Bone mineral density is known to be a heritable, polygenic trait whereas genetic variants contributing to lean mass variation remain largely unknown. We estimated the shared SNP heritability and performed a bivariate GWAS meta-analysis of total-body lean mass (TB-LM) and total-body less head bone mineral density (TBLH-BMD) regions in 10,414 children. The estimated SNP heritability is 43% (95% CI: 34–52%) for TBLH-BMD, and 39% (95% CI: 30–48%) for TB-LM, with a shared genetic component of 43% (95% CI: 29–56%). We identify variants with pleiotropic effects in eight loci, including seven established bone mineral density loci: WNT4, GALNT3, MEPE, CPED1/WNT16, TNFSF11, RIN3, and PPP6R3/LRP5. Variants in the TOM1L2/SREBF1 locus exert opposing effects TB-LM and TBLH-BMD, and have a stronger association with the former trait. We show that SREBF1 is expressed in murine and human osteoblasts, as well as in human muscle tissue. This is the first bivariate GWAS meta-analysis to demonstrate genetic factors with pleiotropic effects on bone mineral density and lean mass.
he interaction between skeletal muscle and bone has long been perceived as being mechanical, where bone provides an attachment site for muscles and muscles apply forces to bone. Such muscle-derived forces drive an adaptive response in bone affecting its morphology, structure, and strength as portrayed by the mechanostat theory. Whenever mechanical stimuli exceed a set point, bone tissue is added at the location, where it is mechanically necessary. As muscle mass is a key determinant of bone mineral density (BMD) variation, it is expected that these traits are phenotypically highly correlated. Recent studies show that the coupling between these two tissues is much more complex. It extends beyond the well-established mechanical interaction, already beginning during embryonic development, when osteoblasts and muscle cells share a common mesenchymal precursor. Indeed, there is constant paracrine crosstalk between bone and muscle throughout life, with both sharing common genetic determinants exerting pleiotropic effects.

BMD, measured by dual-energy X-ray absorptiometry (DXA), is commonly used in clinical practice to assess bone health in young populations and to diagnose osteoporosis and determine fracture risk in the older populations. Conveniently, lean mass, which is a good proxy for skeletal muscle mass, can also be derived from the same whole-body DXA scans. A previous study in postmenopausal women showed DXA lean mass has a high correlation with skeletal muscle mass of the whole body measured by magnetic resonance imaging (MRI). Furthermore, another study in peri-pubertal children showed a 0.98 correlation between DXA derived lean mass in the mid third femur and skeletal muscle mass of the same region measured with MRI.

Heritability studies have demonstrated that between 50 and 85% of BMD variation can be explained by genetic factors, whereas heritability estimates for muscle phenotypes, such as lean mass, grip strength, arm flexion, and leg strength, range between 30 and 65%. Genome-wide association studies (GWAS) have been successful in revealing the genetic architecture of BMD, identifying >60 different loci robustly associated with the trait at different skeletal sites. In contrast, very few robust associations have been reported to date by GWAS of lean mass. GWAS usually evaluate one trait at a time, however, recent methods enable multivariate analyses to be performed. Multivariate methodologies offer several advantages over univariate GWAS including, increased power (when traits are genetically correlated), reduction of the multiple testing burden and enhanced biological insight (in the case of pleiotropy).

Therefore, this approach is well-suited to study the genetic influence over the bone-muscle unit. We applied a bivariate GWAS approach for total-body lean mass (TB-LM) and BMD less head region (TBBLH-BMD) in four pediatric cohorts leading to the identification of variants in eight different loci with statistical evidence for pleiotropic effects on both traits. Seven of these are established BMD loci, while the 17p11.2 locus is reported for the first time as associated with musculoskeletal traits. Our functional follow-up points to SREBF1 as the most likely gene underlying the association signal. SREBP1, the product of SREBF1 is implicated in osteoblast and myoblast differentiation by previous studies. Our findings shed a light into the role of SREBF1 in the crosstalk between bone and muscle during development.

Results

SNP heritability and genetic correlation. Our study includes data from 10,414 participants from four different pediatric cohorts (Table 1, Supplementary Note). In all four studies TBBLH-BMD and TB-LM phenotypic correlation (ρ) adjusted for sex, age, and fat percent was significant (P < 0.01) and of similar magnitude, Generation R (ρ=0.44), ALSPAC (ρ=0.45), BMD-CS (ρ=0.49), and COPSAC (ρ=0.42). GCTA calculated SNP heritability and genetic correlation for adjusted TBBLH-BMD and TB-LM were estimated for the Generation R (N = 3027) and ALSPAC (N = 4820) studies, where there was sufficient statistical power for the analysis (Table 2). Significant SNP heritability estimates ranged between 30 and 45% for both traits in these two cohorts, with a genetic correlation of ~30% in both studies. Similar results were derived using LD-score methodology, which estimated the heritability of TBBLH-BMD to be 43% (CI: 34–52%), 39% (CI: 30–48%) for TB-LM and a shared genetic component of 43% (CI: 29–56%).

Univariate and bivariate GWAS of lean mass and BMD. Univariate GWAS meta-analysis identified variants associated at the genome-wide significant (GWS, 5×10−8) level with TBBLH-BMD mapping to four different loci all of which have been previously associated with BMD in adult and/or children. Notwithstanding, twin studies have calculated the additive genetic correlation of BMD and lean mass to range from 30 to 45%. GWAS usually evaluate one trait at a time, however, recent methods enable multivariate analyses to be performed. Multivariate methodologies offer several advantages over univariate GWAS including, increased power (when traits are genetically correlated), reduction of the multiple testing burden and enhanced biological insight (in the case of pleiotropy). Therefore, this approach is well-suited to study the genetic influence over the bone-muscle unit. We applied a bivariate GWAS approach for total-body lean mass (TB-LM) and BMD less head region (TBBLH-BMD) in four pediatric cohorts leading to the identification of variants in eight different loci with statistical evidence for pleiotropic effects on both traits. Seven of these are established BMD loci, while the 17p11.2 locus is reported for the first time as associated with musculoskeletal traits. Our functional follow-up points to SREBF1 as the most likely gene underlying the association signal. SREBP1, the product of SREBF1 is implicated in osteoblast and myoblast differentiation by previous studies. Our findings shed a light into the role of SREBF1 in the crosstalk between bone and muscle during development.

Table 1 Anthropometric characteristics of study participants

|               | Generation R | ALSPAC | BMD-CS | COPSAC |
|---------------|--------------|--------|--------|--------|
| Mean (SD)     | n = 4071     | n = 5251 | n = 821 | n = 273 |
| Age, years    | 6.21±0.32    | 9.94±0.32 | 8.74±1.91 | 6.89±0.72 |
| Women (%)     | 2035±1998    | 2673±50% | 634±77% | 144±53% |
| CEU ancestry  | 2171±53.32   | 5251±100% | 1.32±0.12 | 271±100% |
| Height (m)    | 1.19±0.63    | 1.4±0.64 | 1.32±0.12 | 1.24±0.61 |
| Weight (kg)   | 23.08±1.08   | 34.7±7.41 | 30.78±8.76 | 24.6±4.25 |
| TBBLH BMD (g cm⁻²) | 0.555±0.05 | 0.777±0.053 | 0.67±0.10 | 0.583±0.05 |
| TBBLH BMC (g) | 528.5±104.21 | 891.92±181.85 | 726.12±205.92 | 2932±591.04 |
| TB Lean Mass (g) | 16,393±2284 | 24,553±3184 | 22,360±5830 | 17,460±2600 |
| TB Fat Mass (g) | 5862±2328 | 8561±5108 | 7590±3580 | 7010±2490 |
bias (i.e., population stratification, cryptic family relatedness or genotyping errors) (Supplementary Fig. 1).

On the basis of the covariate adjusted TBLH-BMD/TB-LM bivariate analysis, the genomic inflation factor \( \lambda \) was 1.08. In contrast to the two univariate analyses, the bivariate meta-analysis identified eight different GWAS signals (Fig. 1). Four of them were not found to be GWAS in the TBLH-BMD analysis mapping to the 4q22.1 MEPE locus; the 13q14.11 TNFSF11 locus; the 14q2.12 RN3 loci, all known BMD loci, and the 17p11.2 TOM1L2/SREBF1 locus (Table 3, Supplementary Fig. 2). Signals in the 2q24.3, 11q13.2, and 17p11.2 loci presented nominal evidence of association (\( P < 0.05 \)) with both traits (Table 3, Supplementary Data 2). Detailed description of all these known BMD loci can be found in the Supplementary Note.

The signal on 17p11.2 (lead SNP rs7501812, \( P = 1.4 \times 10^{-10} \)) was the only GWAS signal where association was stronger with TB-LM than with TBLH-BMD, mapping to a region not previously associated with neither trait. The region underlying this signal (Fig. 2) extends along an LD-block harbouring several genes including MYO15A, LRRCA8, MRR33B, C17orf39 [GID4], DRG2, RAI1, SREBF1, TOM1L2, ATPAF2, the latter seven all shown to be expressed in skeletal muscle. Detailed information for the genes residing at 17p11.2 is provided in Supplementary Table 1. All GWS SNPs in this region yielded nominally significant opposite effect for the coded allele in TBLH-BMD as compared to TB-LM, despite the positive correlation between these traits.

Although using different phenotypes than the ones used in our analysis, Genetic bivariate strategies in adult populations have approached the bone/lean mass relationship. A GWAS bivariate analysis of bone size and appendicular lean mass in Chinese and European individuals \(^{31}\) reported a potential association signal mapping to the GLYAT gene (11q12.1) arising from a low frequency (MAF < 0.05) variant not present in our meta-analysis. In addition, a linkage study reported a significant signal (LOD score = 4.86) mapping to the 15q13 locus and multiple suggestive signals (LOD score < 3) in 7p22, 7q21, 7q32, and 13q11.32. In our bivariate GWAS meta-analysis none of these regions (including GLYAT) contained significant associated SNPs (\( P < 9.0 \times 10^{-6} \) after multiple correction, 5537 SNPs tested). Additionally, we applied a bivariate analysis to the summary statistics of previously reported univariate GWAS meta-analyses of BMD and lean mass traits in adults .\(^ {20, 27} \) (Supplementary Data 3). We only found evidence of genetic variants exerting pleiotropic effects on both traits in the 11q13.2 locus. However, consistent with our findings in children, variants in the 17p11.2 locus (led by rs7501812) showed opposite association with lumbar spine BMD and TB-LM at the margin of reaching statistical significance (\( P < 0.07 \)).

Table 2 Trait SNP heritability and genetic correlation TBLH-BMD/TB-LM in pediatric populations

| Study       | N   | TBLH-BMD Heritability |              | TB-LM Heritability |              | Genetic correlation TBLH-BMD/ TB-LM |
|-------------|-----|----------------------|--------------|-------------------|--------------|-----------------------------------|
|             |     | \( h^2 \) | SE   | \( P \) | \( h^2 \) | SE   | \( P \) | \( h^2 \) | SE   | \( P \) |
| Generation R| 3028| 0.311   | 0.12 | 0.004  | 0.400 | 0.12 | 3 \( \times 10^{-14} \) | 0.299 | 0.21 | 0.126 |
| ALSPAC      | 4820| 0.437   | 0.07 | 6 \( \times 10^{-10} \) | 0.325 | 0.12 | 3 \( \times 10^{-6} \) | 0.323 | 0.12 | 0.016 |

Only unrelated individuals (e.g., no two individuals in the analysis were closer than third degree cousins in either of the two studies) were included in the analysis.

As most of the association signal in 17p11.2 arises from non-coding variants, we reviewed the possible regulatory annotation of these SNPs by using data from the ENCODE and ROADMAP EPIGENOMICS projects through the UCSC browser (Fig. 3). Shadowed areas correspond to CTCF (CCCTC-binding factor)-associated areas of chromatin interaction by looping. As these experiments are not available in musculoskeletal cells, we examined interactions in MCF-7 and K562 cells, for which the CTCF binding sites coincide with those predicted to exist in the musculoskeletal cells. Three selected interacting regions are shown: the first one brings together an intronic region of TOM1L2 and the 3′ region of SREBF1; the second one, contains two regions within TOM1L2; and the third one, intronic regions from TOM1L2 and ATPAF2. The topological associated domains (TADs) in the region are in line with these results, evidencing areas of complex chromatin structures (Supplementary Fig. 3).

The SNP with the strongest bivariate association with TBLH-BMD and TB-LM in this locus, rs7501812, is also a cis-eQTL variant found to regulate the expression of SREBF1, C17orf39 [GID4], TOM1L2 and ATPAF2 (FDR < 0.05) in whole blood, based on expression data from 5311 non-transformed peripheral blood samples and publicly available in the Blood eQTL browser . Nonetheless, SREBF1 expression, represented by two probes in the data set, showed the highest correlation with rs7501812. The G-allele, associated with higher TBLH-BMD (beta = 0.043 SD; \( P = 2.0 \times 10^{-3} \)) and lower TB-LM (beta = −0.056 SD; \( P = 5.5 \times 10^{-3} \)) in our study, decreased the expression of the two SREBF1 probes in whole blood. By interrogating the browser per gene, rather than per SNP, we generally found that alleles from the bivariate GWAS SNPs associated with higher TBLH-BMD.
and lower TB-LM in the region, are associated with decreased expression of \textit{SREBF1}, \textit{TOM1L2}, and \textit{C17orf39} [GID4] but increased expression of \textit{ATPAF2} in whole blood (Supplementary Data 4). We applied the Summary Mendelian Randomization (SMR) approach\textsuperscript{40} to prioritize genes underlying the GWS signal in the 17p12 region. Results from this approach combining Mendelian randomization and heterogeneity in dependent instruments (HEIDI) tests using the Blood eQTL browser data set\textsuperscript{39}, indicated that \textit{SREBF1} is the most likely gene driving the associations with TB-LM ($P_{\text{SMR}}=1.2\times10^{-4}$; $P_{\text{HEIDI}}=0.23$) and TBLH-BMD ($P_{\text{SMR}}=2.9\times10^{-3}$; $P_{\text{HEIDI}}=0.79$) (Supplementary Fig. 4). Expression data from GTEx also showed a highly negative significant association of rs7501812-G allele with the expression of \textit{SREBF1} in whole blood ($P=2.4\times10^{-14}$). In contrast, GTEx data in skeletal muscle showed a positive correlation for rs7501812-G with \textit{SREBF1} expression ($P=6.2\times10^{-4}$). Furthermore, rs4925114 ($P=3.5\times10^{-5}$) and rs11654081 ($P=6.0\times10^{-6}$) major alleles (LD ~ 0.75 with rs7501812-G), significantly upregulated the expression of \textit{SREBF1} in skeletal muscle (FDR<0.05) (Supplementary Fig. 5).

**Functional characterization of \textit{SREBF1}.** Leveraging mouse calvaria-derived cells, we assessed gene expression during osteoblastogenesis, which revealed that \textit{Rail}, \textit{Tomi1l2}, \textit{Atpf2}, and \textit{Srebf1} were expressed in this cell type without major changes from pre-osteoblast to its mature stage (Supplementary Fig. 6). In contrast, \textit{Lrrc48} was not expressed at all in the pre- or mature osteoblast, as determined by RNAseq profiles. We also assessed expression profiles during human mesenchymal stem cell (hMSC)
dystrophy, and this mechanism cannot be ruled out as key in genes despite being expressed in bone. To note, SREBP-1 has also a separate group of postmenopausal women (correlation (P<0.05) with donor phenotypes for either of the eight loci associated with both traits. Seven of the identified loci were associated with BMD in previous GWAS of adult and paediatric populations. The 17p11.2 region, marked by a long stretch of LD harbouring multiple genes in the region including RA11, LRRC48, MIR33B, C17orf39, DRG2, MYO15A, SREBF1, TOM1L2, and ATPAF2. Different lines of evidence, arising from in-silico and in-vivo functional follow-up suggest that TOM1L2 or SREBF1 are the main candidate genes underlying the bivariate association signal; confirming the existence of pleiotropic effects on BMD and lean mass arising from the 17p11.2 region.

SNP heritability estimated both by GCTA and LD score regression confirm that TBLH-BMD and TB-LM are moderately high heritable traits. In addition, we found that more than one-third of the tagged additive genetic effect is shared between TBLH-BMD and TB-LM. These estimates are lower than those derived from previous twin studies one of them estimating the heritability for TBLH-BMD to be between 80 and 90%, between 70 and 88% for TB-LM and with a genetic correlation (shared heritability) of 46%.

This could be due to an influence of the family based heritability estimates but also a consequence of the amount, and type, of markers being surveyed, limited to the genotyped SNPs and the underlying variants that they tag. Thus, if the selected markers are not sufficiently correlated (LD tagging) with the true genetic variants explaining the trait-variance, SNP heritability would certainly be less than true heritability.

The univariate approach for TBLH-BMD identified four GWS signals, all of them reported before in BMD GWAS at different skeletal sites. However, we did not find any GWS result for TB-LM in our meta-analysis in pediatric populations. This is in line with the scarce number of loci identified with lean mass variation in adults. To date, only five loci have been identified as associated with TB-LM in an study comprising >80,000 adults.

It is worth noting that lean mass in children has a higher proportion of organ tissue vs. skeletal mass as compared to adults. This could hamper the identification of loci influencing muscle mass and claims for new larger efforts in children. Yet, implementing a bivariate meta-analytical approach we identified eight associated loci, three of which were nominally associated (P<0.05) with both TBLH-BMD and TB-LM. Therefore, the bivariate analysis is an alternative approach to diminish such power limitations. Our bivariate approach showed particularly high statistical power to identify variants, in which the coded allele exerted an opposite effect on TBLH-BMD as compared to TB-LM. Our meta-analysis did not replicate potential pleiotropic signals previously reported in adults. Nevertheless, the bivariate analysis of summary statistics of published BMD and lean mass effects in adults we did, identified variants with strong evidence for pleiotropic effects in the 11q13.2 locus. The lack of replication in the other regions may well be a consequence of differential genetic effects during the life course, e.g., the

| Locus | SNP | CHR | Position | A1 | A2 | EAF | Beta TBLH-BMD | P TBLH-BMD | Beta TBLM | P TBLM | P bivariate |
|-------|-----|-----|----------|----|----|-----|--------------|-----------|-----------|---------|-------------|
| 1p36.12 | rs6684375 | 1 | 22579021 | T | C | 0.19 | 0.0842 | 2.42E-06 | -0.0273 | 0.14 | 4.69E-09 |
| 2q24.3 | rs6726821 | 2 | 16628360 | G | T | 0.45 | -0.0769 | 2.95E-08 | -0.0104 | 0.48 | 3.56E-08 |
| 4q22.1 | rs7672749 | 4 | 89017308 | A | G | 0.09 | 0.0822 | 4.35E-04 | -0.0591 | 0.01 | 3.29E-08 |
| 7q31.31 | rs917727 | 7 | 120805815 | T | C | 0.29 | 0.1289 | 5.29E-17 | -0.0042 | 0.77 | 3.07E-20 |
| 11q33.1 | rs12284933 | 11 | 68076065 | A | G | 0.24 | -0.0969 | 1.85E-09 | -0.0699 | 1.49E-05 | 2.19E-09 |
| 13q44.1 | rs9525638 | 13 | 42026577 | C | T | 0.42 | 0.0756 | 6.42E-08 | -0.0006 | 0.97 | 1.35E-08 |
| 14q2.2 | rs754388 | 14 | 92185163 | G | C | 0.18 | -0.0862 | 3.26E-06 | 0.0210 | 0.24 | 3.36E-08 |
| 17p11.2 | rs7501812 | 17 | 17691632 | G | A | 0.41 | 0.0431 | 0.0002 | -0.0563 | 5.35E-05 | 1.44E-10 |

Estimates were derived from 10,414 children participants of four different pediatric studies worldwide. Beta coefficients and allele frequencies (EAF) are reported for the A1 allele.
accumulation of environmental factors attenuating effects in the elderly. Overall, the role of protein products of genes such as LR5P, TNSF11, WNT16, MEPE, and WNT4-mapping to the identified loci-in bone metabolism is long-established. However, none of the genes in the identified loci has been yet implicated in muscle metabolism and opens up new opportunities to study pleiotropic effects and the interplay between bone and muscle.

We further focused on the 17p11.2 locus, which is novel in the spectrum of musculoskeletal traits GWAS, and which is the only associated region where the signal was stronger for TB-LM than for TBLH-BMD. Despite the positive correlation between TB-LM and TBLH-BMD, variants in this region exerted opposite effects on these traits, probably facilitating their discovery in a bivariate analysis powerful setting. Moreover, this region was recently described as a QTL for BMD phenotypes in rats. The authors attributed the association signal to the lead SNP of the bivariate association (rs7501812). It is the strongest one. Also, SMR implicates SREBF1 as the most likely candidate gene underlying the 17p11.2 signal. Yet, TOM1L2 or RAI1 probes were not included in the analysis as they lacked a GWS cis-eQTL. Such stringency in the SMR inclusion criteria results from the basic assumption of the instrument used in the Mendelian randomisation approach, i.e., the SNP should be strongly associated with exposure. Further, the e-QTL effects identified in whole blood may not constitute good proxies for eQTLs exerting tissue-specific effects. However, there is no eQTL data available from other tissues (particularly bone and muscle) with comparable statistical power to the one provided by the large-scale eQTL study in whole blood used here.

In addition, rs4925114 and rs11654081, both in high LD with rs7501812, show the strongest correlation with the expression of SREBF1 in skeletal muscle. Interestingly, the same allele in either of these three SNPs correlates with lower SREBF1 expression in whole blood and higher SREBF1 expression in skeletal muscle. Unfortunately, the GTEx eQTL database has no data available for bone related cells. Our ex-vivo models in murine osteoblasts established that RAI1, Atpaf2, Srebfl, and Tom1l2 were expressed in bone. With the exception of Srebfl, knockout mice for these genes all present with a skeletal phenotype. While Srebfl KO models have been reported in the literature, we have no certainty that bone phenotyping was performed on these mice.

We also confirmed the expression of SREBF1 in human MSC-derived osteoblasts. Further, analysis of expression profiles in muscle biopsies from post-menopausal women showed a correlation between the expression of RAI1 and TOM1L2 and muscle thickness, and noteworthy, a negative correlation between SREBF1 expression and bone density parameters. Altogether, SREBF1 and TOM1L2 are the strongest candidates underlying the genetic association signal. The role of active SREBP-1, SREBF1 product, exerting opposing effects on osteoblast and myoblast differentiation has been well documented. However, despite the fact that the potential role of TOM1L2 in the musculoskeletal system remains to be elucidated, miRNAs predicted to regulate its expression have been shown to be involved in osteogenic

Fig. 2 Regional association plot for the bivariate meta-analysis of TBLH-BMD and TB-LM displaying the 17p11.2 locus. Genetic coordinates are as per Hapmap phase II-CEU
Overall, there is compelling evidence that SREBF1 exerts opposite effects on the differentiation of myocytes and osteoblasts. These opposite effects are paralleled in the SNPs associated with TBLH-BMD/TB-LM in our bivariate approach, and therefore, it is likely that SREBF1 is underlying the association signal. The G-allele from rs7501812, shown to downregulate the expression of SREBF1 in whole blood and upregulate it in skeletal muscle, is actually associated with higher TBLH-BMD and lower TB-LM (Fig. 5). Whereas, the expression of SREBF1 in muscle biopsies correlated with lower BMD parameters in a direction opposite to what is predicted by the SREBP-1 function in bone and muscle metabolism, the number of expression profiles included in analysis is small; also, this unexpected correlation could arise from tissue-specific effects. The eQTL analysis in GTEx showed that SREBF1 is ubiquitously expressed and suggests that SNPs in 17p11.2 exert apparent opposite effect in blood and muscle. In addition, there are cellular processes that cannot be captured by our analyses and can be the origin of some inconsistencies seen at the association level. For instance, SREBF1 must be proteolytically released from a transmembrane site in the Golgi complex to become active; then this activated fragment needs to be transported to the nucleus in order to exert its effect.

In summary, in this first large-scale bivariate GWAS meta-analysis of TBLH-BMD and TB-LM in pediatric cohorts, we identified eight GW loci (in/near WNT4, GALNT3, MEPE, CEPD1/WNT16, TNSF11, RIN3, TOM1L2/SREBF1, and PPP6R3/LRP5). Although for most of these loci the association with TBLH-BMD is the dominant association responsible for their significance, we report in the 17p11.2 locus, a stronger association with TB-LM compared to TBLH-BMD. SREBF1 arises as a compelling candidate to underlie this association, as this transcription factor is known to play an opposite role in osteoblast and myoblast differentiation; in line with the finding derived through our bivariate approach. Nevertheless, the level of evidence gathered does not allow us to discard fully that other genes in the region, particularly TOM1L2, may be implicated in these processes, and additional functional assessments are required. Alternatively, this study demonstrates that a considerable increase in power can be achieved when using a bivariate methodology as compared to testing each component of a multivariate phenotype individually, particularly in those cases where the effect of the associated variant goes in opposite directions in the involved differentiation and skeletal muscle development. In addition, the knockout mice of this gene observed with kyphosis. In summary, in this first large-scale bivariate GWAS meta-analysis of TBLH-BMD and TB-LM in pediatric cohorts, we identified eight GW loci (in/near WNT4, GALNT3, MEPE, CEPD1/WNT16, TNSF11, RIN3, TOM1L2/SREBF1, and PPP6R3/LRP5). Although for most of these loci the association with TBLH-BMD is the dominant association responsible for their significance, we report in the 17p11.2 locus, a stronger association with TB-LM compared to TBLH-BMD. SREBF1 arises as a compelling candidate to underlie this association, as this transcription factor is known to play an opposite role in osteoblast and myoblast differentiation; in line with the finding derived through our bivariate approach. Nevertheless, the level of evidence gathered does not allow us to discard fully that other genes in the region, particularly TOM1L2, may be implicated in these processes, and additional functional assessments are required. Alternatively, this study demonstrates that a considerable increase in power can be achieved when using a bivariate methodology as compared to testing each component of a multivariate phenotype individually, particularly in those cases where the effect of the associated variant goes in opposite directions in the involved differentiation and skeletal muscle development. In addition, the knockout mice of this gene observed with kyphosis.

In summary, in this first large-scale bivariate GWAS meta-analysis of TBLH-BMD and TB-LM in pediatric cohorts, we identified eight GW loci (in/near WNT4, GALNT3, MEPE, CEPD1/WNT16, TNSF11, RIN3, TOM1L2/SREBF1, and PPP6R3/LRP5). Although for most of these loci the association with TBLH-BMD is the dominant association responsible for their significance, we report in the 17p11.2 locus, a stronger association with TB-LM compared to TBLH-BMD. SREBF1 arises as a compelling candidate to underlie this association, as this transcription factor is known to play an opposite role in osteoblast and myoblast differentiation; in line with the finding derived through our bivariate approach. Nevertheless, the level of evidence gathered does not allow us to discard fully that other genes in the region, particularly TOM1L2, may be implicated in these processes, and additional functional assessments are required. Alternatively, this study demonstrates that a considerable increase in power can be achieved when using a bivariate methodology as compared to testing each component of a multivariate phenotype individually, particularly in those cases where the effect of the associated variant goes in opposite directions in the involved differentiation and skeletal muscle development.
phenotypes. Moreover, we show that this methodology may help discover previously unknown mechanisms of pleiotropic effects of genetic variants.

Methods

Study populations. TBLH-BMD and TB-LM were measured in four pediatric population-based studies: the Generation R Study, the Avon Longitudinal Study of Parents and their Children (ALSPAC), the Bone Mineral Density in Childhood Study (BMDCS), and the Copenhagen Prospective Studies on Asthma in Childhood (COPSAC) cohort (Supplementary Note). All participants underwent TBDXA scans; measurements were conducted by well-trained research assistants and daily quality control assurance was performed. Before the scan procedure, participants were asked to take off their shoes, heavy clothes, and metal accessories. As recommended by the International Society for Clinical Densitometry, TBLH-BMD was the measurement used in the analysis by all four studies. TB-LM and total body fat mass (TB-FM) were derived from the same scans. In addition, all individuals included in this study had genome-wide array data input to the HapMap Phase II reference panel (build 36 release 22) assessed as best guess genotypes, all markers with MAF < 0.05 were excluded from the analysis. The study was approved by the ALSPAC Law and Ethics committee (ALSPAC), The Ethics Committee for Copenhagen and the Danish Data Protection Agency (COPSAC), the Medical and Ethical Review Committee of the Eramus University Medical Center (Generation R), and the institutional review boards of Children’s Hospital of Los Angeles (Los Angeles, CA), Cincinnati Children’s Hospital Medical Center (Cincinnati, OH), Crehighton University (Omaha, NE), Children’s Hospital of Philadelphia (Philadelphia, PA), and Columbia University (New York, NY) (BMDCS). Written informed consent was provided by all parents or custodians of the participants. A detailed description of each of these study populations is provided in Supplementary Note.

Statistical methods. TBLH-BMD and TB-LM measurements were adjusted by age, gender, height, fat percent (TB-FM/weight), and study specific covariates (genetic principal components and measurement center, when applicable). Standardized residuals were then generated. Phenotypic correlation (r) between these variables was evaluated by a Pearson correlation test in each study independently.

SNP heritability and genetic correlation. To characterize the extent to which common genetic variants determine pediatric BMD and lean mass, and the shared genetic etiology of these traits, we applied two recently proposed approaches for estimating SNP heritability and genetic correlation based on genome-wide sharing genetic etiology of these traits, we applied two recently proposed approaches for common genetic variants determine pediatric BMD and lean mass, and the shared standardized residuals were then generated. Phenotypic correlation (r) between these variables was evaluated by a Pearson correlation test in each study independently.

Univariate and bivariate GWAS. Analysis of the two quantitative traits was performed in each study individually based on best guess data using the "qtcommand" available in PLINK. This routine enables the user to obtain, in addition to the effect estimates, the means and standard deviations of both TBLH-BMD and TB-LM standardized residuals stratified by genotype. Meta-analysis of 2,276,811 SNPs, present in at least two of the studies, was performed using an inverse variance weighing. In the meta-analysis approach, we used a newly proposed method for bivariate meta-analysis, which is based on calculating the within-study covariances of the outcome specific estimates. The algorithm uses a general approach and has been proposed for both, discrete and continuous outcomes, but in the case of continuous outcomes, such as the one encountered here, a direct extension of the method proposed by Wei and Higgins is applied. We used the additive model of inheritance (per-allele mean difference of the quantitative phenotype) for both outcomes and the method of moments for estimation with the mvmeta command in Stata. Source code and details are given in the webpage of the Department of Computer Science and Biomedical Informatics of the University of Thessaly, Greece. The overall bivariate association test was obtained by a standard Wald-type statistic (chi-square on 2 d.f.) that tests the null hypothesis that a particular SNP is not associated with either one of the outcomes. Genome-wide significance (GWS) was defined as a P < 5×10⁻⁵. Manhattan plots were generated in Easystrap.

Adult bivariate analysis based on summary statistics of GWAS in musculoskeletal traits. We assessed the evidence of association of the GWS markers identified here (149 SNPs, Supplementary Data 3), in a bivariate analysis of previous powered GWS meta-analyses of lean mass (total body (N = 38,000)) and BMD (lumbar spine and femoral neck (N = 32,000)) performed in adult and elderly population. We chose to use a recently introduced method that performs bivariate GWS allowing for mixed directions of effect known as empirical-weighted linear-combined test statistics (eLC), implemented in a C++ ELX package and publicly available. Briefly, eLC directly combines correlated Z test statistics (calculated as Z values derived from the univariate GWS meta-analyses) with a weighted sum of univariate test statistics to empirically maximize the overall association signal and also to account for the phenotypical correlations the traits. In this case, we ran two analyses one incorporating the data from femoral neck BMD and total body lean mass meta-analyses, and another with data from lumbar spine and total body lean mass meta-analyses. Markers with nominal effect in both BMD and lean mass traits and with bivariate P <= 0.05 were deemed to exert pleiotropic effects in adults.

Gene Annotation and eQTL Analyses. We identified all variants from the 1000 Genomes Project Phase 1 reference panel in strong linkage disequilibrium (LD) (CEU r² > 0.8) with the lead SNPs in the novel identified susceptibility loci, by bivariate analysis, using HaploReg for functional annotation. First, these proxy SNPs were scrutinized in the GWAS catalogue to ascertain possible association with other traits. Second, we identified non-synonymous variants that were interrogated for likely downstream functional consequences using SIFT and PolyPhen-2 databases. Third, we interrogated two public databases, i.e., miRNASNP and PolymiRTs56 for presence of any of the associated SNPs in miRNA genes or miRNA-binding sites within 3′UTRs of target genes. Other information about miRNAs, such as miRNA sequence, host gene, conservation and putative targets, were obtained from miBase (release 21) and TargetScan v7.1 websities. Fourth, variants were also evaluated for overlap with regions of predicted regulatory function generated by the Encyclopedia of DNA Elements (ENCODE) Project and the ROADMAP epigenomic project including: regions of enhancer activity, DNA I hypersensitivity, local histone modifications or proteins bound to these regulatory sites in cell lines of potential interest as HSM (Skeletal Muscle Myoblast), HSUMtude (Skeletal Muscle myotubes differentiated from the HSM), SKMC (skeletal muscle cells) and when present in the different assays, Osteobl (Osteoblasts). In addition, evidence of chromatin interactions was further explored by Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET) data from the ENCODE project in both MCF-7 and K562 oncologic cell lines. The UCSC Genome Browser was used for visualizing ENCODE data tracks indicative of regulatory function. Finally, topologically associated domains (TADs) for K562
cells containing the associated SNPs were retrieved from a recent Hi-C interactions browser.  

The eQTL analysis was completed using recently published eQTL data sets from a meta-analysis of the transcriptional profiles from the peripheral blood cells of 5311 Europeans and the GTEx database (GTEx v6p) based on 1641 samples across 43 tissues from 175 individuals.

**Summary data-based mendelian randomization.** We used the recently developed SMR approach in order to identify the most likely gene in the lead SNP region, which has a complicated LD pattern. In short, the SMR approach uses the genetic variant as the instrumental variable to test for the causative effect of the gene expression (the exposure) on the phenotype of interest. This analysis is divided in two steps. A transcriptome association study, where the effect of gene expression on the phenotype of interest is expressed as a function of the GWAS summary statistics and the GWS eQTLs reported by Westra et al. in whole blood. Then, a test for heterogeneity in dependent instruments (HEIDI) is applied; in principle, if the same causal underlying variant affects gene expression and the trait then, all SNPs in high LD with the causal variant will present the same SMR calculated effect. Therefore, testing for heterogeneity in the cis-eQTL region will be equivalent to testing if there is only one causal variant. We used Bonferroni correction to account for multiple testing on 394 probes with GWAS evidence for cis-eQTL effects in chromosome 17; this resulted in a significance threshold of $1.27 	imes 10^{-10}$ (0.05/394). We used a conservative HEIDI threshold $\text{P}=0.05$, to exclude signals with high heterogeneity for different traits. Plotting was performed in R including only the most significant probe per gene in the SMR analysis and using the lean mean GWAS results.

**Functional analyses of SREBF1.** Aiming to corroborate the functional or regulatory mechanism underlying the association between SREBF1 and LM and BMD, we used three different lines of evidence. The expression of SREBF1 was investigated in bone and immune cells differentiating to osteoblasts. In addition, the expression of SREBF1 and its neighbouring genes TOM1L2 and ATPF2 were examined in muscle biopsies from postmenopausal women. Methods specific to each analysis are described below.

**Murine pre-osteoblasts.** All procedures and use of mice for the neonatal osteoblast expression studies were approved by the Jackson Laboratory Animal Care and Use Committee in accordance with NIH guidelines for the care and use of laboratory animals. Pre-osteoblast-like cells were isolated from neonatal calvaria from C57BL/6j mice expressing cyan florescent protein (CFP) and RNA was collected at nine time points post differentiation, every other day for 18 days, starting 2 days after the cells were first exposed to an osteoblast differentiation cocktail. miRNA arrays for each time point were generated by Next Generation High throughput RNA sequencing (RNAseq), using an Illumina HiSeq 2000.

**Human mesenchymal stem cells.** Human bone marrow derived mesenchymal stem cells (hMSC), Lonza Group Ltd., Basel, Switzerland) were seeded in 12-well plates (5x10^6 cells per cm^2) and differentiated into osteoblasts (using α-Mem pH 7.5, 10% heat inactivated foetal calf serum (FCS), 100 μM Dexamethasone and 10 mM β-glycerophosphate). As mentioned in the dataset provided by the company, cells were authenticated by FACS analyses for the presence of surface markers CD105, CD166, CD29, and CD44 and the absence of CD14, CD34 and CD45. In addition, osteogenic, adipogenic and chondrogenic differentiation was shown by alizarin red S staining, oil-red-O staining and collagen II staining, respectively. The human MSCs were tested negative for mycoplasma, both by the laboratory analyses, and those who participated after full clinical examination including the DEXA sampling and procedures were according to the Law of Biobanking in Norway. All studies, validation, and recommendation were obtained by the Norwegian Regional Ethics Committee (REK no-2010/3539 and 2008/233 REC ser-nt D), and all sampling and procedures were according to the Law of Biobanking in Norway. All women who volunteered received a full clinical examination including the DEXA and laboratory analyses, and those who participated after fulfilling the inclusion criteria took pride in contributing to the aim of the study. Written informed consent was provided by all women.

In muscle, gene expression profiles were generated from human thigh muscle biopsies donated by healthy (T-score >1) postmenopausal women ($n=18$), from whom anthropometric measurements and DXA data were also available. Expression profiling was performed using an Affymetrix HG U133 2.0 plus array. The Affymetrix Cell Genome Arrays were imported into the Cell Signaling Network (Partek Inc., St. Louis, MO, USA) and normalized using the RMA (Robust Multichip Average) algorithm. Gene expression patterns were further adjusted for batch effects and eQTLs.

**References**

1. Robling, A. G. Is bone’s response to mechanical signals dominated by muscle forces? Med. Sci. Sports. Exerc. 41, 2044–9 (2009).

2. Li, S. L., Wagner, R., Holm, K., Lehotsky, J. & Zinaman, M. J. Relationship of bone density and bone loss in postmenopausal women with physical activity: a large, long-term prospective study. J. Bone. Miner. Res. 13, 841–5 (1998).

3. Karsent, G. & Olson, E. N. Bone and muscle: interactions beyond mechanical. Am. J. Hum. Biol. 18, 23–38 (2006).

4. Cianferotti, L. & Brandi, M. L. Muscle-bone interactions: basic and clinical aspects. J. Bone. Miner. Res. 137, 23–38 (2002).

5. Karsenti, D. & Kiel, D. P. Genetics of the musculoskeletal system: a pleiotropic paradigm of inter-organ communication. J. Cell. 164, 248–56 (2016).

6. Brotto, M. & Bonewald, L. Bone and muscle: interactions beyond mechanical. Bone 80, 109–14 (2015).

7. Canferrotti, L. & Brandi, M. L. Muscle-bone interactions: basic and clinical aspects. Endocrine. 45, 165–177 (2014).

8. Karasik, D. & Kiel, D. P. Genetics of the musculoskeletal system: a pleiotropic approach. J. Bone. Miner. Res. 23, 788–802 (2008).

9. Chen, Z. et al. Dual-energy X-ray absorptiometry is a valid tool for assessing skeletal muscle mass in older women. J. Nutr. 137, 2775–80 (2007).

10. Bridge, P. et al. Prediction of appendicular skeletal and fat mass in children: excellent concordance of dual-energy X-ray absorptiometry and magnetic resonance imaging. J. Pediatr. Endocrinol. Metab. 22, 795–804 (2009).

11. Ralston, S. H. & Uitterlinden, A. G. Genetics of osteoporosis. Endocr. Rev. 31, 629–662 (2010).

12. Silvonen, K., Magnusson, P. K. E., Tynelius, P., Kaprio, J. & Rasmussen, F. Heritability of body size and muscle strength in young adulthood: a study of one million Swedish men. Genet. Epidemiol. 32, 341–349 (2008).

13. Arden, N. K. & Spector, T. D. Genetic influences on muscle strength, lean body mass, and bone mineral density: a twin study. J. Bone. Miner. Res. 12, 2076–81 (1997).

**Data availability.** Summary statistics of all meta-analyses are available at the GEnetic Factors for OSteoporosis Consortium (GEFOS) website (http://www.gefos.org/?q=content/data-release-2017). Expression data from bone biopsies is available at the European Bioinformatics Institute (EMBL-EBI) ArrayExpress repository, ID: E-MEXP-1618. All other data supporting the findings of this study are available from the corresponding author upon request.

Received: 3 December 2016 Accepted: 1 June 2017
NATURE COMMUNICATIONS

14. Styrkarsdottir, U. et al. Sequence variants in the PTCH1 gene associate with spine bone mineral density and osteoporotic fractures. Nat. Commun. 7, 10129 (2016).
15. Zheng, H. F. et al. Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. Nature. 526, 112–7 (2015).
16. Rivadeneira, F. et al. Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. Nat. Genet. 41, 1199–206 (2009).
17. Medina-Gomez, C. et al. Meta-analysis of genome-wide scans for total body BMD in children and adults reveals allelic heterogeneity and age-specific effects at the WNT16 locus. PLoS Genet. 8, e1002718 (2012).
18. Richards, J. B. et al. Bone mineral density, osteoporosis, and osteoporoctic fractures: a genome-wide association study. Lancet 371, 1505–12 (2008).
19. Kemp, J. P. et al. Phenotypic dissection of bone mineral density reveals skeleton-specificity and facilitates the identification of novel loci in the genetic regulation of bone mass attainment. PLoS Genet. 10, e1004423 (2014).
20. Estrada, K. et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. Nat. Genet. 44, 491–501 (2012).
21. Kohler, C. L. D. et al. Meta-analysis of genome-wide studies identifies WNT16 and ESRR1 SNPs associated with bone mineral density in premenopausal women. J. Bone Miner. Res. 28, 547–58 (2013).
22. Zheng, H. F. et al. WNT16 influences bone mineral density, cortical bone thickness, bone strength, and osteoporotic fracture risk. PLoS Genet. 8, e1002745 (2012).
23. Duncan, E. L. et al. Genome-wide association study using extreme truncation selection identifies novel genes affecting bone mineral density and fracture risk. PLoS Genet. 7, e1001372 (2011).
24. Styrkarsdottir, U. et al. Nonsense mutation in the LGR4 gene is associated with hand bone density. PLoS Genet. 9, e1003520 (2013).
25. Zhang, L. et al. Multistage genome-wide association meta-analyses identify 14 loci associated with risk of fracture. Nat. Genet. 45, 983–95 (2013).
26. Alam, I. et al. High-resolution genome screen for bone mineral density in heterogeneous stock rat. J. Bone Miner. Res. 29, 1619–26 (2014).
27. Kemp, J. P. et al. Phenotypic dissection of bone mineral density reveals skeleton-specificity and facilitates the identification of novel loci in the genetic regulation of bone mass attainment. PLoS Genet. 10, e1004423 (2014).
28. Wang, X., Sato, R., Brown, M. S., Hua, X. & Goldstein, J. L. SREBP-1, a membrane-bound transcription factor released by sterol-regulated proteolysis. Cell 77, 53–62 (1994).
29. Esposito, M. A. & Villarroya-Beltri, C. SREBF1, a master regulator of lipoprotein metabolism. J. Lipid Res. 54, 121–35 (2013).
30. Westra, H. J. et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat. Genet. 45, 1238–U195 (2013).
31. Xu, J., Li, L., Yun, H. F. & Han, Y. S. MR138 promotes smooth muscle cells proliferation and migration in db/db mice through down-regulation of SIRT1. Biochem. Biophys. Res. Commun. 463, 1159–1164 (2015).
32. Koutsououdi, A., Mastroyiannopoulos, N. P., Furling, D., Uney, J. B. & Physiotou, L. Expression of miR-1, miR-133a, miR-133b and miR-206 increases during development of human skeletal muscle. BMC Dev. Biol. 11, 34 (2011).
33. Gorski, J. P. et al. Inhibition of proprotein convertase SKI-1 blocks transcription of key extracellular matrix genes regulating osteoblastic mineralization. J. Biol. Chem. 286, 1836–49 (2011).
ALSPAC Study: We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses. GWAS data was generated by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/2/113/2) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors, and JPK and DME will serve as guarantors for the contents of this paper. JPK is funded by a Wellcome Trust 4-year PhD studentship in molecular, genetic, and life course epidemiology (WT083431MA). This work is supported by a Medical Research Council program grant (MC_UU_12013/4 to D.M.E). D.M.E. is supported by an Australian Research Council Future Fellowship (FT13010709).

The Generation R Study: We gratefully acknowledge the contribution of children and parents, general practitioners, hospitals, midwives and pharmacies in Rotterdam. The generation and management of GWAS genotype data for the Generation R study was done at the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, The Netherlands. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera and Maryke Peters for their help in creating, managing and QC of the GWAS database. The musculoskeletal research of the Generation R study is partly supported by the European Commission grant HEALTH-F2-2008-201865-GEFOS. The general design of Generation R Study is made possible by financial support from the Erasmus Medical Center, Rotterdam, the Erasmus University Rotterdam, the Netherlands Organization for Health Research and Development (ZonMw), the Netherlands Organisation for Scientific Research (NWO), the Ministry of Health, Welfare and Sport. Additionally, the Netherlands Organization for Health Research and Development supported authors of this manuscript (ZonMw 907.00303, ZonMw 916.10159, ZonMw VIDI 016.136.361 to V.W.V.J., and ZonMw VIDI 016.136.367 to F.R.)

COPSAC: We express our gratitude to the participants of the COPSAC2000, COP-SAC2010 and COPSAC-REGISTRY cohort study for all their support and commitment. We also acknowledge and appreciate the unique efforts of the COPSAC research team. SAC2010 and COPSAC-REGISTRY cohort study for all their support and commitment. COPSAC: We express our gratitude to the participants of the COPSAC2000, COP-SAC2010 and COPSAC-REGISTRY cohort study for all their support and commitment. The COPSAC team is supported by ZonMw 907.00303, ZonMw 916.10159, ZonMw VIDI 016.136.361 to V.W.V.J., and ZonMw VIDI 016.136.367 to F.R. (MC_UU_12013/4 to D.M.E). D.M.E. is supported by an Australian Research Council Future Fellowship (FT13010709).

U.S. CHARGE Consortium focused on a variety of disease phenotypes. Building on GWAS for NHLBI-diseases: the CHARGE Consortium: Results from the meta-analysis of GWAS for lean body mass and Aging Research on Genomic Epidemiology (CHARGE, http://depts.washington.edu/chargeco/wiki/Main_Page), which is an international consortium of cohort studies focused on a variety of disease phenotypes. Building on GWAS for NHLBI-diseases: the U.S. CHARGE Consortium' was funded by the NIH (the American Recovery and Reinvestment Act of 2009, ARRA, S0C2HH102419). DPK was supported by a grant from the National institute on Arthritis, Musculoskeletal and Skin Diseases (R01 AR41398).

Author contributions
Conceived and designed the experiments: C.M.G., J.P.K., J.H.T., D.M.E., and F.R. Performed the experiments: C.M.G., J.P.K., E.K., A.C., C.G.B., T.S.A., D.H.M.H., M.G., S.D., D.P.K., C.A.-B., S.R., J.v.d.P. Analyzed the data: C.M.G., J.P.K., N.L.D., E.K., A.C., C.G.B., E.E., M.G., S.D., D.P.K., B.C.J.v.E., C.A.-B., S.R., J.v.d.P., J.H.T., P.G.B., D.M.E., F.R., Y.-H.H. Contributed reagents/materials/analysis tools: N.L.D., C.G.B., M.G., E.E., B.C.J.v.E., C.A.-B., S.R., K.M.G., T.R., J.v.d.P., P.G.B. Wrote first draft of the paper: C.M.G., F.R. Contributed and approved final version of the paper: C.M.G., J.P.K., N.L.D., E.K., A.C., B.S.Z., K.B., C.G.B., T.S.A., H.B., E.E., D.H.M.H., L.F.B., J.P.G., M.G., Y.-H.H., S.D., D.G.D., M.T.M., D.P.K., B.C.J.v.E., C.A.-B., S.R., K.M.G., T.R., D.K., J.v.d.P., V.W.V.J., A.G.U., J.H.T., S.F.A.G., D.M.E., P.G.B., F.R.

Additional information
Supplementary Information accompanies this paper at doi:10.1038/s41467-017-00108-3.

Competing interests: E.K. is now working as medical specialist at Novo Nordisk A/S. All other authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017