This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Published in final edited form as:

**Title:** Defining the role of common variation in the genomic and biological architecture of adult human height.

**Authors:** Wood AR et al.


In the absence of a copyright statement, users should assume that standard copyright protection applies, unless the article contains an explicit statement to the contrary. In case of doubt, contact the journal publisher to verify the copyright status of an article.
Defining the role of common variation in the genomic and biological architecture of adult human height

A full list of authors and affiliations appears at the end of the article.

Abstract

Using genome-wide data from 253,288 individuals, we identified 697 variants at genome-wide significance that together explain one-fifth of heritability for adult height. By testing different numbers of variants in independent studies, we show that the most strongly associated ~2,000, ~3,700 and ~9,500 SNPs explained ~21%, ~24% and ~29% of phenotypic variance. Furthermore, all common variants together captured the majority (60%) of heritability. The 697 variants clustered in 423 loci enriched for genes, pathways, and tissue-types known to be involved in growth and together implicated genes and pathways not highlighted in earlier efforts, such as signaling by fibroblast growth factors, WNT/beta-catenin, and chondroitin sulfate-related genes. We identified several genes and pathways not previously connected with human skeletal growth, including mTOR, osteoglycin and binding of hyaluronic acid. Our results indicate a genetic architecture for human height that is characterized by a very large but finite number (thousands) of causal variants.

Height is a classical polygenic trait that has provided general insights into the genetic architecture of common human traits and diseases, and into the prospects and challenges of different methods used to identify genetic risk factors. Studies consistently estimate that the additive genetic contribution to normal variation in adult height (“narrow sense heritability”) is approximately 80% \(^1\)\(^-\)\(^3\). Previous analysis of genome-wide association studies (GWAS) of adult height showed that common variants together account for 50% of this heritable contribution to height variation\(^4\)\(^,\)\(^5\). The most recent GWAS of adult height identified 180 loci, which together highlighted many genes relevant to human skeletal growth that had not been implicated in previous studies\(^6\). Common variants in these loci, however, only accounted for 10% of the phenotypic variation (~12% of heritability). Here, we report results from a GWAS meta-analysis of adult height in 253,288 individuals of European ancestry. We show that additive contributions of fewer than 10,000 SNPs (at \(P<5\times10^{-3}\)) can account for 36% of the heritability of adult height. Variants reaching genome-wide significance (\(P<5\times10^{-8}\)) in this larger study (697 SNPs) clustered in loci, were substantially enriched for regulatory variants, and implicated multiple known and previously unknown genes and pathways relevant to growth. More broadly, our results provide evidence that increasing GWAS sample sizes to the order of 100,000s, now plausible for many common
traits, will likely continue to identify the variants and loci that close the “missing heritability” gap, whilst improving knowledge of the biology of those traits.

Results

The overall analysis strategy is illustrated in Supplementary Figure 1. We first performed a GWAS meta-analysis of adult height using summary statistics from 79 studies consisting of 253,288 individuals of European ancestry (Online Methods). We identified 697 SNPs that reached genome-wide significance ($P < 5 \times 10^{-8}$) using an approximate conditional and joint multiple-SNP (COJO) analysis in GCTA8 (Online Methods) which takes linkage disequilibrium (LD) between SNPs into account (Supplementary Table 1; Supplementary Figs. 2–3). The 697 SNPs clustered in 423 loci, with a locus defined as one or multiple jointly associated SNPs located within ±1Mb of each other. Most of these 697 SNPs are uncorrelated although those in close physical proximity (e.g. < 1Mb) may be in partial LD (see Supplementary Table 1 for LD between adjacent pairs of the 697 SNPs). The clustering of signals was non-random (empirical enrichment of 1.4 fold, $P < 1 \times 10^{-4}$) with 90, 26 and 31 loci containing 2, 3 and ≥4 signals respectively, (Supplementary Note and Supplementary Tables 1 and 2). We observed strong evidence of clustering of association signals within loci across a range of locus sizes, from 100kb to 1.25Mb, but the clustering was almost entirely driven by variants within 250kb of index SNPs (Supplementary Note and Supplementary Table 2). As shown in Figure 1 and Supplementary Figure 4, in some loci, multiple signals cluster tightly around a single gene, whereas in other cases, the clustering of associated variants is likely due to multiple different height-related genes being in close proximity.

Of the 697 SNPs, 403 were represented on the Metabochip array9. Using data from 80,067 individuals genotyped on the Metabochip array from 37 independent studies, we observed very strong evidence of concordance of effect sizes between the Metabochip and GWAS samples ($P = 1.9 \times 10^{-160}$); and >99% of variants were directionally consistent between Metabochip and GWAS (Online Methods, Supplementary Note, and Supplementary Table 3).

We observed a large genome-wide ‘inflation’ factor of the test statistic for association even after we corrected each study’s test statistics by its individual inflation factor (single λGC = 1.94). At least two phenomena could have contributed to this observation. First, as described previously10, highly polygenic models of inheritance are expected to increase the genomic inflation factor to levels comparable to what we observe. Second, height is particularly susceptible to confounding by population ancestry (stratification), which can also lead to inflation of the test statistics. We addressed these possibilities by comparing our results with those obtained using more stringent corrections for stratification (linear mixed models), and with results obtained in subsets of studies in which a purely family-based analysis was feasible, and by performing a within-family prediction analysis which partitioned the variance in the genetic predictor into the contributions of true associations and population stratification.

Our linear mixed model (LMM) analyses, performed in a subset of 15 individual studies comprising 59,380 individuals, provided strong evidence that the inflated statistics were
driven predominantly by the highly polygenic nature of the trait. This approach utilizes a genomic relationship matrix (GRM) calculated through genome-wide SNP data to correct for distant relatedness between all pairs of individuals within a study. This resulted in a single $\lambda_{GC}$ of 1.20. This value was entirely consistent with the single $\lambda_{GC}$ of 1.20 obtained from the standard GWAS analysis of the same individuals and a single $\lambda_{GC}$ of 1.94 obtained from the full 253,288 individuals (Supplementary Table 4). Because this approach may be overly conservative for a strongly genetic and highly polygenic trait, each study additionally repeated the analyses for each chromosome using a GRM generated from the remaining 21 chromosomes, or in the case of the largest study (WGHS) repeating the analysis for all odd numbered chromosomes using a GRM generated from the even numbered chromosomes and vice versa. The single $\lambda_{GC}$ inflation factor for this analysis, 1.23, was also entirely consistent with the standard GWAS results (Online Methods, Supplementary Note, and Supplementary Table 4).

Our family based analyses also provided strong evidence that the inflated statistics are driven predominantly by the highly polygenic nature of height. We assessed whether variants that reached genome-wide significance after single GC correction replicated in family-based analyses of up to 25,849 samples (effective sample size 14,963, using methods that are immune to stratification (Online Methods, Supplementary Note, and Supplementary Tables 5 and 6). We identified genome-wide significant associations from a meta-analysis that excluded the family-based samples, and tested these associations for replication in the family-based samples; a lower rate of replication than expected could be due to inflation of effect sizes in the discovery sample from the “winner’s curse” and/or stratification. Of 416 genome-wide significant SNPs representing multiple signals selected after exclusion of family-based studies, 371 SNPs had a consistent direction of effect (compared with 208 expected by chance, and 400 expected in the absence of any inflation of estimated effect sizes), and 142 replicated with $P<0.05$ (compared with 21 expected by chance, and 210 expected in the absence of effect size inflation; Supplementary Table 5). These analyses (particularly the directional consistency) shows that most of the loci represent true associations, but also shows that there is a modest inflation in the effect size estimates, due to stratification and/or the winner’s curse. To distinguish between these possibilities, we repeated this analysis, substituting for the family-based samples a random set of studies with similar total effective sample size. The number of replicating loci was only slightly lower in the family-based cohorts than in the random samples (Supplementary Table 5, 12–17 fewer replications attributable to stratification at different $P$-value thresholds). This indicates that most of the modest inflation in effect estimates is due to the winner’s curse, that a small amount of inflation is due to residual stratification, and that few (upper limit ~15–25; Supplementary Note and Supplementary Table 5) if any of the loci that reach genome-wide significance after single GC correction are likely to be complete false positives due to stratification (that is, no real association whatsoever with height).

**Variance explained by SNPs at different significance levels**

Having established that single GC correction is sufficient to identify SNPs that are likely to be truly associated with height, we next performed a series of analyses using GWAS data from five independent validation studies to quantify the fraction of phenotypic variance
explained by SNPs selected from the GCTA-COJO analyses \(^7\) of the meta-analysis data, which excluded data from the validation studies, at a range of statistical thresholds, and to quantify the accuracy of predicting height using these selected SNPs (Online Methods). We first developed a new method that uses within-family prediction to partition the variance of the SNP-based predictor into components due to real SNP effects, errors in estimating SNP effects, and population stratification (Online Methods), and applied the method to data on full-sib pairs from three of the five validation studies (Online Methods). Consistently across the three studies, all the partitioned variance components increased as a less stringent significance level was used for SNP selection in the discovery sample and the error variance increased more dramatically than the genetic variance when more SNPs selected at a less significance level were included in the predictor (Fig. 2a–c). We demonstrated the partitioning of variance due to population stratification by the within-family prediction analyses with and without adjusting for principal components (PCs) (Supplementary Fig. 5). The results again confirmed that the impact of population stratification on the top associated SNPs was minor and demonstrated that the variation in the predictor due to true SNP effect, estimation error and population stratification was quantifiable. We next inferred, using these partitioned variance components from the within-family prediction analysis, how well different selected sets of SNPs would predict height in independent samples. We showed that the observed prediction accuracy (squared correlation between phenotype and predictor, \(R^2\)) in five different population-based cohorts was highly consistent with the values inferred from the within-family based analyses, with prediction accuracy peaking at \(\sim 17\%\) using the \(\sim 1,900\) SNPs reaching \(P<5\times10^{-5}\) (Fig. 2d). Finally we estimated variance explained by the selected SNPs in population-based studies using the GCTA-GREML method \(^4,8\) (Fig. 2e). The results showed that \(\sim 670\) SNPs at \(P<5\times10^{-8}\) and \(\sim 9,500\) SNPs at \(P<5\times10^{-3}\) captured \(\sim 16\%\) and \(\sim 29\%\) of phenotypic variance respectively (Table 1), which was also consistent with the estimates inferred from the within-family prediction analysis. As shown in equation \([19]\), prediction \(R^2\) is not equal to the variance explained but a function of the variance of true SNP effects and the error variance in estimating SNP effects, in the absence of population structure. This is demonstrated in Figure 2, where at thresholds below genome-wide significance, variance explained is higher than the prediction accuracy, because the latter is deflated both by imprecise estimates of effect sizes (estimation errors) and by inclusion of SNPs that are not associated with height. The estimate of variance explained by all the HapMap3 (ref. 11) SNPs without SNP selection was \(\sim 50\%\) (Table 1), consistent with previous estimates \(^4,5\). Thus, a group of \(\sim 9,500\) SNPs (representing \(<1\%\) of common SNPs) selected at \(P<5\times10^{-3}\), explained \(\sim 29\%\) of phenotypic variance. Since \(\sim 50\%\) of phenotypic variance is explained by all common SNPs, the selected set of SNPs, despite being limited to \(<1\%\) of common SNPs, accounts for the majority of variance attributable to all common SNPs (29/50 = 60\%). This set of \(\sim 9,500\) SNPs strongly clustered with the newly established height loci: 1,704 (19\%) variants were located within 250kb of one of the 697 genome-wide associated SNPs, suggesting that a substantial fraction of “missing heritability” is within already identified loci. This clustering of additional variants within identified loci was confirmed in a parallel analysis based on two left-out studies where we observed that SNPs in closer physical proximity with the top associated SNPs explained disproportionately more variance (Online Methods and Supplementary Fig. 6).
Larger GWAS identifies new biologically relevant genes and pathways

Having shown that ~1% of variants can account for the majority of heritability attributable to common variation, we next considered whether the expanded set of height-associated variants could be used to identify the genomic features and biological pathways of most relevance to normal variation in adult height. To test whether our GWAS could implicate new biology, we used established and novel approaches to test whether the height-associated loci were enriched for functionally relevant variants, genes, pathways, and tissues.

As with the 180 variants identified in our previous analysis, the 697 variants were non-randomly distributed with respect to functional and putatively functional regions of the genome (Online Methods). We observed that height associated variants were enriched for non-synonymous SNPs (nsSNPs) (empirical enrichment of 1.2 fold, \( P=0.02 \)), cis-regulatory effects in blood (empirical enrichment of 1.5 fold, \( P=0.03 \)), a curated list of genes that underlie monogenic syndromes of abnormal skeletal growth\(^1\) (empirical enrichment 1.4 fold, \( P=0.013 \)), associations with apparently unrelated complex traits in the NHGRI GWAS catalog (empirical enrichment 2.6 fold, \( P<1\times10^{-4} \)) and functional chromatin annotations in multiple tissues and cell types (empirical enrichment 1.8 fold, \( P<1\times10^{-3} \)) (Supplementary Note and Supplementary Tables 7–11).

The greater resolution of height associated variants provided by increased sample size, combined with improved gene prioritization and gene set enrichment approaches, identified multiple new tissues, gene sets and specific genes that are highly likely to be involved in the biology of skeletal growth. Specifically, using a variety of established and novel pathway methods, we identified ~3 times as many enriched pathways and prioritized ~5 times as many genes (including genes newly prioritized in previously identified loci) compared to results derived from identical pathway methods to the previous GWAS of 133,000 individuals (Table 2).

We first focused on existing pathway and gene prioritization methods: (1) MAGENTA\(^1\), a method designed to identify gene sets enriched in GWAS data, and (2) GRAIL\(^1\), which uses published literature to highlight connections between likely relevant genes within GWAS loci. As expected, the GRAIL and MAGENTA analyses confirmed several previously identified gene sets and pathways clearly relevant to skeletal growth, but in the larger sample they also provided evidence for additional known and novel genes, gene sets and protein complexes not identified in our previous smaller study (for example, FGF signaling, WNT signaling, osteoglycin, and other genes related to bone or cartilage development) (Supplementary Tables 12–13 and Supplementary Fig. 7).

To obtain more detailed insight into height biology, we applied DEPICT, a novel data-driven integrative method that uses gene sets reconstituted based on large scale expression data to prioritize genes and gene sets, and also to identify tissues enriched in highly expressed genes from associated loci (Pers et al. in preparation; Online Methods and Supplementary Note). The DEPICT analysis highlighted 2,330 reconstituted gene sets (after pruning for high levels of redundancy). These gene sets both confirmed and extended the MAGENTA and GRAIL findings, and identified novel pathways not identified in our previous height GWAS (for example regulation of beta-catenin, biology related to...
glycosaminoglycans such as chondroitin sulfate and hyaluronic acid, and mTOR signaling) (Supplementary Table 14). Gene sets identified based on 327 strictly novel height variants (>1Mb from the 180 known variants loci) highly resembled gene sets highlighted by the already known 180 loci (Spearman’s rank correlation coefficient between gene set enrichment Z-scores $r=0.91$, $P=2\times10^{-16}$). Thus, the variants discovered through increased sample size continued to highlight specific and relevant growth-associated gene sets, while the combined analysis of both old and new loci provided the additional power needed to identify new gene sets (Table 3 and Supplementary Table 14).

The DEPICT analysis also prioritized tissues and individual genes. We found that genes within associated height loci were enriched for expression in tissues related to chondrocytes (cartilage, joint capsule, synovial membrane, and joints; $P<5.5\times10^{-9}$, FDR<0.001), and other musculoskeletal, cardiovascular, and endocrine tissue-types (FDR<0.05) (Fig. 3; Supplementary Fig. 8; Supplementary Table 15). We also showed that a subset of the 697 height associated SNPs that represented lead cis-eQTLs in blood defined 75 genes that were collectively enriched for expression in cartilage ($P=0.008$) (Supplementary Note and Supplementary Table 8).

We used DEPICT to prioritize 649 genes (at FDR<0.05) within height-associated loci (Table 3 and Supplementary Table 16). Of these 649 genes, 202 genes (31%) were either significant in the GRAIL analysis (Supplementary Tables 13 and 16) and/or overlapped with a list of abnormal skeletal growth syndromes that we assembled from the OMIM database (n=40; Supplementary Tables 9 and 16). Many other newly prioritized genes had additional supporting evidence (Supplementary Table 16), including specific expression in the growth plate, and/or connections to relevant pathways (for example: GLI2 and LAMA5 [hedgehog signaling]; FRS2 [FGF signaling]; AXIN2, NFATC1, CTNNB1, FBXW11, WNT4, WNT5A and VANGL2 [WNT/beta-catenin signaling]; SMAD3 and MTOR [TGF-beta and/or mTOR signaling]; WWP2/miR140, IBSP, SHOX2 and SP3 [required in mice for proper bone and cartilage formation]; CHYS1, DSE and PCOLCE2 [glycosaminoglycan/collagen metabolism]; SCARA3, COP22, TBX18, CRISPLD1 and SLIT3 [differential expression in growth plate and predicted to be in highly relevant pathways]).

DEPICT also prioritizes genes that are new candidates for playing a role in skeletal growth. The genes newly and strongly implicated in this study included not only genes with obvious relationships to skeletal biology, such as SOX5 and collagen genes, but also genes that have no clear published connection to skeletal growth, and likely represent as yet unknown biology (Table 3 and Supplementary Table 16). DEPICT strongly prioritized genes that do not have published annotations related to growth-related pathways but are predicted to be in gene sets that are both enriched in the associated loci and clearly connected to growth. These include genes newly predicted to be in pathways related to cartilage or bone development (FAM101A, CRISPLD1 and the noncoding RNA LINC00476), collagen or extracellular matrix (GLT8D2, CCDC3, and ZCCHC24), histone demethylation (ATAD2B and TSTD2) and other genes predicted to have skeletal phenotypes but not currently annotated as belonging to relevant pathways (ARSJ, PSKH1, COP22, ADAMTS17 and the microRNA cluster MIR17HG). Of note, mutations in both ADAMTS17 and MIR17HG have been identified as causes of syndromic short stature in humans.
As suggested by the prioritization of ADAMTS17 and MIR17HG, it is possible that some of the newly highlighted genes may also underlie new syndromes of abnormal skeletal growth. As a further proof of principle, the second entry on our list of prioritized genes (Table 3 and Supplementary Table 16), CHSY1, was not a known monogenic gene in the OMIM database when we assembled our list, but mutations in this gene have since been shown to cause a syndrome including brachydactyly and short stature. Thus, the novel DEPICT method, applied to the larger GWAS data set, not only identified similar biology to GRAIL and MAGENTA but also implicated a large number of additional genes, gene sets and pathways that that are likely important in skeletal biology and human growth.

**Discussion**

By performing a large GWAS study on adult height, a highly heritable polygenic trait, we have provided answers to several current questions of relevance to the genetic study of polygenic diseases and traits. First, we showed that by conducting larger GWAS, we can identify SNPs that explain a substantial proportion of the heritability attributable to common variants. As hypothesized by Yang et al. (2010), the heritability directly accounted for by variants identified by GWAS and inferred by whole-genome estimation approaches are converging with increasing sample size. The variance explained by genome-wide significant SNPs has increased from 3–5% with discovery samples of ~25,000 (ref. 19) to 10% with a discovery sample size of ~130,000 (ref. 6) to 16% with a discovery sample size of 250,000 (this study), and the variance explained from all captured common SNPs is ~50% (Supplementary Fig. 9). The variance explained by genome-wide significant SNPs on a chromosome is also proportional to its length, consistent with the conclusion made by Yang et al. (Supplementary Fig. 9). Our new results show that ~21%, ~24% and ~29% of phenotypic variance in independent validation samples is captured by the best ~2,000, ~3,700 and ~9,500 SNPs respectively selected in the discovery samples (Table 1), and that the correlation between actual and predicted height in independent samples from the same population has increased to 0.41 (maximum prediction $R^2 = 0.41^2 = 0.17$, Fig. 2d). The results are consistent with a genetic architecture for human height that is characterized by a very large but finite number (thousands) of causal variants, located throughout the genome but clustered in both a biological and genomic manner. Such a genetic architecture may be described as pseudo-infinitesimal, and may characterize many other polygenic traits and diseases. There is also strong evidence of multiple alleles at the same locus segregating in the population and for associated loci to overlap with Mendelian forms, suggesting a large but finite genomic mutational target for height, and effect sizes ranging from minute ($<1\text{mm} ; <0.01 \text{SDs}$) to gigantic ($>300\text{mm} ; >3 \text{SDs}$, in the case of monogenic mutations).

It has been argued that the biological information emerging from GWA studies will become less relevant as sample sizes increase, because as thousands of associated variants are discovered, the range of implicated genes and pathways will lose specificity and cover essentially the entire genome. If this were the case, then increasing sample sizes would not help to prioritize follow up studies aimed at identifying and understanding new biology, and the associated loci would blanket the entire genome. Our study provides strong evidence to the contrary: the identification of many 100’s and even 1000’s of associated variants can continue to provide biologically relevant information. In other words, the variants identified...
in larger sample sizes both display a stronger enrichment of pathways clearly relevant to skeletal growth and prioritize many additional new and relevant genes. Furthermore, the associated variants are often non-randomly and tightly clustered (typically separated by <250 kb), resulting in the frequent presence of multiple associated variants in a locus. The observations that genes and especially pathways are now beginning to be implicated by multiple variants suggests that the larger set of results retain biological specificity but that at some point, a new set of associated variants will largely highlight the same genes, pathways and biological mechanisms as have already been seen. This endpoint (which we have not clearly reached for height) could be considered analogous to reaching “saturation” in model organism mutagenesis screens, where new alleles typically map to previously identified genes\textsuperscript{21}.

We have identified a large number of gene sets and pathways that are enriched for associations with height. Although the number of gene sets and pathways is large, many are overlapping and likely represent multiple annotations of a much smaller set of core biological mechanisms. We also highlight individual genes within associated loci as being relevant to skeletal growth, including candidates for contributing to syndromes of abnormal skeletal growth; for example, we strongly implicated \textit{CHSY1}, recently identified as an underlying cause of a monogenic syndrome with short stature and brachydactyly\textsuperscript{17,18}. The lists of prioritized genes and pathways should therefore provide a rich trove of data for future studies of skeletal growth; to facilitate such studies, we have made our results (including genome-wide association results and complete list of highlighted genes and pathways) publicly available. Based on the results of large genetic studies of height, we anticipate that increasing the number of associated loci for other traits and diseases could yield similarly rich lists that would generate new biological hypotheses and motivate future research into the basis of human biology and disease.

\textbf{URLs}


\textbf{ONLINE METHODS}

\textbf{Genome-wide association study meta-analysis}

We combined height summary association statistics from 79 genome-wide association (GWA) studies in a meta-analysis of 253,288 individuals using the same methods and studies as previously described\textsuperscript{6} and additional studies as described in Supplementary Tables 17–19. A total of 2,550,858 autosomal SNPs were meta-analyzed using inverse-variance fixed effects method using METAL\textsuperscript{22}.

\textbf{GCTA-COJO: conditional and joint multiple SNPs analysis}

We used GCTA-COJO analysis\textsuperscript{7,8} to select the top associated SNPs. This method uses the summary statistics from the meta-analysis and LD correlations between SNPs estimated from a reference sample to perform a conditional association analysis\textsuperscript{7}. The method starts
with an initial model of the SNP that shows the strongest evidence of association across the whole genome. It then implements the association analysis conditioning on the selected SNP(s) to search for the top SNPs one-by-one iteratively via a stepwise model selection procedure until no SNP has a conditional $P$-value that passes the significance level. Finally, all the selected SNPs are fitted jointly in the model for effect size estimation. We used 6,654 unrelated individuals from the ARIC cohort as the reference sample for LD estimation. There were ~3.0M SNPs included in the original meta-analysis. We included in this analysis only the SNPs (~2.48M) on HapMap2 and with sample size > 50,000. We used the genome-wide significance level $P<5\times10^{-8}$ (as reported in Supplementary Table 1).

**Metabochip replication**

We combined height summary association statistics from 37 independent studies genotyped using Illumina’s Metabochip array in a meta-analysis of 80,067 individuals of European ancestry (Supplementary Tables 20–22). Each study tested association between each genotyped SNP and the same QC procedures, height transformations, adjustment, and inheritance model as described for the GWA analysis. Genomic control correction was applied to results for each study prior to meta-analysis, using a set of 4,427 SNPs associated with QT interval to control study-specific inflation factors. We used the inverse-variance fixed effects meta-analysis method.

**Validation – linear mixed model (LMM) based association analysis**

Each of 15 studies (59,380 individuals) used genome-wide SNP information to calculate a genomic relationship matrix (GRM) for all pairs of individuals and used this to correct association statistics for cryptic relatedness and population stratification. Each study used a linear mixed model as implemented in the software EMMAX. Meta-analysis was performed as described for the standard GWAS and using a single GC correction. Each study additionally repeated the analyses for each chromosome using a GRM generated from the remaining 21 chromosomes, or in the case of the largest study (WGHS) repeating the analysis for all odd numbered chromosomes using a GRM generated from the even numbered chromosomes and vice versa. Each study then combined association results from the 22 or 2 parts of the genome into one set of data and we repeated the single GC meta-analysis.

**Validation – within family (transmission) association analyses**

A pure transmission based analysis was performed in seven cohorts for SNPs representing 416 signals of association (Supplementary Note), selected after repeating meta-analysis excluding these studies, with single GC correction. Filtering of low imputation quality SNPs in the studies was followed by inverse variance method of meta-analysis of the family based results. Because of the presence of related individuals, family based studies have lower power at a given sample size. For each study, we calculated the effective sample size (the size of a sample of unrelated individuals that would have the equivalent power; see Supplementary Note and Winkler et al. 24). Estimation of winner’s curse in our data set was performed by repeating the meta-analysis excluding either the family-based studies or excluding random sets of studies from GIANT matched by effective sample size to the
family based studies. Independent genome-wide significant loci were selected from each meta-analysis. Power for replication in the excluded samples was estimated at different \( P \)-value thresholds and the deficit in replications (number of replications expected minus number observed) was calculated. The contribution of the winner’s curse to the deficit in replications was estimated as the average deficit across the three sets of random non-family-based cohorts. By subtracting this from the deficit observed for the family-based cohorts, we estimated the lack of replication that could be attributed to stratification (either inflation of effect size for true associations, or false positive associations).

### Variance and heritability explained

We used GCTA-COJO analysis (Online Methods) to select the top associated SNPs at a range of stringent significance levels \((5 \times 10^{-3}, 5 \times 10^{-4}, 5 \times 10^{-5}, \ldots, 5 \times 10^{-8})\) for estimation and prediction analyses. We then quantified the variance explained by those selected SNPs using a three-stage analysis, i.e. within-family prediction, GCTA-GREML analysis and population based prediction, in five validation studies (B-PROOF, FRAM, QIMR, TwinGene and WTCCC-T2D). To avoid sample overlap, we repeated the main GWAS meta-analysis and the multiple-SNP analysis five times, each time excluding one of the five validation studies. This approach ensured complete independence between data used to discover SNPs, and data used to estimate how much variance in height these SNPs explained and how well they predicted height. For the within-family prediction analyses, we selected 1,622, 2,758 and 1,597 pairs of full sibs from the QIMR, TwinGene and FRAM cohorts, respectively, with one sib pair per family. For the whole-genome estimation and prediction analyses, we used GCTA-GRM\(^8\) to estimate the genetic relatedness between individuals and selected unrelated individuals with pairwise genetic relatedness <0.025 in each of the five studies, i.e. B-PROOF \((n = 2,555)\), FRAM \((n = 1,145)\), QIMR \((n = 3,627)\), TwinGene \((n = 5,668)\) and WTCCC-T2D \((n = 1,914)\).

### Within-family prediction analysis

We used the SNPs selected from GCTA-COJO analysis to create a genetic predictor (also called “genetic profile score”) for each of all the full sibs using PLINK\(^{25}\). We then adjusted the genetic predictor by the first 20 principal components (PCs) generated from the principal component analysis (PCA)\(^{26}\). By comparing the predictors within and between families, we partitioned the variance in the predictor analysis into components due to real SNP effects \((V_g)\), errors in estimating SNP effects \((V_e)\), and population structure \((C_g + C_e)\), as described in the Online Methods below.

We calculated the weighted average of each of the four (co)variance components over the three cohorts by their sample size, i.e. \(\Sigma_i (V_g(i) / n_i) / \Sigma_i (n_i)\) with the subscript \(i\) indicating the cohort and \(n\) being the sample size. From the results of these partitioning analyses within families we can infer what the prediction \(R^2\) (Equation 19 in Online Methods below) and what the proportion of variance explained by SNPs (i.e. \(V_g / V_P\) with \(V_P\) being the phenotypic variance) would be in a sample of unrelated individuals when using the same set of SNPs. We then tested these inferred values in unrelated samples.
**GCTA-GREML analysis**

We performed the GREML analysis\(^4\) in GCTA\(^8\) to estimate the variance explained by the selected SNPs \((h^2_g)\) in each of the five validation studies. This method fits the effects of a set of SNPs simultaneously in a model as random effects and estimates the genetic variance captured by all the fitted SNPs without testing the significance of association of any single SNPs. We combined the estimates of \(h^2_g\) from the five studies by the inverse-variance approach, i.e. \(\Sigma_i (h^2_g)/SE_i^2)/\Sigma_i (1/SE_i^2)\).

**Population-based prediction analysis**

We created a genetic predictor using the selected SNPs for the unrelated individuals in each of the five validation studies. We then calculated the squared correlation \((R^2)\) between phenotype and predictor in each validation study, and calculated the weighted average of the prediction \(R^2\) by the sample size across the five studies, i.e. \(\Sigma_i (R^2_i n_i)/\Sigma_i (n_i)\).

**Theory and method to partition the variance in a genetic predictor**

Under the assumption of an additive genetic model, the phenotype of a quantitative trait can be written as

\[
y + g + \varepsilon = \sum_i x_i b_i + \varepsilon \quad [1]
\]

where \(y\) is the trait phenotype, \(g\) is the total genetic effect of all SNPs, \(x\) is an indicator variable for SNP genotypes, \(b\) is the SNP effect, and \(\varepsilon\) is the residual.

From this model, the additive genetic variance is

\[
\text{var}(g) = \sum_i \text{var}(x_i) b_i^2 + \sum_i \sum_{j(i \neq j)} \text{cov}(x_i, x_j) b_i b_j \quad [2]
\]

with the first component being the expected value of additive genetic variance under linkage equilibrium (LE) and second component being the deviation from the expected value could be caused by linkage disequilibrium (LD), population structure or selection\(^27\).

Considering a pair of full siblings in a family, the additive genetic covariance between the sibs is

\[
\text{cov}(g_1, g_2) = \text{cov}(\sum_i x_{1i} b_i, \sum_i x_{2i} b_i) = \sum_i \text{cov}(x_{1i}, x_{2i}) b_i^2 + \sum_i \sum_{j(i \neq j)} \text{cov}(x_{1i}, x_{2j}) b_i b_j \quad [3]
\]

For full sibs,

\[
\text{cov}(x_{1i}, x_{2j}) = \frac{1}{2} \text{var}(x_i),
\]
\[ \text{cov}(x_{1i}, x_{2j}) = \frac{1}{2} \text{cov}(x_i, x_j) \text{ for SNPs that are in LD, and} \]
\[ \sum_i \sum_j \text{cov}(x_{1i}, x_{2j}) \hat{b}_i \hat{b}_j = \sum_i \sum_j \text{cov}(x_i, x_j) \hat{b}_i \hat{b}_j \text{ for SNPs that are not in LD (as shown by both empirical and simulation results)} \]

Let
\[ V_g = \sum_i \text{var}(x_i) \hat{b}_i^2 + \sum_i \sum_{j(i \neq j)} \text{cov}(x_i, x_j) \hat{b}_i \hat{b}_j \text{ (SNPs are in LD), and} \]
\[ C_g = \sum_i \sum_{j(i \neq j)} \text{cov}(x_i, x_j) \hat{b}_i \hat{b}_j \text{ (SNPs are not in LD but correlated due to population structure)} \]

Therefore, the genetic variance is
\[ \text{var}(g) = V_g + C_g \quad [4] \]

The genetic covariance between a pair of full-sibs is
\[ \text{cov}(g_1, g_2) = \frac{1}{2} V_g + C_g \quad [5] \]

If we take a set of SNPs with their effects estimated from GCTA-COJO analysis (Online Methods), and create a predictor using these SNPs in an independent validation sample, we can write the predictor as
\[ \hat{g} = \sum_i x_i \hat{b}_i \quad [6] \]

where \( \hat{b} \) is the estimate of \( b \) with \( \hat{b} = b + e \) with \( e \) being the error in estimating \( b \).

If we assume \( b \) and \( e \) are independent and denote
\[ V_e = \sum_i \text{var}(x_i) e_i^2 \text{ and} \]
\[ C_e = \sum_i \sum_{j(i \neq j)} \text{cov}(x_i, x_j) e_i e_j \text{, the variance of the predictor is} \]
\[ \text{var}(\hat{g}) = \text{var}(\sum_i x_i \hat{b}_i) = \sum_i \text{var}(x_i) \hat{b}_i^2 + \sum_i \sum_{j(i \neq j)} \text{cov}(x_i, x_j) \hat{b}_i \hat{b}_j \]
\[ = \sum_i \text{var}(x_i) \hat{b}_i^2 + \sum_i \text{var}(x_i) e_i^2 + \sum_i \sum_{j(i \neq j)} \text{cov}(x_i, x_j) \hat{b}_i \hat{b}_j + \sum_i \sum_{j(i \neq j)} \text{cov}(x_i, x_j) e_i e_j \quad [7] \]
\[ = V_g + V_e + C_g + C_e \]

The covariance between the predictors of a pair of full-sibs is
\[ \text{cov}(\hat{g}_1, \hat{g}_2) = \text{cov}(\sum_i x_{1i} \hat{b}_i, \sum_i x_{2j} \hat{b}_j) = \frac{1}{2} \sum_i \text{var}(x_i) \hat{b}_i^2 + \sum_i \sum_{j(i \neq j)} \text{cov}(x_{1i}, x_{2j}) \hat{b}_i \hat{b}_j \]
\[ = \frac{1}{2} V_g + \frac{1}{2} V_e + C_g + C_e \quad [8] \]

The covariance between the true phenotype and the predictor of a same individual is
\[ \text{cov}(y, \hat{g}) = \text{cov}(g + \varepsilon, g + \varepsilon) = \text{var}(g) = V_g + C_g \]  

[9]

The covariance between the true phenotype of one sib and the predictor of the other sib is

\[
\text{cov}(y_1, \hat{g}_2) = \text{cov}(g_1 + \varepsilon_1, g_2 + \varepsilon_2) = \frac{1}{2} \sum_i \text{var}(x_i) b_i^2 + \sum_{i (i \neq j)} \text{cov}(x_i, x_j) b_i b_j = \frac{1}{2} V_g + C_g
\]  

[10]

If we define \( \Delta \hat{g} = \hat{g}_1 - \hat{g}_2 \) and \( \Delta y = y_1 - y_2 \),

\[
\text{var}(\Delta \hat{g}) = \text{var}(\hat{g}_1) + \text{var}(\hat{g}_2) - 2 \text{cov}(\hat{g}_1, \hat{g}_2) = V_g + V_e
\]  

[11]

\[
\text{cov}(\Delta y, \Delta \hat{g}) = \text{cov}(y_1, \hat{g}_1) + \text{cov}(y_2, \hat{g}_2) - \text{cov}(y_1, \hat{g}_2) - \text{cov}(y_2, \hat{g}_1) = V_g
\]  

[12]

We therefore can calculate these four parameters as

\[
V_g = \text{cov}(\Delta y, \Delta \hat{g})
\]  

[13]

\[
V_e = \text{var}(\Delta \hat{g}) - V_g
\]  

[14]

\[
C_g = \text{cov}(y, \hat{g}) - V_g
\]  

[15]

\[
C_e = 2 \text{cov}(\hat{g}_1, \hat{g}_2) - \text{var}(\hat{g}) - C_g
\]  

[16]

where \( V_g \) can be interpreted as the variance explained by real SNP effects, \( C_g \) is the covariance between predictors attributed to the real effects of SNPs that are not in LD but correlated due to population stratification, \( V_e \) is the accumulated variance due to the errors in estimating SNP effects, and \( C_e \) is the covariance between predictors attributed to errors in estimating the effects of SNPs that are correlated due to population stratification.

To assess the prediction accuracy, we usually perform a regression analysis of the real phenotype against the predictor, i.e.

\[ y = \beta_0 + \hat{g} \beta_1 + \varepsilon \]  

[17]

so that the regression slope is actually

\[ \beta = \text{cov}(y, \hat{g}) / \text{var}(\hat{g}) = (V_g + C_g) / (V_g + V_e + C_g + C_e) \]  

[18]

with the regression \( R^2 \) being...
In the absence of population structure,

\[ R^2 = \frac{(V_g + C_g)^2}{(V_g + V_e + C_g + C_e)} \]  \[19\]

Variance explained by SNPs in proximity to the top associated SNPs

We performed analyses to quantify the variance explained by SNPs in close physical proximity to the top associated SNPs in 9,500 unrelated individuals (pairwise genetic relatedness < 0.025) from a combined dataset of the QIMR and TwinGene cohorts. As in previous analyses, to avoid sample overlap between discovery and validation studies, we repeated the discovery meta-analysis excluding the QIMR and TwinGene cohorts, and identified 643 genome-wide significant SNPs from the GCTA-COJO analysis of the summary statistics using ARIC data for LD estimation. We used GCTA-GREML analysis\(^4,8\) to quantify the phenotypic variance explained by all the common SNPs (MAF > 0.01) within 100Kb, 500Kb or 1Mb of the 643 genome-wide significant SNPs. We show in Supplementary Figure 6a that there are 104K, 423K and 745K SNPs within 100Kb, 500Kb and 1Mb of the top associated SNPs, which explain 20.8\% (s.e. = 1.3\%)\(^8\), 25.7\% (s.e. = 1.8\%)\(^8\) and 29.5\% (s.e. = 2.2\%) of phenotypic variance, respectively. We then applied a regression-based approach\(^28\) to adjust for LD between SNPs. The estimates of variance explained after LD-adjustment were slightly higher than those without adjustment, and the ratio of between the estimates with and without LD-adjustment was consistently ~1.05 regardless of the window size (Supplementary Fig. 6a). However, the difference is small.

We then sought to investigate whether or not there is an enrichment of additional association signals at the top associated loci. We varied the window size from 20Kb to 50Kb, 100Kb, 150Kb, 200Kb, 300Kb, 400Kb, 500Kb, 750Kb and 1Mb, and fitted a two-component model in GCTA-GREML analysis, with the first component being the top associated SNPs and the second component being the rest of SNPs within the window. We found that the per-SNP variance explained excluding the top SNPs (variance explained by the second component divided by the number of SNPs included in this component) decreased with the size of window (Supplementary Fig. 6b), implying that SNPs in closer physical proximity to the top associated SNPs tend to explain disproportionally more variance.

Enrichment of associated SNPs in ENCODE regions, loci containing OMIM genes, eQTLs and nsSNPs

To identify putative causal variants among the height-associated markers, we explored whether the height-associated SNPs were in strong LD ($r^2$>0.8) with non-synonymous coding variants in 1000 Genomes Project CEU Phase 1 data, showed an effect on whole blood gene expression levels, were located within ENCODE-annotated regions, were within loci harboring monogenic growth genes, or had previously been associated with other complex traits in NHGRI GWAS catalog ($P<5\times10^{-8}$) (Supplementary Tables 7–11). To estimate the empirical assessment of enrichment for listed features we used 10,000
permutations of random sets of SNPs matched to the pruned (LD $r^2 > 0.1$) 628 height-associated SNPs by the number of nearby genes (within a distance of LD $r^2 > 0.5$), physical distance to nearest gene, and minor allele frequency.

**Enrichment of genes in associated loci in known and novel pathways**

**Data-Driven Expression-Prioritized Integration for Complex Traits (DEPICT) analysis**—The DEPICT method (T.H.P. *et al.*, unpublished data; see Geller *et al.* for an earlier application of DEPICT) relies on pre-computed predictions of gene function based on a heterogeneous panel of 77,840 expression arrays (Fehrmann *et al.*, manuscript in review; ref. 30), 5,984 molecular pathways (based on 169,810 high-confidence experimentally derived protein-protein interactions), 2,473 phenotypic gene sets (based on 211,882 gene-phenotype pairs from the Mouse Genetics Initiative (see URLs)), 737 Reactome pathways, 5,083 Gene Ontology terms, and 184 KEGG pathways. The method leverages these predictions to extend the functional annotations of genes, including genes that previously had only a few or no functional annotations. DEPICT facilitates the analysis of GWAS data by (1) assessing whether genes in associated loci are enriched in tissue-specific expression, (2) identifying reconstituted gene sets that are enriched in genes from associated loci, and (3) systematically identifying the most likely causal gene(s) at a given locus (see Supplementary Note for a more detailed description of DEPICT). In order to run DEPICT, we first clumped the summary statistics from the meta-analysis using 500kb flanking regions, $r^2 > 0.1$, and excluded SNPs with $P \geq 5 \times 10^{-8}$, which resulted in 628 SNPs. We then mapped genes to each of the 628 best-associated SNPs. For a given SNP, this was accomplished by including all genes that resided within LD $r^2 > 0.5$ boundaries of that SNP, and always including the nearest gene, to its locus gene set. We used a locus definition that was calibrated using the GWAS data for height levels presented in this paper and optimized capture of known monogenic genes for those traits. We merged overlapping loci, and excluded loci that mapped near or within the major histocompatibility complex locus (chromosome 6, location: 20 to 40 Mb), which resulted in a list of 566 non-overlapping loci that were used as input to DEPICT. HapMap Project Phase II CEU genotype data was used for all LD calculations.

**GRAIL and MAGENTA analysis**—The GRAIL algorithm was run using the LD pruned ($r^2 > 0.1$) 628 SNPs without correcting for gene size, and using text-mining data up to December 2006 (default setting). MAGENTA was run with the single genomic control adjusted summary statistics as input using default settings and excluding the HLA region.

**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

**Authors**

Andrew R Wood, Tonu Esko, Jian Yang, Sailaja Vedantam, Tune H Pers, Stefan Gustafsson, Audrey Y Chu, Karol Estrada, Jian'an Luan, Zoltán Kutalik, Najaf Amin, Martin L Buchkovich, Damien C Croteau-Chonka, Felix R Day, Yanan Duan, Tove Fall, Rudolf Wood et al.

Nat Genet. Author manuscript; available in PMC 2015 May 01.

Nat Genet. Author manuscript; available in PMC 2015 May 01.
Amouye1,2, Folkert W Asselbergs2,2, Themistocles L Assimes1,2, Murielle Bochud1,1,2, Bernhard O Boehm1,3,4, Eric Boerwinkle2,5, Erwin P Bottinger5, Claude Bouchard2,6, Stéphane Cauchy1,3,4, John C Chambers4,8,9,26, Stephen J Chanock7,9, Richard S Cooper1,7, Paul IW de Bakker7,2,28,3,28,4, George Dedoussis1,10, Luigi Ferrucci7,9, Paul W Franks4,6,9,25, Philippe Froguel6,2,8,9,13, Leif C Groop1,17,29,1,28,25, Christopher A Haiman1,2, Anders Hamsten6,8, M Geoffrey Hayes1,2, Jennie Hu8,17,22,22, David J. Hunter2,20,25,25,26, Kristian Hveem1,2,8, J Wouter Jukema7,2,20,28,27, Robert C Kaplan2,28,9, Mika Kivimäki23,9, Diana Kuh1,2,4, Markku Laakso2,29, Yongmei Liu2,8,9, Nicholas G Martin6,10, Winfried März50,161,29, Mads Melbye43,123, Susanne Moebus3, Patricia B Munroe1,15, Inger Njolstad1,3, Ben A Oostra1,8,130,29,1, Colin NA Palmer1,15, Nancy L Pedersen22,2, Markus Perola2,5,2,107, Louis Péрусse2,248,270, Ulrike Peters4,7, Joseph E Powell6,7, Chris Power224, Thomas Quertermous1,123, Rainer Rauramaa1,16,24,1, Eva Reinmaa2,2, Paul M Ridker1,11,29,2, Fernando Rivadeneira12,57,58, Jerome I Rotter2,29,3, Timo E Saaristo2,29,29,2, Danish Saleheen19,29,29,27, David Schlessinger2,28,9, P Eline Slagboom3,8,39, Harold Snieder122, Tim D Spector55, Konstantin Strauch25,153, Michael Stumvoll91,92, Jaakko Tuomilehto52,29,3,00,30,1, Matti Uusitupa30,2,30,1, Pim van der Harst2,23,5,2,28,0, Henry Völzke77,12,4,1, Mark Walker3,4, Nicholas J Wareham1,4, Hugh Watkins2,4,6, H-Erich Wichmann3,05,3,06,3,07, James F Wilson45, Pieter Zanen1,3, Panos Deloukas2,9,29,4,9,30, Iris M Heid3,2,35, Cecilia M Lindgren4,2,4, Karen L Mohlke1,19, Elizabeth K Speliotes1,3,0, Unnur Thorsteinsdottir1,174,3,11, Inês Barroso2,29,3,12,3,13, Caroline S Fox1,18, Kari E North48,3,14, David P Strachan3,15, Jacques S. Beckmann1,16,17,3,16, Sonja I Berndt2,9,2, Michael Boehnke2,25, Ingrid B Borecki1,4, Mark I McCarthy2,4,6,1,31,2,17, Andres Metspalu1,2,2, Kari Stefansson1,174,3,11, André G Uitterlinden12,57,58, Cornelia M van Duijn18,57,58,130, Lude Franke2,5, Cristen J Willer3,18,3,19,3,20, Alkes L. Price4,2,26,3,21, Guillaume Lettre2,26,2,69, Ruth JF Loos1,4,5,4,14,3,22, Michael N Weedon1,19, Erik Ingelsson9,10,24, Jeffrey R O’Connell2,26,6, Goncalo R Abecasis2,5,6, Daniel I Chasman1,11,29,2,1, Michael E Goddard3,23,3,24,3, Peter M Visscher6,7,1, Joel N Hirschhorn3,4,5,3, and Timothy M Frayling1,4

Affiliations

1Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter EX1 2LU, UK 2Estonian Genome Center, University of Tartu, Tartu 51010, Estonia 3Division of Endocrinology, Genetics and Basic and Translational Obesity Research, Boston Children’s Hospital, Boston, MA 02115, USA 4Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge 02142, MA, USA 5Department of Genetics, Harvard Medical School, Boston, MA 02115, USA 6Queensland Brain Institute, The University of Queensland, Brisbane 4072, Australia 7The University of Queensland Diamantina Institute, The Translation Research Institute, Brisbane 4012, Australia 8Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Lyngby 2800, Denmark 9Science for Life Laboratory, Uppsala University, Uppsala 75185, Sweden 10Department of Medical Sciences, Molecular Epidemiology, Uppsala University, Uppsala 75185, Sweden 11Division of Preventive Medicine, Brigham and Women’s Hospital, Boston, MA 02215, USA 12Department
of Internal Medicine, Erasmus Medical Center, 3015GE Rotterdam, The Netherlands 13 Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA 14 MRC Epidemiology Unit, University of Cambridge, Institute of Metabolic Science, Addenbrooke’s Hospital, Hills Road, Cambridge, CB2 0QQ, UK 15 Institute of Social and Preventive Medicine (IUMSP), Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne 1010, Switzerland 16 Swiss Institute of Bioinformatics, Lausanne 1015, Switzerland 17 Department of Medical Genetics, University of Lausanne, Lausanne 1005, Switzerland 18 Genetic Epidemiology Unit, Department of Epidemiology, Erasmus University Medical Center, 3015 GE Rotterdam, The Netherlands 19 Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA 20 Channing Division of Network Medicine, Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA 02115, USA 21 Division of Statistical Genomics, Department of Genetics Washington University School of Medicine, St. Louis, MO, USA 22 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 17177, Sweden 23 Department of Genetics, University Medical Center Groningen, University of Groningen, 9700 RB Groningen, The Netherlands 24 Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK 25 Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA 26 Montreal Heart Institute, Montreal, Quebec H1T 1C8, Canada 27 Institute of Molecular and Cell Biology, University of Tartu, Tartu 51010, Estonia 28 Istituto di Ricerca Genetica e Biomedica (IRGB), Consiglio Nazionale delle Ricerche, Cagliari, Sardinia 09042, Italy 29 Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK 30 Institute for Medical Informatics, Biometry and Epidemiology (IMIBE), University Hospital Essen, Essen, Germany 31 Clinical Epidemiology, Integrated Research and Treatment Center, Center for Sepsis Control and Care (CSCC), Jena University Hospital, Jena, Germany 32 Department of Genomic Epidemiology, Institute of Epidemiology and Preventive Medicine, University of Regensburg, D-93053 Regensburg, Germany 33 Harvard School of Public Health, Department of Nutrition, Harvard University, Boston, MA 2115, USA 34 HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA 35 Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, D-85764 Neuherberg, Germany 36 Telethon Institute for Child Health Research, Centre for Child Health Research, The University of Western Australia, Western Australia 6008, Australia 37 Section on Growth and Development, Program in Developmental Endocrinology and Genetics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA 38 Netherlands Consortium for Healthy Aging (NCHA), Leiden University Medical Center, Leiden 2300 RC, The Netherlands 39 Department of Molecular Epidemiology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands 40 Department of Human Genetics, Leiden University Medical Center, 2333 ZC Leiden, The Netherlands 41 Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School
of Medicine, Baltimore, MD 21205, USA 42Cardiology, Department of Specialties of Internal Medicine, Geneva University Hospital, Geneva 1211, Switzerland 43Department of Epidemiology Research, Statens Serum Institut, Copenhagen DK-2300, Denmark 44Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110, USA 45Centre for Population Health Sciences, University of Edinburgh, Teviot Place, Edinburgh, EH8 9AG, Scotland, UK 46Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford OX3 9DU, UK 47Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA 48Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA 49William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, EC1M 6BQ UK 50Vth Department of Medicine (Nephrology, Hypertensiology, Endocrinology, Diabetology, Rheumatology), Medical Faculty of Mannheim, University of Heidelberg, Germany 51Department of Internal Medicine II, Ulm University Medical Centre, D-89081 Ulm, Germany 52National Institute for Health and Welfare, FI-00271 Helsinki, Finland 53Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI USA 54The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA 55Department of Twin Research and Genetic Epidemiology, King’s College London, London SE1 7EH, UK 56Department of Cardiology, University Medical Center Groningen, University of Groningen, 9700RB Groningen, The Netherlands 57Netherlands Consortium for Healthy Aging (NCHA), 3015GE Rotterdam, The Netherlands 58Department of Epidemiology, Erasmus Medical Center, 3015GE Rotterdam, The Netherlands 59Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892, USA 60QIMR Berghofer Medical Research Institute, Queensland 4006, Australia 61Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford OX3 7LJ, UK 62Department of Genomics of Common Disease, School of Public Health, Imperial College London, Hammersmith Hospital, London, UK 63Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA 64Department of Clinical Sciences, Genetic & Molecular Epidemiology Unit, Lund University Diabetes Center, Skåne University Hospital, Malmö 205 02, Sweden 65Department of Public Health and Clinical Medicine, Unit of Medicine, Umeå University, Umeå 901 87, Sweden 66Department of Odontology, Umeå University, Umeå 901 85, Sweden 67University of Eastern Finland, FI-70210 Kuopio, Finland 68Atherosclerosis Research Unit, Center for Molecular Medicine, Department of Medicine, Karolinska Institutet, Stockholm 17176, Sweden 69Division of Biostatistics, Washington University School of Medicine, St. Louis, MO 63110, USA 70Translational Gerontology Branch, National institute on Aging, Baltimore MD 21225, USA 71Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, D-17475 Greifswald, Germany 72Department of Cardiology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands 73Department of Gerontology and Geriatrics, Leiden University Medical Center, 2300 RC Leiden,  

Nat Genet. Author manuscript; available in PMC 2015 May 01.
The Netherlands 74Experimental Cardiology Laboratory, Division Heart and Lungs, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands 75Department of Medical Genetics, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands 76Department of Endocrinology, University of Groningen, University Medical Center Groningen, Groningen, 9700 RB, The Netherlands 77DZHK (Deutsches Zentrum für Herz-Kreislaufforschung – German Centre for Cardiovascular Research), partner site Greifswald, D-17475 Greifswald, Germany 78Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, D-17475 Greifswald, Germany 79Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA 80Core Genotyping Facility, SAIC-Frederick, Inc., NCI-Frederick, Frederick, MD 21702, USA 81CNRS UMR 8199, F-59019 Lille, France 82European Genomic Institute for Diabetes, F-59000 Lille, France 83Université de Lille 2, F-59000 Lille, France 84Ealing Hospital NHS Trust, Middlesex UB1 3HW, UK 85Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK 86School of Health and Social Studies, Dalarna University, Falun, Sweden 87PathWest Laboratory Medicine of Western Australia, NEDLANDS, Western Australia 6009, Australia 88Geriatric Unit, Azienda Sanitaria Firenze (ASF), Florence, Italy 89Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX, USA 90Department of Medical Sciences, Endocrinology, Diabetes and Metabolism, Uppsala University, Uppsala 75185, Sweden 91IBF Adiposity Diseases, University of Leipzig, D-04103 Leipzig, Germany 92Department of Medicine, University of Leipzig, D-04103 Leipzig, Germany 93LifeLines, University Medical Center Groningen, University of Groningen, 9700 RB Groningen, The Netherlands 94Department of Pharmacology and Therapeutics, University College Cork, Cork, Ireland 95Department of Statistics & Biostatistics, Rutgers University, Piscataway, N.J. USA 96Department of Genetics, Rutgers University, Piscataway, N.J. USA 97Department of Biology, Norwegian University of Science and Technology, Trondheim, Norway 98Genome Technology Branch, National Human Genome Research Institute, NIH, Bethesda, MD 20892, USA 99Clinical Trial Service Unit, Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK 100Division of Endocrinology, Diabetes and Metabolism, Ulm University Medical Centre, D-89081 Ulm, Germany 101Department of Radiology, University Medical Center Utrecht, Utrecht, The Netherlands 102Department of Biomedical Informatics, Vanderbilt University, Nashville, TN 37232, USA 103Department of Human Nutrition, Wageningen University, Wageningen, The Netherlands 104Department of Dietetics-Nutrition, Harokopio University, Athens, Greece 105Medical Research Institute, University of Dundee, Ninewells Hospital and Medical School, Dundee DD1 9SY, UK 106Department of Internal Medicine B, University Medicine Greifswald, D-17475 Greifswald, Germany 107Institute for Molecular Medicine, University of Helsinki, FI-00014 Helsinki, Finland 108Laboratory of Epidemiology and Population Sciences, National Institute on Aging, NIH, Bethesda, MD 20892, USA 109Department of Public Health and Caring Sciences, Geriatrics, Uppsala University, Uppsala 75185, Sweden 110Kaiser
Wood et al. Page 22

Permanente, Division of Research, Oakland, CA 94612, USA 111Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, D-85764 Neuherberg, Germany 112German Center for Diabetes Research (DZD), Neuherberg, Germany 113Department of Medicine III, University Hospital Carl Gustav Carus, Technische Universität Dresden, D-01307 Dresden, Germany 114Department of Public Health and Clinical Medicine, Unit of Nutritional Research, Umeå University, Umeå 90187, Sweden 115Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands 116Kuopio Research Institute of Exercise Medicine, Kuopio, Finland 117MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, EH4 2XU, Scotland, UK 118National Heart, Lung, and Blood Institute, the Framingham Heart Study, Framingham MA 01702, USA 119Department of Neurology, Boston University School of Medicine, Boston, MA 02118, USA 120Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, 2300 RC Leiden, The Netherlands 121Faculty of Psychology and Education, VU University Amsterdam, Amsterdam, The Netherlands 122Department of Epidemiology, University Medical Center Groningen, University of Groningen, 9700 RB Groningen, The Netherlands 123Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA 124Institute for Community Medicine, University Medicine Greifswald, D-17475 Greifswald, Germany 125Division of Medical Genetics, Department of Biomedicine, University of Basel, Basel, Switzerland 126Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany 127Institute of Human Genetics, University of Bonn, Bonn, Germany 128Department of Public Health and General Practice, Norwegian University of Science and Technology, Trondheim 7489, Norway 129Hannover Unified Biobank, Hannover Medical School, Hannover, D-30625 Hannover, Germany 130Center for Medical Systems Biology, Leiden, The Netherlands 131Department of Pulmonary Physiology and Sleep Medicine, NEDLANDS, Western Australia 6009, Australia 132School of Medicine and Pharmacology, University of Western Australia, CRAWLEY 6009, Australia 133Uppsala University, Department of Immunology, Genetics & Pathology, SciLifeLab, Rudbeck Laboratory, SE-751 85, Uppsala, Sweden 134Department of Haematology, University of Cambridge, Cambridge CB2 0PT, UK 135NHS Blood and Transplant, Cambridge CB2 0PT, UK 136Department of Medicine, University of Oulo, Oulo, Finland 137Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA 138Unit of Periodontology, Department of Restorative Dentistry, Periodontology and Endodontology, University Medicine Greifswald, D-17475 Greifswald, Germany 139Department of Internal Medicine I, Ulm University Medical Centre, D-89081 Ulm, Germany 140Institute of Human Genetics, Helmholtz Zentrum München - German Research Center for Environmental Health, D-85764 Neuherberg, Germany 141Department of Medical Sciences, Cardiovascular Epidemiology, Uppsala University, Uppsala 75185, Sweden 142Centre for Bone and Arthritis Research, Department of Internal Medicine and Clinical Nutrition, Institute of Medicine, Sahlgrenska Academy, University of

Natr Genet. Author manuscript; available in PMC 2015 May 01.
Gothenburg, Gothenburg 413 45, Sweden 143 The Genetics of Obesity and Related Metabolic Traits Program, The Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA 144 Steno Diabetes Center A, S, Gentofte DK-2820, Denmark 145 Service of Nephrology, Department of Medicine, Lausanne University Hospital (CHUV), Lausanne 1005, Switzerland 146 School of Social and Community Medicine, University of Bristol, Bristol BS8 2BN, UK 147 Tropical Metabolism Research Unit, Tropical Medicine Research Institute, The University of the West Indies, Mona, Kingston 7, Jamaica 148 Global Health Institute, Department of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland 149 Institute of Microbiology, University Hospital and University of Lausanne, Lausanne 1011, Switzerland 150 The Center for Observational Research, Amgen, Inc., Thousand Oaks, CA 91320, USA 151 Center for Evidence-based Healthcare, University Hospital Carl Gustav Carus, Technische Universität Dresden, D-01307 Dresden, Germany 152 Department of Medicine I, University Hospital Grosshadern, Ludwig-Maximilians-Universität, D-81377 Munich, Germany 153 Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, D-85764 Neuherberg, Germany 154 Deutsches Forschungszentrum für Herz-Kreislauferkrankungen (DZHK) (German Research Centre for Cardiovascular Research), Munich Heart Alliance, D-80636 Munich, Germany 155 Department of Respiratory Medicine, Sir Charles Gairdner Hospital, NEDLANDS, Western Australia 6009, Australia 156 Department of Epidemiology and Biostatistics, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands 157 Department of Internal Medicine, Division of Endocrinology and Metabolism, Medical University of Graz, 8036 Graz, Austria 158 Diabetology-Endocrinology-Nutrition, AP-HP, Bichat Hospital, F-75018 Paris, France 159 INSERM, U872, Centre de Recherche des Cordeliers, F-75006 Paris, France 160 Paris Diderot University, F-75018 Paris, France 161 Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz 8036, Austria 162 Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA 163 Deutsches Herzzentrum München, Technische Universität München, D-80636 Munich, Germany 164 National Cancer Institute, Bethesda, MD, USA 165 Department of Sociology, University of Helsinki, Helsinki FI-00014, Finland 166 EMGO Institute for Health and Care Research, VU University, 1081BT Amsterdam, The Netherlands 167 Department of Psychiatry, Neuroscience Campus, VU University Amsterdam, Amsterdam, The Netherlands 168 Icelandic Heart Association, Kopavogur 201, Iceland 169 University of Iceland, Reykjavik 101, Iceland 170 Molecular & Cellular Therapeutics, Royal College of Surgeons in Ireland, 123 St Stephens Green, Dublin 2, Ireland 171 Institute of Cardiovascular and Medical Sciences, Faculty of Medicine, University of Glasgow, Glasgow G12 8TA, UK 172 Department of Medical Sciences, Molecular Medicine, Uppsala University, Uppsala 75144, Sweden 173 Department of Public Health Sciences, Stritch School of Medicine, Loyola University of Chicago, Maywood, IL 61053, USA 174 deCODE Genetics, Amgen inc., Reykjavik 101, Iceland 175 Department of Ocylog, University of Cambridge, Cambridge CB2 0Q, UK

Nat Genet. Author manuscript; available in PMC 2015 May 01.
Department of Internal Medicine section of Geriatrics, Academic Medical Center, Amsterdam, The Netherlands

Department of Child and Adolescent Psychiatry, Psychology, Erasmus University Medical Centre, 3000 CB Rotterdam, The Netherlands

Department for Health Evidence, Radboud University Medical Centre, 6500 HB Nijmegen, The Netherlands

Department of Genetics, Radboud University Medical Centre, 6500 HB Nijmegen, The Netherlands

Department of Clinical Chemistry, Ulm University Medical Centre, D-89081 Ulm, Germany

Deutsches Forschungszentrum für Herz-Kreislauferkrankungen (DZHK) (German Research Centre for Cardiovascular Research), partner site Hamburg/Lubeck/Kiel, Lubeck, Germany

Institut für Integrative und Experimentelle Genomik, Universität zu Lübeck, D-23562 Lübeck, Germany

Department of Community Medicine, Faculty of Health Sciences, UiT The Arctic University of Tromsø, Tromsø, Norway

MRC Unit for Lifelong Health and Ageing at UCL, London WC1B 5JU, UK

Department of Epidemiology and Public Health, EA3430, University of Strasbourg, Faculty of Medicine, Strasbourg, France

Department of Internal Medicine, University Medical Center Groningen, University of Groningen, 9700RB Groningen, The Netherlands

Pathology and Laboratory Medicine, The University of Western Australia, Western Australia 6009, Australia

Cedars-Sinai Diabetes and Obesity Research Institute, Los Angeles, CA, USA

Department of Prosthetic Dentistry, Gerostomatology and Dental Materials, University Medicine Greifswald, D-17475 Greifswald, Germany

Biological Psychology, VU University Amsterdam, 1081BT Amsterdam, The Netherlands

Institute of Social and Preventive Medicine (IUMSP), Centre Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, Switzerland

Ministry of Health, Victoria, Republic of Seychelles

Laboratory Medicine, Hospital of Desio, department of Health Sciences, University of Milano, Bicocca, Italy

Clinical Pharmacology Unit, University of Cambridge, Addenbrooke’s Hospital, Hills Road, Cambridge CB2 2QQ, UK

Clinical Pharmacology and Barts and The London Genome Centre, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, UK

Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville TN 37203, USA

Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN 37232, USA

Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA

Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

Division of Cardiovascular Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands

Clinic of Cardiology, West-German Heart Centre, University Hospital Essen, Essen, Germany

Department of General Practice and Primary Health Care, University of Helsinki, FI-00290 Helsinki, Finland

Unit of General Practice, Helsinki University Central Hospital, Helsinki 00290, Finland

Department of Internal Medicine, University of Pisa, Pisa, Italy

CNR Institute of Clinical Physiology, University of Pisa, Pisa, Italy

Department of Wood et al. Nat Genet. Author manuscript; available in PMC 2015 May 01.
Cardiology, Toulouse University School of Medicine, Rangueil Hospital, Toulouse, France
Robertson Center for Biostatistics, University of Glasgow, Glasgow, UK
NorthShore University HealthSystem, Evanston, IL, University of Chicago, Chicago, IL, USA
Service of Therapeutic Education for Diabetes, Obesity and Chronic Diseases, Geneva University Hospital, Geneva CH-1211, Switzerland
Vanderbilt University School of Medicine, Department of Medicine, Pharmacology, Pathology, Microbiology and Immunology, Nashville, Tennessee, USA
Leeds MRC Medical Bioinformatics Centre, University of Leeds, UK
Institute of Biomedical & Clinical Science, University of Exeter, Barrack Road, Exeter, EX2 5DW
Department of Psychiatry, Washington University School of Medicine, St. Louis, MO 63110, USA
Center for Biomedicine, European Academy Bozen, Bolzano (EURAC), Bolzano 39100, Italy
Affiliated Institute of the University of Lübeck, D-23562 Lübeck, Germany
Division of Genomic Medicine, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA
Institute of Cardiovascular Science, University College London, WC1E 6BT, UK
Department of Vascular Medicine, Academic Medical Center, Amsterdam, The Netherlands
Centre for Cardiovascular Genetics, Institute Cardiovascular Sciences, University College London, London WC1E 6JJ, UK
Cardiovascular Genetics Division, Department of Internal Medicine, University of Utah, Salt Lake City, Utah 84108, USA
School of Population Health and Sansom Institute for Health Research, University of South Australia, Adelaide 5000, Australia
South Australian Health and Medical Research Institute, Adelaide, Australia
Centre for Paediatric Epidemiology and Biostatistics, UCL Institute of Child Health, London WC1N 1EH, UK
National Institute for Health and Welfare, FL-90101 Oulu, Finland
MRC Health Protection Agency (HPE) Centre for Environment and Health, School of Public Health, Imperial College London, UK
Unit of Primary Care, Oulu University Hospital, FI-90220 Oulu, Finland
Biocenter Oulu, University of Oulu, FI-90014 Oulu, Finland
Institute of Health Sciences, FI-90014 University of Oulu, Finland
Hjelt Institute Department of Public Health, University of Helsinki, FI-00014 Helsinki, Finland
Department of Forensic Molecular Biology, Erasmus MC, 3015GE Rotterdam, The Netherlands
UKCRC Centre of Excellence for Public Health (NI), Queens University of Belfast, Northern Ireland
Faculty of Medicine, Institute of Health Sciences, University of Oulu, Oulu, Finland
Unit of General Practice, Oulu University Hospital, Oulu, Finland
Department of Urology, Radboud University Medical Centre, 6500 HB Nijmegen, The Netherlands
Imperial College Healthcare NHS Trust, London W12 0HS, UK
National Heart and Lung Institute, Imperial College, London W12 0NN, UK
Department of Epidemiology and Public Health, UCL London, WC1E 6BT, UK
Department of Medicine, Kuopio University Hospital and University of Eastern Finland, FI-70210 Kuopio, Finland
Department of Physiology, Institute of Biomedicine, University of Eastern Finland, Kuopio Campus, Kuopio, Finland
Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital and University of Eastern Finland, Kuopio, Finland
Department of Clinical Chemistry, Fimlab Laboratories and School of Medicine, Toulouse University School of Medicine, Rangueil Hospital, Toulouse, France
Medicine University of Tampere, FI-33520 Tampere, Finland
Department of Health Sciences, University of Milano, I 20142, Italy
Fondazione Filarete, Milano I 20139, Italy
Division of Nephrology and Dialysis, San Raffaele Scientific Institute, Milano I 20132, Italy
Università Vita-Salute San Raffaele, Milano I 20132, Italy
Institut Universitaire de Cardiologie et de Pneumologie de Québec, Faculty of Medicine, Laval University, Quebec, QC G1V 0A6, Canada
Institute of Nutrition and Functional Foods, Laval University, Quebec, QC G1V 0A6, Canada
Department of Biostatistics, University of Washington, Seattle, WA 98195, USA
Department of Surgery, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands
Department of Biostatistics, University of Liverpool, Liverpool L69 3GA, UK
Department of Pediatrics, University of Iowa, Iowa City, Iowa IA 52242, USA
Institute of Epidemiology II, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany, D-85764 Neuherberg, Germany
Department of Neurology, General Central Hospital, Bolzano 39100, Italy
Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA
Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, FI-20521 Turku, Finland
Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, FI-20521 Turku, Finland
Human Genomics Laboratory, Pennington Biomedical Research Center, Baton Rouge, LA 70808, USA
Center for Systems Genomics, The Pennsylvania State University, University Park, PA 16802, USA
Croatian Centre for Global Health, Faculty of Medicine, University of Split, 21000 Split, Croatia
Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester LE3 9QP, UK
National Institute for Health Research (NIHR) Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester, LE3 9QP, UK
South Carelia Central Hospital. 53130 Lappeenranta, Finland
Paul Langerhans Institute Dresden, German Center for Diabetes Research (DZD), Dresden, Germany
International Centre for Circulatory Health, Imperial College London, London W2 1PG, UK
Program for Personalized and Genomic Medicine, and Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, MD 21201, USA
Geriatric Research and Education Clinical Center, Vetrans Administration Medical Center, Baltimore, MD 21201, USA
HUCHART Heart and Lungcenter, Department of Medicine, Helsinki University Central Hospital, FI-00290 Helsinki, Finland
Université de Montréal, Montreal, Quebec H1T 1C8, Canada
Department of Kinesiology, Laval University, Quebec, QC G1V 0A6, Canada
Dipartimento di Scienze Farmacologiche e Biomolecolari, Università di Milano & Centro Cardiologico Monzino, IRCCS, Milan 20133, Italy
Department of Food Science and Nutrition, Laval University, Quebec, QC G1V 0A6, Canada
The electronic medical records and genomics (eMERGE) consortium
Myocardial Infarction Genetics (MIGen) Consortium
Membership to this consortium is provided below
Population Architecture using Genomics and Epidemiology Consortium
The LifeLines Cohort Study, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
Institut Pasteur de Lille; INSERM, U744; Université de Lille 2;
F-59000 Lille, France 279 Department of Cardiology, Division Heart and Lungs, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands 280 Durrer Center for Cardiogenetic Research, Interuniversity Cardiology Institute Netherlands-Netherlands Heart Institute, 3501 DG Utrecht, The Netherlands 281 Lee Kong Chian School of Medicine, Imperial College London and Nanyang Technological University, Singapore, 637553 Singapore, Singapore 282 Health Science Center at Houston, University of Texas, Houston, TX, USA 283 Department of Medicine, Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA 284 Department of Epidemiology, University Medical Center Utrecht, Utrecht, The Netherlands 285 Lund University Diabetes Centre and Department of Clinical Science, Diabetes & Endocrinology Unit, Lund University, Malmö 221 00, Sweden 286 Harvard School of Public Health, Department of Epidemiology, Harvard University, Boston, MA 2115, USA 287 Interuniversity Cardiology Institute of the Netherlands (ICIN), Utrecht, the Netherlands 288 Albert Einstein College of Medicine. Department of epidemiology and population health, Belfer 1306, NY 10461, USA 289 Center for Human Genetics, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA 290 Synlab Academy, Synlab Services GmbH, Mannheim, Germany 291 Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands 292 Harvard Medical School, Boston, MA 02115, USA 293 Institute for Translational Genomics and Population Sciences, Los Angeles BioMedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA 294 Finnish Diabetes Association, Kirjoniementie 15, FI-33680 Tampere, Finland 295 Pirkanmaa Hospital District, Tampere, Finland 296 Center for Non-Communicable Diseases, Karachi, Pakistan 297 Department of Medicine, University of Pennsylvania, Philadelphia, USA 298 Laboratory of Genetics, National Institute on Aging, Baltimore, MD 21224, USA 299 Instituto de Investigacion Sanitaria del Hospital Universario LaPaz (IdiPAZ), Madrid, Spain 300 Diabetes Research Group, King Abdulaziz University, Jeddah, Saudi Arabia 301 Centre for Vascular Prevention, Danube-University Krems, 3500 Krems, Austria 302 Department of Public Health and Clinical Nutrition, University of Eastern Finland, Finland 303 Research Unit, Kuopio University Hospital, Kuopio, Finland 304 Institute of Cellular Medicine, Newcastle University, Newcastle NE1 7RU, UK 305 Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, D-85764 Munich, Germany 306 Klinikum Grosshadern, D-81377 Munich, Germany 307 Institute of Epidemiology I, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany, D-85764 Neuherberg, Germany 308 Department of Pulmonology, University Medical Center Utrecht, Utrecht, The Netherlands 309 King Abdulaziz University, Jeddah 21589, Saudi Arabia 310 Department of Internal Medicine, Division of Gastroenterology, and Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109 311 Faculty of Medicine, University of Iceland, Reykjavik 101, Iceland 312 University of Cambridge Metabolic Research Laboratories, Institute of Metabolic Science, Addenbrooke’s Hospital, Cambridge CB2 0QQ, UK 313 NHIR Cambridge
Acknowledgments

A full list of acknowledgments appears in the Supplementary Note.

References


Nat Genet. Author manuscript; available in PMC 2015 May 01.
13. Segre AV, Groop L, Mootha VK, Daly MJ, Altshuler D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. PLoS Genet. 2010; 6


**AUTHOR CONTRIBUTIONS**

*Steering Committee Overseeing the Consortium*

Writing Group (wrote, edited and commented on manuscript)

Data preparation group (checked and prepared data from contributing cohorts for meta-analyses)

Height meta-analyses group (GWAS and Metabochip) (analyses specific to the manuscript)

Mixed linear model analyses

Large lambda group

Family transmission analyses

Variance, heritability, and prediction analyses

Biological Enrichment and Pathway analyses

ENCODE working group
M.L.B., G.L. (chair), K.S.L.
Gene expression (eQTL) working group


Other Contributions

(DEPICT) R.F., L. Franke, J. Karjalainen, T.H.P.

Project Design, Management and Coordination of Contributing Studies

Previous GWAS studies


New GWAS studies


Metabochip studies

Genotyping of Contributing Studies

**Previous GWAS studies**

(AGES) A.V. Smith; (B58C T1D CONTROLS) W.L.M.; (B58C WTCCC) W.L.M.; (CAPS) H. Grönberg; (CROATIA-Vis) C. Hayward; (EGCUT) M. Nelis; (EPIC-Norfolk) N.J.W.; (FENLAND) N.J.W.; (Finnish Twin Cohort) J. Kaprio; (KORA S3) T.I., M. Müller-Nurasyid; (MICROS) A.A.H; (NFBC66) M.J; (ORCADES) A.F.W.; (PLCO) S.J.C.; (RSI) K.E., C. Medina-Gomez, F. Rivadeneira, A.G.U.; (SASBAC) P. Hall; (SHIP) A. Hannemann, M. Nauck; (WGHS) D.I.C., L.M.R.; (WTCCC-CAD) A.S.H. N.J.S.; (WTCCC-T2D) A.T.H, M.I.M.; (Young Finns Study (YFS)) T.L., O.T.R.

**New GWAS studies**


**Metabochip studies**

Phenotype Coordination of Contributing Studies

Previous GWAS studies


New GWAS studies


Metabochip studies

Data Analysis

Previous GWAS studies

(AGES) A.V. Smith; (ARIC) K.L. Monda, K.E.N.; (B58C T1D CONTROLS) D.P.S.;
(B58C WTCCC) D.P.S.; (CAPS) E.I.; (CHS) R.C.K., B.M.; (COLAUS) S. Bergmann, Z.K.;
(CROATIA-Vis) C. Hayward; (deCODE) V. Steinhorsdottir, G.T.; (EGCUT) M. Nelis;
(EPIC-Norfolk) J.H.Z.; (FENLAND) J. Luan; (FRAM) L.A.C., N.L.H.; (FUSION) C.J.W.;
(GerMIFS II) C.W.; (H2000) N.E.; (HPFS) L.Q.; (NHS) L.Q. (NSPHS) A. Johansson;
(PLCO) S.I.B.; (RSI) K.E., C. Medina-Gomez, F. Rivadeneira, A.G.U.; (RUNMC) S.H.V.;
(SardiNIA) S. Sanna; (SASBAC) E.I.; (SEARCH) J.P.T.; (SHIP) A. Teumer; (WGHS)
D.I.C., L.M.R, A.Y.C; (WTCCC-T2D) A.P.M., T. Ferreira; A. Mahajan, R.M.

New GWAS studies

(ATCG) P.I.W.d.B., P.J.M., S.R.; (Athero-Express Biobank Studies) S.W.v.d.L.; (B-
PROOF) S.v.D.; (BHS) M.C.; (BLSA) T.T.; (CLHNS) D. Croteau-Chonka, (DESIR) S.
(ERF) N.A., A.D.; (FamHS) M.F.F.; (GOOD) M.C.; (HBCS) N.E.; (Health
ABC) M.A.N.; (HERITAGE Family Study) C. Bouchard, M. A. Sarzynski, D.C.R., T.R.,
T.K.R, Y.J.S., (HYPERGENES) S. Lupoli; (InCHIANTI) D.P., T.T., A.R.W.; (IPM) J. Jeff,
V. Lotay, Y. Lu; (LifeLines) I.M.N., J.V.V.V.; (LLS) M. Beekman, J.J.H.; (LOLIPOP)
W.Z; (MGS) J. Shi, (NELSON) S.R., J.V.S; (PLCO2) S.I.B., Z.W.; (PREVEND) P.v.d.H.,
I.M.L., N.V.; (PROCARDIS) A. Goel; (PROSPER/PHASE) I.F., B.M.B., S.T.; (QFS) J.
Blangero, L.P.; (QIMR) G. Hemani, D.R.N., J.E.P.; (RISC) D.P., A.R.W.; (RSII) K.E., C.
Medina-Gomez, F. Rivadeneira, A.G.U.; (RSIII) K.E., C. Medina-Gomez, F. Rivadeneira,
A.G.U.; (SHIP-TREND) A. Teumer; (SORBS) R.M., (TRAILS) H. Snieder; (TWINGENE)
E.I., S.G.; (TwinsUK) M. Mangino; (WGHS) D.I.C., L.M.R.

Metabochip Studies

(ADVANCE) D. Absher, T.L.A, L.L.W.; (AMCPAS) S. Kanoni; (ARIC) S. Buyske, A.E.J.,
K.E.N.; (B1958C) T. Ferreira; (BHS) D. Anderson; (CARDIOGENICS) S. Kanoni;
(DESIR) S. Cauchi, L.Y.; (DGE DietGeneExpression) I.H.C.; (DIAGEN) A.U.J., G.M.;
(DILGOM) K.K.; (DUNDEE) T. Ferreira; (EAS) J.L.B., R.M.F.; (EGCUT) T.E., K.F.,
E.M.; (ELY) J. Luan; (EMIL (SWABIA)) B.O.B.; (EPIC-Norfolk) J. Luan. (FBPP) A.C.,
G.B.E.; (FENLAND) J. Luan; (GLACIER) F. Renstrom, D. Shungin; (GxE) C.D.P., (HNR)
S. Pechlivanis, A. Scherag; (IMPROVE) L. Folkersen, R.J.S.; (KORA S3) J.S.R.; (KORA
S4) E.A.; (Leipzig) A. Mahajan, I.P.; (LURIC) G. Delgado, T.B.G., M.E.K., S. Pilz, H.
Scharnag; (MEC) U.L., F.R.S.; (METSIM) A. Stancáková; (NSHD) A.W., J. Luan;

*Nat Genet*. Author manuscript; available in PMC 2015 May 01.
Figure 1. Regional association plots for loci with multiple association signals
Panels a to d highlight examples of multiple signals after approximate conditional joint multiple-SNP analysis GCTA-COJO analysis. SNPs are shaded and shaped based on the index SNP with which they are in strongest LD ($r^2 > 0.4$). Panels a–c show the majority of signals clustering in and around a single gene (ACAN, ADAMTS17, PTCH1, respectively) whereas panel d shows the multiple signals clustering through proximity.
Figure 2. Quantifying the variance explained by height associated SNPs at different levels of significance

The SNPs were selected from the approximate conditional and joint multiple SNPs association analysis using GCTA-COJO analysis with the target cohort being excluded from the meta-analysis. a, b, c, Partitioning the variance in the SNP-derived genetic predictor using a within-family analysis. The SNP-based predictor was adjusted by the first 20 PCs. The four variance-covariance components $V_g$, $V_e$, $C_g$ and $C_e$ are defined in Online Methods. d, Accuracy of predicting phenotype by the genetic predictor in unrelated individuals. The prediction $R^2$ shown on the y-axis is the squared correlation between phenotype and predictor. The SNP-based predictor was adjusted by the first 20 PCs. The solid line is the average prediction $R^2$ weighted by sample size over the five cohorts. The dashed line is the prediction accuracy inferred from the within-family prediction analysis (Equation 19 in Online Methods). e, The variance explained by the SNPs was estimated by the whole-genome estimation method in GCTA. The phenotype was adjusted by the first 20 PCs. Each error bar represents the standard error of the estimate. The estimates from all the five cohorts (B-PROOF, FRAM, QIMR, TwinGene and WTCCC-T2D) were averaged by the inverse-variance approach. The dashed line is the variance explained inferred from the within-family prediction analysis. In panels d and e, the number shown in each column is the number of SNPs used in the analysis.
Figure 3. Tissue enrichment combined with pruned gene set network

Genes within genome-wide significant height associated loci enriched for several relevant tissue annotations as well as gene sets. **a**, Genes in associated loci tended to be highly expressed in tissues related to chondrocytes and osteoblasts (cartilage, joints, and spine), and other musculoskeletal, cardiovascular and endocrine tissue-types. The analysis was conducted based on the DEPICT method and 37,427 human microarray samples. Tissue annotations are sorted by physiological system and significance. Significantly enriched (FDR<0.05) tissues are color-coded in black. **b**, Significantly enriched reconstituted gene sets ($P$-value<$1\times10^{-11}$, FDR<$1\times10^{-5}$) identified by DEPICT. Nodes represent reconstituted gene sets and are color-coded by statistical significance. Edge thickness between nodes is proportional to degree of gene overlap as measured by the Jaccard index. Nodes with gene overlap greater than 25% were collapsed into single meta nodes and marked by blue borders. **c**, reconstituted gene sets comprised by the Chordate Embryonic Development meta node,
which represented several gene sets relevant to human height (e.g. ossification, embryonic skeletal system development, and limb development).
Table 1

Estimates of variance explained by SNPs selected at different significance levels

The SNPs were selected by an approximate conditional and joint multiple-SNP analysis (GCTA-COJO) of the summary statistics from the meta-analysis. The target cohort for variance estimation was excluded from the meta-analysis.

<table>
<thead>
<tr>
<th>Threshold</th>
<th>QIMR (n = 3,924)</th>
<th>FRAM (n = 1,145)</th>
<th>TwinGene (n = 5,668)</th>
<th>WTCCC-T2D (n = 1,914)</th>
<th>B-PROOF(n = 2,555)</th>
<th>a Weighted average</th>
<th>b Pred.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#SNP</td>
<td>$h^2_g$</td>
<td>SE</td>
<td>#SNP</td>
<td>$h^2_g$</td>
<td>SE</td>
<td>#SNP</td>
</tr>
<tr>
<td>SE-8</td>
<td>675</td>
<td>0.164</td>
<td>0.016</td>
<td>656</td>
<td>0.190</td>
<td>0.040</td>
<td>670</td>
</tr>
<tr>
<td>SE-7</td>
<td>887</td>
<td>0.187</td>
<td>0.017</td>
<td>862</td>
<td>0.210</td>
<td>0.045</td>
<td>866</td>
</tr>
<tr>
<td>SE-6</td>
<td>950</td>
<td>0.212</td>
<td>0.020</td>
<td>1891</td>
<td>0.183</td>
<td>0.060</td>
<td>1918</td>
</tr>
<tr>
<td>SE-5</td>
<td>3754</td>
<td>0.248</td>
<td>0.024</td>
<td>3671</td>
<td>0.239</td>
<td>0.080</td>
<td>3689</td>
</tr>
<tr>
<td>SE-3</td>
<td>9693</td>
<td>0.297</td>
<td>0.035</td>
<td>9403</td>
<td>0.171</td>
<td>0.126</td>
<td>9548</td>
</tr>
<tr>
<td>HM3</td>
<td>1.08M</td>
<td>0.473</td>
<td>0.086</td>
<td>1.06M</td>
<td>0.313</td>
<td>0.291</td>
<td>1.12M</td>
</tr>
</tbody>
</table>

*The estimates from all the five cohorts were averaged by the inverse-variance approach i.e. $\sum_i (h^2_g(i)/SE^2(i))/\sum_i (1/SE^2(i))$; *b* the predicted variance explained by the selected SNPs ($V_g$) from the within-family prediction analysis; *c* SNPs from HapMap3 project11.
Table 2
Comparison of prioritized variants, loci, biology and variance explained from GWASs on human stature with 130,000 individuals (previously published in Lango Allen et al., 2010) and with 250,000 individuals (this paper).

<table>
<thead>
<tr>
<th>SNP based comparisons</th>
<th>Height GWAS with 130,000 samples (Lango Allen et al., Yang et al.)*</th>
<th>Height GWAS with 253,288 samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>GWAS significant SNPs</td>
<td>199</td>
<td>697</td>
</tr>
<tr>
<td>Genomic loci# (±− 1Mb)</td>
<td>180</td>
<td>423</td>
</tr>
<tr>
<td>Loci# with multiple signals</td>
<td>19</td>
<td>147</td>
</tr>
<tr>
<td>Secondary associations in loci#</td>
<td>19</td>
<td>273</td>
</tr>
</tbody>
</table>

**Biological annotation (DEPICT at FDR < 0.05)**

| Prioritised genes | 92 | 649 |
| Loci& with prioritised gene | 74 (43%) | 422 (75%) |
| Pruned gene sets and protein-protein complexes% | 813 | 2,330 |
| Tissues and cell-types | 5 | 43 |

**Variance explained**

| GWAS significant SNPs | 10% | 16% |
| Deep list of SNPs at 1×10⁻³ | 13% | 29% |
| All common SNPs | 45% ** | 50% |

**Heritability explained**

| GWAS significant SNPs | 12.5% | 20% |
| Deep list of SNPs at 1×10⁻³ | 16% | 36% |
| All common SNPs | 56% ** | 62.5% |

* Counts, numbers and estimates for Lango Allen et al. are taken from respective publication.

# Genomic loci defined by distance: ±− 1Mb from index height SNP

& Genomic loci defined by LD: r² > 0.5 from index height SNP

% After clamping of similar gene sets and pathways

## Significantly prioritized novel human growth associated genes

The table lists 20 genes prioritized by DEPICT. Genes are ranked by the number of lines of supporting evidence and the DEPICT $P$-value (Supplementary Table 16). Because 20 of the 30 top-ranked genes were in a curated list of genes known to cause syndromes of skeletal\textsuperscript{12}, these “OMIM genes” are not shown here. The top fifteen genes with prior literature support (based on GRAIL) are shown, followed by the top five novel genes. Each gene is accompanied by the significantly enriched reconstituted gene sets in which it appears in (DEPICT gene set enrichment analysis). Abbreviations; (GO – Gene Ontology; MP– Mice Phenotypes from Mouse Genome Informatics database; InWeb – protein-protein interaction complexes; KEGG and REACTOME databases).

<table>
<thead>
<tr>
<th>Locus (height SNP)</th>
<th>Gene symbol</th>
<th>New locus</th>
<th>Prioritization $P$-value</th>
<th>Levels of biological annotation</th>
<th>Top ranking reconstituted gene sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs10784128</td>
<td>FRS2</td>
<td>N</td>
<td>$1.0\times10^{-16}$</td>
<td>7</td>
<td>PI 3K cascade (REACTOME, $P=6.2\times10^{-13}$); Chronic Myeloid Leukemia (KEGG, $P=1.6\times10^{-12}$); Response To Fibroblast Growth Factor Stimulus (GO, $P=5.4\times10^{-11}$);</td>
</tr>
<tr>
<td>rs2166898</td>
<td>GLI2</td>
<td>Y</td>
<td>$4.4\times10^{-16}$</td>
<td>7</td>
<td>Growth Factor Binding (GO, $P=2.6\times10^{-14}$); Regulation Of Osteoblast Differentiation (GO, $P=2.2\times10^{-13}$); WNT-Protein Binding (GO, $P=1.9\times10^{-12}$);</td>
</tr>
<tr>
<td>rs526896-rs9327705</td>
<td>TBX4</td>
<td>N</td>
<td>$9.9\times10^{-9}$</td>
<td>7</td>
<td>Short Mandible (MP, $P=3.3\times10^{-19}$); Respiratory System Development (GO, $P=3.1\times10^{-17}$); Abnormal Ulna Morphology (MP, $P=1.9\times10^{-11}$);</td>
</tr>
<tr>
<td>rs16860216</td>
<td>SOX8</td>
<td>N</td>
<td>0.016</td>
<td>7</td>
<td>Small Thoracic Cage (MP, $P=6.9\times10^{-14}$); Short Ribs (MP, $P=2.7\times10^{-8}$); Short Sternum (MP, $P=6.5\times10^{-7}$);</td>
</tr>
<tr>
<td>rs1199734</td>
<td>LATS2</td>
<td>Y</td>
<td>$1.0\times10^{-16}$</td>
<td>6</td>
<td>Partial Lethality Throughout Fetal Growth And Development (MP, $P=1.2\times10^{-18}$); Growth Factor Binding (GO, $P=2.6\times10^{-15}$); TGFB1 protein complex (InWeb, $P=6.3\times10^{-12}$);</td>
</tr>
<tr>
<td>rs1232101</td>
<td>PDS5B</td>
<td>N</td>
<td>$1.0\times10^{-16}$</td>
<td>6</td>
<td>Chromatin Binding (GO, $P=6.4\times10^{-17}$); Nuclear Hormone Receptor Binding (GO, $P=2.4\times10^{-12}$); RBBP4 protein complex (InWeb, $P=1.3\times10^{-13}$); WNT16 protein complex (InWeb, $P=1.9\times10^{-9}$);</td>
</tr>
<tr>
<td>rs6746356</td>
<td>SP3</td>
<td>Y</td>
<td>$1.0\times10^{-16}$</td>
<td>6</td>
<td>BCOR protein complex (InWeb, $P=2.7\times10^{-13}$); ARF2 protein complex (InWeb, $P=4.5\times10^{-7}$); Intracellular Steroid Hormone Receptor Signaling Pathway (GO, $P=5.0\times10^{-9}$);</td>
</tr>
<tr>
<td>rs3923086</td>
<td>AXIN2</td>
<td>Y</td>
<td>$2.2\times10^{-16}$</td>
<td>6</td>
<td>Signaling By Transforming Growth Factor Beta (KEGG, $P=3.8\times10^{-15}$); WNT Receptor Signaling Pathway (GO, $P=6.9\times10^{-14}$); Polydactyly (MP, $P=1.5\times10^{-12}$);</td>
</tr>
<tr>
<td>rs3790086</td>
<td>LTBP1</td>
<td>N</td>
<td>$1.3\times10^{-13}$</td>
<td>6</td>
<td>Abnormal Skeleton Morphology (MP, $P=1.1\times10^{-13}$); TGF Beta Signaling Pathway (KEGG, $P=3.8\times10^{-15}$); Growth Factor Binding (GO, $P=2.6\times10^{-14}$);</td>
</tr>
<tr>
<td>rs3034172</td>
<td>WNT5A</td>
<td>Y</td>
<td>$4.3\times10^{-12}$</td>
<td>6</td>
<td>Partial Lethality Throughout Fetal Growth And Development (MP, $P=1.2\times10^{-18}$); Tissue Morphogenesis (GO, $P=4.1\times10^{-20}$); Abnormal Skeleton Morphology (MP, $P=1.1\times10^{-15}$);</td>
</tr>
<tr>
<td>rs3915129</td>
<td>CTNNB1</td>
<td>Y</td>
<td>$3.5\times10^{-12}$</td>
<td>6</td>
<td>AR protein complex (InWeb, $P=8.9\times10^{-13}$); TCEB1 protein complex (InWeb, $P=1.5\times10^{-13}$); GTF2I protein complex (InWeb, $P=4.6\times10^{-11}$);</td>
</tr>
<tr>
<td>Locus (height SNP)</td>
<td>Gene symbol</td>
<td>New locus</td>
<td>Prioritization P-value</td>
<td>Levels of biological annotation</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------</td>
<td>-----------</td>
<td>------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>rs12330322</td>
<td>BMP2</td>
<td>N</td>
<td>5.6x10^{-10}</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Top ranking reconstituted gene sets</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Transcription Factor Binding (GO, P=4.7x10^{-26}); Complete Embryonic Lethality During Organogenesis (MP, P=4.9x10^{-21}); Short Mandible (MP, P=3.3x10^{-19})</td>
<td></td>
</tr>
<tr>
<td>rs10958476-rs6999671</td>
<td>BMP6</td>
<td>N</td>
<td>2.9x10^{-8}</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Small Basisphenoid Bone (MP, P=8.9x10^{-13}); TGF Beta Signaling Pathway (KEGG, P=3.8x10^{-15}); Growth Factor Binding (GO, P=2.6x10^{-13})</td>
<td></td>
</tr>
<tr>
<td>rs568914</td>
<td>SOX5</td>
<td>Y</td>
<td>4.6x10^{-7}</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Disproportionate Dwarf (MP, P=1.8x10^{-13}); Abnormal Cartilage Morphology (MP, P=1.9x10^{-13}); Short Limbs (MP, P=2.8x10^{-13})</td>
<td></td>
</tr>
<tr>
<td>rs17807185</td>
<td>WNT4</td>
<td>Y</td>
<td>4.6x10^{-7}</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Morphogenesis Of An Epithelium (GO, P=2.3x10^{-17}); Gland Development (GO, P=5.4x10^{-16}); Basal Cell Carcinoma (KEGG, P=1.5x10^{-12})</td>
<td></td>
</tr>
<tr>
<td>Novel genes without prior evidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs8042424</td>
<td>CHSY1</td>
<td>N</td>
<td>1.0x10^{-16}</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Abnormal Cartilage Morphology (MP, P=9.9x10^{-13}); Abnormal Bone Ossification (MP, P=2.1x10^{-15}); Signaling by Transforming Growth Factor Beta (REACTOME, P=5.9x10^{-5})</td>
<td></td>
</tr>
<tr>
<td>rs7652177</td>
<td>FNDC3B</td>
<td>N</td>
<td>1.0x10^{-16}</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Abnormal Spongiotrophoblast Layer Morphology (MP, P=3.2x10^{-16}); Decreased Length Of Long Bones (MP, P=2.7x10^{-12}); ITGB1 protein complex (InWeb, P=5.2x10^{-8})</td>
<td></td>
</tr>
<tr>
<td>rs7284476</td>
<td>TRIOBP</td>
<td>Y</td>
<td>1.0x10^{-16}</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Negative Regulation Of Cell Proliferation (GO, P=4.3x10^{-17}); Abnormal Vitelline Vasculature Morphology (MP, P=1.7x10^{-15}); Beta-Catenin Binding (GO, P=3.0x10^{-5})</td>
<td></td>
</tr>
<tr>
<td>rs2149163-rs3927536</td>
<td>BNC2</td>
<td>N</td>
<td>1.0x10^{-16}</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Short Ulna (MP, P=4.7x10^{-13}); Abnormal Joint Morphology (MP, P=8.6x10^{-11}); Regulation Of Chondrocyte Differentiation (GO, P=2.9x10^{-6})</td>
<td></td>
</tr>
<tr>
<td>rs3790086</td>
<td>WWP2</td>
<td>Y</td>
<td>1.0x10^{-16}</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cartilage Development (GO, P=2.0x10^{-19}); Chondrocyte Differentiation (GO, P=3.0x10^{-15}); Signaling By Platelet-Derived Growth Factor (REACTOME, P=4.8x10^{-10})</td>
<td></td>
</tr>
</tbody>
</table>