Nat Genet. 2019;51(2):258-66.

16. Kim SK. Identification of 613 new loci associated with heel bone mineral density and a polygenic risk score for bone mineral density, osteoporosis and fracture. PLoS One. 2018;13(7):e0200785.
17. Forgetta V K-BJ, Forest M, Durand A, Bhatnagar S, Kemp J, Morris JA, Kanis JA, Kiel DP, McCloskey EV, Rivadeneira F, Johannson H, Harvey N, Cooper C, Evans DM, Pineau J, Leslie WD, Greenwood CMT, Brent Richards J. B. Machine Learning to Predict Osteoporotic Fracture Risk from Genotypes [Internet]. 2018. Available from: https://www.biorxiv.org/content/early/2018/09/12/413716.
18. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 2018;50(9):1219-24.
19. Theriault S, Lali R, Chong M, Velianou JL, Natarajan MK, Pare G. Polygenic Contribution in Individuals With Early-Onset Coronary Artery Disease. Circ Genom Precis Med. 2018;11(1):e001849.
20. Hartikka H, Makitie O, Mannikko M, Doria AS, Daneman A, Cole WG, et al. Heterozygous mutations in the LDL receptor-related protein 5 (LRP5) gene are associated with primary osteoporosis in children. J Bone Miner Res. 2005;20(5):783-9.
21. Kampe AJ, Costantini A, Makitie RE, Jantti N, Valta H, Mayranpaa M, et al. PLS3 sequencing in childhood-onset primary osteoporosis identifies two novel disease-causing variants. Osteoporos Int. 2017;28(10):3023-32.
22. Saarinen A, Mayranpaa MK, Lehesjoki AE, Makitie O. Low-density lipoprotein receptor-related protein 5 (LRP5) variation in fracture prone children. Bone. 2010;46(4):940-5.
23. Costantini A, Skarp S, Kampe A, Makitie RE, Pettersson M, Mannikko M, et al. Rare Copy Number Variants in Array-Based Comparative Genomic Hybridization in Early-Onset Skeletal Fragility. Front Endocrinol (Lausanne). 2018;9:380.
24. Mayranpaa MK, Viljakainen HT, Toiviainen-Salo S, Kallio PE, Makitie O. Impaired bone health and asymptomatic vertebral compressions in fracture-prone children: a case-control study. J Bone Miner Res. 2012;27(6):1413-24.
25. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12(3):e1001779.
26. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016;48(10):1279-83.
27. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203-9.
28. Rosendahl J, Valkama S, Holmlund-Suila E, Enlund-Cerullo M, Hauta-Alus H, Helve O, et al. Effect of Higher vs Standard Dosage of Vitamin D3 Supplementation on Bone Strength and Infection in Healthy Infants: A Randomized Clinical Trial. JAMA Pediatr. 2018;172(7):646-54.
29. Helve O, Viljakainen H, Holmlund-Suila E, Rosendahl J, Hauta-Alus H, Enlund-Cerullo M, et al. Towards evidence-based vitamin D supplementation in infants: vitamin D intervention in infants (VIDI) - study design and methods of a randomised controlled double-blinded intervention study. BMC Pediatr. 2017;17(1):91.

30. Saari A, Sankilampi U, Hannila ML, Kiviniemi V, Kesseli K, Dunkel L. New Finnish growth references for children and adolescents aged 0 to 20 years: Length/height-for-age, weight-for-length/height, and body mass index-for-age. Ann Med. 2011;43(3):235-48.
31. Loh PR, Danecek P, Palamara PF, Fuchsberger C, Y AR, H KF, et al. Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet. 2016;48(11):1443-8.
32. Durbin R. Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). Bioinformatics. 2014;30(9):1266-72.
33. Wang C, Zhan X, Liang L, Abecasis GR, Lin X. Improved ancestry estimation for both genotyping and sequencing data using projection procrustes analysis and genotype imputation. Am J Hum Genet. 2015;96(6):926-37.
34. Wang C, Zhan X, Bragg-Gresham J, Kang HM, Stambolian D, Chew EY, et al. Ancestry estimation and control of population stratification for sequence-based association studies. Nat Genet. 2014;46(4):409-15.
35. Khera AV, Chaffin M, Wade KH, Zahid S, Brancale J, Xia R, et al. Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. Cell. 2019;177(3):587-96 e9.
36. Warrington NM, Kemp JP, Tilling K, Tobias JH, Evans DM. Genetic variants in adult bone mineral density and fracture risk genes are associated with the rate of bone mineral density acquisition in adolescence. Hum Mol Genet. 2015;24(14):4158-66.
37. McCloskey EV, Kanis JA, Oden A, Harvey NC, Bauer D, Gonzalez-Macias J, et al. Predictive ability of heel quantitative ultrasound for incident fractures: an individual-level meta-analysis. Osteoporos Int. 2015;26(7):1979-87.
38. Medina-Gomez C, Kemp JP, Trajanoska K, Luan J, Chesi A, Ahluwalia TS, et al. Life-Course Genome-wide Association Study Meta-analysis of Total Body BMD and Assessment of Age-Specific Effects. Am J Hum Genet. 2018;102(1):88-102.

Figure Legends

Figure 1. Flow diagram with the inclusion criteria to the study across the 3 pediatric fracture cohorts. OI: osteogenesis imperfecta; QC: genotype quality control; PCA: principal component analysis; *COL1A1*: Collagen type I alpha 1 chain; *COL1A2*: Collagen type I alpha 2 chain; *LRP5*: Low-density lipoprotein receptor-related protein 5; *PLS3:* Plastin 3; *WNT1*: Wnt Family Member 1; CNVs: copy number variants; ISCD: International Society for Clinical Densitometry

Figure 2. Comparison of the distribution of gSOS in the juvenile fracture cohort vs in UK Biobank.

Density plot of the distribution of gSOS in 131 individuals of the juvenile fracture cohort (in pink), projected on the distribution of gSOS in 80,027 subjects from UK Biobank (in blue). gSOS: genetically predicted speed of sound; UKBB: UK Biobank

Table 1. Cohort characteristics.

	Canadian cohort	Finnish cohort "fracture-prone"	Finnish cohort "primary osteoporosis"
N*	60	53	18
Age	11.4 (1-21.2)	10.3 (4.3-16.8)	-
Females	22 (37%)	16 (30%)	-
zBMD	-1.82 (1.3)	-0.46 (1.0)	-1.84 (1.34)
Vertebral fracture	11 (18%)	14 (26%)	13 (72%)
Any long bone fracture	54 (90%)	49 (92%)	16 (89%)
Long bone fracture N	2.5 (0-12)	3.0 (0-7)	2.0 (0-7)
OI gene panel	negative	not performed	negative

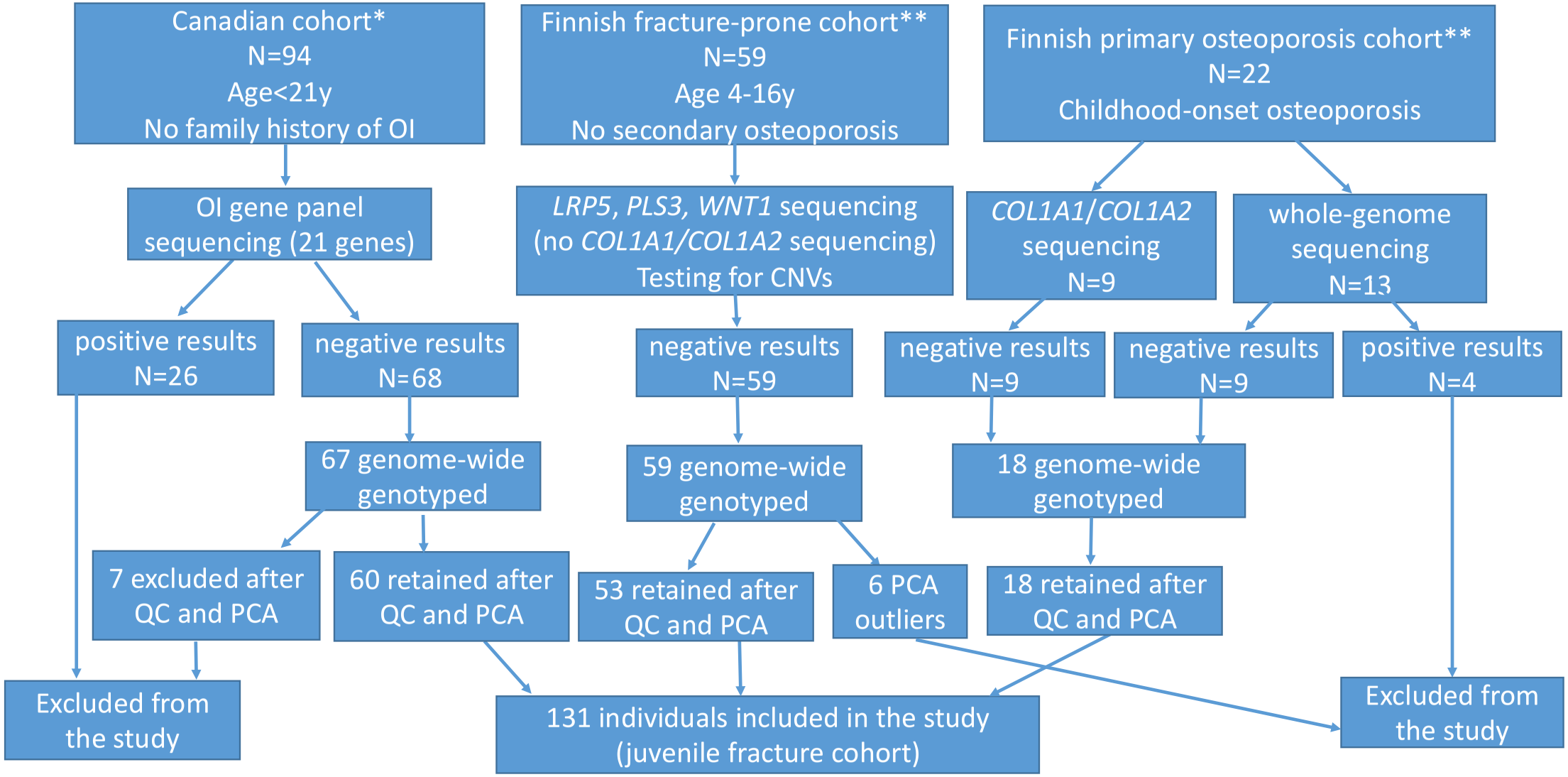
*N of individuals included in the gSOS analysis after QC and excluding outliers at PCA

Values for zBMD are expressed as means and SD. Age and long bone fracture N are expressed as median and range. zBMD: lumbar bone mineral density z-score; OI: osteogenesis imperfecta; PCA: principal component analysis

Table 2. Results of the main analysis and sensitivity analyses.

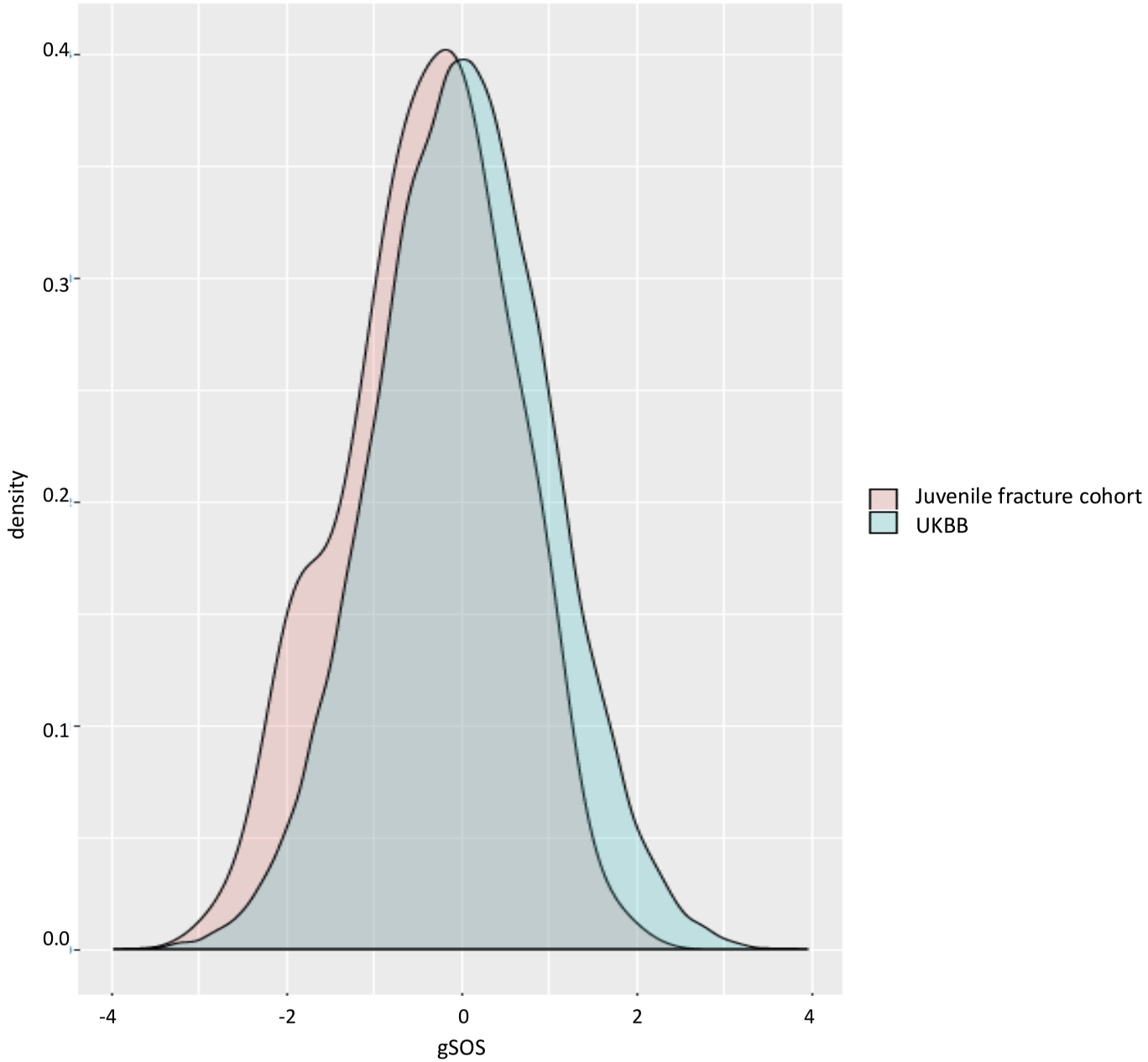
Main analysis with gSOS estimates in the fracture cohorts and in the VIDI cohort and their comparison to UK Biobank. Sensitivity analyses with gSOS estimates in subsets of the main cohort, and in the main cohort excluding low-info score SNPs from gSOS, and their comparison to UK Biobank. Values are expressed in means and standard deviations. gSOS: genetically predicted speed of sound; SNP: single nucleotide polymorphism; VIDI: Vitamin D Intervention in Infants study

Cohort	gSOS	p-value for comparison with UK Biobank
Main analysis		
UK Biobank (n=80,027)	0.00 (1.00)	-
VIDI (n=898)	0.01 (1.00)	0.69
All juvenile fracture (n=131)	-0.47 (1.00)	1.1×10^{-5}
Canadian cohort (n=60)	-0.82 (0.90)	3.7×10^{-9}
Finnish "fracture prone" cohort (n=53)	-0.04 (1.06)	0.77
Finnish "primary osteoporosis" cohort (n=18)	-0.54 (1.01)	0.04
Sensitivity analyses		
"Suspicion of monogenic disease" subset (n=78)	-0.76 (1.06)	5.3×10^{-10}
"Childhood osteoporosis" subset (n=57)	-0.67 (1.07)	1.9×10^{-5}
All juvenile fracture excluding low-info score SNPs(n=131)	-0.40 (1.10)	7.8×10^{-5}



*Significant fracture history defined as at least one of the following: ≥ 1 long-bone fractures of the lower extremities; ≥ 2 long-bone fractures of the upper extremities; ≥1 vertebral compression fracture.

**Significant fracture history defined using the ISCD criteria: ≥ 1 vertebral fracture ; or ≥ 2 long-bone fractures before age 10y; or ≥ 3 long-bone fractures before age 19y


0.4
0.3
0.2
0.1
0.0
density
-4
-2
0
2
4
gSOS
Juvenile fracture cohort
UKBB

Supplemental Tables and Figures

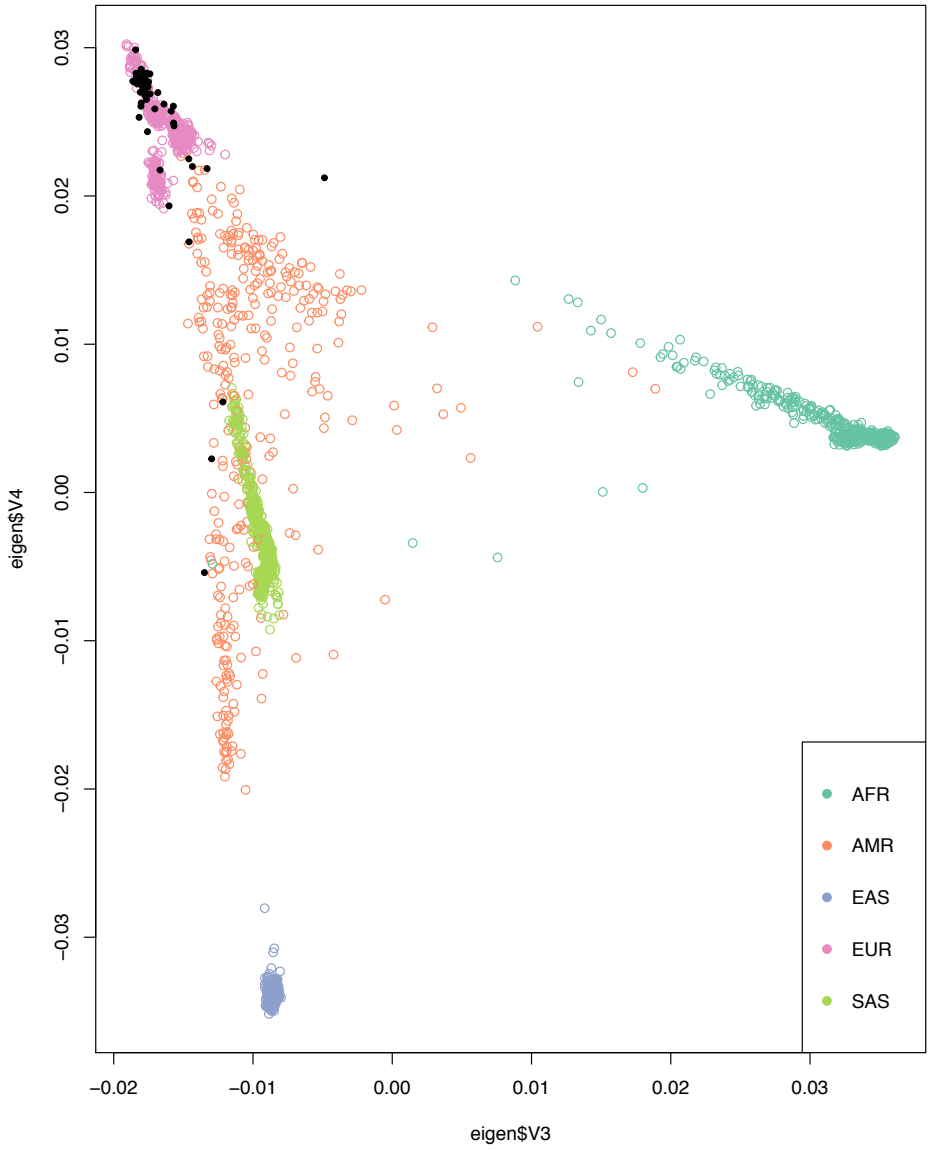
Supplemental Table 1. Cohort characteristics of the UK Biobank control cohort

	UK Biobank control cohort
Sample size	80,027
Age (mean/SD)	56.77 (8.03)
Females (counts (%))	43.7 (55%)
SOS (mean/SD)	-0.02 (0.99)
Prevalent fracture (counts (%))	8,244 (10.3%)

SOS: speed of sound from heel ultrasound

Supplemental Figure 1. Principal component analysis in the Canadian cohort.

Plot depicts comparison of the top 2 principal components (eigen$V4 vs eigen$V3) from the analysis of the 65 participants (in black dots) projected on 2,504 individuals from 1000 Genomes, that cluster in 5 ancestries, with the pink dots representing the European cluster. AFR: African; AMR: Ad Mixed American; EAS: East Asian; EUR: European; SAS: South Asian

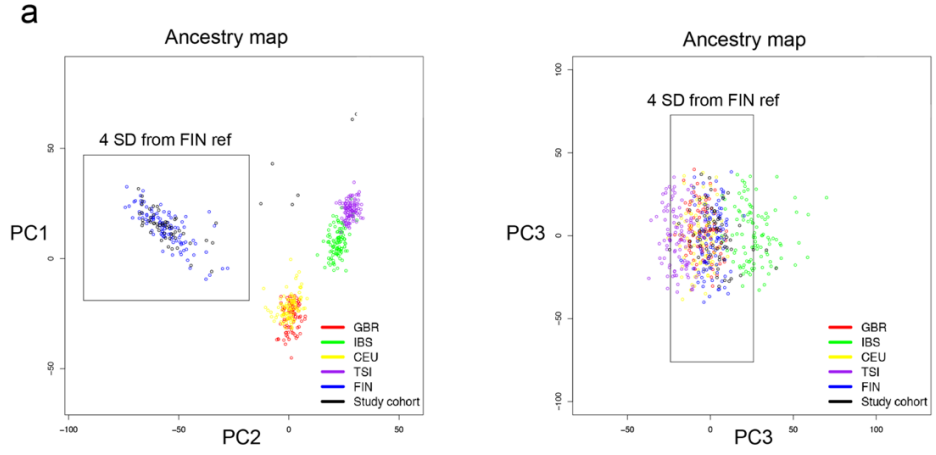


Supplemental Figure 2. Principal component analysis of the two Finnish cohorts.

Using the European subset of the 1000 genomes phase 3 population, including 99 Finnish individuals, a 4-dimensional reference ancestry map was created. The subjects in the two study cohorts have then been projected onto the reference map (black dots). The black boxes represent the space within 4SD of the 99 Finnish individuals. **a.** Fracture prone cohort. **b.** Primary osteoporosis cohort. PC1: principal component 1; PC2: principal component 2; PC3: principal component 3; GBR: British in England and Scotland; IBS: Iberian Population in Spain; CEU: Utah Residents (CEPH) with Northern and Western European Ancestry; TSI: Toscani in Italia; FIN: Finnish in Finland

Fracture prone children - 59 samples

a



Primary osteoporosis - 18 samples

b