The role of common genetic variation in model polygenic and monogenic traits

Submitted by

Hana Langa Allen

of the Peninsula College of Medicine and Dentistry Graduate School to the Universities of Exeter and Plymouth as a thesis for the degree of Doctor of Philosophy in Molecular Genetics, July 2010

I certify that all material in this thesis which is not my own work has been identified and that no unchanged or acknowledged material has previously been submitted and approved for the award of a degree by this or any other University

……..

ABSTRACT

The aim of this thesis is to explore the role of common genetic variation, identified through genome-wide association (GWA) studies, in human traits and diseases, using height as a model polygenic trait, type 2 diabetes as a model common polygenic disease, and maturity onset diabetes of the young (MODY) as a model monogenic disease.

The wave of the initial GWA studies, such as the Wellcome Trust Case-Control Consortium (WTCCC) study of seven common diseases, substantially increased the number of common variants associated with a range of different multifactorial traits and diseases. The initial excitement, however, seems to have been followed by some disappointment that the identified variants explain a relatively small proportion of the genetic variance of the studied trait, and that only few large effect or causal variants have been identified. Inevitably, this has led to criticism of the GWA studies, mainly that the findings are of limited clinical, or indeed scientific, benefit.

Using height as a model, Chapter 2 explores the utility of GWA studies in terms of identifying regions that contain relevant genes, and in answering some general questions about the genetic architecture of highly polygenic traits.

Chapter 3 takes this further into a large collaborative study and the largest sample size in a GWA study to date, mainly focusing on demonstrating the biological relevance of the identified variants, even when a large number of associated regions throughout the genome is implicated by these associations. Furthermore, it shows examples of different features of the genetic architecture, such as allelic heterogeneity and pleiotropy.

Chapter 4 looks at the predictive value and, therefore, clinical utility, of variants found to associate with type 2 diabetes, a common multifactorial disease that is increasing in prevalence despite known environmental risk factors. This is a disease where knowledge of the genetic risk has potentially substantial clinical relevance.

Finally, Chapter 5 approaches the monogenic-polygenic disease bridge in the direction opposite to that approached in the past: most studies have investigated genes mutated in monogenic diseases as candidates for harboring common variants predisposing to related polygenic diseases. This chapter looks at the common type 2 diabetes variants as modifiers of disease onset in patients with a monogenic but clinically heterogeneous disease, maturity onset diabetes of the young (MODY).

TABLE OF CONTENTS

Abstract 2

Table of contents 4

Acknowledgements 6

Author's declaration 8

Chapter 1: Introduction 9

- The genetic architecture of common complex disease 12
- Finding susceptibility genes for common complex disease: past approaches 15
- The era of the genome-wide association study 16
- Recent findings in genome-wide association studies 18
- Height as a model polygenic trait 23
- What benefits can GWAS findings bring to clinical practice 23
- Conclusions 26
- Aims of the Thesis 27

Chapter 2: Genome-wide association analysis identifies 20 loci that influence adult height 33

- Supplementary Information 65

Chapter 3: Hundreds of variants influence human height and cluster within genomic loci and biological pathways 94

- Supplementary Information 119

Chapter 4: Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk 216

- Supplementary Information 240

Chapter 5: Polygenic risk variants for type 2 diabetes susceptibility modify age at diagnosis in monogenic HNF1A diabetes 247

- Supplementary Information 268

Chapter 6: Discussion And Conclusions 273
Common variants have small effects and explain small proportion of the heritable component 275
Insights into genetic architecture and biological mechanisms 277
Disease prediction 278
Future directions 280
Conclusions 283

Appendix I: List of Publications 287

ACKNOWLEDGEMENTS

This thesis would not have been possible without my two great supervisors, Dr. Michael Weedon and Prof. Timothy Frayling. Their scientific knowledge and professionalism are truly outstanding, and my own academic achievements would not have been possible without their unreserved support, encouragement, advice and guidance. Right from the start of the thesis they showed great belief in me, and I can only hope that I have met their high expectations. They have also created a great working environment within the Genetics of Complex Traits group, and I have enjoyed immensely the privilege of being part of this successful team. I am particularly thankful to John Perry and Rachel Freathy, for their analytical and statistical help whenever I needed it, and Kirstie Parnell, for help with all lab-related matters.

I am indebted to Prof. Andrew Hattersley, the leader of the brilliant Molecular Genetics group, for so sincerely promoting individual group members, including myself. His encouragement has helped me build confidence in my own work and ability, and develop professionally. The discussions we had over the years turned quickly from daunting and challenging, to exciting and inspiring. For all things monogenic, invaluable has been help from Prof. Sian Ellard and other members from the molecular genetic diagnostic lab. I am especially grateful to Beverley Shields, Andrew Parish and Kevin Colclough for helping me with MODY samples and sequencing, without you guys I would have been lost!

I have made many life-long friends and colleagues over the last three years, too many to mention them all here, but I would like say a big thank you to my office colleagues Bob Morse, Claire Bennett, Emma Taylor, Anna Murray and John Perry for sharing lots of laughs, and occasional disappointments, with me. My PhD was funded by the Peninsula Medical School, and I am grateful to Prof. Angela Shore, the director of the Institute of Biomedical and Clinical Sciences, for giving me her whole-hearted support.

Above all, my greatest inspiration has always been my family. I would like to dedicate this PhD to my parents, who gave me all the opportunities in life they

possibly could, and my late grandmother, whose endless kindness and positive attitude will always remain my moral compass. I am also grateful for having such considerate and witty brother and sister to share life experiences with. Finally, I am very lucky to have the most wonderful husband and friend – thank you for the many years of love and support in every way imaginable, for sharing all the good moments and bad with me, and for making me so happy.

AUTHOR'S DECLARATION

I was involved in study design, analysis and manuscript preparation for all of the studies that are included as chapters in this thesis. Some of the studies include analyses performed by other authors; however, in each case I had a major role and was the first or joint first author on the paper. My specific contributions for each chapter are listed below.

Chapter 1: I wrote the original review article that forms most of the Introduction. Michael Weedon commented on and edited parts of the manuscript. The article briefly describes the Wellcome Trust Case-Control Consortium (WTCCC) study, and I was the main Exeter analyst of the WTCCC-T2D data.

Chapter 2: I was involved in study design, led the statistical analyses, and co-wrote the manuscript.

Chapter 3: I had considerable general and specific contributions to this large international collaborative study. I was an active member of the central analysis group and contributed to many analytical decisions made during the course of the study. I quality-checked hundreds of files submitted by over 50 different research groups, performed all main meta-analyses and several sub-analyses (secondary signals meta-analysis, overlap with other traits and diseases, population stratification, overlap with CNVs), and co-wrote the manuscript. I also wrote a Perl script for extracting genotype dosages from certain imputation files, which facilitated conditional analyses in several contributing cohorts.

Chapter 4: I designed and led the study, performed all statistical analyses and wrote the manuscript.

Chapter 5: I designed and led the study, performed all statistical analyses and wrote the manuscript. I also set up a collaboration with a Norwegian group who kindly sent me their raw data, which I then combined with my own and analysed.

CHAPTER 1: INTRODUCTION

Adapted from published review article:

Hana Lango and Michael N Weedon: What will Whole Genome Searches for Susceptibility Genes for Common Complex Disease Offer to Clinical Practice? *Journal of Internal Medicine* (2008); 263: 16-27

Many human diseases have a genetic component. Some diseases are caused entirely by a genetic mutation, and much success has been had in identifying the genes that, when mutated, cause these monogenic disorders [1]. Over 1500 genes for monogenic diseases such as cystic fibrosis and maturity onset diabetes of the young (MODY) have been identified (http://www.ncbi.nlm.nih.gov/sites/entrez?db=OMIM). Identifying these monogenic disease genes has led to deep insights into the biology of these and closely related diseases, and has led to the development of therapeutic measures [2]. Monogenic diseases, however, are relatively rare within populations, explaining only a small percentage of the overall disease burden.

In the developed world the majority of disease results from common, but complex disorders such as diabetes, obesity, and cancer. Environmental and lifestyle factors such as diet, exercise and smoking are important risk factors for the development of these diseases; however, twin, family, admixture and migration studies have also demonstrated a large genetic component to an individual's disease risk. Heritably estimates for some common diseases are given in **Table 1**. A heritability of 50% suggests that half of the variation in disease risk for an individual in a population can be explained by genetic variation.

Success in identifying the genes and variants which explain the genetic component of common complex disease has been slow, but recent advances in the understanding of the genetic architecture of complex traits and diseases, together with advances in high-throughput genotyping technology, has led to a new era of genetic analyses – genome-wide association studies – which are providing novel and important insights into common polygenic disorders. These findings could be of substantial clinical importance in the relatively near future.

Here, we first discuss the genetic architecture of common complex disease. We then describe how we go about finding polygenic disease genes using the genome-wide association study approach. We go on to present some recent, exciting findings in the genetics of complex disease, and introduce height as a

model polygenic trait. Finally, we discuss the potential clinical applications of finding common disease susceptibility variants and implicated genes.

Table 1. Heritability of some common complex diseases. * males, ** females.

Disease	**h^2 (95% CI / SE / SD)**	**References**
Age-related maculopathy Age-related macular degeneration	0.45 (CI 0.35-0.53) 0.46-0.71	3 4
Crohn's disease	1.00 (CI 0.80-1.00	5
Prostate cancer	0.42 (CI 0.29-0.50)	6
Breast cancer	0.27 (CI 0.04-0.41)	6
Type 2 diabetes	0.26 (CI 0.0-0.85)	7
Body mass index	0.54 (SE 0.05) 0.40 (SD 0.075)	8 9
Coronary artery disease (CAD) Death from CAD	0.49 (SE 0.12) 0.57 (CI 0.45–0.69)* 0.38 (0.26–0.50)**	10 11 11
Hypertension	0.80 (SE 0.19)	12

The genetic architecture of common complex disease

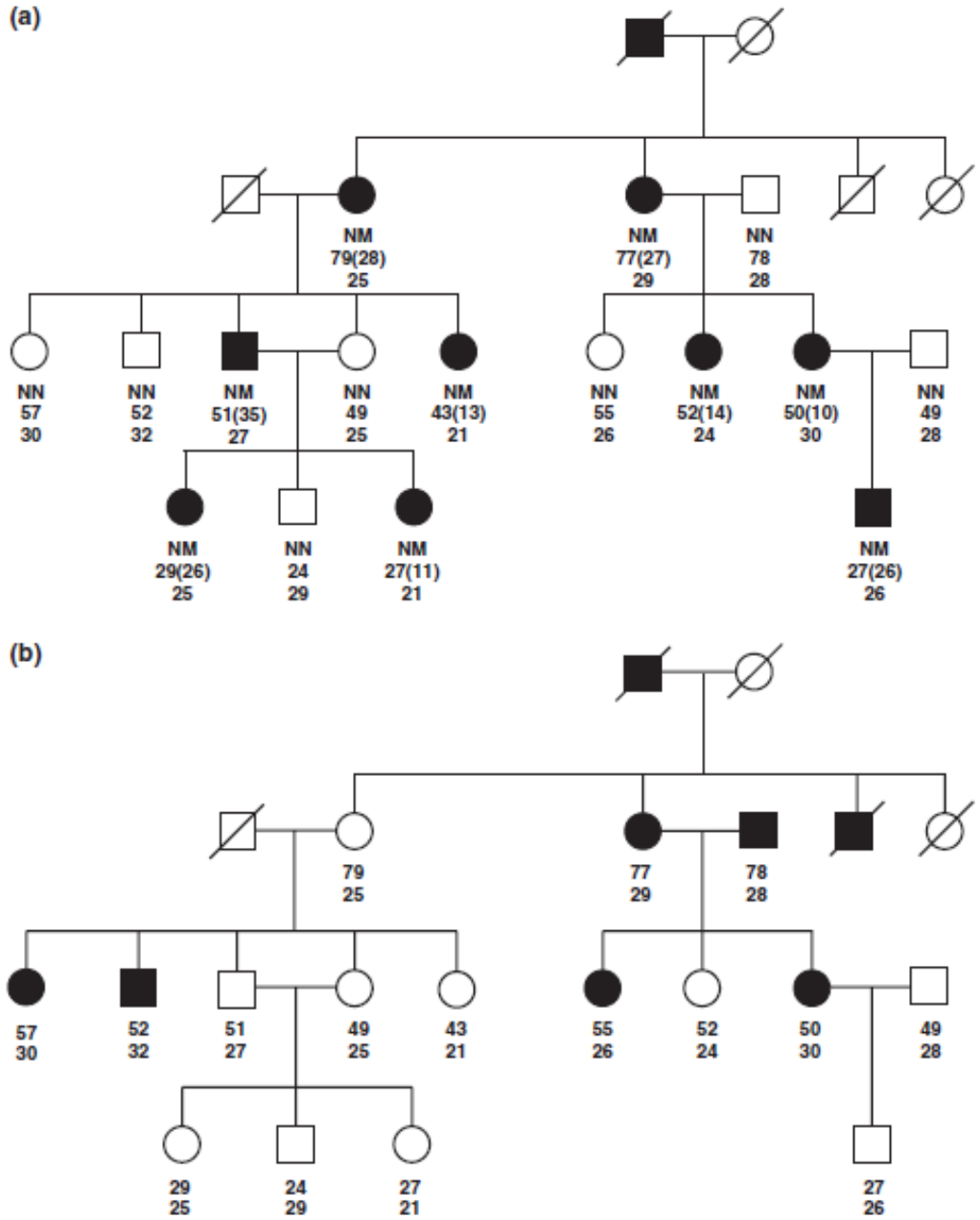
In the search for genetic variants (see **Box 1**) that predispose to common disease, the most powerful strategies depend on the (often unknown) underlying genetic model. For monogenic disorders the genetic model can, most often, be defined as dominant (where a single mutated gene causes the disease) or recessive (where two copies of the disease genes are required). These models produce distinct inheritance patterns in families that allow, with sufficiently large families, the single highly-penetrant causal mutations to be identified through classical linkage studies. This is in contrast to polygenic diseases that strongly cluster in families and are highly heritable, but do not demonstrate simple inheritance patterns. **Figure 1** compares monogenic and polygenic disease inheritance. One explanation for the polygenic inheritance pattern is that many (tens to hundreds) of common genetic variants (minor allele frequency >1% in the population), each with only a modest effect on disease risk (affecting relative risk by < 50%) are responsible for the heritability of polygenic disease. This is the common disease/common variant hypothesis (CDCV) [13].

A number of people have argued against the CDCV hypothesis, suggesting that rare, modest-risk alleles may explain a large proportion of the variation in susceptibility to common disease [14-16], and it is likely that both common and rare alleles are important in polygenic disease. However, given the current near impossibility of reliably detecting effects of rare alleles (owing to sample size and sequencing constraints), studies have focused on finding common disease alleles.

Box 1. The nature of genetic variation

Genetic variation can take the form of chromosomal rearrangements, large-scale deletions or insertions, small-scale deletions or insertions, or single base pair changes. A polymorphism is defined as a genetic variant that has at least two alleles in a population at a frequency of greater than 1%. Single-base pair substitution polymorphisms are referred to as single nucleotide polymorphisms (SNPs). SNPs account for most of the genetic variation of the human genome. There are thought to be 20-30 million SNPs across the human genome. Many of these are catalogued in online databases, and are publicly accessible. With the recent completion of the human genome project, the physical map position of these SNPs is precisely defined. Unlike insertions and deletions, SNPs are not thought to mutate very frequently.

Figure 1: (a) A typical monogenic pedigree (e.g. MODY). (b) A typical polygenic pedigree (e.g. type 2 diabetes). □ = male without the disease; ■ = male with the disease; ○ = female without the disease; ● = female with the disease. NN and NM are the normal and mutation (disease) genotypes, respectively. Age (age at diagnosis) and body mass index are indicated by the top and bottom numbers, respectively.



Finding susceptibility genes for common complex disease: past approaches

Until very recently, the two major strategies that have been used to identify complex disease genes were positional cloning through genome-wide linkage scans, and candidate gene association studies. The linkage approach has been very successful in identifying genes responsible for monogenic diseases following Mendelian pattern of inheritance, but very few linkage studies of diseases with polygenic inheritance patterns have provided reproducible evidence for linkage, and only a small number of disease genes have been identified through this approach [17].

In their 1996 paper, Risch and Merikangas [18] showed that genetic association studies are much more powerful than linkage studies in identifying common variants of modest effects. Since then genetic association studies have become the method of choice for identifying common gene variants predisposing to disease. At their core, genetic association studies of disease are straightforward. In its simplest form, when applied to unrelated individuals, the frequency of a variant (allele) of a SNP is determined in a sample of subjects with a particular disease, and a sample of subjects without the disease. A statistically significant higher frequency of a variant of a SNP (or other genetic variant) in the cases versus the controls suggests that it is associated with a particular disease.

There are around 20-30 million SNPs in the human genome and, until recently, it was not possible to assay such a large number of variants. Instead, many studies used a candidate gene approach and analysed a small number of candidate SNPs in these genes. However, most of these studies used sample sizes that provided insufficient power to detect the association unless the allele had extremely high odds ratio [19]. When an association was detected it was usually a false-positive or, in a few cases, a true-positive with a greatly overestimated risk effect [20, 21]. In type 2 diabetes, for example, years of research had identified only two reproducibly associated susceptibility variants, in *PPARG* [22] and *KCNJ11* [23]

genes. Lessons have been learned from the poor performance of the candidate gene studies (**Box 2**), and the power of the association approach is now being demonstrated with the advent of the large-scale genome-wide association study.

The era of the genome-wide association study

We are now in the era of the genome-wide association study, whereby using the phenomenon of linkage disequilibrium (LD; association between alleles in a population, due to their proximity on a chromosome such that recombination at meiosis has not had a chance to "separate" them) and DNA chips that allow the assaying of several hundred SNPs simultaneously, we are able to evaluate a large proportion of the 20-30 million common genetic variants across the human genome. The Affymetrix 500K chip (Affymetrix, Inc., Santa Clara, CA, USA) captures ~65% of common variation across the genome, and the Illumina 317K chip (Illumina, Inc., San Diego, CA, USA) captures ~75% [24, 25]. Coverage of the genome for these chips can be substantially increased using the statistical technique of imputation, whereby information from a group of typed SNPs and LD information from the HapMap (http://www.hapmap.org) can be used to infer genotypes at untyped SNPs [26]. We now discuss some of the recent exciting findings from genome-wide association studies.

Box 2. Lessons learned from past approaches

Many lessons have been learned from past failing of complex trait genetic studies, the most important of which are:

- A large sample size is crucial. Polygenic variants have small effects, which to detect reliably requires many thousands of subjects. Past studies used just tens to just a few hundred subjects, which will be powered to find nothing but the strongest polygenic effects.
- Statistical significance levels need to be interpreted with much caution. A *P*-value of 0.05 in a genome-wide association study cannot be considered significant; when genetic association studies test hundreds of thousands of variants against many traits a much more stringent threshold is required. *P*-values of $<5x10^{-7}$ and less are required for a SNP to be considered to have genome-wide evidence for association.
- Replication of findings is essential. The multiple hypothesis testing problem makes replication of genome-wide association results (in suitably powered follow-up studies) essential.
- An even more powerful approach than an initial GWA scan followed by replication is to put all available GWA studies together and perform one large meta-analysis of individual studies' summary statistics. Such approach inevitably requires collaboration between research groups that would otherwise compete with each other. The formation of several international consortia has already led to meta-analyses of tens of thousands of samples and to the identification of a number of novel genome-wide associations.
- Comprehensive coverage of the common variation across the genome is required. By focusing on known biological candidates there is less chance of important novel insights into disease pathophysiology. Also, the lack of success of this approach in type 2 diabetes for example, bears testament to the lack of underlying knowledge of common disease pathophysiology.

Recent findings in genome-wide association studies

One of the largest, most comprehensive GWA study to date was carried out by the Wellcome Trust Case-Control Consortium (WTCCC) [27]. The WTCCC study used the Affymetrix 500K GeneChip and examined 3000 shared controls and 2000 cases, all of UK Caucasian ancestry, for seven common complex diseases: bipolar disorder (BD), coronary artery disease (CAD), Crohn disease, hypertension, rheumatoid arthritis (RA), type 1 diabetes and type 2 diabetes [27]. The study identified 25 independent association signals at a stringent level of significance ($P<5x10^{-7}$). Association signals were identified for all diseases except hypertension, where the strongest signal had $P=7.7x10^{-7}$. This initial analysis of the WTCCC study therefore doubled the number of known complex disease genes. However, the WTCCC was primarily a hypothesis generating study, with only the "low hanging fruit" being convincingly identified in this "first pass" analysis: many of the SNPs with *P* values greater than $5x10^{-7}$ will also be disease-causing variants. As will be described in the sections below, follow-up studies in sufficiently powered replication cohorts, and the combination of findings from several GWA scans have confirmed many other complex disease variants.

Type 2 diabetes

It is perhaps the results of the first type 2 diabetes genome-wide scans[28-32] that best illustrate the power of the GWA approach for identifying novel genes that are important in the etiology of complex disease. Following up the results from the initial WTCCC GWA scan [27], we worked closely with the DGI [28] and FUSION [32] groups, who had performed similar studies. We used the combined information from these three studies to prioritise variants for follow-up. Including replication samples, these three studies provided data from 14586 cases and 17968 controls. The combined analyses identified three entirely novel type 2 diabetes susceptibility genes: *CDKAL1* (cyclin-dependent kinase 5 [*CDK5*] regulatory subunit associated protein 1-like 1) (OR 1.12, combined $P = 4.1x10^{-11}$), *IGF2BP2* (insulin-like growth factor 2 binding protein 2) (OR 1.14, combined $P=8.6x10^{-16}$), and *CDKN2A*/*CDKN2B* (cyclin-dependent kinase inhibitor 2 A/B) gene region (OR

1.20, combined $P=7.8x10^{-15}$) and demonstrated that integrating the results from multiple genome-scans can aid the prioritisation of signals for replication, and allow confirmation of genes at appropriate levels of statistical confidence not possible with individual GWA studies.

Other type 2 diabetes GWA studies have also been published. The deCODE study [31] of several European and a Chinese population replicates the association of the *CDKAL1* variant (OR 1.20 in Europeans and 1.25 in Chinese). These four studies also confirm the association of variants near *HHEX* (homeobox, hematopoietically expressed) and *SLC30A8* (solute carrier family 30 [zinc transporter], member 8) genes, originally published by Sladek *et al.* [29]. Importantly, as a positive control, associations for variants in *PPARG* [22], *KCNJ11* [23] and *TCF7L2* [33], originally identified through candidate gene and positional cloning methods, were also seen in the GWA scans, with expected odds ratios.

Of the variants identified through the GWA approach, the two in or near the *CDKN2A/CDKN2B* gene are particularly interesting. *CDKN2A* encodes $P16^{INK4a}$, and is a known tumor-suppressor gene [34]. Mutations of *CDKN2A* cause diverse neoplasias. CDKN2A is an inhibitor of cyclin dependent kinase 4 (CDK4), which is important for beta-cell replication [35]. Overexpression of *Cdkn2a* in mice leads to decreased islet proliferation, while *Cdkn2a* knockout mice demonstrate enhanced islet proliferation and survival after beta-cell ablation [36]. Overexpression of *Cdkn2b* causes islet hypoplasia and diabetes in murine models [37]. Together with the *CDKAL1* association, the *CDKN2A/B* finding implicates the cyclin-dependent kinase pathway in the pathophysiology of type 2 diabetes.

Another interesting feature of the *CDKN2A/B* finding is that, as described below, variants of the *CDKN2A/B* gene have also recently been shown to predispose to myocardial infarction (MI). Determining why a gene predisposes to type 2 diabetes and heart disease may lead to an explanation for the link between these two disorders.

The *CDKN2A/B* finding also highlights the power of GWA studies to identify variants outside described genes: while one of the signals occurs in the *CDKN2A/B* region, the other (much stronger) association signal occurs >200kb from these genes, in a gene desert. This association would not have been picked up by a candidate gene approach. Identifying the mechanism by which this variant (presumably) affects *CDKN2A/B* expression will provide new insights into the regulation of this important gene(s).

The other newly identified type 2 diabetes genes are generally involved in beta-cell development and function, and insulin secretion [28, 30, 32]. For example, the *HHEX* gene is highly expressed in fetal and adult pancreas, and is implicated in pancreatic development [38, 39]. It is a target of WNT signalling pathway, which has been shown to be critical for the development of the pancreas and islets during embryonic growth [40]. Importantly, *TCF7L2* also has an important role in WNT signaling, acting as a nuclear receptor for β-catenin [41]. Together, these findings highlight the importance of the WNT signaling pathway in glucose homeostasis.

Obesity

In addition to the newly identified type 2 diabetes genes described above, the WTCCC study found strong association with *FTO* (fat mass and obesity associated) gene region (OR 1.27, P=2.0x10^{-8}) [30]. This finding, which was the strongest susceptibility locus outside *TCF7L2*, showed strong replication in further 3757 type 2 diabetes cases and 5346 controls from the UK (OR 1.22, P=5.4x10^{-7}) [30]. However, the lack of such strong association in the DGI study [28], which matched cases and controls for BMI, and the FUSION study [32], where there was minimal BMI differences between cases and controls, suggested that the association with type 2 diabetes was caused by the primary effect on adiposity. Indeed, adjustment for BMI in the UK replication samples abolished the type 2 diabetes association (OR 1.03, P=0.44).

This exciting observation lead to the study of association of *FTO* gene variation with BMI and the risk of being overweight and obese in an additional

19424 adults and 10172 children, all of white European origin [42]. In the combined dataset each additional copy of the rs9939609 risk allele is associated with a BMI increase of ~0.4kg/m^2 (P=3x10^{-35}). Individuals homozygous for the A allele (16% of the population) are at a substantially increased risk of being overweight (OR 1.38, P=4x10^{-11}) and obese (OR 1.67, P=1x10^{-14}) compared to those homozygous for the low-risk T allele (37% of the population). This association was observed in children at ages 7-11, but not at birth, and reflects a specific increase in fat mass[42].

FTO is a gene of unknown function in an unknown pathway. It seems to be widely expressed in both fetal and adult tissues, with highest levels in the brain [42]. One possibility therefore is that FTO is an important regulator of appetite. This would be consistent with the role of monogenic obesity genes, such a Leptin, but much work is needed to determine whether this is the case. It is clear though that understanding how variants of the *FTO* gene increase fat-mass will lead to the identification of a new obesity pathway, with implications for drug development and treatments.

Age-related macular degeneration

Age-related macular degeneration (AMD), the main cause of blindness in developed countries, is a chronic, common and complex disease characterised by progressive destruction of retina's central region and drusen formation behind the retina (reviewed in [43]). Currently, there is no broadly effective therapy available. The major environmental risk factor for AMD is smoking (smokers have up to 2.5-fold increased risk of AMD than non-smokers) [44, 45]. One of the first published genome-wide case-control studies was by Klein and colleagues [46]. Using the relatively sparse Affymetrix 100K chip they identified a common variant in the complement factor H gene (*CFH*) as the SNP most strongly associated with AMD [46]. Although this was a small study (96 cases and 50 controls), it increased its power by using enriched samples (severe AMD cases and older controls to increase the probability of them not developing AMD). A more recent case-control candidate gene study replicated the association of *CFH* gene, and confirmed that

individuals homozygous for the most strongly associated risk allele have over 7-fold higher risk for AMD than those homozygous for the non-risk allele [47]. Human CFH is a regulator of the innate complement system that responds to infection by normally attacking only the diseased cells. Observations of activated complement components within drusen of AMD patients, and of strong effects of smoking and age on CFH plasma levels, suggest that AMD may result from abnormal complement activation in an anomalous inflammatory response [46]. Although the *CFH* polymorphisms are non-coding, they may alter the binding of CFH to heparin and C-reactive protein [46]. Furthermore, since CFH is a member of the complement and coagulation cascade pathway, these findings highlight that several different complement and coagulation factors may be potential drug targets and justify further research.

Crohn disease

Crohn disease, most commonly affecting ileum and colon, is a common form of idiopathic inflammatory bowel disease (IBD) where genetic predisposition has been supported by twin studies showing concordance rate of 50% in monozygotic compared to 10% in dizygotic pairs. Previously, years of research effort involving linkage, candidate gene and targeted association studies, identified only two genuinely associated variants, in *CARD15* gene and the *IBD5* haplotype. A recent GWA study by Rioux and colleagues [48] identified and replicated several new susceptibility loci for ileal Crohn disease. The most associated SNP, independently identified by a smaller German study [49], was a non-synonymous amino acid change in ATG16 autophagy-related 16-like 1 (*ATG16L1*) gene. The risk allele is a major allele (it has a frequency of about 52% and 60% in controls and cases, respectively), and individuals carrying one copy are at a 35-45% higher risk of developing the disease than those carrying no *ATG16L1* risk alleles [48, 49]. This SNP is in strong LD (r2=0.97) with the strongest signal in the WTCCC scan for Crohn disease. Autophagy is a constitutive biological process involved in immune pathogen recognition, and the variants in *ATG16L1* gene may alter innate immune control or antigen presentation in the adaptive immune pathways [48]. The WTCCC

study identified four novel association signals, all of which have since been replicated. These map to *IRGM* (immunity-related guanosine triphosphatase), *MST1* (macrophage stimulating 1), *NKX2-3* (NK2 transcription factor related, locus 3), and *PTPN2* (protein tyrosine phosphatase, non-receptor type 2) gene regions. These novel findings highlight that defects in a number of components of innate and adaptive immune pathways, such as those in autophagy and the processing of phagocytosed bacteria, are a major cause of Crohn disease.

Height as a model polygenic trait

Adult human height is a classic, highly heritable polygenic trait, product of many different developmental processes. It is an easily and accurately measured phenotype, with data available in large numbers of individuals from virtually every disease or population cohort used in GWA studies. This makes height an ideal model trait that can be used to study the architecture of polygenic traits in general, as well as answer questions about the biological relevance and clinical usefulness of associated variants with modest effect sizes identified through GWA studies with large sample sizes. Furthermore, the implicated genes should provide important insights into the mechanisms of normal growth and development, and would be excellent candidate genes for mutations responsible for growth and skeletal disorders without known genetic causes.

Two independent common variants have already been identified by GWA studies, in the *HMGA2* [50] and *GDF5-UQCC* [51] gene regions. Both studies used sample size of several thousand individuals, and it was clear that much larger studies would be needed for sufficient power to identify additional common variants for human height.

What benefits can GWAS findings bring to clinical practice

Identifying the genes and pathways involved in predisposing individuals to complex disease is giving us insights into the pathophysiology of these diseases, which may eventually lead to the development of novel treatments. These insights are the most valuable thing to come out of these genetic studies, and will, in time,

without doubt lead to better clinical management and treatment of these diseases. But is there any immediate clinical utility of knowing an individual's genotypes at these disease variants?

Pharmacogenetics

Pharmacogenetics is a study of how genetically determined variation affects an individual's response to drugs. It is well known that adverse side effects and therapeutic failure of drugs may both have a strong genetic component [52]. If a patient's genotypes in the relevant genes are known, it may be possible to truly personalise medication and optimise treatment by selecting the most effective drug and its dose. It has been estimated that adverse drug reactions account for 6.5% of hospital admissions, 4% of hospital bed capacity, and 0.15% fatalities in England, at a projected annual cost to the NHS of £466 million [53]. Although a large portion of this figure can be attributed to prescription errors and accidental overdoses, pharmacogenetics could be used to identify individuals both at a highest and lowest risk of developing adverse effects to particular drugs or doses. Furthermore, clinical efficacy could be much improved if drugs are prescribed only to individuals likely to benefit from them, thus reducing the number of ineffective treatments. Pharmacogenetic testing will be particularly desirable in cases where it proves to be more effective than the current practice of just careful monitoring of a patient's response to a drug/dose.

There are many examples of where this has already happened, such as for monogenic types of diabetes. The definition of six different genetic subtypes in MODY has led to recognition of different clinical phenotypes [54]. Molecular genetics is now commonly used as a diagnostic test, and the diagnosis has a dramatic effect on the treatment decisions in MODY. Patients with glucokinase mutations have life-long mild but stable fasting hyperglycaemia from birth, and require no treatment because they essentially have a glucose sensing defect where glucose is regulated at a higher level [54]. In contrast, patients with *HNF1A* mutation have a progressive hyperglycaemia but are very sensitive to hypoglycaemic effects of sulphonylureas [55]. The correct diagnosis is very important in this case because

many of these patients are misdiagnosed as having type 2 diabetes, in which case the most common pharmacological treatment is metformin, but *HNF1A* patients have a four-fold greater response to sulphonylurea gliclazide than to metformin [55].

The ability to analyse thousands of SNPs in a large number of individuals will be essential for defining genetic heterogeneity of diabetes, which could then be translated into clinical heterogeneity between the patients. Even if a gene does not cause or predispose to a disease, it can still interfere with the drug-targeted metabolic pathway and modify the toxicity (effectiveness and side effects) of the drug and the clinical response. For example, the putative role of *TCF7L2* in beta-cell function has lead to a hypothesis that *TCF7L2* variation may have an effect on glycaemic response to sulphonylureas, but not to metformin [56]. A study that tested this using 747 and 864 metformin users with type 2 diabetes found that, while the tested variant had no effect on metformin response, in the sulphonylurea treatment group only 40% of patients homozygous for the risk allele reached the target HbA1c < 7%, compared to 61% of patients with the other two genotypes (OR 0.46, P<0.001) [56]. This study provides a strong example of pharmacogenetics in type 2 diabetes, where a common variant predicts treatment response. If a patient's genotypes in several of the relevant genes are known, it may be possible to personalise medication and optimise treatment.

Disease prediction

In type 2 diabetes and other complex polygenic diseases the number of confirmed common risk variants is relatively small. This means that, for now, molecular genetics cannot be used as a diagnostic test because family history and environmental and lifestyle influences, such as BMI, smoking, physical activity and social class, have much higher predictive power. One way of assessing predictive power of polygenic variant information is to use the area under the ROC curve (AUC), which measures the discriminatory power of the test [57]. The test is 100% accurate when the AUC is 1, and is no better than chance when AUC is 0.5. It has been estimated that for an AUC of ~0.8 it will be necessary to genotype 20-25 risk variants with allele frequencies greater than 10% and ORs of 1.5 [58] . In

multifactorial diseases where preventative measures exist, coupled with the low cost of genotyping, it may be practically and economically justifiable to identify individuals with the highest risk of developing the disease if the preventative measures are still effective in such high-risk groups. This should certainly be the case in type 2 diabetes where lifestyle factors have a big impact on onset and severity of the disease.

A recent type 2 diabetes case-control study demonstrated that, although individual susceptibility variants only moderately increase the risk of type 2 diabetes and are of limited use in disease prediction, by combining the information from the three replicated risk variants (known at the time) it is possible to identify individuals at significantly greater risk of developing the disease than when a single polymorphism is used [59]. Individuals with all six risk alleles had an OR of 5.71 (95% CI 1.15 to 28.3) compared to those with no risk alleles, and the area under the ROC curve for the three polymorphisms was 0.58. This study, however, only included variants discovered prior to the availability of the GWA approaches, and now it would be interesting to look at the combined predictive value of the much increased number of variants identified through GWA studies.

Conclusions

Genome-wide association studies are a powerful new approach to identifying genes influencing predisposition to common, complex disease. Genome-wide studies are “hypothesis-free” and allow the identification of previously unsuspected genes and pathways in disease aetiology. One of the challenges now is to determine how relevant the new variants are for our understanding, and eventual treatment, of disease. Considering that little is still known about the exact roles of most of these new variants and implicated genes and regions, even about the higher-risk ones such as *TCF7L2* in type 2 diabetes, at the moment they remain just potential targets for predictive testing and drug development.

In the short-term the greatest value of GWAS is likely to come from etiological insights - the effect of variants on disease onset, severity and progression, as well as an individual's response to treatment. In the long-term, there are two major clinical benefits. Firstly, there is a great opportunity to identify important new biological pathways, genes and gene interactions, providing us with more complete picture of disease aetiology and progression, and the best therapeutic targets. Excitingly, there may be cases where there are the same targets for more than one disease, such as the *CDKN2A/B* gene region in type 2 diabetes and cardiovascular disease. Secondly, once enough variants are identified that together explain a clinically useful portion of a disease, it will be possible to identify individuals at most risk of developing the disease. Genetic testing will be particularly useful where effective prevention is available, and if studies show that this genetic knowledge will positively influence individuals' lifestyle choices. The initial success of many genome-wide association studies has certainly brought much excitement to the scientific community, but there will be several years before patients see real benefits in clinical practice.

Aims of the Thesis

The aim of this thesis is to explore the biological and clinical value of common variants identified through genome-wide association studies, by using height as a model polygenic trait, type 2 diabetes as a model common, multifactorial disease, and *HNF1A*-MODY as a model monogenic disease.

References

1. Jimenez-Sanchez, G., Childs, B. & Valle, D. Human disease genes. Nature 409, 853-855 (2001).

2. Pearson, E. & Hattersley, A. Genetic aetiology alters response to treatment in diabetes. Diabetic Medicine 20 (S2), 12 (2003).

3. Hammond, C. J. et al. Genetic influence on early age-related maculopathy: A twin study. Ophthalmology 109, 730-736 (2002).

4. Seddon, J. M., Cote, J., Page, W. F., Aggen, S. H. & Neale, M. C. The US Twin Study of Age-Related Macular Degeneration: Relative Roles of Genetic and Environmental Influences. Archives in Ophthalmology 123, 321-327 (2005).

5. Tysk, C., Lindberg, E., Jarnerot, G. & Floderus-Myrhed, B. Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. Gut 29, 990-996 (1988).

6. Lichtenstein, P. et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. New England Journal of Medicine 343, 78-85 (2000).

7. Poulsen, P., Ohm Kyvik, K., Vaag, A. & Beck-Nielsen, H. Heritability of Type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance - a population-based twin study. Diabetologia 42, 139-145 (1999).

8. Henkin, L. et al. Genetic Epidemiology of Insulin Resistance and Visceral Adiposity: The IRAS Family Study Design and Methods. Annals of Epidemiology 13, 211-217 (2003).

9. Coady, S. A. et al. Genetic Variability of Adult Body Mass Index: A Longitudinal Assessment in Framingham Families. Obesity Res 10, 675-681 (2002).

10. Fischer, M. et al. Distinct Heritable Patterns of Angiographic Coronary Artery Disease in Families With Myocardial Infarction. Circulation 111, 855-862 (2005).

11. Zdravkovic, S. et al. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. Journal of Internal Medicine 252, 247-254 (2002).

12. Robinson, R. F., Batisky, D. L., Hayes, J. R., Nahata, M. C. & Mahan, J. D. Significance of Heritability in Primary and Secondary Pediatric Hypertension. American Journal of Hypertension 18, 917-921 (2005).

13. Reich, D. E. & Lander, E. S. On the allelic spectrum of human disease. Trends in Genetics 17, 502-10 (2001).

14. Pritchard, L. E. et al. Analysis of the CD3 gene region and type 1 diabetes: application of fluorescence-based technology to linkage disequilibrium mapping. Human Molecular Genetics 4, 197-202 (1995).

15. Terwilliger, J. D., Haghighi, F., Hiekkalinna, T. S. & Goring, H. H. A bias-ed assessment of the use of SNPs in human complex traits. Curr Opin Genet Dev 12, 726-34 (2002).

16. Pritchard, J. K. & Cox, N. J. The allelic architecture of human disease genes: common disease-common variant or not? Human Molecular Genetics 11, 2417-23 (2002).

17. Altmuller, J., Palmer, L. J., Fischer, G., Scherb, H. & Wjst, M. Genomewide scans of complex human diseases: True linkage is hard to find. American Journal Of Human Genetics 69, 936-950 (2001).

18. Risch, N. & Merikangas, K. The Future of Genetic Studies of Complex Human Diseases. Science 273, 1516-1517 (1996).

19. Redondo, M. J., Fain, P. R. & Eisenbarth, G. S. in Recent Progress In Hormone Research, Vol 56 (ed. Means, A. R.) 69-89 (The Endocrine Society, Chevy Chase, MD, 2001).

20. Hirschhorn, J. N., Lohmueller, K., Byrne, E. & Hirschhorn, K. A comprehensive review of genetic association studies. Genetics In Medicine 4, 45-61 (2002).

21. Lohmueller, K. E., Pearce, C. L., Pike, M., Lander, E. S. & Hirschhorn, J. N. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nature Genetics 33, 177-182 (2003).

22. Altshuler, D. et al. The common PPARγ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nature Genetics 26, 76-80 (2000).

23. Gloyn, A. L. et al. Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. Diabetes 52, 568-72 (2003).

24. Barrett, J. C. & Cardon, L. R. Evaluating coverage of genome-wide association studies. Nature Genetics 38, 659-662 (2006).

25. Pe'er, I. et al. Evaluating and improving power in whole-genome association studies using fixed marker sets. Nature Genetics 38, 663-667 (2006).

26. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nature Genetics 39, 906 (2007).

27. Burton, P. R. et al. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447, 661-678 (2007).

28. Saxena, R. et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 316, 1331-1336 (2007).

29. Sladek, R. et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 445, 881 (2007).

30. Zeggini, E. et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 316, 1336-1341 (2007).

31. Steinthorsdottir, V. et al. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. Nature Genetics 39, 770-775 (2007).

32. Scott, L. J. et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 316, 1341-1345 (2007).

33. Grant, S. F. A. et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nature Genetics 38, 320-323 (2006).

34. Wolfel, T. et al. A P16(Ink4a)-Insensitive Cdk4 Mutant Targeted By Cytolytic T-Lymphocytes In A Human-Melanoma. Science 269, 1281-1284 (1995).

35. Rane, S. G. et al. Loss of Cdk4 expression causes insulin-deficient diabetes and Cdk4 activation results in beta-islet cell hyperplasia. Nature Genetics 22, 44-52 (1999).

36. Krishnamurthy, J. et al. p16(INK4a) induces an age-dependent decline in islet regenerative potential. Nature 443, 453-457 (2006).

37. Moritani, M. et al. Hypoplasia of endocrine and exocrine pancreas in homozygous transgenic TGF-beta 1. Molecular And Cellular Endocrinology 229, 175-184 (2005).

38. Bogue, C. W., Ganea, G. R., Sturm, E., Ianucci, R. & Jacobs, H. C. Hex expression suggests a role in the development and function of organs derived from foregut endoderm. Developmental Dynamics 219, 84-89 (2000).

39. Bort, R., Martinez-Barbera, J. P., Beddington, R. S. P. & Zaret, K. S. Hex homeobox gene-dependent tissue positioning is required for organogenesis of the ventral pancreas. Development 131, 797-806 (2004).

40. Papadopoulou, S. & Edlund, H. Attenuated Wnt signaling perturbs pancreatic growth but not pancreatic function. Diabetes 54, 2844-2851 (2005).

41. Smith, U. TCF7L2 and type 2 diabetes - we WNT to know - Commentary. Diabetologia 50, 5-7 (2007).

42. Frayling, T. M. et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science 316, 889-894 (2007).

43. Klein, R., Peto, T., Bird, A. & Vannewkirk, M. R. The epidemiology of age-related macular degeneration. American Journal of Ophthalmology 137, 486 (2004).

44. Seddon, J. M., Willett, W. C., Speizer, F. E. & Hankinson, S. E. A Prospective Study of Cigarette Smoking and Age-Related Macular Degeneration in Women. JAMA 276, 1141-1146 (1996).

45. Christen, W. G., Glynn, R. J., Manson, J. E., Ajani, U. A. & Buring, J. E. A Prospective Study of Cigarette Smoking and Risk of Age-Related Macular Degeneration in Men. JAMA 276, 1147-1151 (1996).

46. Klein, R. J. et al. Complement factor H polymorphism in age-related macular degeneration. Science 308, 385-9 (2005).

47. Maller, J. et al. Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. Nature Genetics 38, 1055-1059 (2006).

48. Rioux, J. D. et al. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. Nature Genetics 39, 596 (2007).

49. Hampe, J. et al. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. Nature Genetics 39, 207-211 (2007).

50. Weedon, M.N. *et al.* A common variant of HMGA2 is associated with adult and childhood height in the general population. Nature Genetics 39, 1245-1250 (2007)

51. Sanna, S *et al*. Common variants in the GDF5-UQCC region are associated with variation in human height. Nature Genetics 40, 198 (2008)

52. Meyer, U. A. PHARMACOGENETICS - FIVE DECADES OF THERAPEUTIC LESSONS FROM GENETIC DIVERSITY. Nature Reviews Genetics 5, 669 (2004).

53. Pirmohamed, M. et al. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. BMJ 329, 15-19 (2004).

54. Hattersley, A. T. & Pearson, E. R. Minireview: Pharmacogenetics and beyond: The interaction of therapeutic response, beta-cell physiology, and genetics in diabetes. Endocrinology 147, 2657-2663 (2006).

55. Pearson, E. R., Liddell, W. G., Shepherd, M., Corrall, R. J. & Hattersley, A. T. Sensitivity to sulphonylureas in patients with hepatocyte nuclear factor 1 alpha gene mutations: evidence for pharmacogenetics in diabetes. Diabetic Medicine 17, 543-5 (2000).

56. Pearson, E. R. et al. Pharmacogenetics in Type 2 diabetes: variation in TCF7L2 influences response to sulphonylureas. Diabetic Medicine 24, 5-5 (2007).

57. Janssens, A., Pardo, M. C., Steyerberg, E. W. & van Duijn, C. M. Revisiting the clinical validity of multiplex genetic testing in complex diseases. American Journal Of Human Genetics 74, 585-588 (2004).

58. Janssens, A. et al. Predictive testing for complex diseases using multiple genes: Fact or fiction? Genetics In Medicine 8, 395-400 (2006).

59. Weedon, M. N. et al. Combining Information from Common Type 2 Diabetes Risk Polymorphisms Improves Disease Prediction. PLoS Medicine 3 (2006).

CHAPTER 2: GENOME-WIDE ASSOCIATION ANALYSIS IDENTIFIES 20 LOCI THAT INFLUENCE ADULT HEIGHT

Michael N Weedon[*], Hana Lango[*], Cecilia M Lindgren, Chris Wallace, David M Evans, Massimo Mangino, Rachel M Freathy, John RB Perry, Suzanne Stevens, Alistair S Hall, Nilesh J Samani, Beverly Shields, Inga Prokopenko, Martin Farrall, Anna Dominiczak, Diabetes Genetics Initiative, The Wellcome Trust Case Control Consortium, Toby Johnson, Sven Bergmann, Jacques S Beckmann, Peter Vollenweider, Dawn M Waterworth, Vincent Mooser, Colin N A Palmer, Andrew D Morris, Willem H Ouwehand, Cambridge GEM Consortium, Mark Caulfield, Patricia B Munroe, Andrew T Hattersley, Mark I McCarthy, Timothy M Frayling

* denotes equal contributions

Published: *Nature Genetics* (2008) 40:575

Abstract

Adult height is a model polygenic trait, but there has been limited success in identifying the genes underlying its normal variation. To identify gene variants influencing adult height, we used genome-wide association data from 13,665 individuals and genotyped 39 variants in an additional 16,482 samples. We identified 20 variants associated with adult height ($P<5x10^{-7}$, with 10 reaching $P<1x10^{-10}$). Combined, the 20 SNPs explain ~3% of height variation, with a ~5cm difference between the 6.2% of people with =<17 or fewer "tall" alleles compared to the 5.5% with >=27. The loci identified implicate genes in hedgehog signaling (*IHH, HHIP, PTCH1*), extracellular matrix (*EFEMP1, ADAMTSL3, ACAN*) and cancer (*CDK6, HMGA2, DLEU7*) pathways and provide novel insights into human growth and development processes. Finally, our results provide insights into the genetic architecture of a classic quantitative trait.

Introduction

Adult height is a model polygenic trait. It is the ideal phenotype for genetic studies of quantitative traits in humans, as it is easily and accurately measured, and is highly heritable, with up to 90% of variation in adult height within a population being explained by genetic variation [1-5]. Final adult height is the result of growth and development processes. Identifying genes for human height should therefore provide insights into mechanisms of growth and development, as well as into the genetic architecture of quantitative traits, and how best to dissect them.

Despite its strong heritability, there has been little success in identifying the specific genetic variants that influence height in the general population [5, 6]. Some mutations resulting in extreme stature have been identified, but these are rare and cannot explain normal variation of adult height [6]. Linkage and candidate gene association studies have not identified any robustly associated loci. The advent of genome-wide association (GWA) studies, however, is providing new opportunities for identifying genetic variants influencing adult height.

Recently, using GWA study data from 4,921 individuals we identified the most convincing example of a common variant associated with adult height [7]. The variant was the only one to reach a level of significance suggestive of true association in the GWA study (P=4x10^{-8}), and we confirmed the association in 19,064 adults from four further studies (P=3x10^{-11}). The variant associated with a 0.4cm greater height per allele, explained ~0.3% of the population variation of height, and occurred in the *HMGA2* oncogene. In this study, we extend our analyses to a two-staged design of 13,665 individuals with GWA study data and 16,482 follow-up individuals.

Results

Height loci identified

We used GWA data from five studies that ranged in size from 1,437 to 3,560 people of UK ancestry and a sixth study of 2,978 Scandinavian individuals for which summary height association statistics have been made publicly available (http://www.broad.mit.edu/diabetes/scandinavs/index.html; **Supplementary Table 1)**. All studies were genotyped using the Affymetrix 500K chip. We compared the additive model statistics of 402,951 SNPs that passed quality-control (QC) criteria in at least 4 of the 6 studies to those expected under the null distribution using quantile-quantile (QQ) plots. **Figure 1** shows that the sequential addition of each of the six studies resulted in increased deviation of the observed statistics from the null distribution. The number of independent SNPs (using a cut-off of less than 0.2 for the pairwise linkage disequilibrium statistic r^2) reaching a $P<1x10^{-5}$, was 4 (N=1914), 6 (4892), 12 (6788), 13 (8668), 18 (12228), and 27 (13665) as each study was added in, against the expected <4 under the null distribution.

In the meta-analysis of 13,665 individuals with GWA data, there were many more significant associations than expected by chance. For example, we observed 8 independent signals with a $P<5x10^{-7}$, where we would expect none under the null distribution, and 27 with a $P<1x10^{-5}$, where we would expect < 4. Approximately 23 of these loci are therefore likely to represent true positives. The availability of dense genome-wide SNP data allows us to be confident that these results are not due to population stratification. First, any individuals of non-European ancestry were removed. Second, adjusting for residual population structure using EIGENSTRAT[8] did not affect the distribution of effect sizes (**Supplementary Figure 1** gives individual study QQ plots before and after EIGENSTRAT adjustment). Third, the genomic control inflation factor lambda[9] for the GWA study meta-analysis was only 1.12, despite the large size of the study (there is a strong relationship between sample size and lambda[10]) and the apparently highly polygenic nature of height. Fourth, 12 of the ancestry informative markers (AIMS) described by the WTCCC, which vary substantially in allele

frequency across the UK, did not associate with height (all $P>0.01$; the 13th AIM did not pass QC criteria in this study; **Supplementary Table 2**).

We took forward 39 SNPs into the second stage of our study: the genotyping of an additional 16,482 individuals of European ancestry from four studies (**Supplementary Table 1**). Of these, 27 represented all the independent ($r^2<0.2$) signals with a $P<1x10^{-5}$. Eleven of the SNPs represented independent regions where there was a SNP with a $P<1x10^{-4}$ and a gene within flanking recombination hotspots in which mutations affect length in mouse studies or cause monogenic human phenotypes of extreme stature. GWA data from CoLaus (one of our stage 2 cohorts) became available during the course of our analyses, and we also took forward a SNP representing a region with the strongest association ($P=4x10^{-8}$) from that study. Five of the AIMs with the largest differences in allele frequency across the UK[11] were also genotyped in stage 2 samples.

In the stage 2 analyses, 20 of the 39 SNPs reached a $P<0.005$ (with the same direction of effect as the GWA data), all of which reached a $P<5x10^{-7}$ in a joint analysis across GWA and stage 2 samples. While this is an arbitrary statistical cut-off, we chose to focus on these SNPs for the reasons discussed in [11], and we note that of the SNPs that reached a $P<5x10^{-7}$ in [11] and that have been subjected to replication efforts, all have been confirmed. The majority of the 20 SNPs had *P*-values substantially lower than $5x10^{-7}$: 17 of the SNPs reached a $P<5x10^{-8}$ and 10 reached a $P<1x10^{-10}$ in joint analyses. Of the 19 SNPs that did not reach $P<5x10^{-7}$, 15 had the same direction of effect in stage 2 as in stage 1 ($P=0.02$), suggesting that there are true positives amongst these. The details of the 20 SNPs are presented in **Figure 2** and **Table 1**, and details of the SNPs that did not reach the statistical cut-off are presented in **Supplementary Table 3.** For the 20 SNPs, there was no evidence of heterogeneity across studies when taking into account the number of tests (all $P>0.008$). In both joint and stage 2 only analyses, none of the WTCCC AIMS was associated with height providing further evidence that population stratification is unlikely to have influenced the results (all

P>0.01 see **Supplementary Table 2**). This means the associations are likely to reflect true biological effects on height.

Implicated genes and their function

The correlation between SNPs due to linkage disequilibrium (LD) and the occurrence of many of the 20 SNPs in non-coding regions means we cannot be certain about which genes are involved, but they implicate genes of many different functions in several different pathways and processes. In 10 instances, genes within the region of interest have previously been implicated in the regulation of growth because of known effects from rodent knockouts and/or human syndromes. LD plots for each region are presented in **Supplementary Figure 2,** and **Table 2** shows the genes most probably affected by the associated SNPs, along with the pathways the genes are known to be involved in, and where known, the monogenic syndromes caused by mutations in the associated genes and the phenotypes from knockout mouse models.

In two instances, there is evidence that the SNPs we have identified (or those in LD with them) influence gene expression. We used data from the publicly available "mRNA by SNP Browser 1.0" program described recently [12] to determine if any of the SNPs were associated with mRNA expression levels in lymphocytes. Rs2282978, which associates with height at $P=8x10^{-23}$, and occurs in the 4th intron of the *cyclin-dependent kinase 6 (CDK6)* gene associated with expression of *CDK6* ($P=1x10^{-6}$). Rs1863913, an $r^2 = 1$ proxy for rs10935120 (height $P=7x10^{-8}$), which occurs in intron 2 of *anaphase promoting complex 13* (*ANAPC13*) and 4.4kb upstream of *centrosomal protein 63* (*CEP63*) was associated with *ANAPC13* ($P=9x10^{-18}$) and *CEP63* ($P=4x10^{-12}$) expression. There was no evidence for any of the other SNPs affecting any transcript levels in these lymphoblastoid cell lines.

The genes implicate a number of biological pathways and processes in the normal determination of human height, including: hedgehog signaling (*IHH, HHIP, PTCH1)*; basic cell cycle regulation (*CDK6*, one of the cyclin-dependent kinases implicated in cell cycle progression [13]); extra-cellular matrix (*ADAMTSL3* and

EFEMP1); chromatin rearrangement and polycomb proteins (*HMGA2* and *SCMH1*). Several of the genes are also disrupted in cancers (e.g. *HMGA2, CDK6, DLEU7*) providing further evidence of a link between normal growth and unregulated cell differentiation. For other loci, no gene in the region is an obvious candidate for influencing height, and in one case (rs4549631) only a hypothetical gene, *LOC387103,* is within a 750Kb window of the SNP.

One of the most interesting findings is that rs6060373 (P=2x10^{-17}) is highly correlated (HapMap r^2 = 0.89) with a functional SNP in the *GDF5* gene that has recently been convincingly shown to alter the risk of osteoarthritis [14, 15]. The allele that is associated with higher height associates with a decreased risk of hip and knee osteoarthritis. A plausible explanation of these associations is that the variant influences the "thickness" of a person's cartilage.

Methodological issues

We next performed a series of analyses to address additional important issues about the genetic architecture of human height. Although our results are limited to height, our findings may prove useful in guiding studies of other quantitative traits.

We first tested whether the SNPs representing the 20 loci deviated from an additive model or had different effect sizes in males and females. There was suggestive evidence for deviation from an additive (per-allele) mode of inheritance for two of the variants: rs12735613 (P=0.009) and rs1390401 (P=0.007). There was also suggestive evidence that rs6440003, the most strongly associated SNP in our study, had a greater effect in females (0.12SD [0.09, 0.14]) than males (0.07SD [0.04, 0.09]), P=0.01 (**Table 1**).

Adult height is the result of both growth throughout childhood and loss of height during the ageing process. We therefore assessed the influence of age on the 20 robust associations. We found no evidence that the effects on height were different in individuals under age 50 years compared to those aged >50 years (all P>0.01; similar results were obtained when we used a cut-off of 40 years of age),

or when adjusting for age decade (see **Supplementary Table 4**). This suggests the effects are predominantly on developmental and childhood growth rather than driven by processes involved in loss of height, although studies of more young adults and children are needed to confirm this.

It has often been stated that gene-gene interactions may play a prominent role in complex traits but there are few, if any, empirical data to show this. We looked for any evidence of deviation from an additive model of the joint effects between all possible pairs of the 20 loci. When taking into account the number of tests, we find no strong evidence for deviation from additivity (all $P>0.017$; see **Supplementary Table 5**).

To assess the combined impact of the 20 SNPs on adult height we analysed only the UK stage 2 samples. This removes the bias due to the effect of the "winners curse" [16], which we observed in our data with 17 of the 20 SNPs having a larger effect size in the GWA study compared to our follow-up study ($P=0.003$ in a test against a 50/50 distribution). **Figure 3** shows the linear increase in the average height of individuals with increasing numbers of "tall" alleles, and the normal distribution of the frequency of "tall" alleles. Combined, these 20 SNPs explain ~2.9% of the variance in adult height in the UK stage 2 sample. There is a 0.7SD (~5cm) difference in height between the 6.2% of people with 17 or fewer "tall" alleles compared to the 5.5% of people with 27 or more.

Power and sample size issues are of major importance to the field of complex traits genetics. Our results indicate that many tens of thousands of individuals will be needed to reliably detect a large proportion of the variance in some quantitative traits. In this study, real signals only emerged after combining many individually underpowered GWA studies (**Figure 1** and **Supplementary Figure 3)**. We used the effect sizes observed in the stage 2 samples for each of the 20 SNPs to determine how much power we had to detect the associations in the GWA study; **Figure 4** plots the results. We had low power to detect some of the SNPs. For example, for 4 of the SNPs we had less than 10% power to detect the associations at a $P<1x10^{-5}$ in the GWA study.

Considerable effort and resources have been devoted to identifying regions of the genome that are shared between relatives of similar height more often than expected by chance – the linkage approach to gene identification. We analysed the overlap between linked regions (LOD >2.0, at http://www.genomeutwin.org/stature_gene_map.htm) and our association results [5]. We assumed a linked region to be a 10Mb window around the peak marker for all regions with LOD >2.0. Given the proportion of the genome that these regions cover we would have expected 3.5 ($5.3x10^8$ base-pairs covered by linkage regions / $3.0x10^9$ base-pairs in the human genome) of the 20 SNPs to have occurred in linked regions by chance alone, and we observed 4 (*P*=0.73); for linked regions with LOD scores > 3, the corresponding statistics were: expected 0.80 and observed 1, *P*=0.81. We found no evidence of over-representation of significant associations in linked regions (227 SNPs of 79241 SNPs (0.29%) in LOD > 2 linked regions had *P* values less than 0.001, compared to 892/323710 (0.28%) in non-linked regions, *P*=0.60. For LOD > 3 linked regions the corresponding figures are 48/22036 (0.22%) and 1071/380915 (0.28%), respectively, *P*=0.08).

Discussion

Our results are consistent with Fisher's proposal from 1918 that many variants of individually small effect explain the heritability of height [17]. The 20 robustly associated variants alter height by between ~0.2 and 0.6cm per allele, based on the stage 2 samples, but explain only ~3% of the variation in height within the population.

Some of the remaining heritability of height will be explained by additional SNPs, with small effect. Firstly, we have shown that some of the SNPs which we took forward into stage 2, but that did not reach a $P<5\times10^{-7}$ on joint analyses probably represent true associations (for example, an excess of SNPs showed the same direction of effect in stage 2 as in stage 1, $P=0.02$). Secondly, we observed a large effect of the winners curse [16] and, as such, we had low power to detect some of the SNPs in the GWA part of our study, strongly suggesting that there are many more common variants of a similar effect size to be found. As **Figure 4** shows, identifying these and variants of even smaller effect will require tens of thousands of individuals.

To further investigate whether there are more height associated SNPs to be identified through larger sample size, we compared our results to those presented in the accompanying manuscript from Lettre *et al.* [18] They identify association for several of the loci reported in our study (*ZBTB38*, *HMGA2*, *GDF5*, *HHIP*, *ADAMTSL3*, *CDK6*), and find suggestive association with a SNP at the *FUBP3* locus ($P=8\times10^{-7}$), which we also followed up and found suggestive evidence for ($P=2\times10^{-5}$). *FUBP3* therefore likely represents an additional height gene. We produced a QQ plot for the *P*-values observed in the Lettre *et al.*[18] study for the most-associated 10,000 SNPs from our study, excluding known loci. The deviation of the observed statistics from the null distribution (see **Figure 5)** clearly demonstrates that there are many more height-associated SNPs that remain to be identified from GWA studies. While SNPs will explain some of the residual variation, it is possible that much of the heritability of height will be explained by

rare variants or copy number polymorphisms, which are not captured by the GWA approach.

As we only tested an additive model, and did not perform sex specific analyses on a genome-wide level we were biased away from detecting sex-specific and non-additive effects in this study. However, we did find some weak evidence that our most associated SNP had a stronger effect in females (0.12SD [0.09, 0.14]) than males (0.07SD [0.04, 0.09]), *P*=0.01, although this finding needs to be replicated. Given that final adult height is highly dichotomized by sex, growth trajectories show clear sex differences and sex hormones influence height, further studies are needed to investigate more thoroughly the presence of sex-specific effects. It will also be important to test for non-additive within and between loci effects, and to investigate the role of these and other loci in individuals of non-European ancestry.

We found no overlap between previously reported linkage peaks and the results from our GWA study. The variants we have identified have small effects, and as such it is not surprising that they do not, individually, explain previously observed linkage peaks. It may be that some of the linkage peaks are explained by low-frequency, relatively high penetrance alleles, which would not be captured using the GWA approach. However, our findings do not support the idea that genes with common variants associating with height also contain the type of variant that is readily identifiable through the linkage approach.

A limitation of this study is that we have not fine-mapped the identified loci. However, ten of the loci we have identified contain genes previously known to be involved in growth from rare human syndromes or animal studies, and we have shown that common variation in or around these genes influences normal human growth. Additionally, two of the variants appear to alter expression of nearby genes (*CDK6* and *ANAPC13*). Further fine mapping and functional studies of these and the remaining loci will likely provide completely novel insights into growth and development. Mutations in these regions may also explain some monogenic syndromes for which no genes have currently been identified. The

observation that half of the identified loci contained candidate genes suggests that combining genome-wide with candidate gene approaches may be a productive way for identifying more height loci.

In conclusion, using 13,665 individuals with genome-wide scan data and 16,482 follow-up subjects we have identified 20 regions of the genome, common variation of which influences adult height. The study highlights several important pathways and processes involved in normal growth, and provides insights into the genetic architecture of a classic quantitative trait.

Methods

Study Descriptions

Genome-wide association (Stage 1) samples

Four of the six genome-wide scan studies were part of the UK Wellcome Trust Case Control Consortium (WTCCC) and have been described in detail previously[11]. Briefly, these four studies were the type 2 diabetes (WTCCC-T2D), hypertension (WTCCC-HT), and coronary artery (WTCCC-CAD) disease branches and the national blood service (WTCCC-UKBS) controls. A manuscript describing the cohorts used in the Diabetes Genetics Initiative (DGI) 500K genome-wide association study for type 2 diabetes has already been published [19] and a description of the sample is also available online (http://www.broad.mit.edu/diabetes/). The EPIC Obesity case-cohort study includes 3,847 participants and is nested within the EPIC-Norfolk Study, a population-based cohort study of 25,663 Europid men and women aged 39-79 years recruited in Norfolk, UK between 1993 and 1997. The cases (N = 1,685) were randomly selected from the obese individuals within this cohort and are defined as those with a BMI >30 kg/m^2. The control-cohort consists of 2,566 individuals randomly selected from the EPIC-Norfolk study and thus by design, 381 individuals are part of the control-cohort as well as the case group.

Basic anthropometric data for all genome-wide studies are presented in **Supplementary Table 1**. Extensive quality control steps were taken to exclude poorly performing or non-European descent samples from analyses. For 5 of the 6 GWA studies these are described in detail in [11, 19]. For the EPIC-Obesity study, of the 3,847 participants, 277 were excluded (sample call rate <94%: N = 202, heterozygosity <23% or >30%: 36, >5.0% discordance in SNP pairs with $r^2 = 1$ in HapMap: N = 25, ethnic outlier: N = 8, related individuals (concordance with another DNA is >70.0% and <99.0%, 1 selected based on sample call rate): N = 5, duplicate (concordance with another DNA is >99.0%, 1 selected based on sample

call rate): N = 1), and for 10 individuals no genotype data was available, such that 3560 individuals were included in the analyses.

The WTCCC-T2D, WTCCC-HT, DGI and EPIC-Obesity studies measured height using standard anthropometric techniques. For WTCCC-CAD and WTCCC-UKBS height data was self reported from questionnaires. The lack of evidence of heterogeneity across all studies for the 20 confirmed loci indicates that the inclusion of self-report data has not affected the results appreciably.

All subjects gave written informed consent and the project protocols were approved by the local research ethics committees in the UK.

Stage 2 samples

UKT2D GCC

This study has been described previously [20]. All subjects were of self-reported white European descent, living in the Tayside region of Dundee, UK. Height measurements were made as for the WTCCC samples. This study was approved by the Tayside Medical Ethics Committee and informed consent was obtained from all subjects.

EFSOCH

EFSOCH (Exeter Family Study of Childhood Health) is a prospective study of parents and children from a consecutive birth cohort [21]. Subjects were recruited from a postcode-defined region of Exeter, UK between 2000 and 2004 and were of self-reported white, European descent. Parental height was measured using a stadiometer by the research midwife at 28 weeks gestation. Ethical approval was given by the North and East Devon Local Research Ethics Committee and informed consent was obtained from the parents of the newborns.

BRIGHT

The MRC British Genetics of Hypertension (BRIGHT) study has been described previously [22]. Briefly, severely hypertensive individuals were recruited

from the Medical Research Council General Practice Framework and other family-physician practices in the UK. All subjects were of self-reported white European ancestry up to level of grandparents. Height was measured by using a Marsden ultrasonic height measure; the standard operating procedure for this is described at the MRC BRIGHT study webpage (www.brightstudy.ac.uk).

CoLaus

The CoLaus study has been described in detail previously [23]. Briefly, it is a single-centre, cross-sectional study including a random sample of 6,188 extensively phenotyped European descent subjects (3,251 women and 2,937 men) aged 35 to 75 years living in Lausanne, Switzerland. Height was measured to the nearest 5mm using a Seca® height gauge (Hamburg, Germany).

Statistical methods

Stage 1 analyses

All GWA studies were genotyped using the Affymetrix 500K chip. For the WTCCC studies we used the WTCCC-defined list of 459,446 QC-passed SNPs [11], with additional exclusion criteria of a MAF > 0.01, and a Hardy-Weinberg Equilibrium $P<1x10^{-4}$ for each individual GWA study, in our analyses. For the EPIC-Obesity study, SNPs included in the analyses have passed the following quality control criteria: they (1) were polymorphic (7,532 excluded), (2) have a call rate ≥90% (31,067 excluded), (3) show Hardy-Weinberg Equilibrium with a $P>10^{-6}$ (25,907 excluded), and (4) have a minor allele frequency of ≥5%. The total number of SNPs analysed is 338,830. The DGI data SNP quality control and exclusion criteria are reported in detail elsewhere [19], and resulted in the use of 386,731 SNPs. We note that there is a small familial component to the DGI data, which is not taken into account in the betas and standard errors provided in the publicly available data that was used in our analyses. The extent of the *P*-value inflation that is caused by this is small (genomic control lambda < 1.1), so will have marginal effects on the association results, but we have provided a DGI excluded

result in **Table 1** to demonstrate the robustness of the associations. We report the 402,951 SNPs which passed QC in at least 4 of the 6 GWA studies.

Individual level genotype data were available from only one GWA study (WTCCC-T2D); only summary height association statistics were available for the other studies. For each GWA study, summary statistics, assuming an additive inheritance model, from linear regression using Z-scores (described below) were generated using PLINK [24] (WTCCC-T2D, WTCCC-UKBS, WTCCC-CAD, DGI), SAS/Genetics 9.1 (SAS Institute Inc., Cary, NC, USA; EPIC-Obesity Study) or using R (http://www.R-project.org; WTCCC-HT).

Stage 2 analyses

For each stage 2 study the associations between genotype and height z-score were examined using linear regression (described below). Stage 2 analyses were performed in Stata/SE 9.1 for Windows (StataCorp LP, Texas, USA) for all studies, except for CoLaus, which were performed using PLINK [24].

Z-score generation

Height was normally distributed in all cohorts. For the WTCCC GWA studies, UKT2D GCC, BRIGHT and EFSOCH studies, sex-specific height Z-scores were generated within each study. Details for the DGI are available at http://www.broad.mit.edu/diabetes/. For EPIC-Obesity, height z-scores were created by sex and age decades (<50, 50<60, 60<70, ≥70). For the CoLaus study, height was corrected using a linear model, regressing height simultaneously onto age, sex, ancestry principal components [8] and grandparental birthplaces. The residuals were rescaled to have variance 1, and then were used as a “corrected” phenotype.

Meta-Analysis

Meta-analysis statistics were generated using the inverse-variance meta-analysis method assuming fixed effects. The Q test was used to test for between-

study heterogeneity. We used Stata/SE 9.1 for Windows (StataCorp LP, Texas, USA) for all meta-analysis calculations.

Eigenstrat Analyses

For the GWA study, EIGENSTRAT [8] was run in each individual study on the full set of markers (~400,000 SNPs). Within each study, similar results were obtained when using the first three principal components or the first ten principal components.

Individual level data analyses in Stage 2 samples

All individual level data analyses were performed in Stata/SE 9.1 for Windows (StataCorp LP, Texas, USA). To test for a deviation from an additive mode of inheritance for each of the 39 SNPs that we took forward into stage 2, we performed a likelihood ratio test of the additive regression model against the full 2df model.

To test for a difference in effect size between sexes, we performed a likelihood ratio test of the additive model against a model that also included a sex by genotype interaction term. To test for an influence of age on the effect size of we compared a regression model including dichotomized age (<50 and >= 50) and genotype to a model that also included a dichotomized age by genotype interaction term. We also performed the same analysis, but using 40 years as a cut-off and age deciles rather than dichotomized age.

For the gene-gene interaction analyses, we assumed additive effects within-loci, and compared a joint effects model to a model containing an interaction term using likelihood ratio tests.

Combined results plot

For the combined effect analyses we used only Stage 2 UK subjects to reduce the effect of the "winners curse"[25]. We only used subjects that had been successfully genotyped at each of the 20 SNPs that reached a $P<5x10^{-7}$, and

grouped subjects by the total number of "tall" alleles that they carried. The mean height (estimated by multiplying the Z-score effect size by 6.82cm, the average SD of adult height across the cohorts used in this study) and frequency were then plotted using SigmaPlot for Windows Version 10.0 (Systat Software, Germany).

Quantile-quantile plots and power results

Quantile-quantile plots were generated using Stata/SE 9.1 for Windows. The 95% concentration bands, which are the approximate 95% confidence intervals around the null distribution were generated as described by [26].

Quanto was used for the power calculation [27]. To assess the impact of the "winners curse" we performed a binomial distribution test of the number of times the stage 1 result was greater than the stage 2 result, compared to that expected under the null of 50%.

Linkage analyses

We used linkage data from the website provided by Perola *et al.*[5] which describes all reports in the literature that achieved LOD scores > 2 for height. Where a peak marker (or markers) was reported we called a 10Mb window around the marker (or markers) a "linked region". Where no peak marker was reported we used the reported DeCode cM coordinates to determine the linked region. To compare the observed number of occasions that one of the 20 "real" SNPs occurred in a linked region to that expected under the null distribution, we took the total number of base-pairs in non-overlapping linked regions and divided it by the number of base-pairs in the human genome (http://genome.ucsc.edu from NCBI build 36.1 statistics). The expected number of times that the 20 real SNPs occurred in linked regions is then 20 x (base-pairs in linked regions / total number of base-pairs in the human genome). We used a Poisson test to determine the significance of the difference in the number of confirmed SNPs observed under linkage peaks compared to the expected number. We did this for SNPs with LODs > 3 and those > 2.

To determine if there was any over-representation of all associations at $P<0.001$ in linked regions we compared the proportions of these SNPs occurring in linked regions to those not occurring in linked regions. Again we performed this for a LOD > 2 and a LOD > 3 cut-off.

Stage 2 genotyping

Genotyping of the UKT2D GCC, BRIGHT and EFOSCH samples was performed by KBiosciences (Hoddesdon, UK) using their own novel system of fluorescence-based competitive allele-specific PCR (KASPar). Details of assay design are available from the KBiosciences website (http://www.kbioscience.co.uk). The CoLaus study is a GWA study (for which GWA data were not available in time for this study to be involved in stage one) and is described in detail elsewhere (Firman *et al.* Submitted).

Acknowledgements

Michael Weedon is a Vandervell Foundation Research Fellow. Cecilia Lindgren is a Nuffield Dept of Medicine Scientific Leadership Fellow. Rachel Freathy is funded by a Diabetes UK research studentship. Sven Bergmann is supported by the Giorgi-Cavaglieri Foundation and the Swiss National Science Foundation (Grant # 3100AO-116323/1), which also supports Jacques S. Beckmann (grant #310000-112552/1). We would like to thank Murielle Bochud, Zoltán Kutalik, Gerard Waeber, Kijoung Song and Xin Yuan for their contribution to the Lausanne study. The WTCCC CAD cohort collection was supported by grants from the British Heart Foundation, Medical Research Council, and National Health Service Research & Development. NJS holds a Chair supported by the British Heart Foundation. We thank the Wellcome Trust for funding. CW is funded by the British Heart Foundation (grant number; FS/05/061/19501). The BRIGHT study is supported by the Medical Research Council (grant number; G9521010D) and the British Heart Foundation (grant number PG02/128).

Author Contributions

MNW, HL, CML, CW, DME, MM, JRBP, SS, IP, members of the DGI, WTCCC, the GEM consortium, SB, TJ, and DMW were responsible for analyzing, quality control checking, and cleaning the data from the individual GWA studies. CW, RMF, BS, MNW and HL were responsible for analysis of the stage 2 samples. MNW performed the meta-analyses. ASH and NJS are principal investigators from the WTCCC-CAD study. MC and MF are principal investigators from the WTCCC-HT study. WHO is principal investigator of the WTCCC-UKBS study. ATH and MIM are principal investigators for the WTCCC-T2D study. JSB, PV, VM are principal investigators of the CoLaus study. MC, MF, AD, PBM are principal investigators on the BRIGHT study. ATH is principal investigator of the EFSOCH study. CNAP and ADM are principal investigators of the Tayside UKT2D-GCC study. MNW, HL, ATH, MIM and TMF wrote the manuscript. ATH, MIM, MNW, and TMF designed and led the study. All authors read and approved the final manuscript.

Figure 1. QQ plots for the 402,951 SNPs from the genome-wide association meta-analysis as more studies are added in. A: N = 1914 (WTCCC-T2D); B: N = 4892 (adding DGI); C: N=6788 (adding WTCCC-HT); D: N=8668 (adding WTCCC-CAD); E: N=12228 (adding EPIC-Obesity); F: N=13665 (adding WTCCC-UKBS). Blue line is the observed P values. The black line is the expected line under the null distribution. The grey bands are 95% concentration bands which are an approximation to the 95% confidence intervals around the expected line.

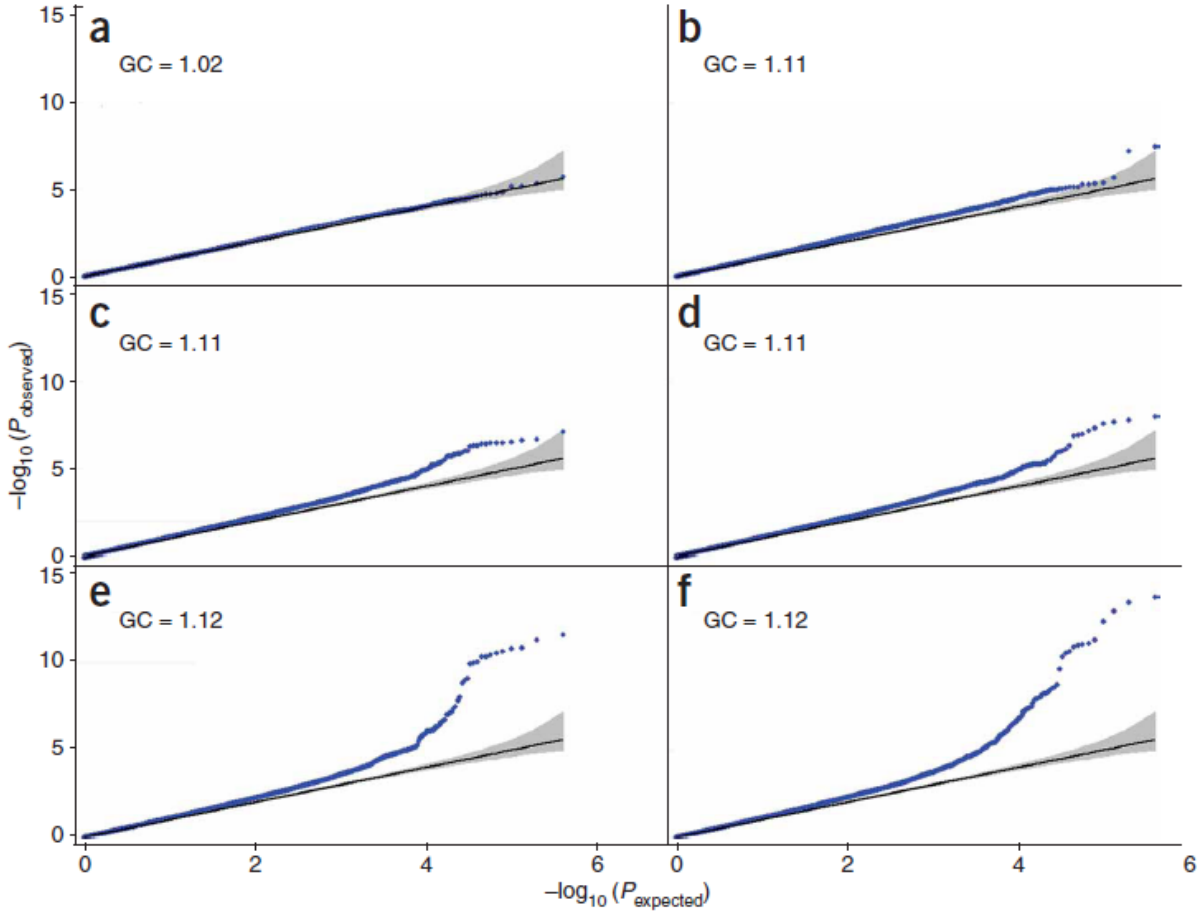


Figure 2. Manhattan plot for the 402,951 SNPs from the stage 1 genome-wide association meta-analysis of the WTCCC-T2D, DGI, WTCCC-HT, WTCCC-CAD, EPIC-Obesity and WTCCC-UKBS studies. The red dots are the SNPs that reached a $P<5x10^{-7}$ in a joint analysis of stage 1 and stage 2 samples.

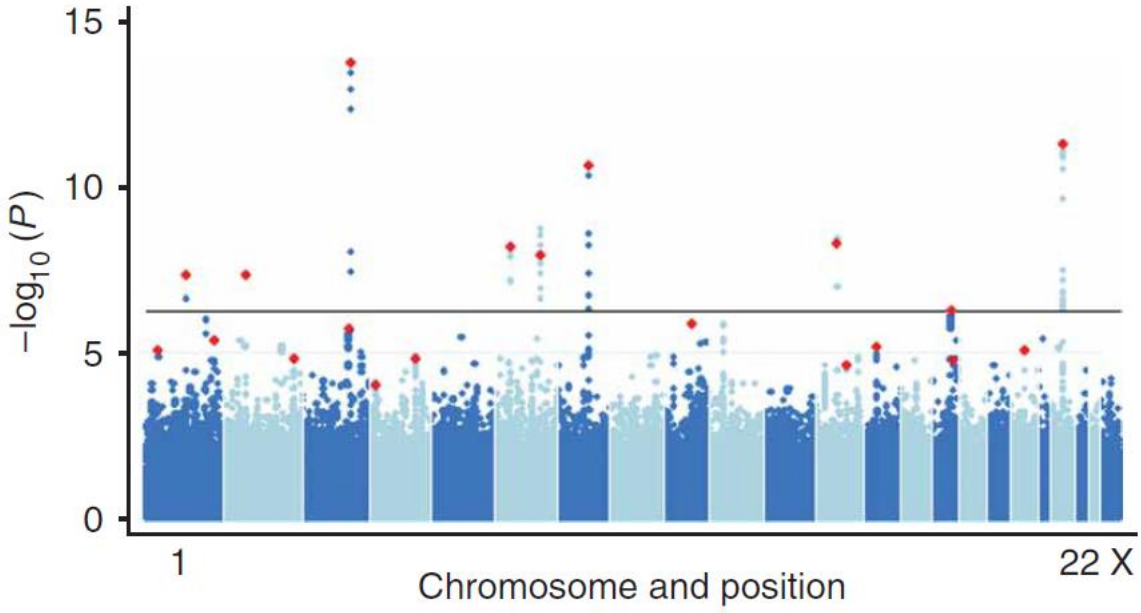


Figure 3. The combined impact of the 20 SNPS with a *P*<5x10^{-7}. Subjects have been classified according to the number of height increasing alleles at each of the 20 SNPS, and the mean height for each group is plotted (blue dots). The black line is a linear regression line through these points. The grey bars represent the proportion of the sample, with increasing numbers of 'tall' alleles. The approximate height difference in cm, was obtained by multiplying the mean Z-score height for each group, by 6.82cm (the approximate average SD of height across the samples used in this study).

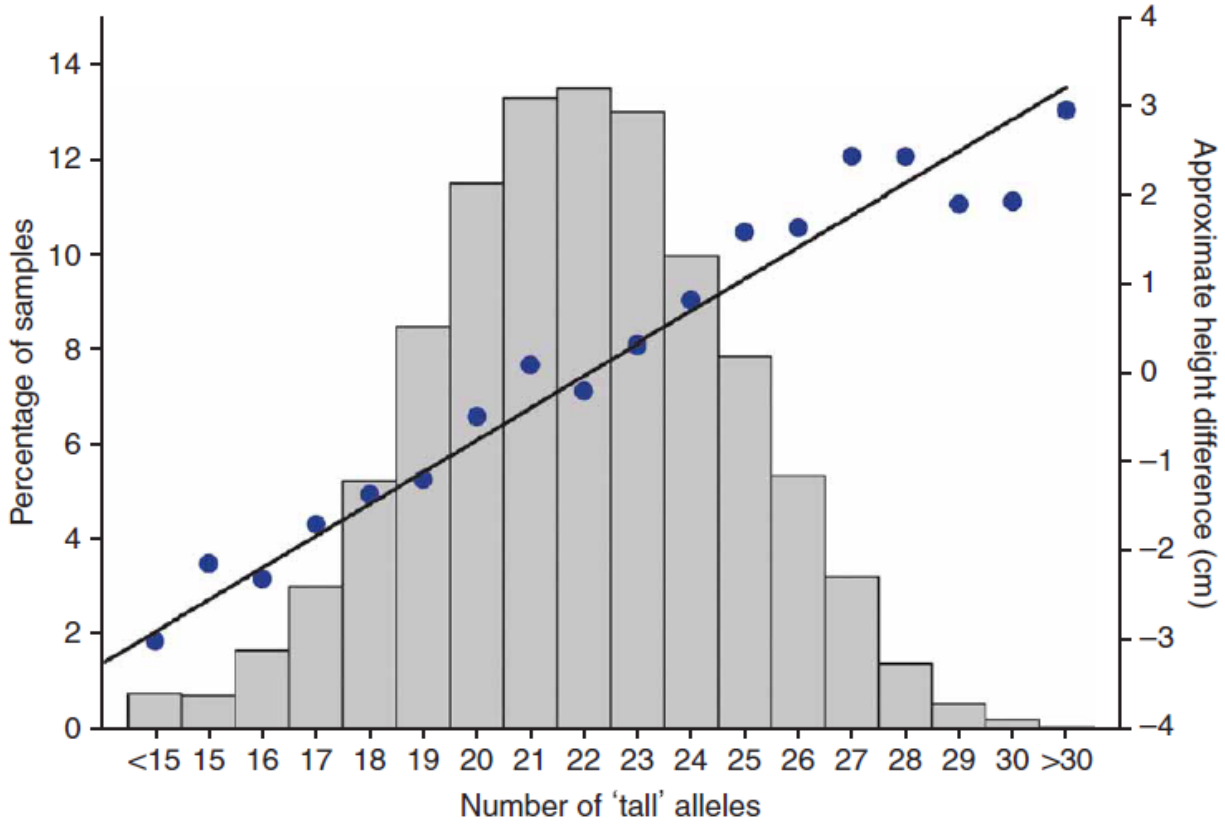


Figure 4 (a) The power of the genome-wide study to identify the variants that had a *P*<5x10^{-7} in the joint analysis at *P*<1x10^{-5} using the effect size estimates from the follow-up samples only and, (b) the sample size required o identify these variants using the effect size estimates from the follow-up samples only at a *P*<5x10^{-7} with 80% power. Effect sizes ranged from 0.083SD with minor allele frequency ~0.44 for rs6440003 to 0.033SD with minor allele frequency of 0.35 for rs8099594.

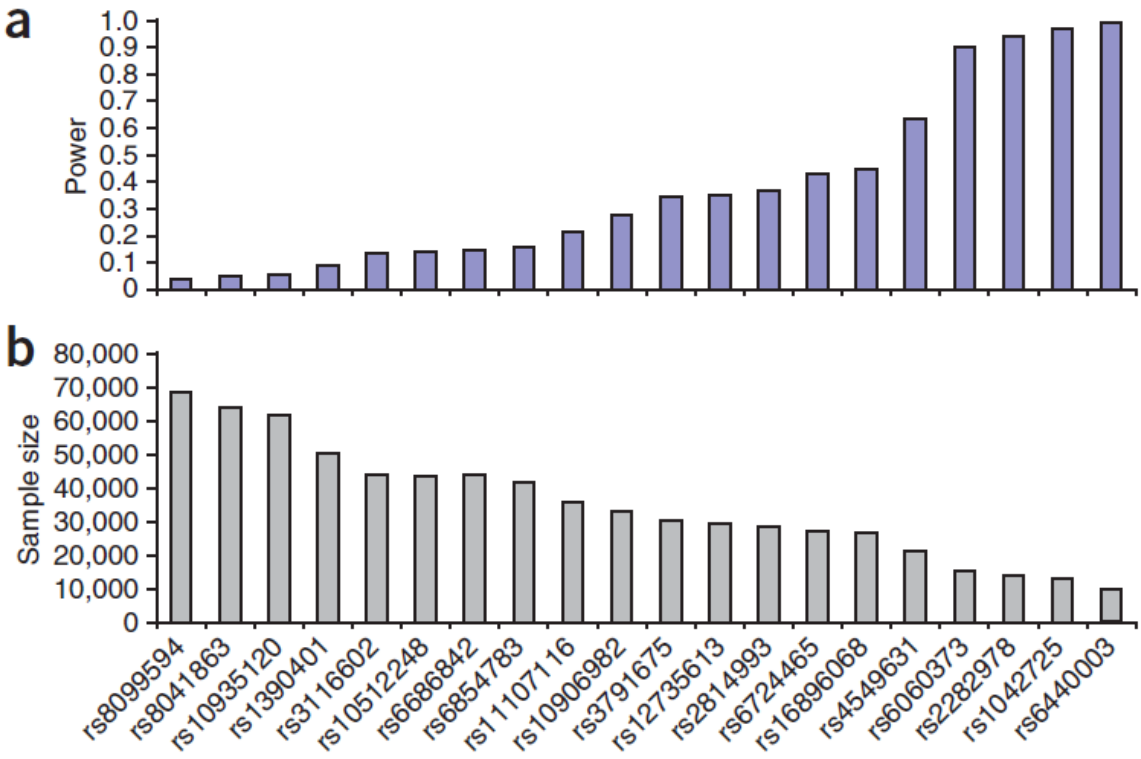


Figure 5. QQ plot for the *P*-values from the accompanying Lettre *et al.*[18] study of the most associated 10,000 SNPs from our study (excluding the DGI component to make the observations independent), including (dark blue dots) and excluding known loci (light blue dots). The black line is the expected line under the null distribution. The grey band is the 95% concentration bands which are an approximation to the 95% confidence intervals around the expected line.

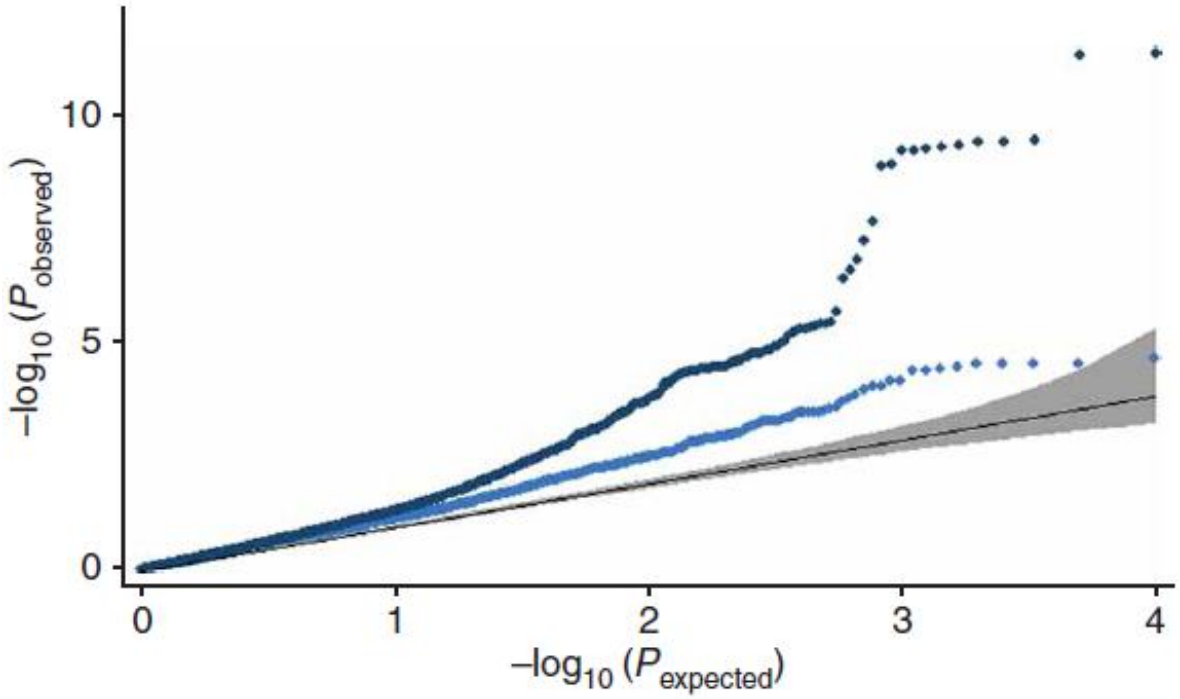


Table 1. Results for the 20 SNPs taken forward into stage 2 that reached *P*<5x10^{-7} in joint analyses. The results are ordered by the joint-analyses *P* value. Chromosome positions are based on NCBI build 125. The alleles all refer to the positive strand. Betas are per each additional copy of allele 1. Minor allele frequency (MAF) based on the minor allele (bold and underlined in the alleles column) in the WTCCC-T2D study. R^2 (% variation explained) is for follow-up sample only, and does not include CoLaus. The additive model test and sex test *P* values do not include data from DGI or CoLaus. * r^2 = 1 proxies used in the stage 2 studies because of assay design issues. Candidate gene is given when monogenic human and/or mouse phenotypes and/or expression results clearly implicate a gene. An overall *P*-value excluding DGI is given because of the small related component of DGI and to provide evidence independent from the accompanying manuscript by Lettre *et al.*[18]

SNP	Candidate Gene	Chromosome (position)	Alleles (1/2)	MAF	Additive model test *P*	Sex test *P*	Male SD difference (95% CI)	Female SD Difference (95% CI)	R^2	GWA Study *P*	Follow-up *P*	Hetero-geneity *P*	Overall *P* excluding DGI	Overall *P*
rs6440003	*ZBTB38*	3 (142576907)	**A**/G	0.44	0.80	0.01	0.07 (0.04, 0.09)	0.12 (0.09, 0.14)	0.32	$1.3x10^{-14}$	$8.7x10^{-12}$	0.52	$2.7x10^{-23}$	$1.8x10^{-24}$
rs2282978 / rs42046 *	*CDK6*	7 (91898623)	**C**/T	0.33	0.14	0.69	0.09 (0.06, 0.12)	0.08 (0.05, 0.11)	0.28	$5.0x10^{-11}$	$5.1x10^{-13}$	0.98	$3.1x10^{-21}$	$7.8x10^{-23}$
rs1042725	*HMGA2*	12 (64644614)	**C**/T	0.49	0.70	0.34	0.05 (0.03, 0.08)	0.07 (0.05, 0.10)	0.25	$5.9x10^{-9}$	$8.6x10^{-11}$	0.50	$1.1x10^{-14}$	$2.5x10^{-18}$
rs6060373	*GDF5*	20 (33377622)	A/**G**	0.38	0.17	0.70	-0.08 (-0.11, -0.05)	-0.07 (-0.10, -0.04)	0.21	$2.2x10^{-12}$	$1.6x10^{-7}$	0.27	$2.0x10^{-15}$	$1.7x10^{-17}$
rs16896068	*LCORL*	4 (17621109)	**A**/G	0.16	0.31	0.99	-0.07 (-0.11, -0.03)	-0.07 (-0.11, -0.03)	0.12	$1.0x10^{-4}$	$2.5x10^{-10}$	0.06	$2.0x10^{-13}$	$2.4x10^{-13}$
rs4549631	*LOC387103*	6 (127008001)	**C**/T	0.50	0.62	0.85	0.06 (0.03, 0.08)	0.05 (0.03, 0.08)	0.11	$1.2x10^{-8}$	$4.6x10^{-6}$	0.47	$2.9x10^{-11}$	$4.7x10^{-13}$
rs3791675	*EFEMP1*	2 (56022960)	C/**T**	0.23	0.43	0.34	0.09 (0.05, 0.12)	0.06 (0.03, 0.10)	0.12	$7.1x10^{-8}$	$6.0x10^{-6}$	0.54	$1.5x10^{-12}$	$2.2x10^{-12}$
rs2814993	*C6orf106*	6 (34726871)	**A**/G	0.15	0.18	0.87	0.09 (0.05, 0.13)	0.10 (0.06, 0.14)	0.20	$8.9x10^{-9}$	$5.7x10^{-5}$	0.04	$4.0x10^{-11}$	$4.1x10^{-12}$
rs10512248	*PTCH1*	9 (95339258)	**G**/T	0.31	0.14	0.10	0.05 (0.02, 0.07)	0.08 (0.05, 0.11)	0.19	$1.5x10^{-6}$	$6.0x10^{-6}$	0.82	$1.0x10^{-9}$	$4.2x10^{-11}$
rs12735613	*SPAG17*	1 (118596015)	**A**/G	0.24	0.0090	0.02	-0.08 (-0.11, -0.05)	-0.03 (-0.06, 0.00)	0.09	$3.4x10^{-8}$	$8.2x10^{-5}$	0.51	$2.0x10^{-9}$	$4.4x10^{-11}$
rs11107116	*SOCS2*	12 (92480972)	G/**T**	0.23	0.047	0.73	-0.04 (-0.07, -0.01)	-0.05 (-0.08, -0.02)	0.06	$2.5x10^{-5}$	$5.6x10^{-6}$	0.41	$2.3x10^{-8}$	$5.6x10^{-10}$
rs6854783 / rs2055059 *	*HHIP*	4 (146000684)	A/**G**	0.43	0.17	0.50	0.06 (0.03, 0.08)	0.04 (0.01, 0.017)	0.10	$1.2x10^{-5}$	$3.2x10^{-5}$	0.24	$2.2x10^{-8}$	$2.1x10^{-9}$
rs1390401	*ZNF678*	1 (224104685)	A/**G**	0.18	0.0067	0.34	0.04 (0.01, 0.08)	0.07 (0.03, 0.10)	0.09	$4.3x10^{-6}$	$2.0x10^{-4}$	0.58	$1.4x10^{-6}$	$5.4x10^{-9}$
rs3116602	*DLEU7*	13 (50009356)	**G**/T	0.21	0.88	0.02	-0.04 (-0.07, 0.00)	-0.09 (-0.12, -0.06)	0.07	$5.6x10^{-6}$	$1.8x10^{-4}$	0.82	$6.1x10^{-9}$	$6.8x10^{-9}$
rs6686842	*SCMH1*	1 (41199964)	C/**T**	0.44	0.30	0.97	-0.05 (-0.08, -0.02)	-0.05 (-0.08, -0.02)	0.14	$8.6x10^{-6}$	$3.3x10^{-4}$	0.57	$4.9x10^{-7}$	$1.7x10^{-8}$
rs10906982	*ADAMTSL3*	15 (82371586)	A/**T**	0.48	0.33	0.92	0.05 (0.02, 0.07)	0.04 (0.02, 0.07)	0.07	$5.4x10^{-7}$	$2.1x10^{-3}$	0.57	$5.3x10^{-7}$	$1.7x10^{-8}$
rs6724465	*IHH*	2 (219769351)	**A**/G	0.10	0.96	0.85	-0.06 (-0.10, -0.02)	-0.05 (-0.10, -0.01)	0.04	$3.1x10^{-5}$	$2.8x10^{-4}$	0.52	$2.2x10^{-6}$	$2.1x10^{-8}$
rs10935120	*ANAPC13 or CEP63*	3 (135715790)	**A**/G	0.33	0.10	0.63	-0.06 (-0.09, -0.03)	-0.05 (-0.08, -0.02)	0.10	$2.2x10^{-6}$	$3.1x10^{-3}$	0.57	$8.7x10^{-7}$	$7.3x10^{-8}$
rs8041863	*ACAN*	15 (87160693)	**A**/T	0.47	0.90	0.21	0.04 (0.01, 0.06)	0.06 (0.03, 0.09)	0.03	$2.2x10^{-5}$	$8.6x10^{-4}$	0.02	$4.9x10^{-9}$	$8.1x10^{-8}$
rs8099594	*DYM*	18 (45245158)	A/**G**	0.35	0.69	0.53	0.05 (0.02, 0.08)	0.04 (0.01, 0.07)	0.01	$7.8x10^{-6}$	$4.1x10^{-3}$	0.008	$1.6x10^{-8}$	$3.1x10^{-7}$

Table 2. Candidate genes in the 20 loci, together with monogenic syndromes and mouse models associated with the genes. Candidate gene is given when monogenic human and/or mouse phenotypes and/or expression results clearly implicate a gene, otherwise nearest gene is given, unless no gene within 500kb window around SNP. Information on each gene was obtained either from the OMIM (http://www.ncbi.nlm.nih.gov/sites/entrez?db=OMIM) or Jackson lab (http://www.jax.org/) websites. * Details are from Uniprot summaries.

SNP	Candidate or nearest gene(s)	Monogenic syndrome caused by mutation in gene	Knockout mouse phenotype	Details*
rs6440003	*Zinc finger and BTB domain-containing protein 38 (ZBTB38)*	-	-	Transcription factor.
rs2282978	*Cyclin-dependent Kinase-6 (CDK6)*	-	15% smaller embryos	Involved in the control of the cell cycle. Interacts with D-type G1 cyclins.
rs1042725	*High-mobility group A2 (HMGA2)*	Disruption causes tall stature, extreme bone and dental overgrowth, and multiple lipomas.	Pygmy mice	Belongs to the non-histone chromosomal high mobility group (HMG) protein family. HMG proteins function as chromatin architectural factors.
rs6060373	*Growth Differentiation Factor 5 (GDF5)*	Chondrodysplasia (abnormally short and deformed limbs); brachydactyly (short digits) DuPan syndrome; multiple synostoses syndrome.	Homozygous null mutants demonstrate skeleton defects, such as reduced or absent limb bones and joints.	Involved in bone formation. Also known as Cartilage-derived morphogenetic protein 1.

rs16896068	*Ligand-dependent nuclear receptor corepressor-like protein (LCORL)*	-	-	May act as transcription activator.
rs4549631	*LOC387103*	-	-	Not known
rs3791675	*EGF-Containing Fibulin-like extracellular matrix protein 1 (EFEMP1)*	Doyne honeycomb retinal dystrophy; no obvious skeletal defects.	Normal phenotype	Extra-cellular matrix. Belongs to the fibulin family.
rs2814993	*C6orf106*	-	-	Not known
rs10512248	*Patched, drosophilia, homolog of, 1 (PTCH1)*	Gorlin syndrome (basal cell carcinoma); holoprosencephaly.	Homozygous null mice die during embryogenesis, heterozygotes larger than normal, with hind limb defects.	Hedgehog signalling. Acts as a receptor for sonic hedgehog (SHH), Indian hedgehog (IHH) and desert hedgehog (DHH).
rs12735613	*Sperm associated antigen 17 (SPAG17)*	-	-	Not known
rs11107116	*Suppressor of cytokine signaling 2 (SOCS2)*	-	Homozygous null mice grow more rapidly. Males are 40% heavier than wild-type littermates; The increase in weight results from general increase in visceral organ weight and long bone length.	SOCS family proteins form part of a classical negative feedback system that regulates cytokine signal transduction. SOCS2 appears to be a negative regulator in the growth hormone/IGF1 signaling pathway.

rs6854783	*Hedgehog interacting protein (HHIP)*	-	Ectopic expression in transgenic mice results in severe skeletal defects similar to those observed in IHH mutants.	Hedgehog signalling. Modulates hedgehog signaling through direct interaction with members of the hedgehog family including SHH, IHH and DHH.
rs1390401	*Zinc finger protein 678 (ZNF678)*	-	-	Transcription factor. Belongs to the Krueppel C2H2-type zinc-finger protein family by similarity.
rs3116602	*Deleted in lymphocytic leukaemia, 7 (DLEU7)*	-	-	Not known
rs6686842	*Sex comb on midleg homolog 1* (*SCMH1)*	-	Homozygous null mice present with multiple defects including of skeleton.	Polycomb protein. A constituent of the mammalian Polycomb repressive complexes 1 involved in chromatin modifications
rs10906982	*ADAMTS-like protein 3 (ADAMTSL3)*	-	-	Extra-cellular matrix. Strongly similar to members of the ADAMTS family but lacks metalloprotease and disintegrin-like domains.
rs6724465	*Indian hedgehog (IHH)*	Brachydactyly; acrocapitofemoral dysplasia (cone-shaped ends of hand and hip bones).	Homozygous null mice display impaired chondrocyte proliferation and maturation, resulting in dwarfism and numerous skeletal abnormalities.	Hedgehog signalling. Intercellular signal essential for a variety of patterning events during development. Binds to the patched (PTC) receptor.

rs10935120	*Anaphase promoting complex subunit 13 (ANAPC13)*	-	-	Cell cycle. Component of the anaphase promoting complex/cyclosome (APC/C), a cell cycle-regulated E3 ubiquitin ligase that controls progression through mitosis and the G1 phase of the cell cycle.
rs8041863	*Aggrecan (ACAN)*	Autosomal dominant spondyloepiphyseal dysplasia type Kimberley, characterised by severe, premature osteoarthritis.	Homozygous mutants are dwarfed at birth.	Extra-cellular matrix. A member of the aggrecan/versican proteoglycan family. Part of the extra-cellular matrix in cartilaginous tissue.
rs8099594	*Dymeclin (DYM)*	Autosomal recessive disorder characterized by abnormal skeletal development and mental retardation.	-	May have a role in process of intracellular digestion of proteins or in proteoglycan metabolism.

References

1. Macgregor, S., Cornes, B.K., Martin, N.G. & Visscher, P.M. Bias, precision and heritability of self-reported and clinically measured height in Australian twins. *Human Genetics* **120**, 571-580 (2006).

2. Preece, M.A. The genetic contribution to stature. *Hormone Research* **45**, 56-58 (1996).

3. Silventoinen, K., Kaprio, J., Lahelma, E. & Koskenvuo, M. Relative effect of genetic and environmental factors on body height: Differences across birth cohorts among Finnish men and women. *American Journal Of Public Health* **90**, 627-630 (2000).

4. Silventoinen, K. *et al.* Heritability of adult body height: A comparative study of twin cohorts in eight countries. *Twin Research* **6**, 399-408 (2003).

5. Perola, M. *et al.* Combined Genome Scans for Body Stature in 6,602 European Twins: Evidence for Common Caucasian Loci. *PLoS Genetics* **3**, e97 (2007).

6. Palmert, M.R. & Hirschhorn, J.N. Genetic approaches to stature, pubertal timing, and other complex traits. *Molecular Genetics and Metabolism* **80**, 1 (2003).

7. Weedon, M.N. *et al.* A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nat Genet* **39**, 1245-1250 (2007).

8. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-9 (2006).

9. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-1004 (1999).

10. Freedman, M.L. *et al.* Assessing the impact of population stratification on genetic association studies. *Nat Genet* **36**, 388-93 (2004).

11. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661 (2007).

12. Dixon, A.L. *et al.* A genome-wide association study of global gene expression. *Nat Genet* **39**, 1202-7 (2007).

13. Malumbres, M. & Barbacid, M. Mammalian cyclin-dependent kinases. *Trends Biochem Sci* **30**, 630-41 (2005).

14. Southam, L. *et al.* An SNP in the 5'-UTR of GDF5 is associated with osteoarthritis susceptibility in Europeans and with in vivo differences in allelic expression in articular cartilage. *Hum Mol Genet* **16**, 2226-32 (2007).

15. Miyamoto, Y. *et al.* A functional polymorphism in the 5' UTR of GDF5 is associated with susceptibility to osteoarthritis. *Nat Genet* **39**, 529-33 (2007).

16. Lohmueller, K.E., Pearce, C.L., Pike, M., Lander, E.S. & Hirschhorn, J.N. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* **33**, 177 (2003).

17. Fisher, R.A. The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh*, 399-433 (1918).

18. Lettre, G. *et al.* Genome-wide association studies identify 10 novel loci for height and highlight new biological pathways in human growth. *In Press*. *Nature Genetics* (2008).

19. Saxena, R. et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 316, 1331-1336 (2007).

20. Zeggini, E. *et al.* Replication of Genome-Wide Association Signals in U.K. Samples Reveals Risk Loci for Type 2 Diabetes. *Science*, 1142364 (2007).

21. Knight, B., Shields, B.M. & Hattersley, A.T. The Exeter Family Study of Childhood Health (EFSOCH): study protocol and methodology. *Paediatric And Perinatal Epidemiology* **20**, 172-179 (2006).

22. Caulfield, M. *et al.* Genome-wide mapping of human loci for essential hypertension. *Lancet* **361**, 2118-23 (2003).

23. Marques-Vidal, P. *et al.* Prevalence and characteristics of vitamin or dietary supplement users in Lausanne, Switzerland: the CoLaus study. *Eur J Clin Nutr* (2007).

24. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).

25. Stirling, W.D. Enhancements To Aid Interpretation Of Probability Plots. *Statistician* **31**, 211-220 (1982).

26. Gauderman, W.J. Sample size requirements for association studies of gene-gene interaction. *Am J Epidemiol* **155**, 478-84 (2002).

SUPPLEMENTARY INFORMATION

Supplementary Table 1. Characteristics of samples used in this study.

		N	% Male	Males: age at study (yrs; mean, SD)	Females: age at study (yrs; mean, SD)	Males: average height (cm; mean, SD)	Females: average height (cm; mean, SD)
a) GWAS	WTCCC-T2D	1914	58.2	59.0 (9.9)	57.9 (10.4)	175.5 (7.0)	161.5 (6.6)
	WTCCC-HT	1896	39.9	56.5 (10.9)	57.8 (11.1)	174.3 (7.4)	161.3 (6.4)
	WTCCC-CAD	1880	79.4	60.0 (8.0)	60.3 (8.5)	173.8 (6.9)	159.8 (6.7)
	WTCCC-UKBS	1437	48.2	45.4 (11.8)	41.4 (12.6)	178.2 (6.7)	164.7 (6.5)
	EPIC-Obesity	3560	45.7	59.8 (9.0)	58.8 (8.9)	173.7 (6.7)	160.9 (6.0)
	DGI (T2D)	1511	50.6	63.1 (10.3)	65.4 (10.5)	174.3 (6.4)	161.1 (6.2)
	DGI (Controls)	1467	48.6	58.4 (10.5)	59.2 (10.3)	175.6 (6.2)	162.4 (5.9)
b) Replication samples	UKT2D-GCC	6698	54.0	61.8 (10.7)	60.6 (11.7)	175.2 (6.9)	161.2 (6.7)
	EFSOCH	1929	49.5	32.9 (6.0)	30.4 (5.2)	177.9 (6.6)	165.0 (6.3)
	BRIGHT	2446	39.0	58.7 (9.7)	59.0 (9.1)	175.6 (7.0)	162.3 (6.6)
	CoLaus	5409	47.1	52.9 (10.8)	53.9 (10.7)	175.0 (7.4)	162.6 (6.7)

Supplementary Table 2. Results for 12 UK ancestry informative markers described by the WTCCC. The beta refers to the effect of each additional copy of allele 1. Only five of the AIMS were taken into the replication cohorts. Chromosome positions are based on NCBI build 125. The alleles all refer to the positive strand. *rs6644913 is also described as an AIM by the WTCCC, but did not pass quality control criteria.

SNP	Chromosome	Alleles (1/2)	Stage 1 beta	Stage 1 *P*	Stage 2 beta	Stage 2 *P*	Overall beta	Overall *P*
rs1042712	2q21	C/G	-0.038	0.04	-0.002	0.89	-0.017	0.32
rs7696175	4p14	C/T	0.010	0.43	0.009	0.40	0.009	0.81
rs1460133	4q28	C/T	0.002	0.90	-	-	-	-
rs9378805	6p25	A/C	-0.005	0.67	0.015	0.16	0.006	0.45
rs3873375	6p21	C/T	0.016	0.21	-	-	-	-
rs11790408	9p24	G/T	0.012	0.35	-	-	-	-
rs12295525	11p15	C/T	0.020	0.33	-	-	-	-
rs12797951	11q13	G/T	-0.004	0.79	-	-	-	-
rs10774241	12p13	A/G	0.007	0.64	-0.013	0.35	-0.004	0.90
rs17449560	14q12	C/G	0.023	0.25	-	-	-	-
rs3760843	19q13	A/T	0.027	0.08	-	-	-	-
rs2143877	20q12	A/G	0.008	0.55	0.004	0.76	0.006	0.52

Supplementary Table 3. Results for the SNPs taken forward into stage 2 which did not reach a *P*<5x10^{-7} in joint analysis. The results are ordered by the joint-analyses *P* value. Chromosomes positions are based on NCBI build 125. The alleles all refer to the positive strand. Betas are per each additional copy of allele 1. Minor allele frequency (MAF) based on the minor allele (bold and underlined in the alleles column) in the WTCCC-T2D study. R^2 is for follow-up sample only, and does not include CoLaus. The additive model test, and sex test P values do not include data from DGI or CoLaus. * Failed in Broad, EPIC-Obesity and CoLaus.

SNP	Chromosome (position)	Alleles (1/2)	MAF	Additive model test *P*	Sex test *P*	Male SD difference (95% CI)	Female SD Difference (95% CI)	R^2 (%)	GWA Study *P*	Follow-up *P*	Overall *P*
rs11049407*	12 (28225631)	A/**G**	0.30	0.24	0.82	0.06 (0.02, 0.09)	0.05 (0.02, 0.08)	0.06	2.5x10^{-5}	1.0x10^{-2}	4.1x10^{-6}
rs508521	6 (116853934)	**A**/G	0.28	0.12	0.63	-0.05 (-0.08, -0.02)	-0.06 (-0.09, -0.03)	0.05	3.3x10^{-6}	4.9x10^{-2}	5.1x10^{-6}
rs509035	3 (173646151)	**A**/G	0.31	0.20	0.92	0.04 (0.01, 0.07)	0.04 (0.01, 0.07)	0.02	9.4x10^{-6}	4.8x10^{-2}	8.5x10^{-6}
rs6598287	15 (98359299)	C/**T**	0.40	0.68	0.79	0.05 (0.02, 0.07)	0.04 (0.01, 0.07)	0.03	4.0x10^{-6}	0.08	1.1x10^{-5}
rs7030440	9 (130470174)	**A**/G	0.33	0.86	0.83	-0.03 (-0.06, 0.00)	-0.03 (-0.06, 0.00)	0.01	4.8x10^{-6}	0.11	2.0x10^{-5}
rs450902	12 (118723761)	**A**/G	0.34	0.85	0.17	0.05 (0.02, 0.08)	0.02 (-0.01, 0.05)	0.03	2.6x10^{-5}	0.12	5.3x10^{-5}
rs4934353	10 (89379558)	A/**G**	0.24	0.76	0.48	-0.03 (-0.06, 0.00)	-0.05 (-0.08, -0.01)	0.02	8.3x10^{-5}	7.0x10^{-2}	6.3x10^{-5}
rs2096196	1 (200664650)	C/**T**	0.44	0.78	0.05	0.02 (-0.01, 0.04)	0.05 (0.03, 0.08)	0.01	1.3x10^{-6}	0.40	9.3x10^{-5}
rs2043314	19 (35805410)	C/**T**	0.48	0.18	0.60	-0.04 (-0.07, -0.01)	-0.03 (-0.06, 0.00)	0.01	2.6x10^{-6}	0.28	9.4x10^{-5}
rs7567851	2 (178510227)	**C**/G	0.08	0.33	0.91	0.06 (0.01, 0.11)	0.05 (0.00, 0.10)	<0.01	5.8x10^{-6}	0.15	1.1x10^{-4}
rs17001086	19 (10978436)	**C**/T	0.11	0.41	0.40	-0.04 (-0.08, 0.00)	-0.07 (-0.11, -0.03)	0.04	7.8x10^{-5}	0.13	1.8x10^{-4}
rs4527833	8 (130830826)	C/**T**	0.48	0.71	0.91	0.03 (0.00, 0.06)	0.03 (0.00, 0.05)	<0.01	7.8x10^{-6}	0.40	3.8x10^{-4}
rs12625434	20 (18369779)	**C**/T	0.50	0.05	0.89	-0.03 (-0.06, 0.00)	-0.03 (-0.06, 0.00)	<0.01	8.5x10^{-6}	0.39	5.6x10^{-4}
rs211389	10 (32334017)	**G**/T	0.12	0.72	0.73	-0.03 (-0.07, 0.01)	-0.04 (-0.08, 0.01)	<0.01	1.1x10^{-6}	0.92	7.3x10^{-4}

rs3806089	6 (71027464)	**C**/T	0.10	0.98	0.66	-0.04 (-0.08, 0.00)	-0.05 (-0.10, -0.01)	<0.01	$3.3x10^{-5}$	0.69	$1.4x10^{-3}$
rs1556263	9 (113544161)	C/**T**	0.47	0.007	0.13	0.04 (0.02, 0.07)	0.01 (-0.01, 0.04)	<0.01	$7.0x10^{-6}$	0.83	$3.8x10^{-3}$
rs12539316	7 (72422549)	A/**G**	0.28	0.97	0.92	-0.03 (-0.06, 0.00)	-0.03 (-0.06, 0.00)	<0.01	$1.4x10^{-5}$	0.22	$4.4x10^{-3}$
rs17082799	5 (92344239)	A/**G**	0.04	0.22	0.98	-0.03 (-0.11, 0.04)	-0.04 (-0.11, 0.04)	<0.01	$6.6x10^{-6}$	0.22	0.055
rs4130172	2 (43028076)	C/**T**	0.37	0.91	0.67	0.03 (0.00, 0.06)	0.04 (0.00, 0.07)	<0.01	$4.1x10^{-6}$	0.23	0.059

Supplementary Table 4. Effect of age on height associations for all 39 SNPs taken into stage 2. The results do not include data from DGI and CoLaus. Effect size estimates are given per additional height-increasing allele. Decade cut offs were <=30, 31 to 40, 41 to 50, 51 to 60, 61 to 70 and > 70 years of age.

SNP	SNP overall SD difference (95% CI)	SNP + age (decade) SD difference (95% CI)	Age test *P* (cut-off 40yrs)	Age test *P* (cut-off 50yrs)
rs6440003	0.090 (0.07, 0.11)	0.088 (0.07, 0.11)	0.34	0.71
rs2282978	0.086 (0.06, 0.11)	0.084 (0.06, 0.10)	0.94	0.54
rs1042725	0.063 (0.04, 0.08)	0.063 (0.04, 0.08)	0.39	0.43
rs6060373	0.076 (0.06, 0.10)	0.076 (0.06, 0.10)	0.50	0.97
rs16896068	0.071 (0.05, 0.10)	0.069 (0.04, 0.09)	0.08	0.05
rs4549631	0.056 (0.04, 0.08)	0.054 (0.04, 0.07)	0.30	0.79
rs3791675	0.075 (0.05, 0.10)	0.074 (0.05, 0.10)	0.48	0.47
rs2814993	0.096 (0.07, 0.12)	0.096 (0.07, 0.12)	0.13	1.00
rs10512248	0.063 (0.04, 0.08)	0.062 (0.04, 0.08)	0.44	0.38
rs12735613	0.059 (0.04, 0.08)	0.059 (0.04, 0.08)	0.93	0.20
rs11107116	0.045 (0.02, 0.07)	0.047 (0.02, 0.07)	0.88	0.41
rs6854783	0.049 (0.03, 0.07)	0.050 (0.03, 0.07)	0.78	0.65
rs1390401	0.053 (0.03, 0.08)	0.048 (0.02, 0.07)	0.32	0.89
rs3116602	0.061 (0.04, 0.08)	0.061 (0.04, 0.08)	0.32	0.31
rs6686842	0.051 (0.03, 0.07)	0.050 (0.03, 0.07)	0.64	0.68
rs10906982	0.046 (0.03, 0.06)	0.045 (0.03, 0.06)	0.11	0.43
rs6724465	0.058 (0.03, 0.09)	0.059 (0.03, 0.09)	0.86	0.07
rs10935120	0.055 (0.03, 0.07)	0.052 (0.03, 0.07)	0.74	0.92
rs8041863	0.048 (0.03, 0.07)	0.048 (0.03, 0.07)	0.48	0.29
rs8099594	0.045 (0.02, 0.07)	0.047 (0.03, 0.07)	0.36	0.51
rs11049407	0.054 (0.03, 0.08)	0.054 (0.03, 0.08)	0.22	0.06
rs508521	0.055 (0.03, 0.08)	0.056 (0.04, 0.08)	0.23	0.87
rs509035	0.039 (0.02, 0.06)	0.041 (0.02, 0.06)	0.62	0.99
rs6598287	0.044 (0.03, 0.06)	0.047 (0.03, 0.07)	0.30	0.69
rs7030440	0.031 (0.01, 0.05)	0.030 (0.01, 0.05)	0.45	0.28
rs450902	0.036 (0.02, 0.06)	0.037 (0.02, 0.06)	0.12	0.17
rs4934353	0.037 (0.02, 0.06)	0.037 (0.02, 0.06)	0.96	0.83
rs2096196	0.034 (0.02, 0.05)	0.036 (0.02, 0.05)	0.10	0.24
rs2043314	0.036 (0.02, 0.06)	0.037 (0.02, 0.06)	0.47	0.73
rs7567851	0.057 (0.02, 0.09)	0.060 (0.03, 0.09)	0.24	0.75
rs17001086	0.055 (0.03, 0.09)	0.056 (0.03, 0.09)	0.93	0.99
rs4527833	0.027 (0.01, 0.05)	0.027 (0.01, 0.05)	0.39	0.36

rs12625434	0.031 (0.01, 0.05)	0.027 (0.01, 0.05)	0.48	0.93
rs211389	0.031 (0.00, 0.06)	0.032 (0.00, 0.06)	0.39	0.64
rs3806089	0.046 (0.01, 0.08)	0.052 (0.02, 0.08)	0.28	0.79
rs1556263	0.030 (0.01, 0.05)	0.028 (0.01, 0.05)	1.00	0.53
rs12539316	0.030 (0.01, 0.05)	0.033 (0.01, 0.05)	0.82	0.48
rs17082799	0.035 (-0.02, 0.09)	0.026 (-0.03, 0.08)	0.71	0.85
rs4130172	0.030 (0.01, 0.05)	0.031 (0.01, 0.05)	0.47	0.65

Supplementary Table 5. Interaction results for each of the 20 SNPs that reached *P*<5x10^{-7} in joint analyses. The beta refers to the multiplicative effect of each additional copy of the "tall" alleles at each of the pairs of SNPs.

SNP1	SNP2	Interaction beta (95% CI)	Interaction P
rs4549631	rs16896068	-0.046 (-0.084, -0.008)	0.017
rs1042725	rs10906982	0.032 (0.005, 0.059)	0.021
rs2814993	rs6060373	0.048 (0.007, 0.09)	0.022
rs2814993	rs4549631	0.045 (0.004, 0.085)	0.030
rs6686842	rs16896068	0.041 (0.002, 0.079)	0.039
rs12735613	rs4549631	0.033 (0.002, 0.065)	0.039
rs6686842	rs2282978	-0.030 (-0.061, 0)	0.049
rs3791675	rs1042725	-0.031 (-0.063, 0.001)	0.054
rs6686842	rs1390401	0.034 (-0.001, 0.069)	0.058
rs6854783	rs11107116	-0.031 (-0.064, 0.002)	0.067
rs6440003	rs6854783	-0.025 (-0.052, 0.003)	0.077
rs3791675	rs10906982	-0.029 (-0.06, 0.003)	0.078
rs10935120	rs4549631	0.025 (-0.004, 0.055)	0.090
rs3116602	rs16896068	0.039 (-0.007, 0.085)	0.093
rs6060373	rs16896068	0.033 (-0.006, 0.071)	0.096
rs12735613	rs3791675	0.031 (-0.006, 0.068)	0.097
rs6854783	rs2282978	0.024 (-0.006, 0.054)	0.120
rs3791675	rs2814993	0.036 (-0.01, 0.083)	0.124
rs11107116	rs1390401	0.033 (-0.009, 0.074)	0.125
rs10512248	rs10906982	-0.022 (-0.051, 0.006)	0.128
rs6686842	rs6440003	-0.021 (-0.049, 0.006)	0.131
rs3791675	rs6440003	0.024 (-0.008, 0.056)	0.138
rs8041863	rs6060373	-0.021 (-0.048, 0.007)	0.140
rs6440003	rs3116602	0.024 (-0.009, 0.057)	0.155
rs2282978	rs8099594	-0.022 (-0.053, 0.009)	0.171
rs11107116	rs3116602	0.027 (-0.013, 0.067)	0.183
rs12735613	rs2814993	0.031 (-0.015, 0.078)	0.187
rs3791675	rs11107116	-0.026 (-0.064, 0.013)	0.188
rs2814993	rs3116602	-0.032 (-0.08, 0.016)	0.189
rs11107116	rs6060373	-0.022 (-0.055, 0.011)	0.191
rs6854783	rs4549631	-0.018 (-0.046, 0.009)	0.193
rs6440003	rs8099594	-0.019 (-0.047, 0.01)	0.193
rs6724465	rs10935120	-0.031 (-0.079, 0.016)	0.194
rs2282978	rs3116602	0.022 (-0.014, 0.058)	0.233
rs12735613	rs6440003	-0.019 (-0.05, 0.012)	0.238
rs6724465	rs2814993	0.039 (-0.027, 0.106)	0.244
rs8099594	rs1390401	0.022 (-0.015, 0.058)	0.246
rs3791675	rs16896068	-0.025 (-0.068, 0.018)	0.248
rs10512248	rs8099594	-0.018 (-0.048, 0.013)	0.252
rs4549631	rs2282978	0.017 (-0.012, 0.047)	0.253
rs3791675	rs1390401	0.023 (-0.017, 0.064)	0.258

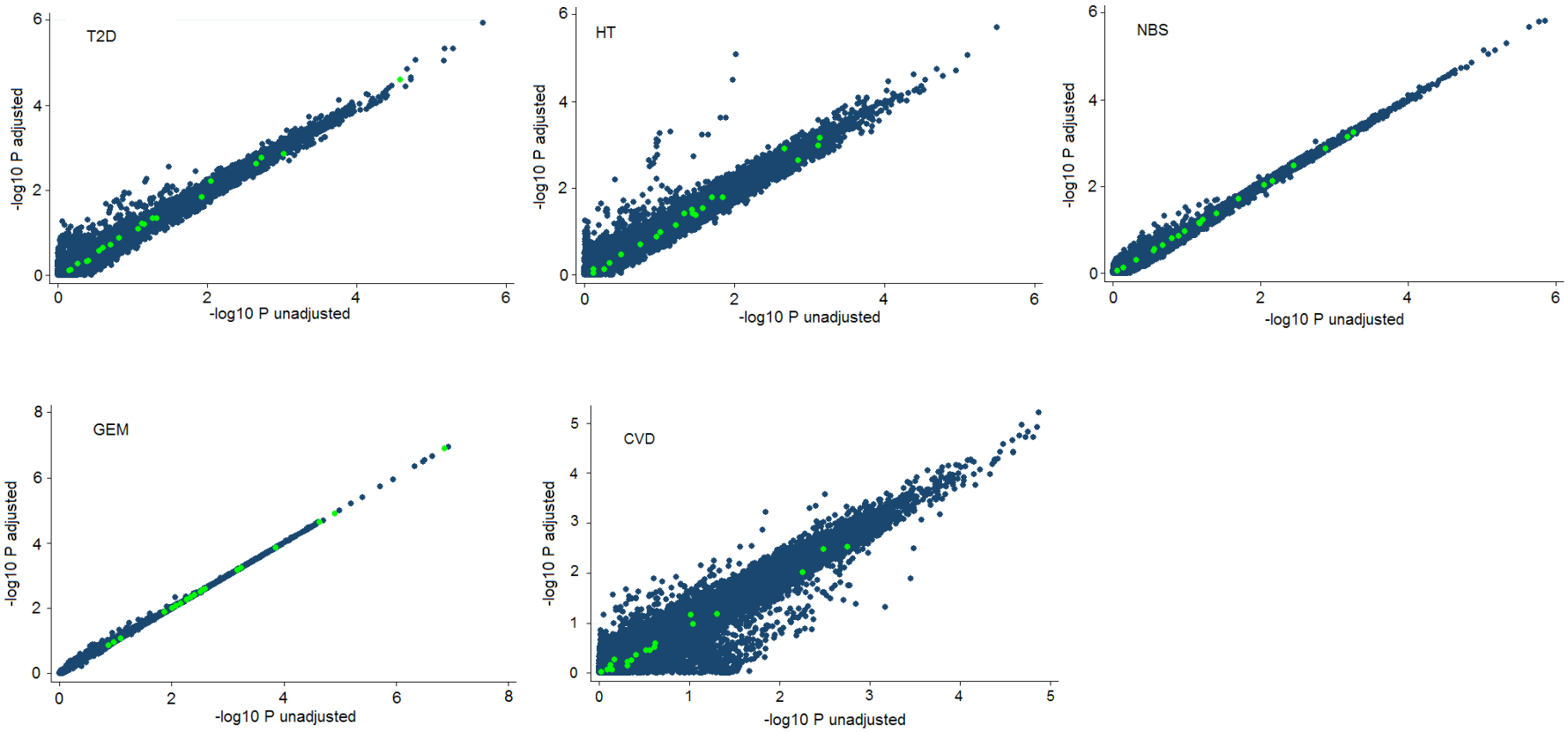
rs2282978	rs6060373	-0.018 (-0.048, 0.013)	0.258
rs1042725	rs1390401	0.020 (-0.015, 0.054)	0.270
rs2814993	rs10906982	-0.022 (-0.063, 0.018)	0.274
rs3116602	rs6060373	0.019 (-0.015, 0.053)	0.274
rs6854783	rs1042725	0.015 (-0.012, 0.043)	0.275
rs4549631	rs8041863	0.015 (-0.012, 0.042)	0.277
rs6440003	rs10512248	-0.016 (-0.045, 0.013)	0.281
rs2282978	rs8041863	0.016 (-0.013, 0.046)	0.282
rs10512248	rs16896068	0.021 (-0.018, 0.061)	0.288
rs6724465	rs1390401	0.031 (-0.026, 0.088)	0.292
rs3116602	rs8041863	-0.017 (-0.051, 0.016)	0.301
rs11107116	rs8099594	-0.018 (-0.052, 0.016)	0.302
rs6854783	rs1390401	0.018 (-0.017, 0.053)	0.308
rs6440003	rs2814993	0.021 (-0.019, 0.061)	0.308
rs4549631	rs1390401	-0.018 (-0.052, 0.017)	0.311
rs10935120	rs6854783	-0.015 (-0.044, 0.014)	0.311
rs8099594	rs16896068	0.020 (-0.019, 0.06)	0.314
rs6686842	rs4549631	0.014 (-0.014, 0.042)	0.326
rs6440003	rs16896068	0.018 (-0.019, 0.056)	0.338
rs6724465	rs10906982	0.022 (-0.023, 0.067)	0.339
rs8041863	rs8099594	0.014 (-0.015, 0.042)	0.339
rs6724465	rs8041863	-0.022 (-0.066, 0.023)	0.340
rs6686842	rs1042725	-0.013 (-0.041, 0.014)	0.345
rs2282978	rs11107116	-0.017 (-0.052, 0.019)	0.349
rs12735613	rs10512248	-0.016 (-0.049, 0.018)	0.360
rs3791675	rs6724465	0.025 (-0.028, 0.078)	0.360
rs10935120	rs16896068	0.018 (-0.022, 0.058)	0.373
rs6440003	rs2282978	0.013 (-0.016, 0.043)	0.373
rs6686842	rs6854783	0.012 (-0.016, 0.041)	0.386
rs8041863	rs16896068	0.016 (-0.021, 0.054)	0.389
rs6440003	rs4549631	-0.012 (-0.039, 0.015)	0.392
rs3791675	rs3116602	-0.017 (-0.055, 0.022)	0.393
rs4549631	rs8099594	-0.012 (-0.041, 0.016)	0.393
rs6724465	rs6060373	-0.020 (-0.067, 0.027)	0.398
rs10935120	rs10906982	-0.012 (-0.041, 0.016)	0.401
rs6440003	rs1390401	-0.015 (-0.049, 0.02)	0.401
rs6686842	rs10906982	0.012 (-0.016, 0.039)	0.409
rs2814993	rs10512248	0.018 (-0.025, 0.06)	0.416
rs6854783	rs6060373	-0.012 (-0.04, 0.017)	0.416
rs10906982	rs16896068	0.015 (-0.022, 0.052)	0.427
rs6686842	rs11107116	0.013 (-0.02, 0.047)	0.430
rs1042725	rs8041863	-0.011 (-0.038, 0.016)	0.438
rs4549631	rs10906982	0.011 (-0.017, 0.038)	0.441
rs6686842	rs10935120	-0.012 (-0.042, 0.018)	0.441
rs4549631	rs1042725	-0.011 (-0.038, 0.017)	0.442
rs4549631	rs11107116	-0.013 (-0.046, 0.02)	0.443

rs12735613	rs10906982	-0.012 (-0.043, 0.019)	0.452
rs12735613	rs16896068	0.016 (-0.027, 0.06)	0.465
rs6854783	rs8099594	0.011 (-0.018, 0.039)	0.473
rs2814993	rs16896068	-0.020 (-0.075, 0.035)	0.474
rs3791675	rs10512248	-0.012 (-0.046, 0.022)	0.486
rs3791675	rs6854783	-0.011 (-0.043, 0.021)	0.491
rs3116602	rs8099594	0.012 (-0.023, 0.047)	0.495
rs6854783	rs10512248	0.010 (-0.019, 0.039)	0.507
rs2282978	rs16896068	-0.014 (-0.054, 0.027)	0.508
rs6440003	rs1042725	0.009 (-0.018, 0.036)	0.509
rs1042725	rs6060373	-0.009 (-0.037, 0.019)	0.509
rs10906982	rs6060373	-0.009 (-0.037, 0.019)	0.518
rs6686842	rs12735613	0.010 (-0.022, 0.042)	0.522
rs10935120	rs1390401	0.011 (-0.025, 0.048)	0.543
rs3791675	rs8099594	-0.010 (-0.044, 0.023)	0.543
rs10512248	rs8041863	-0.009 (-0.038, 0.02)	0.550
rs4549631	rs6060373	0.008 (-0.019, 0.036)	0.553
rs10512248	rs1042725	-0.009 (-0.037, 0.02)	0.554
rs6686842	rs2814993	0.012 (-0.029, 0.053)	0.554
rs6724465	rs3116602	-0.017 (-0.073, 0.04)	0.558
rs3791675	rs6060373	-0.009 (-0.042, 0.023)	0.572
rs8099594	rs6060373	0.008 (-0.021, 0.038)	0.578
rs10512248	rs1390401	0.010 (-0.027, 0.047)	0.594
rs10935120	rs2282978	-0.009 (-0.04, 0.023)	0.597
rs6686842	rs3791675	-0.008 (-0.041, 0.024)	0.609
rs3791675	rs2282978	0.009 (-0.026, 0.044)	0.618
rs11107116	rs10906982	0.008 (-0.024, 0.041)	0.626
rs6724465	rs2282978	-0.012 (-0.061, 0.037)	0.626
rs2814993	rs1042725	0.010 (-0.03, 0.05)	0.627
rs6854783	rs8041863	0.007 (-0.021, 0.034)	0.627
rs12735613	rs2282978	-0.008 (-0.043, 0.026)	0.630
rs2282978	rs10512248	-0.008 (-0.039, 0.023)	0.631
rs2814993	rs1390401	0.012 (-0.039, 0.062)	0.646
rs1042725	rs16896068	-0.009 (-0.046, 0.029)	0.654
rs6854783	rs2814993	0.009 (-0.032, 0.05)	0.659
rs1042725	rs8099594	0.006 (-0.022, 0.035)	0.660
rs10935120	rs8099594	-0.007 (-0.037, 0.024)	0.666
rs10512248	rs11107116	-0.007 (-0.042, 0.027)	0.680
rs10935120	rs6440003	0.006 (-0.023, 0.035)	0.694
rs3116602	rs10906982	-0.007 (-0.04, 0.027)	0.700
rs12735613	rs1390401	-0.008 (-0.047, 0.032)	0.704
rs12735613	rs8099594	0.006 (-0.027, 0.039)	0.710
rs2282978	rs1390401	0.007 (-0.031, 0.045)	0.712
rs12735613	rs3116602	0.007 (-0.031, 0.046)	0.715
rs6060373	rs1390401	0.006 (-0.029, 0.041)	0.721
rs6724465	rs4549631	0.008 (-0.037, 0.054)	0.723

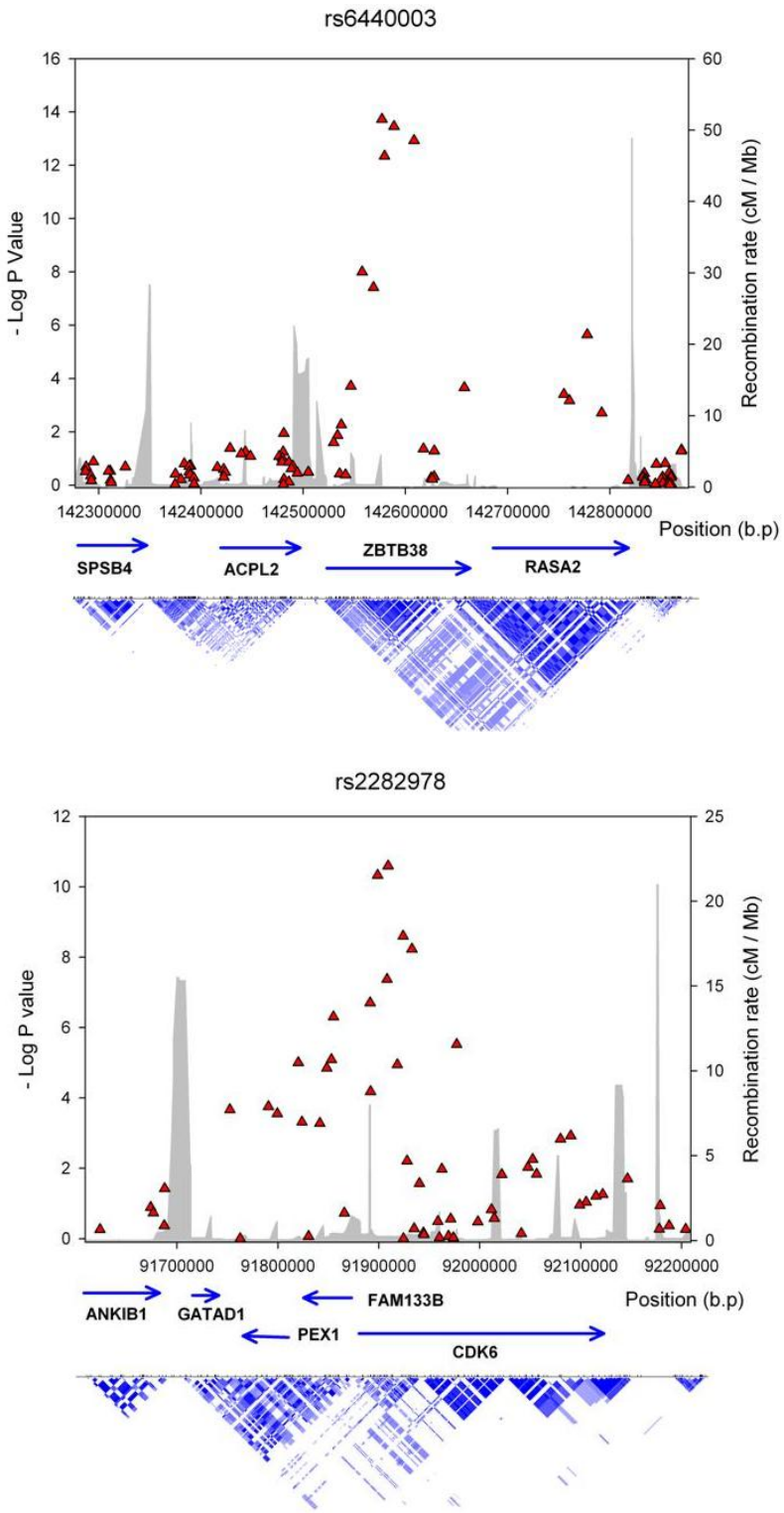
rs10935120	rs8041863	-0.005 (-0.034, 0.024)	0.733
rs8041863	rs1390401	-0.006 (-0.04, 0.029)	0.746
rs10935120	rs10512248	0.005 (-0.026, 0.036)	0.754
rs3116602	rs1390401	-0.006 (-0.048, 0.035)	0.763
rs6686842	rs6060373	0.004 (-0.024, 0.033)	0.766
rs10935120	rs2814993	0.006 (-0.036, 0.049)	0.768
rs12735613	rs6854783	0.005 (-0.027, 0.036)	0.768
rs2282978	rs1042725	0.004 (-0.025, 0.034)	0.769
rs6686842	rs6724465	-0.007 (-0.053, 0.039)	0.770
rs12735613	rs8041863	-0.004 (-0.036, 0.027)	0.782
rs1042725	rs3116602	0.004 (-0.029, 0.038)	0.790
rs10512248	rs6060373	-0.004 (-0.033, 0.025)	0.790
rs12735613	rs6724465	-0.007 (-0.059, 0.045)	0.802
rs6724465	rs6854783	0.006 (-0.04, 0.051)	0.807
rs10906982	rs8041863	0.003 (-0.024, 0.03)	0.819
rs6724465	rs16896068	0.007 (-0.056, 0.07)	0.825
rs6686842	rs10512248	-0.003 (-0.033, 0.026)	0.838
rs10935120	rs11107116	0.003 (-0.031, 0.038)	0.844
rs6854783	rs10906982	0.003 (-0.025, 0.03)	0.845
rs2814993	rs8099594	-0.004 (-0.047, 0.038)	0.845
rs4549631	rs10512248	0.003 (-0.026, 0.032)	0.847
rs10935120	rs6060373	-0.003 (-0.033, 0.027)	0.847
rs12735613	rs6060373	-0.003 (-0.036, 0.029)	0.848
rs6724465	rs11107116	0.005 (-0.048, 0.058)	0.850
rs6686842	rs3116602	0.003 (-0.03, 0.037)	0.850
rs6440003	rs11107116	-0.003 (-0.036, 0.029)	0.852
rs11107116	rs16896068	-0.004 (-0.049, 0.041)	0.853
rs6686842	rs8041863	-0.003 (-0.03, 0.025)	0.854
rs1390401	rs16896068	-0.004 (-0.052, 0.043)	0.857
rs10906982	rs8099594	0.002 (-0.026, 0.031)	0.867
rs3791675	rs4549631	-0.003 (-0.034, 0.029)	0.868
rs12735613	rs1042725	0.003 (-0.029, 0.034)	0.871
rs2814993	rs8041863	0.003 (-0.036, 0.043)	0.871
rs6724465	rs1042725	0.004 (-0.041, 0.049)	0.872
rs6440003	rs6060373	0.002 (-0.026, 0.03)	0.876
rs2282978	rs10906982	0.002 (-0.027, 0.032)	0.878
rs11107116	rs8041863	-0.002 (-0.035, 0.03)	0.887
rs3791675	rs10935120	-0.002 (-0.036, 0.031)	0.892
rs10935120	rs1042725	-0.002 (-0.031, 0.027)	0.895
rs10512248	rs3116602	0.002 (-0.033, 0.037)	0.898
rs4549631	rs3116602	-0.002 (-0.036, 0.031)	0.905
rs1042725	rs11107116	-0.002 (-0.034, 0.031)	0.909
rs2814993	rs2282978	0.003 (-0.041, 0.046)	0.909
rs10935120	rs3116602	-0.002 (-0.037, 0.033)	0.909
rs6686842	rs8099594	-0.002 (-0.031, 0.028)	0.910
rs6440003	rs10906982	-0.002 (-0.029, 0.026)	0.911

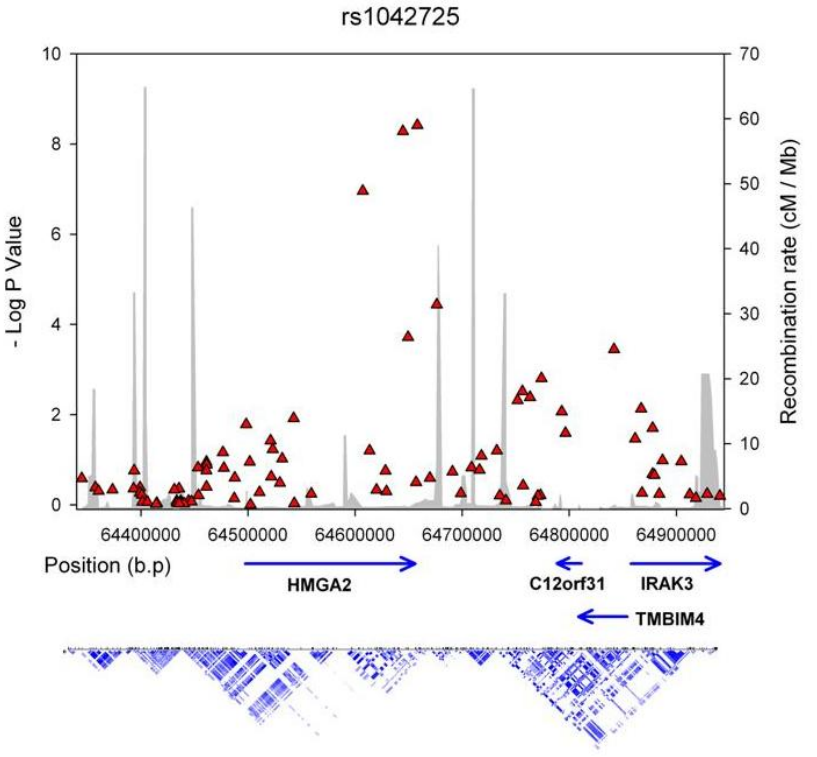
rs6724465	rs8099594	0.002 (-0.045, 0.049)	0.922
rs6440003	rs8041863	-0.001 (-0.028, 0.026)	0.934
rs6724465	rs6440003	0.002 (-0.043, 0.047)	0.943
rs12735613	rs10935120	0.001 (-0.033, 0.035)	0.946
rs6854783	rs16896068	-0.001 (-0.039, 0.037)	0.951
rs12735613	rs11107116	-0.001 (-0.039, 0.037)	0.970
rs2814993	rs11107116	0.001 (-0.047, 0.049)	0.970
rs3791675	rs8041863	0.001 (-0.031, 0.032)	0.973
rs6724465	rs10512248	0.001 (-0.048, 0.049)	0.984
rs10906982	rs1390401	0 (-0.034, 0.035)	0.987
rs6854783	rs3116602	0 (-0.033, 0.034)	0.995

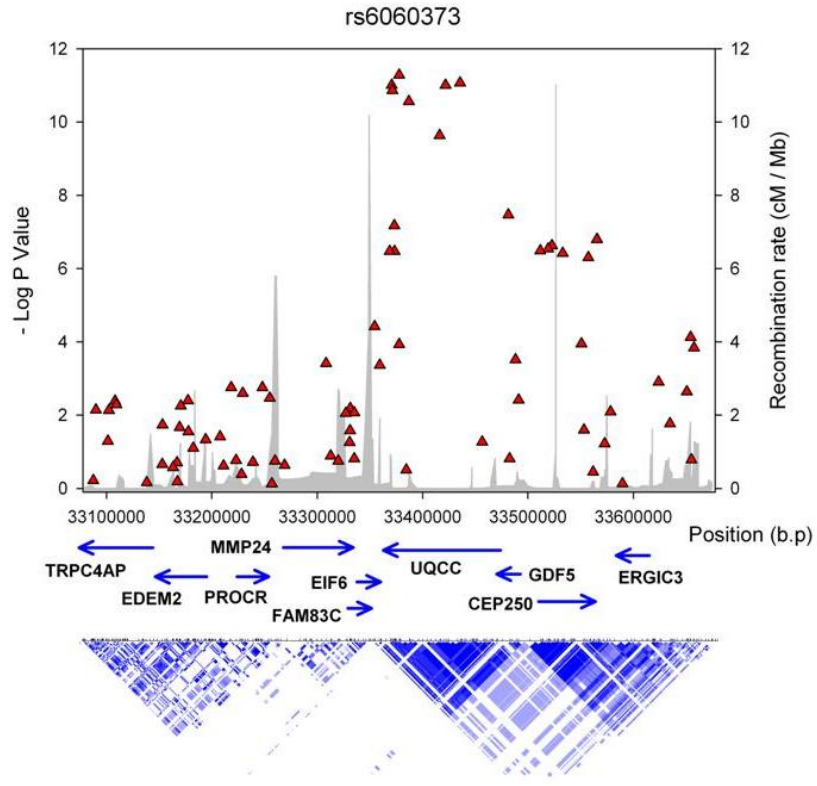
Supplementary Figure 1. A comparison of EIGENSTRAT adjusted *P*-values to unadjusted *P*-values for each individual genome-wide association study. The adjustment is based on the first ten principal components obtained from EIGENSTRAT. Using the first three principal components produced similar results. The green dots represent the 20 SNPs which reached a $P<5x10^{-7}$ in the overall analysis. DGI data were unavailable.

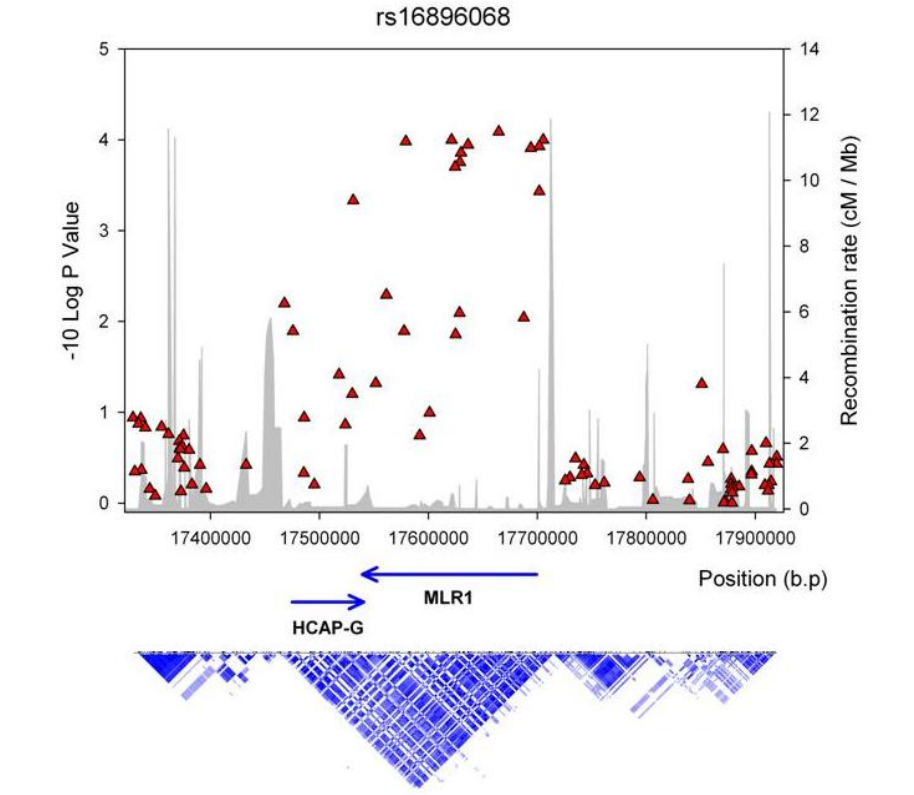


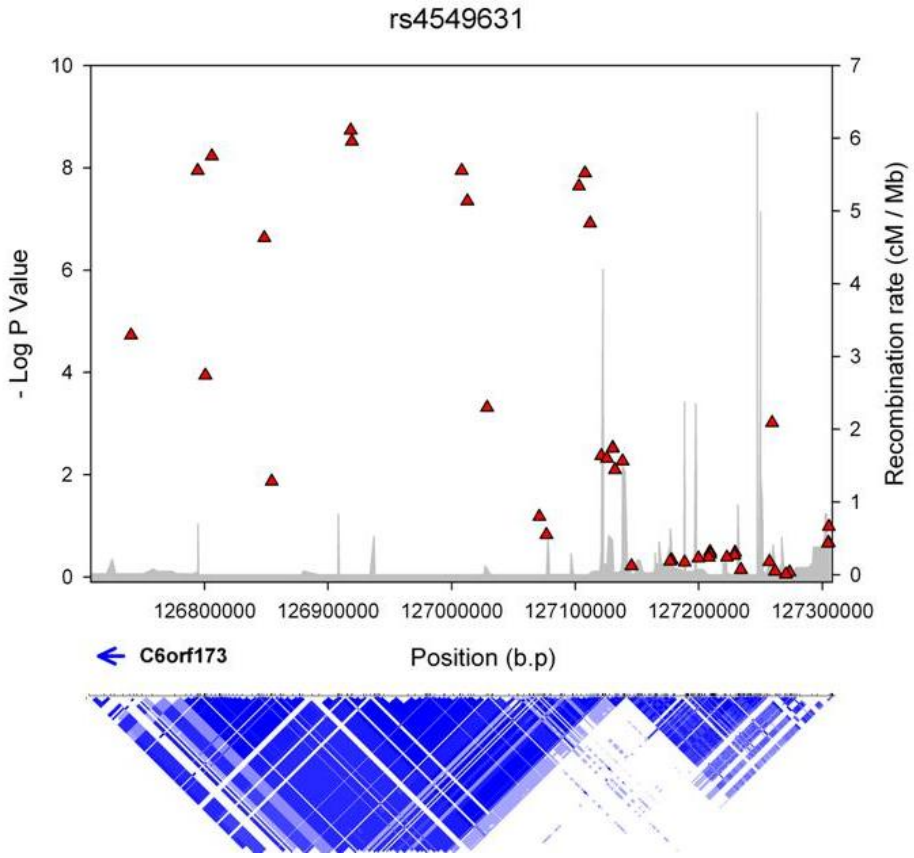
Supplementary Figure 2. LD Plots for the regions around SNPs that have a joint analysis P<5x10^{-7}. Red dots are GWAS association P values. The grey bars represent recombination rates. The LD triangle is based on the r^2 statistic, with darker shading of blue indicating higher r^2 values.

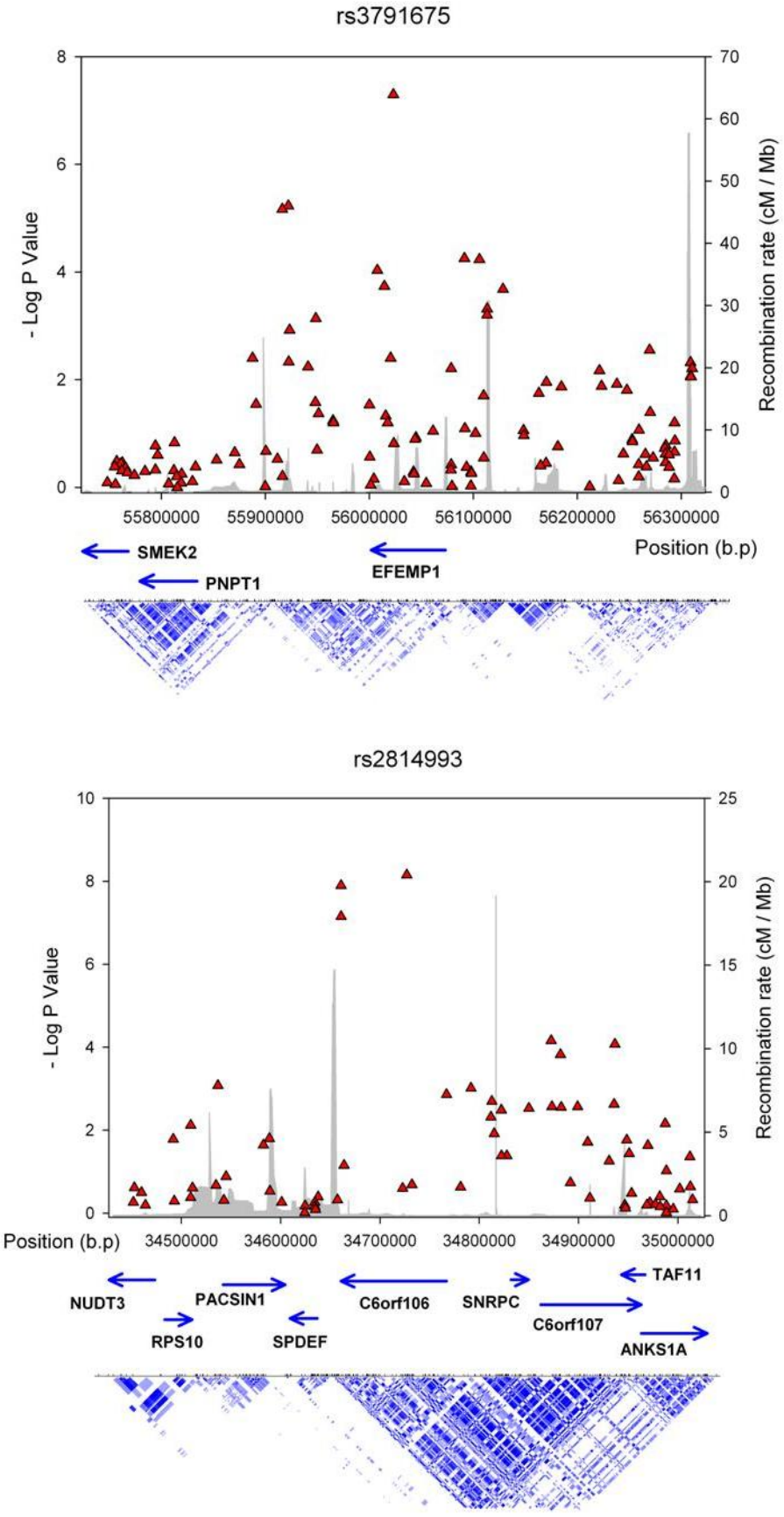


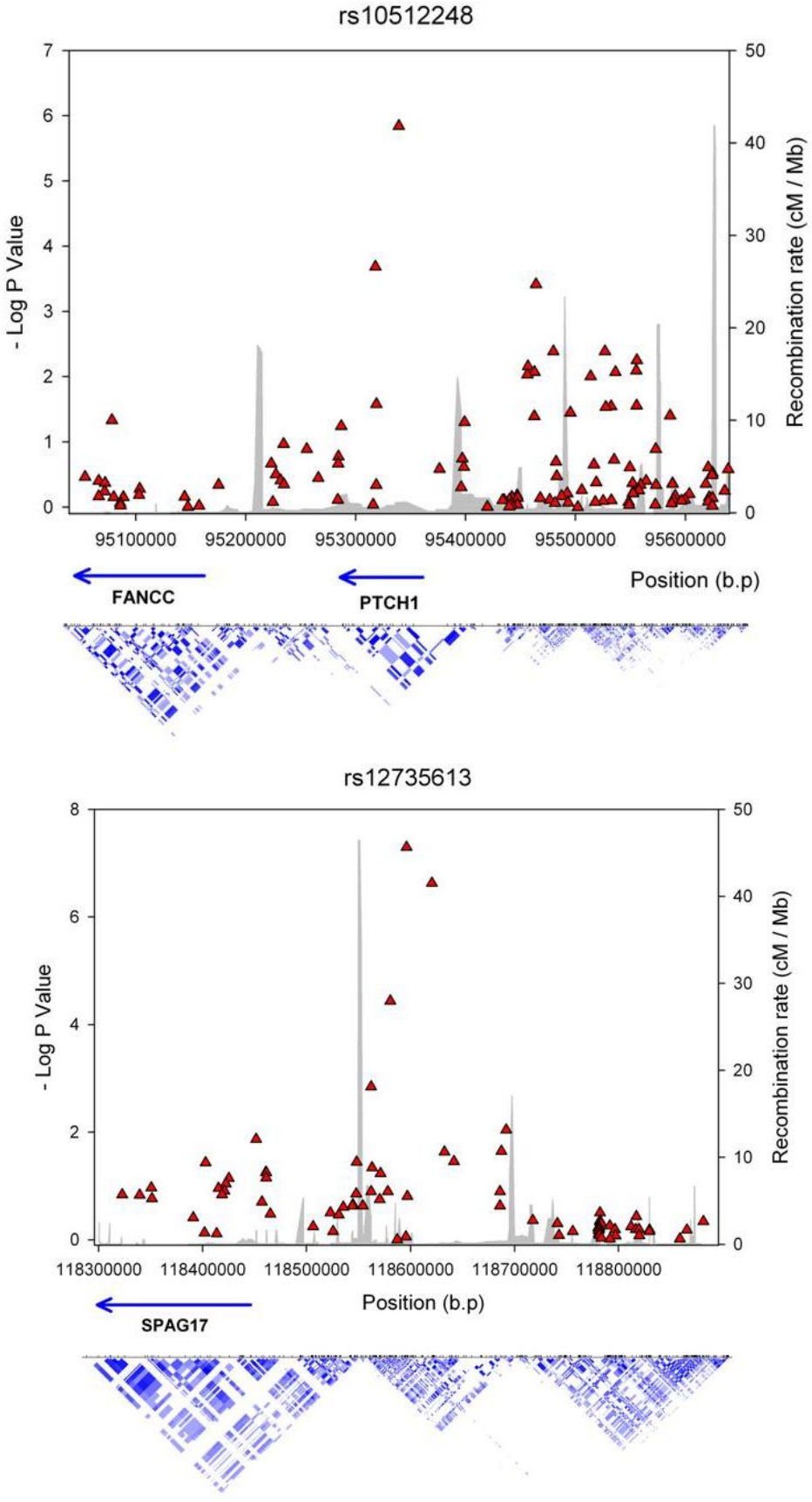

rs1042725
- Log P Value
Recombination rate (cM / Mb)
Position (b.p)
64400000
64500000
64600000
64700000
64800000
64900000
HMGA2
C12orf31
IRAK3
TMBIM4

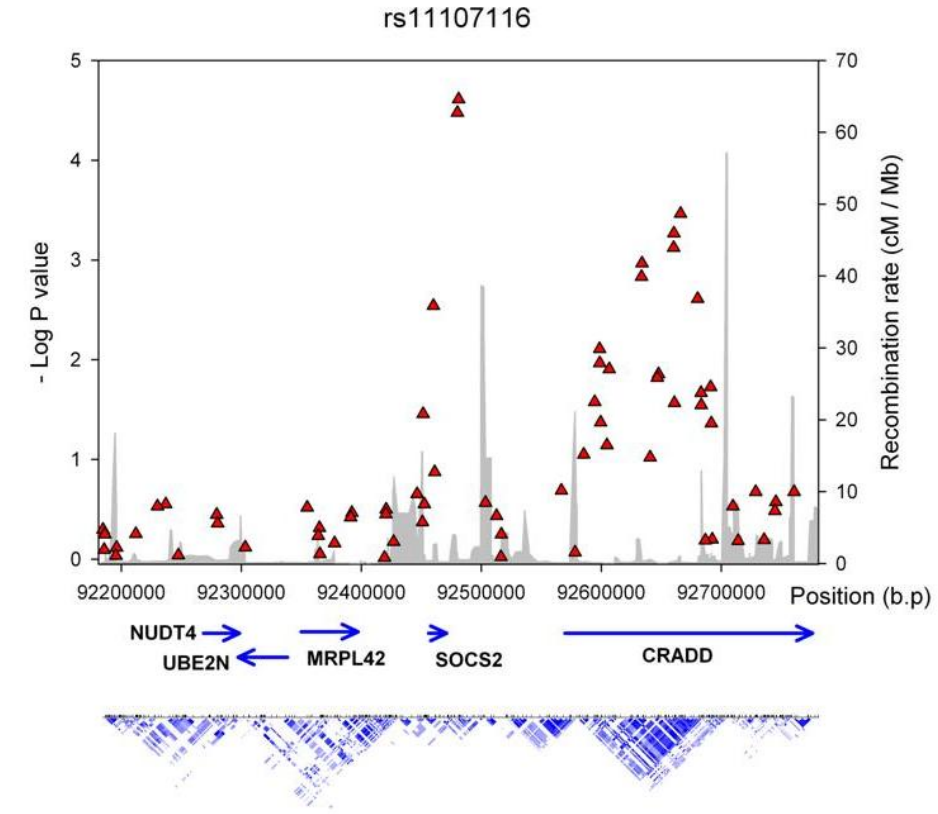

rs6060373
- Log P Value
Recombination rate (cM / Mb)
Position (b.p)
33100000
33200000
33300000
33400000
33500000
33600000
MMP24
TRPC4AP
EDEM2
PROCR
EIF6
FAM83C
UQCC
GDF5
CEP250
ERGIC3

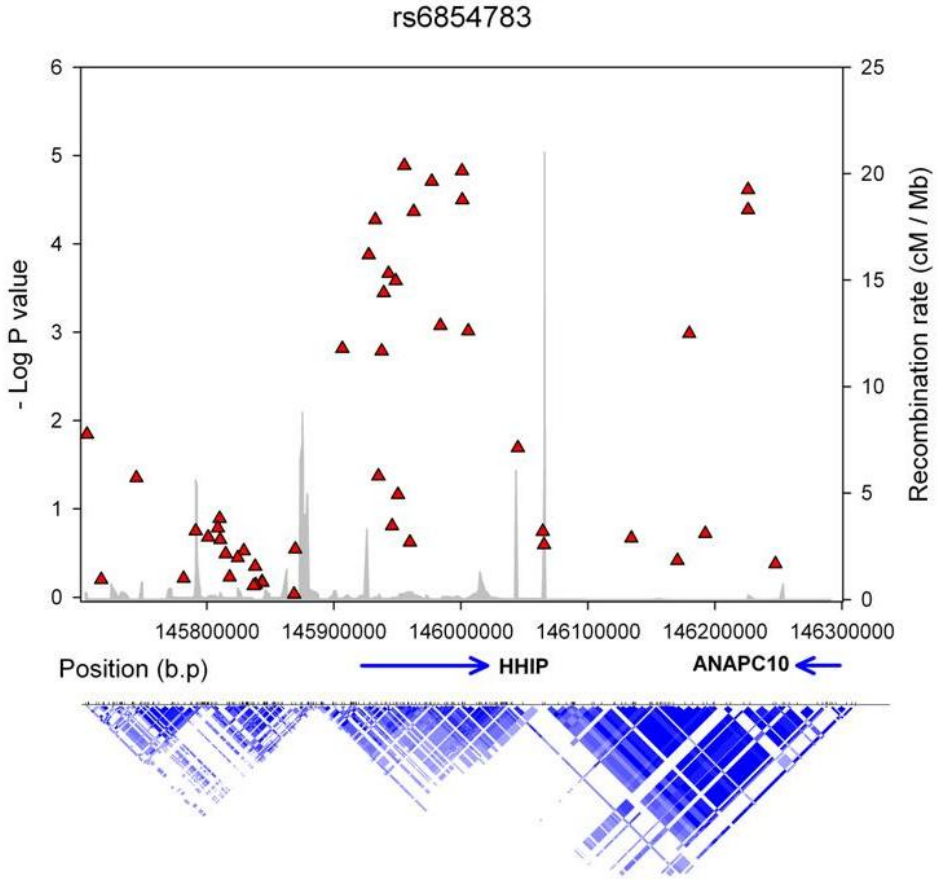

rs16896068
-10 Log P Value
Recombination rate (cM / Mb)
Position (b.p)
MLR1
HCAP-G

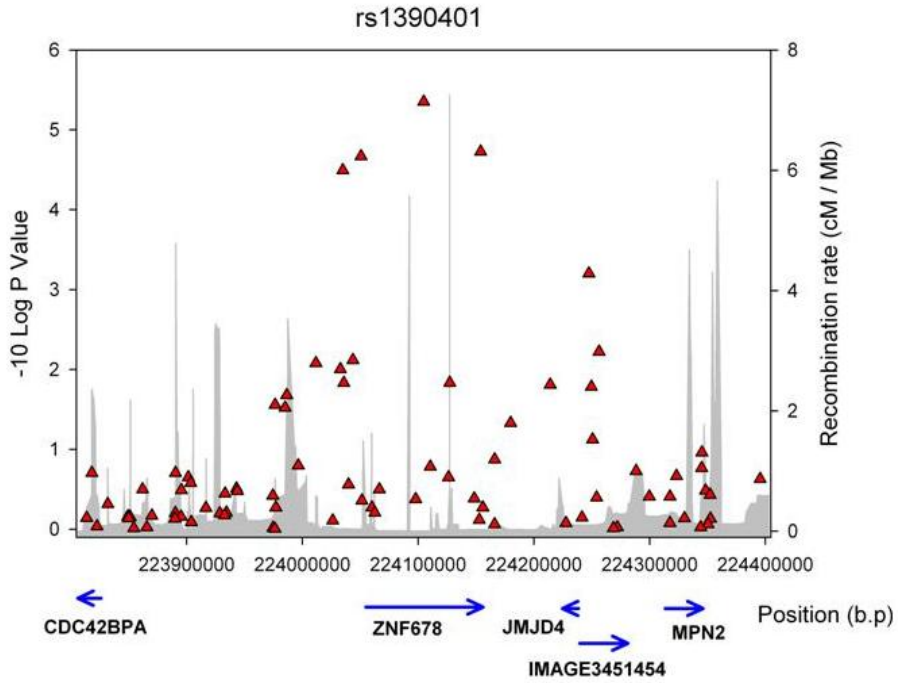

rs4549631
- Log P Value
Recombination rate (cM / Mb)
Position (b.p)
C6orf173

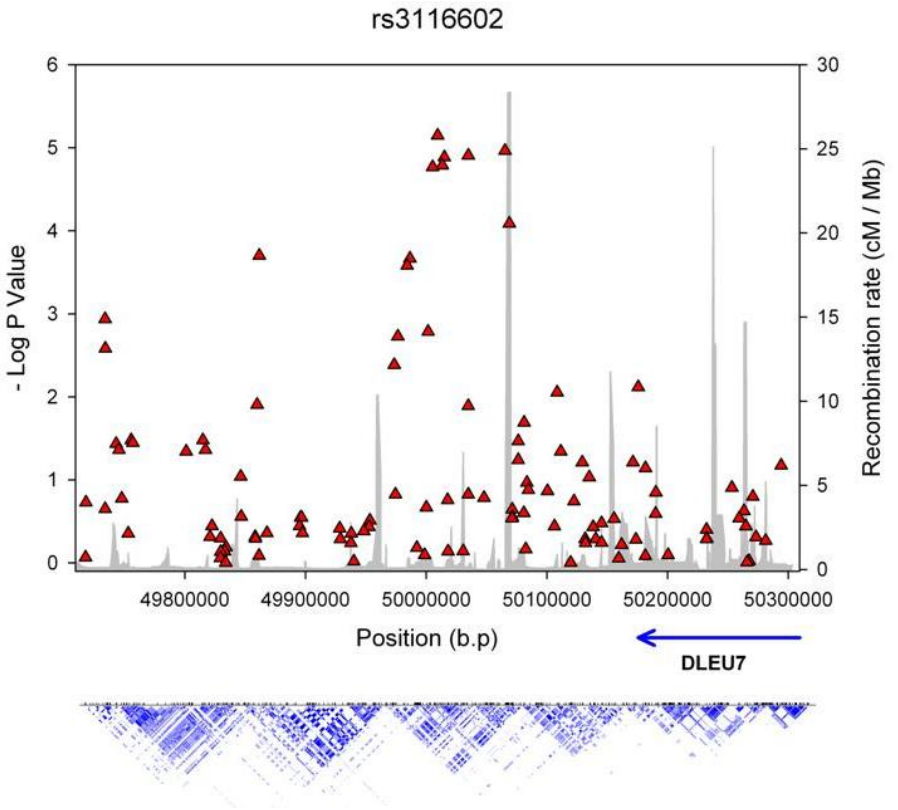

rs3791675
- Log P Value
Recombination rate (cM / Mb)
Position (b.p)
55800000
55900000
56000000
56100000
56200000
56300000
SMEK2
PNPT1
EFEMP1
rs2814993
- Log P Value
Recombination rate (cM / Mb)
Position (b.p)
34500000
34600000
34700000
34800000
34900000
35000000
NUDT3
PACSIN1
RPS10
SPDEF
C6orf106
SNRPC
C6orf107
TAF11
ANKS1A

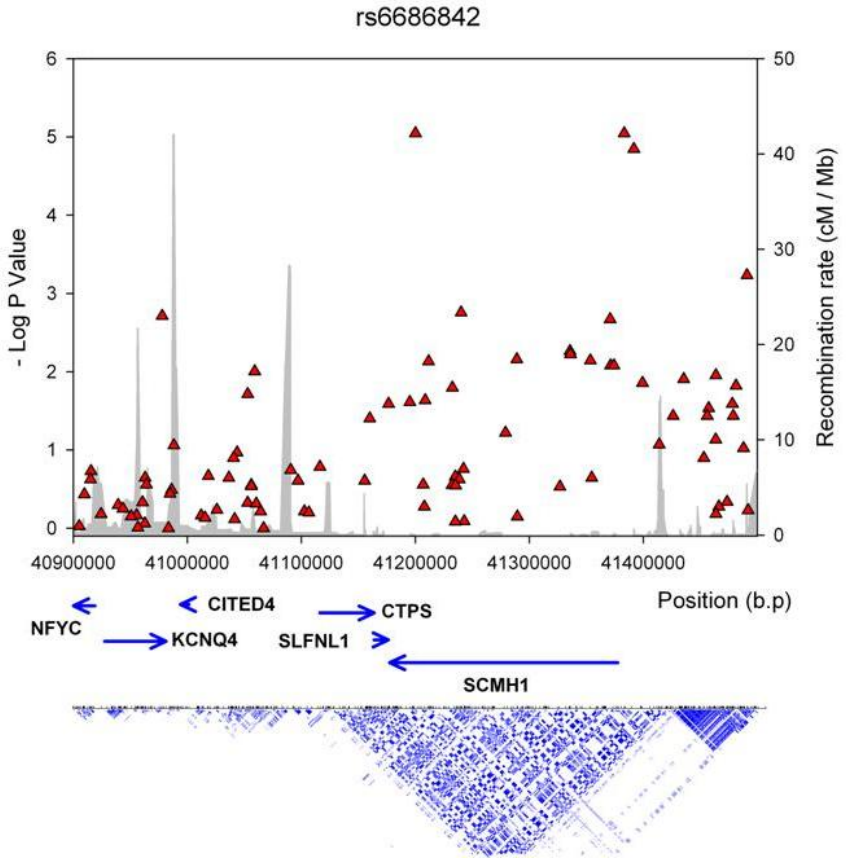

rs10512248
- Log P Value
Recombination rate (cM / Mb)
95100000
95200000
95300000
95400000
95500000
95600000
Position (b.p)
FANCC
PTCH1
rs12735613
- Log P Value
Recombination rate (cM / Mb)
118300000
118400000
118500000
118600000
118700000
118800000
Position (b.p)
SPAG17

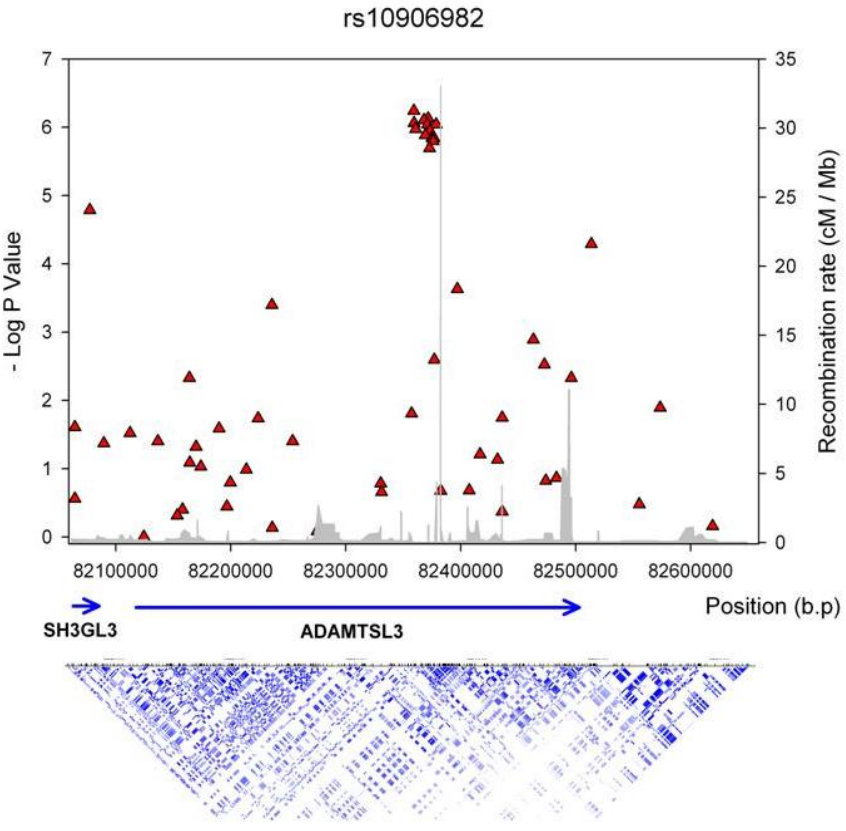

rs11107116
- Log P value
Recombination rate (cM / Mb)
Position (b.p)
92200000
92300000
92400000
92500000
92600000
92700000
NUDT4
UBE2N
MRPL42
SOCS2
CRADD

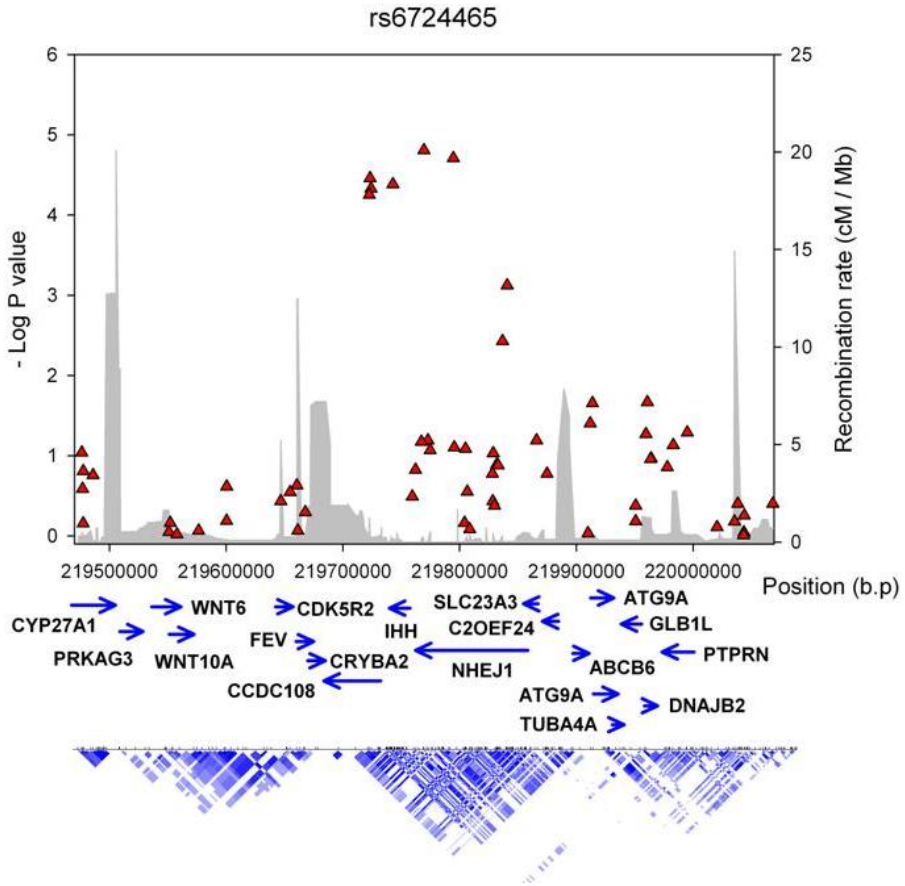

rs6854783
- Log P value
Recombination rate (cM / Mb)
Position (b.p)
145800000
145900000
146000000
146100000
146200000
146300000
HHIP
ANAPC10

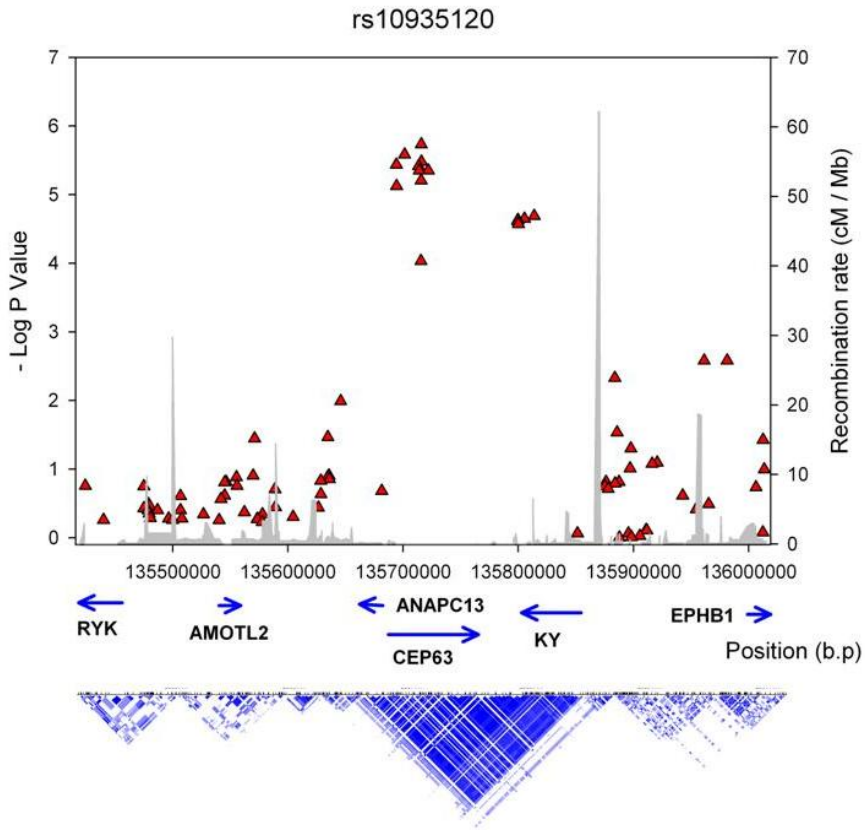

rs1390401
-10 Log P Value
Recombination rate (cM / Mb)
223900000
224000000
224100000
224200000
224300000
224400000
Position (b.p)
CDC42BPA
ZNF678
JMJD4
MPN2
IMAGE3451454

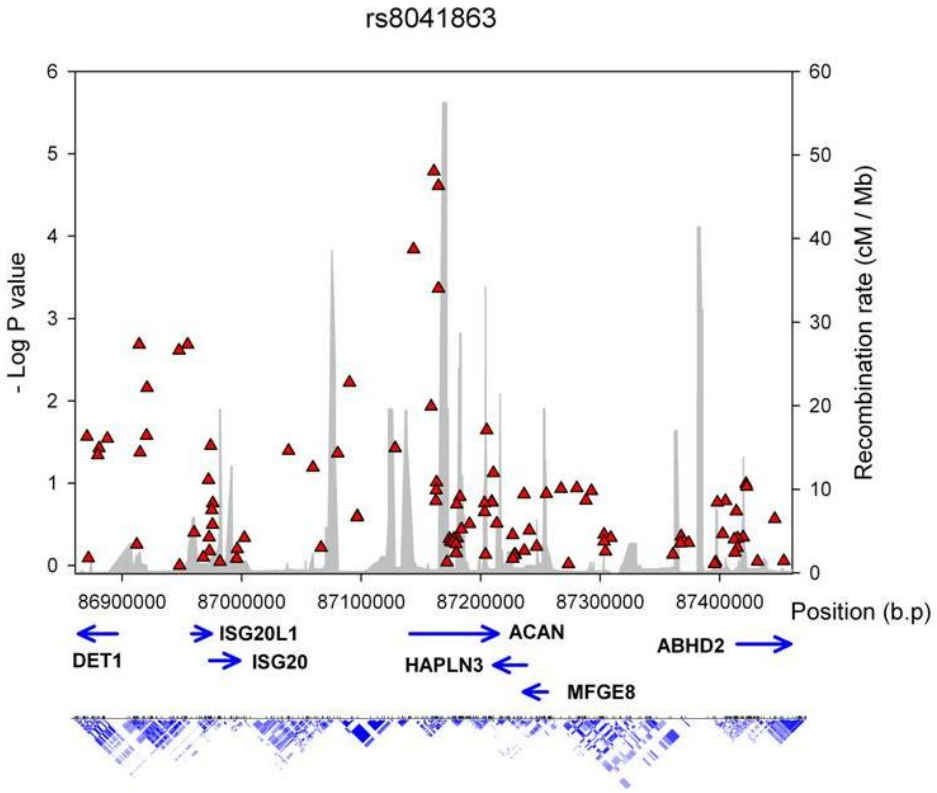

rs3116602
- Log P Value
Recombination rate (cM / Mb)
49800000
49900000
50000000
50100000
50200000
50300000
Position (b.p)
DLEU7

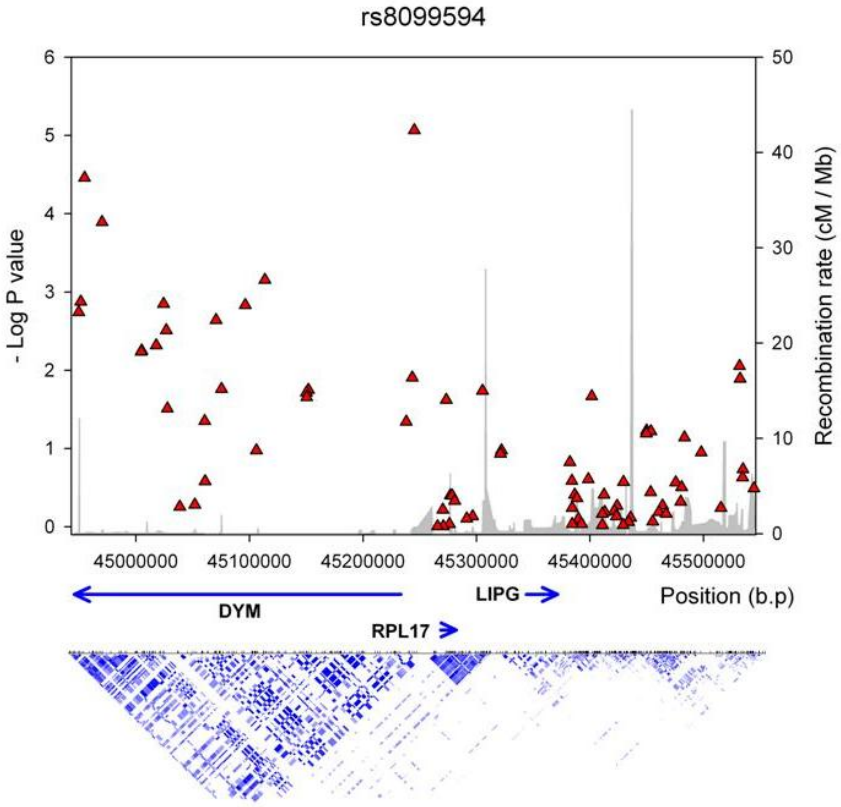

rs6686842
- Log P Value
Recombination rate (cM / Mb)
Position (b.p)
NFYC
CITED4
CTPS
KCNQ4
SLFNL1
SCMH1


rs10906982
- Log P Value
Recombination rate (cM / Mb)
Position (b.p)
SH3GL3
ADAMTSL3

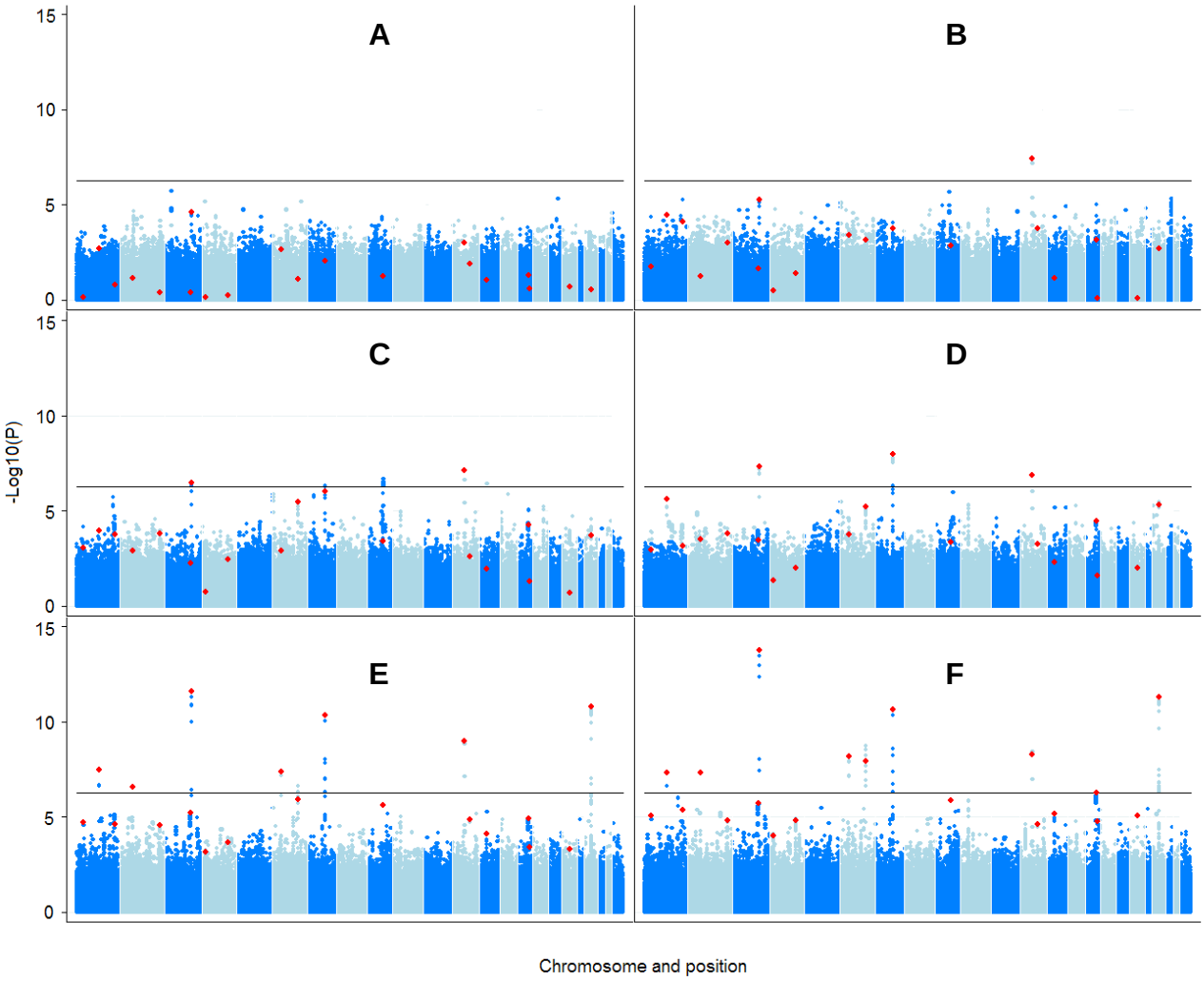

rs6724465
- Log P value
Recombination rate (cM / Mb)
Position (b.p)
219500000
219600000
219700000
219800000
219900000
220000000
CYP27A1
WNT6
CDK5R2
SLC23A3
ATG9A
PRKAG3
WNT10A
FEV
IHH
C2OEF24
GLB1L
CRYBA2
NHEJ1
ABCB6
PTPRN
CCDC108
ATG9A
DNAJB2
TUBA4A


rs10935120
- Log P Value
Recombination rate (cM / Mb)
Position (b.p)
135500000
135600000
135700000
135800000
135900000
136000000
RYK
AMOTL2
ANAPC13
CEP63
KY
EPHB1


rs8041863
- Log P value
Recombination rate (cM / Mb)
Position (b.p)
86900000
87000000
87100000
87200000
87300000
87400000
DET1
ISG20L1
ISG20
ACAN
HAPLN3
MFGE8
ABHD2


rs8099594
- Log P value
Recombination rate (cM / Mb)
Position (b.p)
45000000
45100000
45200000
45300000
45400000
45500000
DYM
LIPG
RPL17

Supplementary Figure 3. Manhattan plots for the 402,951 SNPs from the genome-wide association meta-analysis as more studies are added in. A: N = 1914 (WTCCC-T2D); **B:** N = 4892 (adding DGI); **C:** N=6788 (adding WTCCC-HT); **D:** N=8668 (adding WTCCC-CAD); **E**: N=12228 (adding EPIC-Obesity); **F:** N=13665 (adding WTCCC-UKBS). Red dots represent SNPs that achieved a $P<5x10^{-7}$ in the joint analysis with stage 2 samples. The solid black horizontal line is the $P=5x10^{-7}$ line.



Supplementary Note

Membership of WTCCC, DGI and Cambridge GEMS consortia

Wellcome Trust Case Control Consortium:

Management Committee: Paul R Burton[1], David G Clayton[2], Lon R Cardon[3], Nick Craddock[4], Panos Deloukas[5], Audrey Duncanson[6], Dominic P Kwiatkowski[3,5,] Mark I McCarthy[3,7], Willem H Ouwehand[8,9], Nilesh J Samani[10], John A Todd[2], Peter Donnelly (Chair)[11]

Analysis Committee: Jeffrey C Barrett[3], Paul R Burton[1], Dan Davison[11], Peter Donnelly[11], Doug Easton[12], David Evans[3], Hin-Tak Leung[2], Jonathan L Marchini[11], Andrew P Morris[3], Chris CA Spencer[11], Martin D Tobin[1], Lon R Cardon (Co-chair)[3], David G Clayton (Co-chair)[2]

UK Blood Services & University of Cambridge Controls: Antony P Attwood[5,8], James P Boorman[8,9], Barbara Cant[8], Ursula Everson[13], Judith M Hussey[14], Jennifer D Jolley[8], Alexandra S Knight[8], Kerstin Koch[8], Elizabeth Meech[15], Sarah Nutland[2], Christopher V Prowse[16], Helen E Stevens[2], Niall C Taylor[8], Graham R Walters[17], Neil M Walker[2], Nicholas A Watkins[8,9], Thilo Winzer[8], John A Todd[2], Willem H Ouwehand[8,9]

1958 Birth Cohort Controls: Richard W Jones[18], Wendy L McArdle[18], Susan M Ring[18], David P Strachan[19], Marcus Pembrey[18,20]

Bipolar Disorder (Aberdeen): Gerome Breen[21], David St Clair[21]; (**Birmingham):** Sian Caesar[22], Katherine Gordon-Smith[22,23], Lisa Jones[22]; **(Cardiff):** Christine Fraser[23], Elaine K Green[23], Detelina Grozeva[23], Marian L Hamshere[23], Peter A Holmans[23], Ian R Jones[23], George Kirov[23], Valentina Moskvina[23], Ivan Nikolov[23], Michael C O'Donovan[23], Michael J Owen[23], Nick Craddock[23]; **(London):** David A Collier[24], Amanda Elkin[24], Anne Farmer[24], Richard Williamson[24], Peter McGuffin[24]; **(Newcastle):** Allan H Young[25], I Nicol Ferrier[25]

Coronary Artery Disease (Leeds): Stephen G Ball[26], Anthony J Balmforth[26], Jennifer H Barrett[26], D Timothy Bishop[26], Mark M Iles[26], Azhar Maqbool[26], Nadira Yuldasheva[26], Alistair S Hall[26]; **(Leicester):** Peter S Braund[10], Paul R Burton[1], Richard J Dixon[10],

Massimo Mangino[10], Suzanne Stevens[10], Martin D Tobin[1], John R Thompson[1], Nilesh J Samani[10]

Crohn's Disease (Cambridge): Francesca Bredin[27], Mark Tremelling[27], Miles Parkes[27]; **(Edinburgh):** Hazel Drummond[28], Charles W Lees[28], Elaine R Nimmo[28], Jack Satsangi[28]; **(London):** Sheila A Fisher[29], Alastair Forbes[30], Cathryn M Lewis[29], Clive M Onnie[29], Natalie J Prescott[29], Jeremy Sanderson[31], Christopher G Mathew[29]; **(Newcastle):** Jamie Barbour[32], M Khalid Mohiuddin[32], Catherine E Todhunter[32], John C Mansfield[32]; **(Oxford):** Tariq Ahmad[33], Fraser R Cummings[33], Derek P Jewell[33]

Hypertension (Aberdeen): John Webster[34]; **(Cambridge):** Morris J Brown[35], David G Clayton[2]; **(Evry, France): G** Mark Lathrop[36]; **(Glasgow):** John Connell[37], Anna Dominiczak[37]; **(Leicester):** Nilesh J Samani[10]; **(London):** Carolina A Braga Marcano[38], Beverley Burke[38], Richard Dobson[38], Johannie Gungadoo[38], Kate L Lee[38], Patricia B

Munroe[38], Stephen J Newhouse[38], Abiodun Onipinla[38], Chris Wallace[38], Mingzhan Xue[38], Mark Caulfield[38]; **(Oxford):** Martin Farrall[39]

Rheumatoid Arthritis: Anne Barton[40], The Biologics in RA Genetics and Genomics Study Syndicate (BRAGGS) Steering Committee*, Ian N Bruce[40], Hannah Donovan[40], Steve Eyre[40], Paul D Gilbert[40], Samantha L Hider[40], Anne M Hinks[40], Sally L John[40], Catherine Potter[40], Alan J Silman[40], Deborah PM Symmons[40], Wendy Thomson[40], Jane Worthington[40]

Type [1] Diabetes: David G Clayton[2], David B Dunger[2,41], Sarah Nutland[2], Helen E Stevens[2], Neil M Walker[2], Barry Widmer[2,41], John A Todd[2]

Type [2] Diabetes (Exeter): Timothy M Frayling[42,43], Rachel M Freathy[42,43], Hana Lango Allen[42,43], John R B Perry[42,43], Beverley M Shields[43], Michael N Weedon[42,43], Andrew T Hattersley[42,43]; **(London):** Graham A Hitman[44]; **(Newcastle):** Mark Walker[45]; **(Oxford):** Kate S Elliott[3,7], Christopher J Groves[7], Cecilia M Lindgren[3,7], Nigel W Rayner[3,7], Nicholas J Timpson[3,46], Eleftheria Zeggini[3,7], Mark I McCarthy[3,7]

Tuberculosis (Gambia): Melanie Newport[47], Giorgio Sirugo[47]; **(Oxford):** Emily Lyons[3], Fredrik Vannberg[3], Adrian VS Hill[3]

Ankylosing Spondylitis: Linda A Bradbury[48], Claire Farrar[49], Jennifer J Pointon[48], Paul Wordsworth[49], Matthew A Brown[48,49]

AutoImmune Thyroid Disease: Jayne A Franklyn[50], Joanne M Heward[50], Matthew J Simmonds[50], Stephen CL Gough[50]

Breast Cancer: Sheila Seal[51], Breast Cancer Susceptibility Collaboration (UK)*, Michael R Stratton[51,52], Nazneen Rahman[51]

Multiple Sclerosis: Maria Ban[53], An Goris[53], Stephen J Sawcer[53], Alastair Compston[53]

Gambian Controls (Gambia): David Conway[47], Muminatou Jallow[47], Melanie Newport[47], Giorgio Sirugo[47]; **(Oxford):** Kirk A Rockett[3], Dominic P Kwiatkowski[3,5]

DNA, Genotyping, Data QC and Informatics (Wellcome Trust Sanger Institute, Hinxton): Suzannah J Bumpstead[5], Amy Chaney[5], Kate Downes[2,5], Mohammed JR Ghori[5], Rhian Gwilliam[5], Sarah E Hunt[5], Michael Inouye[5], Andrew Keniry[5], Emma King[5], Ralph McGinnis[5], Simon Potter[5], Rathi Ravindrarajah[5], Pamela Whittaker[5], Claire Widden[5], David Withers[5], Panos Deloukas[5]; **(Cambridge):** Hin-Tak Leung[2], Sarah Nutland[2], Helen E Stevens[2], Neil M Walker[2], John A Todd[2]

Statistics (Cambridge): Doug Easton[12], David G Clayton[2]; **(Leicester):** Paul R Burton[1], Martin D Tobin[1]; **(Oxford):** Jeffrey C Barrett[3], David Evans[3], Andrew P Morris[3], Lon R Cardon[3]; **(Oxford):** Niall J Cardin[11], Dan Davison[11], Teresa Ferreira[11], Joanne Pereira-Gale[11], Ingeleif B Hallgrimsdóttir[11], Bryan N Howie[11], Jonathan L Marchini[11], Chris CA Spencer[11], Zhan Su[11], Yik Ying Teo[3,11], Damjan Vukcevic[11], Peter Donnelly[11]

PIs: David Bentley[5,54], Matthew A Brown[48,49], Lon R Cardon[3], Mark Caulfield[38], David G Clayton[2], Alistair Compston[53], Nick Craddock[23], Panos Deloukas[5], Peter Donnelly[11], Martin Farrall[39], Stephen CL Gough[50], Alistair S Hall[26], Andrew T Hattersley[42,43], Adrian VS Hill[3], Dominic P Kwiatkowski[3,5], Christopher G Mathew[29], Mark I McCarthy[3,7], Willem H

Ouwehand[8,9,] Miles Parkes[27], Marcus Pembrey[18,20], Nazneen Rahman[51], Nilesh J Samani[10], Michael R Stratton[51],[52,] John A Todd[2], Jane Worthington[40]

1 Genetic Epidemiology Group, Department of Health Sciences, University of Leicester, Adrian Building, University Road, Leicester, LE1 7RH, UK; 2 Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge, Wellcome Trust/MRC Building, Cambridge, CB2 0XY, UK; 3 Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK; 4 Department of Psychological Medicine, Henry Wellcome Building, School of Medicine, Cardiff University, Heath Park, Cardiff CF14 4XN, UK; 5 The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK; 6 The Wellcome Trust, Gibbs Building, 215 Euston Road, London NW1 2BE, UK; 7 Oxford Centre for Diabetes, Endocrinology and Medicine, University of Oxford, Churchill Hospital, Oxford, OX3 7LJ, UK; 8 Department of Haematology, University of Cambridge, Long Road, Cambridge, CB2 2PT, UK; 9 National Health Service Blood and Transplant, Cambridge Centre, Long Road, Cambridge, CB2 2PT, UK; 10 Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Groby Road, Leicester, LE3 9QP, UK; 11 Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK; 12 Cancer Research UK Genetic Epidemiology Unit, Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 8RN, UK; 13 National Health Service Blood and Transplant, Sheffield Centre, Longley Lane, Sheffield S5 7JN, UK; 14 National Health Service Blood and Transplant, Brentwood Centre, Crescent Drive, Brentwood, CM15 8DP, UK; 15 The Welsh Blood Service, Ely Valley Road, Talbot Green, Pontyclun, CF72 9WB, UK; 16 The Scottish National Blood Transfusion Service, Ellen's Glen Road, Edinburgh, EH17 7QT, UK; 17 National Health Service Blood and Transplant, Southampton Centre, Coxford Road, Southampton, SO16 5AF, UK; 18 Avon Longitudinal Study of Parents and Children, University of Bristol, 24 Tyndall Avenue, Bristol, BS8 1TQ, UK; 19 Division of Community Health Services, St George's University of London, Cranmer Terrace, London SW17 0RE, UK; 20 Institute of Child Health, University College London, 30 Guilford St, London WC1N 1EH, UK; 21 University of Aberdeen, Institute of Medical Sciences, Foresterhill, Aberdeen, AB25 2ZD, UK; 22 Department of Psychiatry, Division of Neuroscience, Birmingham University, Birmingham, B15 2QZ, UK; 23 Department of Psychological Medicine, Henry Wellcome Building, School of Medicine, Cardiff University, Heath Park, Cardiff CF14 4XN, UK; 24 SGDP, The Institute of Psychiatry, King's College London, De Crespigny Park Denmark Hill London SE5 8AF, UK; 25 School of Neurology, Neurobiology and Psychiatry, Royal Victoria Infirmary, Queen Victoria Road, Newcastle upon Tyne, NE1 4LP, UK; 26 LIGHT and LIMM Research Institutes, Faculty of Medicine and Health, University of Leeds, Leeds, LS1 3EX, UK; 27 IBD Research Group, Addenbrooke's Hospital, University of Cambridge, Cambridge, CB2 2QQ, UK; 28 Gastrointestinal Unit, School of Molecular and Clinical Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU UK; 29 Department of Medical & Molecular Genetics, King's College London School of Medicine, 8th Floor Guy's Tower, Guy's Hospital, London, SE1 9RT, UK; 30 Institute for Digestive Diseases, University College London Hospitals Trust, London, NW1 2BU, UK; 31 Department of Gastroenterology, Guy's and St Thomas' NHS Foundation Trust, London, SE1 7EH, UK; 32 Department of Gastroenterology & Hepatology, University of Newcastle upon Tyne, Royal Victoria Infirmary, Newcastle upon Tyne, NE1 4LP, UK; 33 Gastroenterology Unit, Radcliffe Infirmary, University of Oxford, Oxford, OX2 6HE, UK; 34 Medicine and Therapeutics, Aberdeen Royal Infirmary, Foresterhill, Aberdeen, Grampian AB9 2ZB, UK; 35 Clinical Pharmacology Unit and the Diabetes and Inflammation Laboratory, University of Cambridge, Addenbrookes Hospital, Hills Road, Cambridge CB2 2QQ, UK; 36 Centre National de Genotypage, 2, Rue Gaston Cremieux, Evry, Paris 91057.; 37 BHF Glasgow Cardiovascular Research Centre, University of Glasgow, 126 University Place, Glasgow, G12 8TA, UK; 38 Clinical Pharmacology and Barts and The London Genome Centre, William Harvey Research Institute, Barts and The London, Queen Mary's School of Medicine, Charterhouse Square, London EC1M 6BQ, UK; 39 Cardiovascular Medicine, University of Oxford, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK; 40arc Epidemiology Research Unit, University of Manchester, Stopford Building, Oxford Rd, Manchester, M13 9PT, UK; 41 Department of Paediatrics, University of Cambridge, Addenbrooke's Hospital, Cambridge, CB2 2QQ, UK; 42

Genetics of Complex Traits, Institute of Biomedical and Clinical Science, Peninsula Medical School, Magdalen Road, Exeter EX1 2LU UK; 43 Diabetes Genetics, Institute of Biomedical and Clinical Science, Peninsula Medical School, Barrack Road, Exeter EX2 5DU UK; 44 Centre for Diabetes and Metabolic Medicine, Barts and The London, Royal London Hospital, Whitechapel, London, E1 1BB UK; 45 Diabetes Research Group, School of Clinical Medical Sciences, Newcastle University, Framlington Place, Newcastle upon Tyne NE2 4HH, UK; 46 The MRC Centre for Causal Analyses in Translational Epidemiology, Bristol University, Canynge Hall, Whiteladies Rd, Bristol BS2 8PR, UK; 47 MRC Laboratories, Fajara, The Gambia; 48 Diamantina Institute for Cancer, Immunology and Metabolic Medicine, Princess Alexandra Hospital, University of Queensland, Woolloongabba, Qld 4102, Australia; 49 Botnar Research Centre, University of Oxford, Headington, Oxford OX3 7BN, UK; 50 Department of Medicine, Division of Medical Sciences, Institute of Biomedical Research, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK; 51 Section of Cancer Genetics, Institute of Cancer Research, 15 Cotswold Road, Sutton, SM2 5NG, UK; 52 Cancer Genome Project, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK; 53 Department of Clinical Neurosciences, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ, UK; 54 PRESENT ADDRESS: Illumina Cambridge, Chesterford Research Park, Little Chesterford, Nr Saffron Walden, Essex, CB10 1XL, UK.

Membership of DGI

Diabetes Genetics Initiative of Broad Institute of Harvard and Massachusetts Institute of Technology, Lund University and Novartis Institutes for BioMedical Research

Authors are listed by site, and within each site alphabetically, with the exception that within each site team leaders are listed at the front, and PI's are at the end.

** NPB, PIWdB, VL, RS and BFV contributed equally and are listed alphabetically.*

Site 1 (Massachusetts General Hospital, Children's Hospital, Harvard Medical School, Broad Institute of Harvard and MIT)

*Noël P. Burtt[1,3], *Paul I.W. de Bakker[1,2,4,5,10], *Richa Saxena[1,3-5,10], *Benjamin F. Voight[1,4], Kristin Ardlie[1], Rachel Barry[2], Brendan Blumenstiel[2], Wendy Brodeur[2], Jody Camarata[2], Nancy Chia[2], Matthew DeFelice[2], Mary Fava[2], Jose C. Florez[1,3-5,9,11], Stacey B. Gabriel[2], Diane Gage[2], Casey Gates[2], John Gibbons[2], Lauren Gianniny[1], Candace Guiducci[1], Rachel Hackett[1], Bob Handsaker[2], Claire Healy[2], Sekar Kathiresan[1,3,7,8,11], Guillaume Lettre[1,3,13,14], Helen N. Lyon[1,3,12,14], Vamsi K. Mootha[3,4], Christopher Newton-Cheh[1,8], Marcia Nizzari[2], Kieu Nguyen[2], Melissa Parkin[2], Shaun Purcell[1,4], Carrie Sougnez[2], Elizabeth K. Speliotes[1,3,11], Aarti Surti[1], Ryan Tewhey[1], Joel N. Hirschhorn[1,3,10,12-14], Mark J. Daly[1,3,4,6,11], David Altshuler[1,3-6,9-11]

Site 2 (Lund University, University of Helsinki)

*Valeriya Lyssenko[15], Peter Almgren[15], Anna Berglund[15], Johan Holmkvist[15], Bo Isomaa[16,19], Esa Laurila[15], Olle Melander[15], Marju Orho-Melander[15], Peter Nilsson[15], Hemang Parikh[15], Marketa Sjögren[15], Malin Svensson[15], Margareta Svensson[15], Kristina Bengtsson[17], Ulf Lindblad[17], Marja-Riitta Taskinen[16], Tiinamaija Tuomi[18,19], Leif Groop[15,18]

Site 3 (Novartis Institutes for BioMedical Research)

Hong Chen[20], Gung-Wei Chirn[20], Qicheng Ma[20], Darrell Ricke[20], Delwood Richardson[20], Jeffrey J. Roix[20], Joanne Meyer[20], Thomas E. Hughes[20]

[1] Program in Medical and Population Genetics, [2]Genetic Analysis Platform, and [3]Metabolic Disease Initiative, Broad Institute of Harvard and Massachusetts Institute of Technology, Seven Cambridge Center, Cambridge, Massachusetts 02142, USA.

[4] Center for Human Genetic Research, [5]Department of Molecular Biology and [6]Medicine, [7]Cardiovascular Disease Prevention Center, [8]Cardiology Division, and [9]Diabetes Unit, Massachusetts General Hospital, 185 Cambridge Street, CPZN-6818, Boston, Massachusetts 02114-2790, USA.

[10] Department of Genetics, [11]Medicine and [12]Pediatrics, Harvard Medical School, Boston, Massachusetts, USA.

[13] Division of Endocrinology and [14]Genetics, Children's Hospital, Boston, Massachusetts, USA

[15] Department of Clinical Sciences, Diabetes and Endocrinology Research Unit, University Hospital Malmö, Lund University, Malmö, Sweden.

[16] Malmska Municipal Health Center and Hospital, Jakobstad, Finland.

[17] Skaraborg Institute, Skövde, Sweden.

[18] Department of Medicine, Helsinki University Hospital, University of Helsinki, Helsinki, Finland.

[19] Folkhälsan Research Center, Helsinki, Finland.

[20] Diabetes and Metabolism Disease Area, Novartis Institutes for BioMedical Research, 100 Technology Square, Cambridge, Massachusetts, USA.

CHAPTER 3: HUNDREDS OF VARIANTS INFLUENCE HUMAN HEIGHT AND CLUSTER WITHIN GENOMIC LOCI AND BIOLOGICAL PATHWAYS

Hana Lango Allen *, Karol Estrada *, Guillaume Lettre[,]*, Sonja Berndt[,]*, Michael N Weedon[,]*, Fernando Rivadeneira[,]*, *et al*., for the GIANT Consortium

* Denotes equal contribution

A slightly modified version of a manuscript published in *Nature* (2010) 467:823

Abstract

Most common human traits and diseases have a polygenic pattern of inheritance: DNA sequence variants at many genetic loci influence phenotype. Genome-wide association (GWA) studies have identified >600 variants associated with human traits [1], but these typically explain small fractions of phenotypic variation, raising questions about the utility of further studies. Here, using 183,727 individuals, we show that hundreds of genetic variants, in at least 180 loci, influence adult height, a highly heritable and classic polygenic trait [2,3]. The large number of loci reveals patterns with important implications for genetic studies of common human diseases and traits. First, the 180 loci are not random, but instead are enriched for genes that are connected in biological pathways (P=0.002), and that underlie skeletal growth defects (P<0.001). Second, the likely causal gene is often located near the most strongly associated variant: in 13 of 21 loci containing a known skeletal growth gene, that gene was closest to the associated variant. Third, at least 19 loci have multiple independently associated variants, suggesting that allelic heterogeneity is a frequent feature of polygenic traits, that comprehensive explorations of already-discovered loci should discover additional variants, and that an appreciable fraction of associated loci may have been identified. Fourth, associated variants are enriched for likely functional effects on genes, being over-represented amongst variants that alter amino acid structure of proteins. Our data explain ~10% of the phenotypic variation in height, and we estimate that unidentified common variants of similar effect sizes would increase this figure to ~16% of phenotypic variation (~20% of heritable variation). Although approaches that more comprehensively survey low frequency variants are needed to fully dissect the genetic architecture of polygenic human traits, our findings indicate that GWA studies can identify large numbers of loci that implicate biologically relevant genes and pathways.

In Stage 1 of our study, we performed a meta-analysis of GWA data from 46 studies, comprising 133,653 individuals of recent European ancestry, to identify common genetic variation associated with adult height. To enable meta-analysis of studies across different genotyping platforms, we performed imputation of 2,834,208 single nucleotide polymorphisms (SNPs) present in the HapMap Phase 2 European-American reference panel [4] (**Supplementary Methods)**. After applying quality control filters, each individual study tested the association of adult height with each SNP using an additive model (**Supplementary Methods**). The individual study statistics were corrected using the genomic control (GC) method [5,6] and then combined in a meta-analysis. We then applied a second GC correction on the meta-analysis statistics, although this approach may be overly conservative when there are many real signals of association (**Supplementary Methods**). We detected 207 loci (defined as 1Mb on either side of the most strongly associated SNP) as potentially associated with adult height ($P<5x10^{-6}$, a threshold at which we would only expect 5 by chance).

To identify loci robustly associated with adult height, we took forward at least one SNP (**Supplementary Methods**) from each of the 207 loci reaching $P<5x10^{-6}$ into an additional 50,074 samples (Stage 2). In the joint analysis of our Stage 1 and Stage 2 studies, SNPs representing 180 loci reached genome-wide significance ($P<5x10^{-8}$; **Figure 1** and **Supplementary Table 1**). Additional tests, including genotyping of a randomly-selected subset of 33 SNPs in an independent sample of individuals from the tails of the height distribution [7], provided further validation of our results, with all but two SNPs showing consistent direction of effect (sign test $P<7x10^{-8}$) (**Supplementary Methods and Supplementary Table 2**). Similarly, 206 of the 207 lead SNPs had consistent directions of effect in Stages 1 and 2 (sign test $P<10^{-62}$), indicating that there are almost certainly additional associated variants amongst the loci that narrowly missed genome-wide significance.

Genome wide association (GWA) studies can be susceptible to false positive associations from population stratification [7]. We therefore performed a

family-based analysis, which is immune to population stratification in 7,336 individuals from two cohorts with pedigree information. Alleles representing 150 of the 180 genome-wide significant loci were associated in the expected direction (sign test $P<6x10^{-20}$; **Supplementary Table 3**). The estimated effects on height were essentially identical in the overall meta-analysis and the family-based sample (**Supplementary Table 3**). Together with several other lines of evidence (**Supplementary Methods**), this indicates that stratification is not substantially inflating the test statistics in our meta-analysis.

The large number of associated loci allowed us to address a number of questions about the genetic architecture of height, which will likely provide useful insights about other polygenic traits. One of the most important issues arising from GWA studies has been the question of missing heritability [8]. To date, common genetic variants have typically explained only a small proportion of the heritable component of phenotypic variation. This is particularly true for height, where >80% of the variation within a given population is estimated to be attributable to additive genetic factors [9,10,11], but over 40 previously published variants explain only 3 to 5% of the variance [12,13,14,15,16,17,18,19]. One possible explanation is that many common variants of small effects contribute to phenotypic variation, and current GWA studies remain underpowered to detect the majority of common variants. Using a set of five studies independent from the GWA study discovery samples, we found that the 180 associated SNPs explained on average 10.5% (range between studies 7.9-11.2%) of the variance in adult height, compared with 4.7% (3.7-5.6%) of the variance explained by the previously known loci (**Supplementary Methods)**. Thus, the identified genome-wide significant loci from the increase in sample size resulted in a doubling of phenotypic variation explained.

We considered the possibility that additional loci not reaching stringent thresholds of genome-wide significance even in this large sample could contribute to additional heritability. We, therefore, tested how much more of the variation in height could be explained by a more inclusive set of variants, using an approach

recently described for schizophrenia [20]. We showed that including SNPs associated with height at lower significance levels (in the range of $0.05>P>5x10^{-8}$) increased the variance explained to 13.3% (range 9.7-16.8%) in the same five study set (**Figure 2a**) (**Supplementary Methods).**

As a separate approach, we used a recently developed method (Park et *al*., submitted) to estimate the total number of independent height-associated variants that are likely to exist with effect sizes similar to the ones described by our study. We obtained this estimate using the distribution of effect sizes observed in Stage 2 and the power to detect an association in Stage 1, given these effect sizes (**Supplementary Methods)**. The cumulative distribution of height loci, including those we identified and others as yet undetected, is shown in **Figure 2b**. We estimate that there are 697 loci (95% confidence interval (CI): 483, 1040) with effects equal or greater than those identified, which together would explain approximately 15.7% of the phenotypic variation in height or 19.6% (95% CI: 16.2-25.6) of height heritability (**Supplementary Table 4**). We estimated that a sample size of 500,000 would be needed to detect 99.6% of these loci at $P<5x10^{-8}$. It is important to note that this figure does not account for variants that have effect sizes smaller than those observed in the current study and, therefore, underestimates the contribution of undiscovered common loci to phenotypic variation.

GWA studies have identified few, if any, examples of non-additive effects. We therefore assessed dominant, recessive, and pairwise interaction effects for the 180 associated variants in a subset of 103,034 analyzed individuals. We found no evidence of deviation from an additive model of inheritance, either for single variants or pairwise between variants (**Supplementary Methods**). In addition, there was no evidence for heterogeneity between sexes for the 180 variants (**Supplementary Table 1**). We also assessed all pairwise joint effects genome-wide, using a smaller subset of 10,500 individuals genotyped at 343,249 SNPs (**Supplementary Methods**) and saw no evidence of genome-wide pairwise interaction (**Supplementary Methods and Supplementary Tables 5 and 6**).

These results therefore do not support a role for non-additive models of inheritance substantially contributing to variation in height.

A further possible source of missing heritability is allelic heterogeneity – the presence of multiple, independent variants influencing a trait at the same locus. Because we only considered the most strongly associated variant within each of the 2Mb loci, it is possible that additional variants in those loci could be contributing to phenotypic variation in height. We performed genome-wide conditional analyses in a subset of the studies used in Stage 1, including a total of 106,336 analyzed individuals. Each participating study repeated the primary GWA analysis but additionally adjusted for the SNPs representing the 207 loci associated at $P<5x10^{-6}$ (**Supplementary Methods**). We then meta-analysed these studies in the same way as for the primary GWA study meta-analysis. In this conditional analysis, 19 SNPs within the 207 loci were associated with height at $P<3.3x10^{-7}$ (a Bonferroni-corrected significance threshold calculated from the ~15% of the genome covered by the 207 2Mb loci; **Supplementary Methods**; **Table 1**, **Figure 3**). The distances of the second signals to the lead SNPs suggested that both are likely to be affecting the same gene, rather than being coincidentally in close proximity. At 17 of 17 loci (excluding two contiguous loci in the *HMGA1* region), the second signal occurred within 500kb, rather than between 500kb and 1 Mb, of this lead SNP (binomial test $P=2x10^{-5}$). Further analyses of allelic heterogeneity may identify additional variants that increase the proportion of variance explained. For example, within the 207 2Mb loci, a total of 48 independent SNPs (19 at $P<3.3x10^{-7}$ plus an additional 29) reached $P<1x10^{-5}$ when we would expect <2 by chance at this level of significance. Finally, this analysis can only detect two independent signals at each locus, and at some loci, there are likely to be multiple independent signals that will require additional rounds of conditional analysis to be robustly identified.

Whilst GWA studies have identified many variants robustly associated with common human diseases and traits, the biological significance of these variants, and the genes on which they act, is often unclear. We used the large number of

height loci to address two further questions of potential relevance to all polygenic traits: the extent to which associated variants are over-represented amongst putatively functional variants, and the extent to which associated loci contain genes of known relevance to the phenotype.

We first tested the overlap between the 180 height-associated variants and missense polymorphisms, the putative functional variants, and copy number variants (CNVs). Height variants were 1.7-fold more likely to be closely correlated ($r^2 \geq 0.8$ in HapMap CEU) with nonsynonymous SNPs ($P=0.004$) (**Supplementary Methods**, **Supplementary Table 7**), but there was no correlation with CNV-tagging SNPs (**Supplementary Methods**). We also noted five loci where the height associated variant was strongly correlated ($r^2>0.8$) with variants associated with other traits and diseases (at $P<5\times10^{-8}$), including bone mineral density, rheumatoid arthritis, type 1 diabetes, psoriasis and obesity, suggesting that these variants have pleiotropic effects on human phenotypes (**Supplementary Methods**; **Supplementary Table 8**).

To address the extent to which height variants cluster near biologically relevant genes, we performed a number of analyses. First, we tested whether the signals of association cluster near genes mutated in human syndromes characterized by abnormal skeletal growth. We limited this analysis to the 652 genes occurring within the recombination hotspot-bounded regions surrounding each of the 180 index SNPs. We showed that the 180 loci associated with variation in normal height contained 21 of 241 genes (8.7%) found to underlie such syndromes (**Supplementary Table 9**), compared to a median of 8 (range 1-19) genes identified in 1,000 matched control sets of regions ($P<0.001$). Interestingly, in 13 of these 21 loci the closest gene to the most associated height SNP in the region is the growth disorder gene, and in 9 of these cases, the most strongly associated height SNP is located within the growth disorder gene itself (**Supplementary Methods; Supplementary Table 10**). These results suggest that GWA studies may provide more clues about the identity of the functional genes at each locus than previously suspected.

Next, we investigated whether significant and relevant biological connections exist between the genes within the 180 loci, using two different approaches. We used the GRAIL text-mining algorithm to search for connectivity between genes near the associated SNPs, based on existing literature [21]. Of the 180 loci, 42 contained genes that were connected by existing literature to genes in the other associated loci (the pair of connected genes appear in articles that share scientific terms more often than expected at $P<0.01$). For comparison, when we used GRAIL to score 1,000 sets of 180 SNPs not associated with height (but matched for number of nearby genes, gene proximity, and allele frequency), we only observed 2 sets with 42 or more loci with a connectivity $P<0.01$, thus providing strong statistical evidence that the height loci are functionally related ($P=0.002$) (**Figure 4a**). For the 42 regions with GRAIL connectivity $P<0.01$, the implicated genes and SNPs are highlighted in **Figure 4b**. The most strongly connected genes include those in the Hedgehog, TGF-beta, and growth hormone pathways.

Finally, we used a novel implementation of gene set enrichment analysis (GSEA) (Meta-Analysis Gene-set Enrichment of variaNT Associations, MAGENTA (Segrè et *al*., in revision)) and the biological process classification from the INGENUITY, KEGG, PANTHER and Gene Ontology databases to perform pathway analysis (**Supplementary Methods**). This analysis revealed 17 different biological pathways and 14 molecular functions nominally enriched ($P<0.05$) for associated genes, many of which lie within the validated height loci. These gene-sets include previously reported[12,14] (e.g. Hedgehog signaling) and novel (e.g. TGF-beta signaling, histones, and growth and development-related) pathways and molecular functions (**Supplementary Table 11**). These results provide complementary evidence for some of the genes and pathways highlighted in the GRAIL analysis. For instance, genes such as *TGFB2* and *LTBP1-3* highlight a role for the TGF-beta signaling pathway in regulating human height, consistent with the recent implication of this pathway in Marfan syndrome [22].

We have identified over a hundred novel loci that influence the classic, polygenic trait of normal variation in human height, bringing the total to 180. Our results have potential implications for studies seeking to understand the genetic component to common diseases and traits. We show that loci identified by GWA studies highlight relevant genes: the 180 loci associated with height are non-randomly clustered within biologically relevant pathways and are enriched for genes that are involved in growth-related processes, that underlie syndromes of abnormal skeletal growth, and that even encode therapies or targets for therapies that modulate growth (*GH1*, *IGF1R*, *CYP19A1*, *ESR1*). The large number of loci with clearly relevant genes suggests that the remaining loci could provide potential clues to important and novel biology.

We provide the strongest evidence yet that the causal gene will often be located near the most strongly associated DNA sequence variant. At the 21 loci containing a known growth disorder gene, that gene was on average 81 kb from the associated variant, and in over half of the loci, it was the closest gene to the associated variant. Despite recent doubts about the benefits of GWA studies [23], this finding suggests that GWA studies are useful mapping tools to highlight genes that merit further study. The presence of multiple variants within associated loci (allelic heterogeneity), could help localize the relevant genes within these loci.

By increasing our sample size to over 100,000 individuals, we identified common variants that account for 10.5% of phenotypic variation, with another ~6% of the phenotypic variance explained by common variants of similar effects yet to be discovered. Although 10.5% of phenotypic variation (13% of genetic variation) accounted for by the 180 loci is larger than predicted under some models [23], this figure suggests that GWA studies, as currently implemented and when applied to realistic sample sizes, may not explain a majority of the estimated 80% contribution of genetic factors to variation in height. This conclusion supports the idea that biological insights, rather than predictive power, will be the main outcome of this initial wave of GWA studies, and that new approaches, which could include sequencing studies or GWA studies targeting variants of lower frequency, will be

needed to account for more of the “missing” heritability. Our finding that many loci exhibit allelic heterogeneity suggests that many as yet unidentified causal variants will map to the loci already identified in GWA studies, and that the fraction of causal *loci* that have been identified could be substantially greater than the fraction of causal *variants* that have been identified.

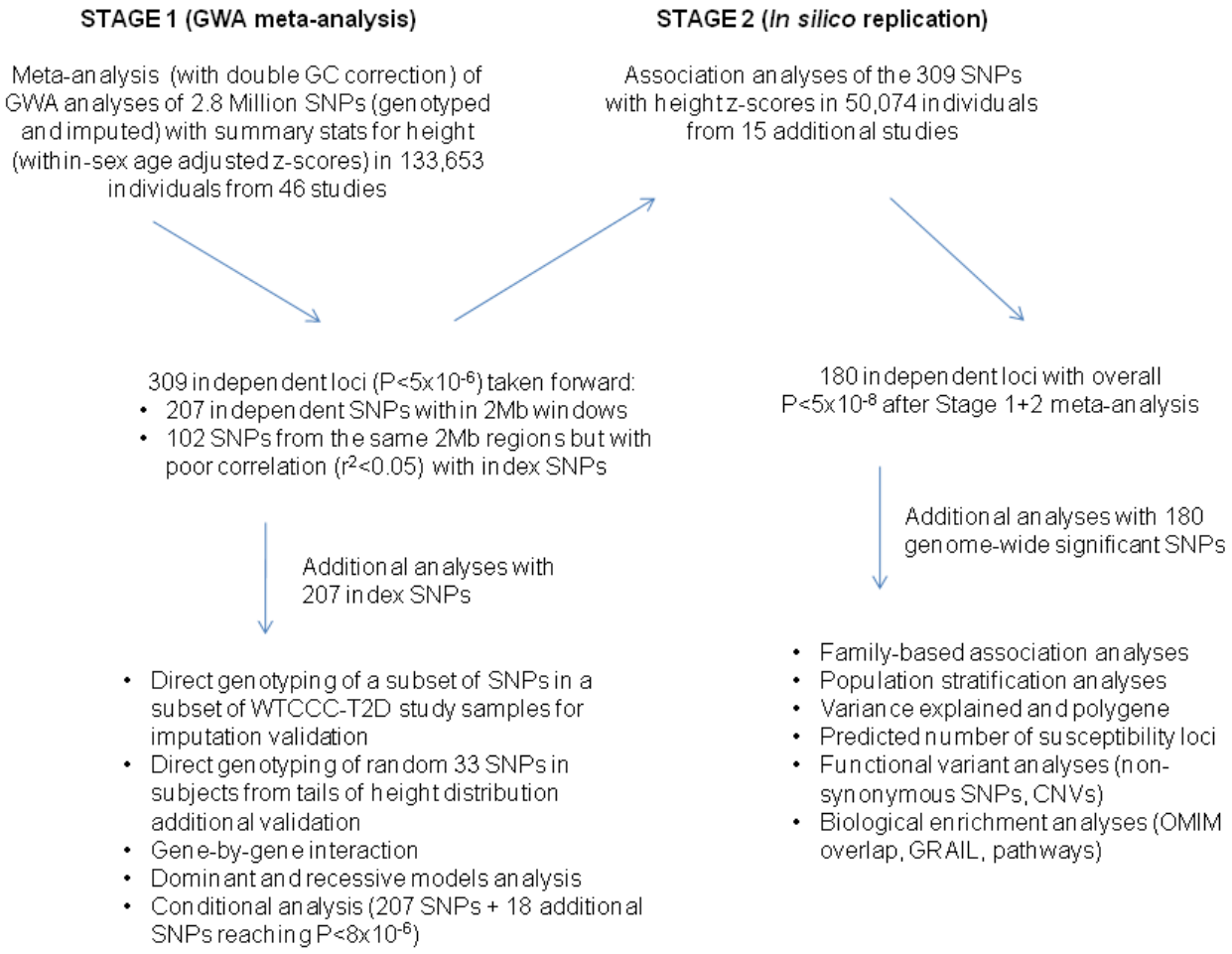
The observed contributions of common and rare variants to heritability will be dependent on the particular phenotype, as well as on the power of future studies to detect both common and rare variants with modest effects. We note that, in our studies, many associated variants are tightly correlated with common nonsynonymous SNPs, which would not be expected if these associated common variants were proxies for collections of rare causal variants, as has been proposed [24]. Although a substantial contribution to heritability by less common and/or quite rare variants may be more plausible, our data do not rule out the regulation of height by a large number of common variants of very small effect.

In summary, our findings indicate that additional approaches, including those aimed at less common or rare variants, will likely be needed to dissect more completely the genetic component to complex human traits. Our results also strongly demonstrate that GWA studies can identify large numbers of loci that together implicate biologically relevant pathways and mechanisms. We envision that thorough exploration of the genes at associated loci through additional genetic, functional, and computational studies will lead to novel insights into human height and other polygenic traits and diseases.

Methods summary

The primary meta-analysis (Stage 1) included 46 GWA studies of 133,653 individuals. The *in-silico* follow up (Stage 2) included 15 studies of 50,074 individuals. All studied individuals were of European ancestry and >99.8% were adults. Details of genotyping, quality control, and imputation methods of each study are given in **Supplementary Methods Tables 1 and 2**. Each study provided summary results of a linear regression of age-adjusted, within-sex Z scores of height against the imputed SNPs, and an inverse-variance meta-analysis was performed in METAL (http://www.sph.umich.edu/csg/abecasis/METAL/). Validation of selected SNPs was performed through direct genotyping in an extreme height panel (N=3,190) using Sequenom iPLeX, and in 492 Stage 1 samples using the KASPar SNP System. Family-based testing was performed using QFAM, a linear regression-based approach that uses permutation to account for dependency between related individuals [25], and FBAT, which uses a linear combination of offspring genotypes and traits to determine the test statistic[26]. We used a previously described method to estimate the amount of genetic variance explained by the nominally associated loci (using significance threshold increments from $P<5\times10^{-8}$ to $P<0.05$) [20]. To predict the number of height susceptibility loci, we took the height loci that reached a significance level of $P<5\times10^{-8}$ in Stage 1 and estimated the number of height loci that are likely to exist based on the distribution of their effect sizes observed in Stage 2 and the power to detect their association in Stage 1. Gene-by-gene interaction, dominant, recessive and conditional analyses are described in detail in **Supplementary Methods**. Empirical assessment of enrichment for coding SNPs used permutations of random sets of SNPs matched to the 180 height-associated SNPs on the number of nearby genes, gene proximity, and minor allele frequency. GRAIL and GSEA methods have been described previously [21] (Segre et *al.*, submitted). To assess possible enrichment for genes known to be mutated in severe growth defects, we identified such genes in the OMIM database (**Supplementary Table 10**), and evaluated the extent of their overlap with the 180 height-associated regions through comparisons

with 1000 random sets of regions with similar gene content (±10%). The design of the study, SNPs taken forward and additional analyses are summarised below.



Author Contributions

Full author contributions and roles are listed in the Supplementary Note.

Acknowledgments

A number of participating studies are members of CHARGE and ENGAGE consortia. We acknowledge funding from the Academy of Finland (104781, 117797, 120315, 121584, 126925, 129269, 129494, 129680, 213506); Affymetrix, Inc for genotyping services (N02-HL-6-4278); Agency for Science, Technology and Research of Singapore (A*STAR); ALF/LUA Gothenburg; Althingi (the Icelandic Parliament); Amgen; AstraZeneca AB; Australian National Health and Medical Research Council (241944, 389875, 389891, 389892, 389938, 442915, 442981, 496739, 496688, 552485, 613672); Australian Research Council (DP0770096); Biocentrum Helsinki; Boston Obesity Nutrition Research Center (DK46200); British Diabetes Association; British Heart Foundation (PG/02/128); British Heart Foundation Centre for Research Excellence, Oxford; CamStrad; Cancer Research UK; Centre for Neurogenomics and Cognitive Research (CNCR-VU); Chief Scientist Office of the Scottish Government (CSO) (CZB/4/279); Council of Health of the Academy of Finland; DIAB Core project of the German Network of Diabetes; Diabetes UK; Donald W. Reynolds Foundation; Emil and Vera Cornell Foundation; Erasmus MC; Estonian Government (SF0180142s08); European Commission (201413, ECOGENE:205419, BBMRI:212111, OPENGENE:245536, ENGAGE:HEALTH-F4-2007-201413, EURODIA:LSHG-CT-2004-518153, EU/WLRT-2001-01254, HEALTH-F2-2008-ENGAGE, HEALTH-F4-2007-201550, LSH-2006-037593, LSHG-CT-2006-018947, LSHG-CT-2006-01947, Procardis:LSHM-CT-2007-037273, POLYGENE:LSHC-CT-2005, QLG1-CT-2000-01643, QLG2-CT-2002-01254, DG XII, Marie Curie Intra-European Fellowship); Eve Appeal; Finish Ministry of Education; Finnish Diabetes Research Foundation; Finnish Diabetes Research Society; Finnish Foundation for Cardiovascular Research; Finnish Medical Society; Finska Läkaresällskapet; Folkhälsan Research Foundation; Fondation LeDucq; Foundation for Life and Health in Finland; Foundation for Strategic Research (SSF); GEN-AU-Programme "GOLD"; Genetic Association InformationNetwork (GAIN); German Bundesministerium fuer Forschung und Technology (01 AK 803 A-H, 01 IG 07015 G); German Federal Ministry of Education and Research (BMBF) (01GS0831); German Ministry for Health, Welfare and Sports; German Ministry of Cultural Affairs; German Ministry of Education, Culture and Science; German National Genome Research Net (NGFN2 and NGFNplus) (01GS0823, 01ZZ0103, 01ZZ0403, 01ZZ9603, 03ZIK012); German Research Council (KFO-152); GlaxoSmithKline; Göteborg Medical Society; Gyllenberg Foundation; Helmholtz Center Munich; Juvenile Diabetes Research Foundation International (JDRF) (U01 DK062418); Karolinska Institute; Knut and Alice Wallenberg Foundation; Lundberg Foundation; March of Dimes (6-FY-09-507); MC Health; Medical Research Council UK (G0000649, G0000934, G0500539, G0600331, G0601261, G9521010D, PrevMetSyn); Microarray Core Facility of the Interdisciplinary Centre for Clinical Research (IZKF) (B27); Mid-Atlantic Nutrition and Obesity Research Center of Maryland (P30 DK072488); Ministry of Health and Department of Educational Assistance (South Tyrol, Italy); Ministry of Science, Education and Sport of the Republic of Croatia (216-1080315-0302); Montreal Heart Institute Foundation; Närpes Health Care Foundation; National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre; NIHR Oxford Biomedical Research Centre; NIHR comprehensive Biomedical Research Centre; National Institutes of Health (263-MA-410953, AA014041, AA07535, AA10248, AA13320, AA13321, AA13326, CA047988, CA49449, CA50385, CA65725, CA67262, CA87969, DA12854, DK062370, DK063491, DK072193, DK079466, DK080145, DK58845, HG002651, HG005214, HG005581, HL043851, HL084729, HL69757, HL71981, K08-AR055688, K23-DK080145, K99-HL094535, M01-RR00425, MH084698, N01-AG12100, N01-AG12109, N01-HC15103, N01-HC25195, N01-HC35129, N01-HC45133, N01-HC55015, N01-HC55016, N01-HC55018 through N01-HC55022, N01-HC55222, N01-HC75150, N01-HC85079 through N01-HC85086, N01-HG65403, R01-AG031890, R01 CA104021, R01-DK068336, R01-DK073490, R01-DK075681, R01-DK075787, R01-HL086694, R01-HL087641, R01-HL087647, R01-HL087652, R01-HL087676, R01-HL087679, R01-HL087700, R01-HL088119, R01-HL59367, R01-MH059160, R01-MH59565, R01-

MH59566, R01-MH59571, R01-MH59586, R01-MH59587, R01-MH59588, R01-MH60870, R01-MH60879, R01-MH61675, R01-MH63706, R01-MH67257, R01-MH79469, R01-MH81800, RL1-MH083268, T32-HG00040, U01-CA098233, U01-GM074518, U01-HG004399, U01-HG004402, U01-HL080295, U01-HL084756, U01-HL72515, U01-MH79469, U01-MH79470, U54-RR020278, UL1-RR025005, Z01-AG00675, Z01-AG007380, Z01-HG000024; contract HHSN268200625226C; ADA Mentor-Based Postdoctoral Fellowship; Pew Scholarship for the Biomedical Sciences); Netherlands Genomics Initiative (NGI)/Netherlands Consortium for Healthy Aging (NCHA) (050-060-810); Netherlands Organisation for Scientific Research (NWO) (Investments nr. 175.010.2005.011, 911-03-012); Netherlands Organization for the Health Research and Development (ZonMw) (10-000-1002); Netherlands Scientific Organization (904-61-090, 904-61-193, 480-04-004, 400-05-717, Center for Medical Systems Biology (NOW Genomics), SPI 56-464-1419) ; NIA Intramural Research Program; Nordic Center of Excellence in Disease Genetics; Novo Nordisk Foundation; Ollqvist Foundation; Oxford NIHR Biomedical Research Centre; Paavo Nurmi Foundation; Perklén Foundation; Petrus and Augusta Hedlunds Foundation; Queensland Institute of Medical Research; Radboud University Nijmegen Medical Centre; Research Institute for Diseases in the Elderly (014-93-015); Royal Swedish Academy of Science; Sahlgrenska Center for Cardiovascular and Metabolic Research (A305:188); Siemens Healthcare, Erlangen, Germany; Signe and Ane Gyllenberg Foundation; Sigrid Juselius Foundation; Social Insurance Institution of Finland; Social Ministry of the Federal State of Mecklenburg-West Pomerania; South Tyrolean Sparkasse Foundation; Stockholm County Council (560183); Support for Science Funding programme; Susan G. Komen Breast Cancer Foundation; Swedish Cancer Society; Swedish Cultural Foundation in Finland; Swedish Foundation for Strategic Research; Swedish Heart-Lung Foundation; Swedish Medical Research Council (K2007-66X-20270-01-3, 8691); Swedish National Cancer Institute; Swedish Research Council; Swedish Society of Medicine; Swiss National Science Foundation (33CSCO-122661); Torsten and Ragnar Söderberg's Foundation; Vandervell Foundation; Västra Götaland Foundation; Wellcome Trust (072960, 075491, 079557, 079895, 083270, 068545/Z/02, 076113/B/04/Z, 076113/C/04/Z, 076113/C/04/Z, 077016/Z/05/Z, 081682/Z/06/Z, 084183/Z/07/Z, 085301/Z/08/Z, 086596/Z/08/Z, 091746/Z/10/Z; WT Research Career Development Fellowship); Western Australian Genetic Epidemiology Resource and the Western Australian DNA Bank (both National Health and Medical Research Council of Australia Enabling Facilities). Detailed list of acknowledgments by study is given in the Supplementary Information.

Table 1: Secondary signals at associated loci after conditional analysis. 19 independent SNPs reached a $P<3.3\times10^{-7}$, a Bonferroni significance threshold based on the proportion of the genome covered by the 207 loci. The results are based on a subset of 106,336 individuals. [a]HapMap CEU phase II release 23 was used; [b]this is the nearest gene or a known skeletal growth disorder gene in the locus (underlined). Positions are based on NCBI build 36. *****nearest conditioned SNP where second signal occurs within 1Mb of two conditioned SNPs.

Second signal SNP	Conditioned SNP	Chr	Second signal SNP position	Distance of conditioned SNP from index SNP (bp)	HapMap[a] r^2	Second signal *P*-value after conditioning	Second signal *P*-value pre-conditioning	Gene[b]
rs2280470	rs16942341	15	87196630	6721	0.009	1×10^{-14}	1×10^{-15}	*ACAN*
rs10859563	rs11107116	12	92644470	141835	0.003	3×10^{-12}	8×10^{-10}	*SOCS2*
rs750460	rs5742915	15	72028559	95127	0.004	4×10^{-12}	7×10^{-08}	*PML*
rs6938239	rs2780226*	6	34791613	484583	0.019	6×10^{-12}	9×10^{-14}	*HMGA1*
rs7652177	rs572169	3	173451771	196650	0.006	7×10^{-11}	1×10^{-11}	*GHSR*
rs7916441	rs2145998	10	80595583	196119	0.112	6×10^{-10}	3×10^{-07}	*PPIF*
rs3792752	rs1173727	5	32804391	61887	0.02	7×10^{-10}	4×10^{-08}	*NPR3*
rs10958476	rs7460090	8	57258362	98355	0.02	1×10^{-09}	5×10^{-13}	*SDR16C5*
rs2353398	rs7689420	4	145742208	45594	0.022	2×10^{-09}	1×10^{-10}	*HHIP*
rs2724475	rs6449353	4	17555530	87056	0.098	2×10^{-09}	8×10^{-16}	*LCORL*
rs2070776	rs2665838	17	59361230	41033	0.15	9×10^{-09}	1×10^{-14}	*GH region*
rs1401796	rs227724	17	52194758	60942	0.005	2×10^{-08}	7×10^{-07}	*NOG*
rs4711336	rs2780226*	6	33767024	540046	0.111	3×10^{-08}	5×10^{-08}	*HMGA1*
rs6892884	rs12153391	5	170948228	187815	0	4×10^{-08}	2×10^{-05}	*FBXW11*
rs1367226	rs3791675	2	55943044	21769	0.204	4×10^{-08}	0.1245	*EFEMP1*
rs2421992	rs17346452	1	170507874	187964	0.019	5×10^{-08}	1×10^{-05}	*DNM3*
rs225694	rs7763064	6	142568835	270147	0.001	1×10^{-07}	2×10^{-06}	*GPR126*
rs10187066	rs12470505	2	219223003	393610	0.022	2×10^{-07}	5×10^{-08}	*IHH*
rs879882	rs2256183	6	31247431	241077	0.016	2×10^{-07}	8×10^{-08}	*MICA*

Figure Legends

Figure 1. 180 loci associated with adult height variation. (A) Karyogram displaying the genome location of the 180 height SNPs identified from the primary meta-analysis (green) and the 19 secondary signals (red) discovered in the conditional analysis to be associated with height. The closest genes to the SNPs (gray) are followed by a MIM (blue) label if the gene underlies a skeletal growth-related Mendelian disorder described in OMIM. (B) Quantile-quantile plot of SNPs after Stage 1 GIANT GC-corrected meta-analysis (black), after removal of SNPs near 47 loci previously shown to associated with height in Caucasians (blue), and after removal of SNPs near 180 loci shown to associate with height in this study (red). All SNPs near (2Mb window) or in linkage disequilibrium (r^2 ≥0.01) with the 47 or 180 index height SNPs were excluded to draw the blue and red distributions, respectively.

Figure 2. Phenotypic variance explained by common variants. (A) Variance explained is higher when SNPs not reaching genome-wide significance are included in the prediction model. The y-axis represents the proportion of variance explained at different *P*-value thresholds from Stage 1 meta-analysis. Results are given for five studies that were not part of the GWA discovery set (Stage 1). *Proportion of variation explained by 180 SNPs reaching genome-wide significance in Stage 1+2 meta-analysis. (B) Cumulative number of susceptibility loci expected to be discovered, including those we have already identified and others as yet undetected, by the expected percentage of phenotypic variation explained and sample size required assuming a GC correction utilized. The projections are based on loci that achieved a significance level of $P<5x10^{-8}$ in the initial scan and the distribution of their effect sizes in Stage 2. The dotted red line corresponds to expected phenotypic variance explained by the 110 loci that reached genome-wide significance in Stage 1, were replicated in Stage 2 and had at least 1% power.

Figure 3. Example regional association plots of loci with secondary signals from conditional analysis before and after conditioning. The plots are centered on the conditioned SNP (shown as the yellow diamond) at the locus. The secondary signal SNP is highlighted as the pink diamond. r^2 is based on the CEU HapMap II samples. The blue line and right hand Y axis represent CEU HapMap II based recombination rates. LD Plots were created by LocusZoom (http://csg.sph.umich.edu/locuszoom/).

Figure 4. Loci associated with height contain genes related to each other. (A) 180 SNPs associated with adult height variation. The y-axis plots GRAIL *P*-values on a log scale. The histogram on the left side corresponds to the distribution of GRAIL *P*-values for 1,000 sets of 180 matched SNPs. The scatter plot on the right represents GRAIL results for the 180 height SNPs found in this study (blue dots). The black horizontal line marks the median of the GRAIL *P*-values (*P*=0.14). The top 10 keywords linking the genes were: ‘growth’, ‘kinase’, ‘factor’, ‘transcription’, ‘signaling’, ‘binding’, ‘differentiation’, ‘development’, ‘insulin’, ‘bone’. (B) Graphical representation of the connections between SNPs and corresponding genes for the 42 SNPs with GRAIL *P*<0.01. Thicker and redder lines imply stronger literature-based connectivity.

Figure 1A: Genome locations of the 199 height SNPs associated with height.

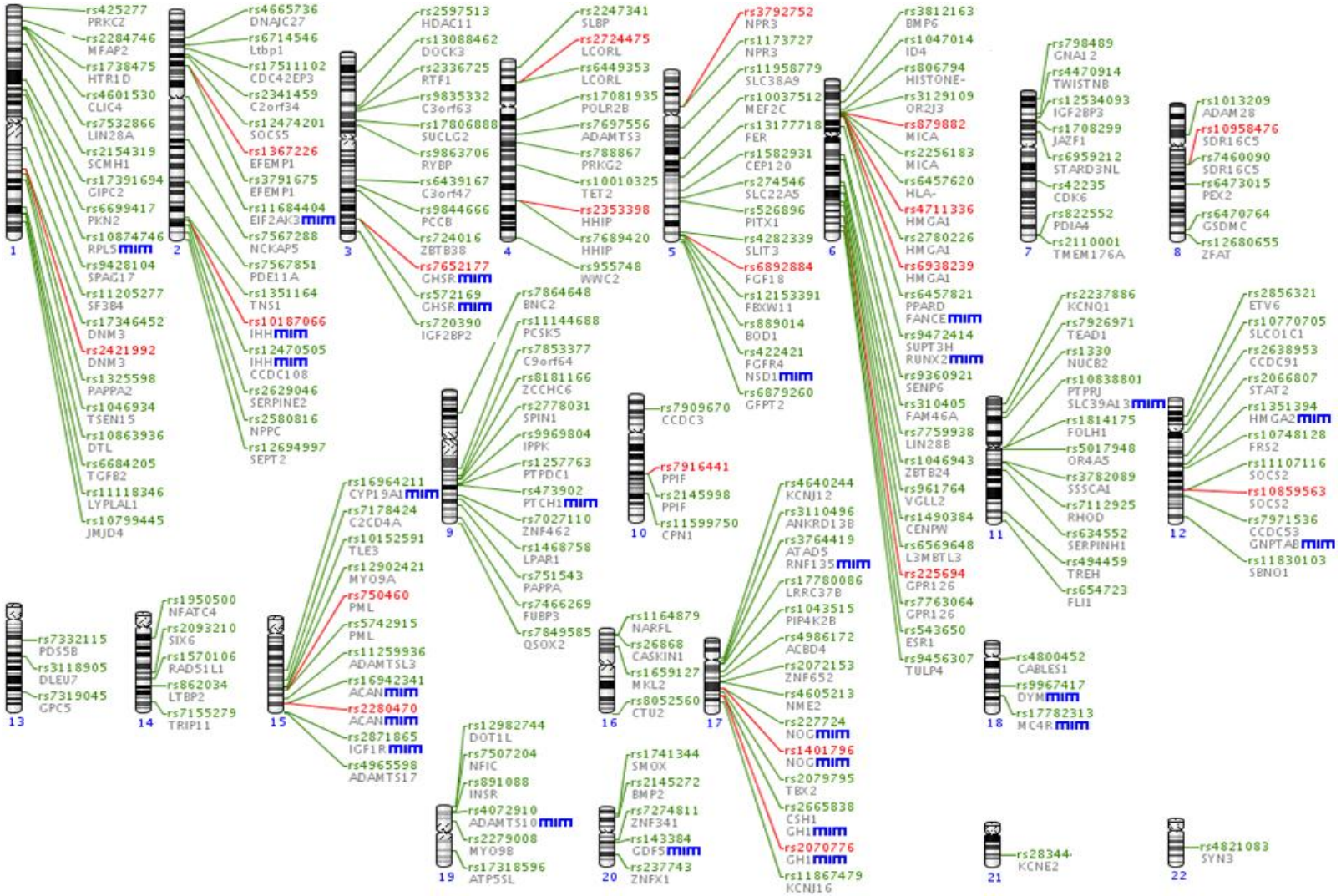


Figure 1B: Quantile-quantile plot for height.

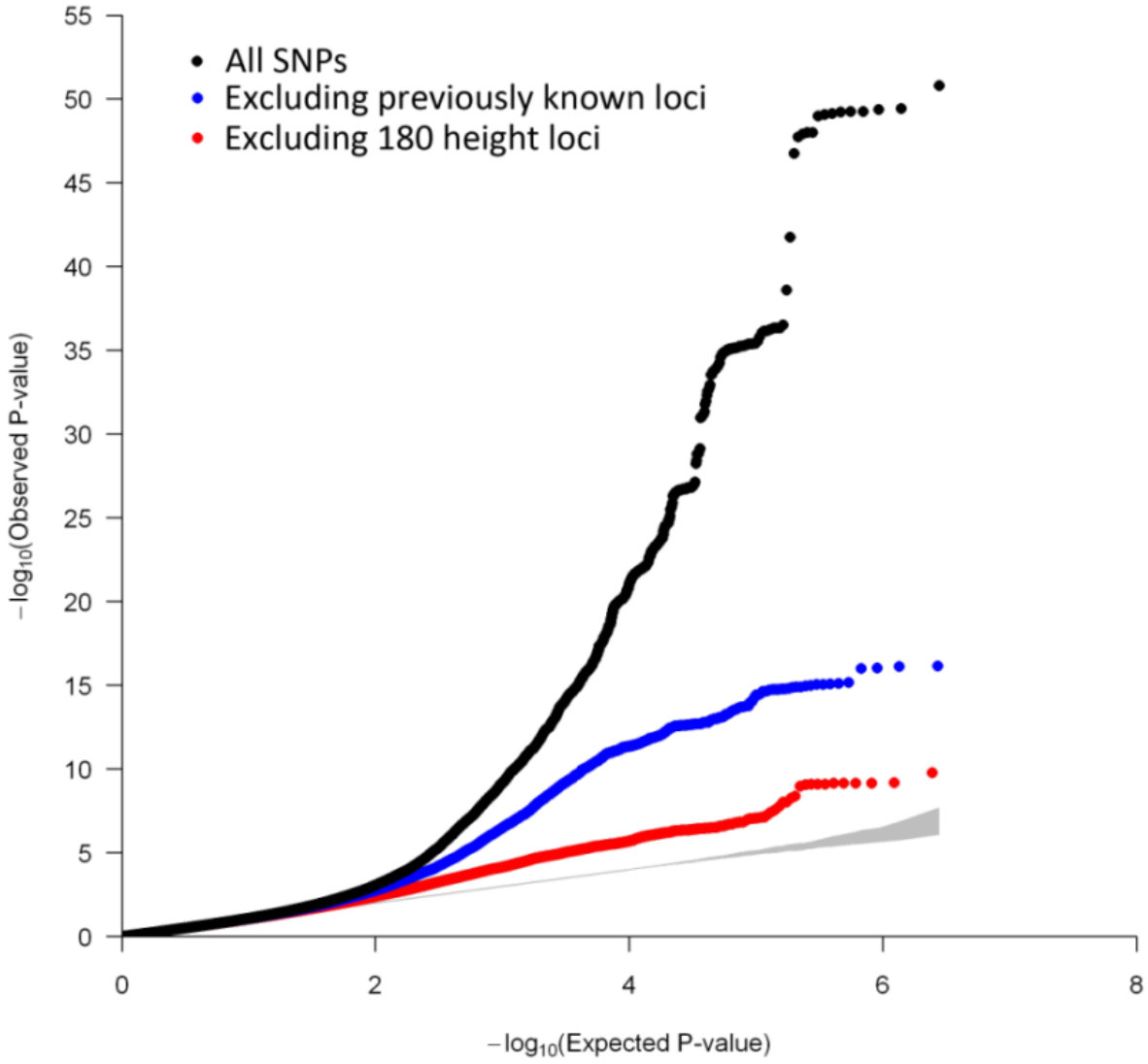


Figure 2A: Proportion of phenotypic variance explained.

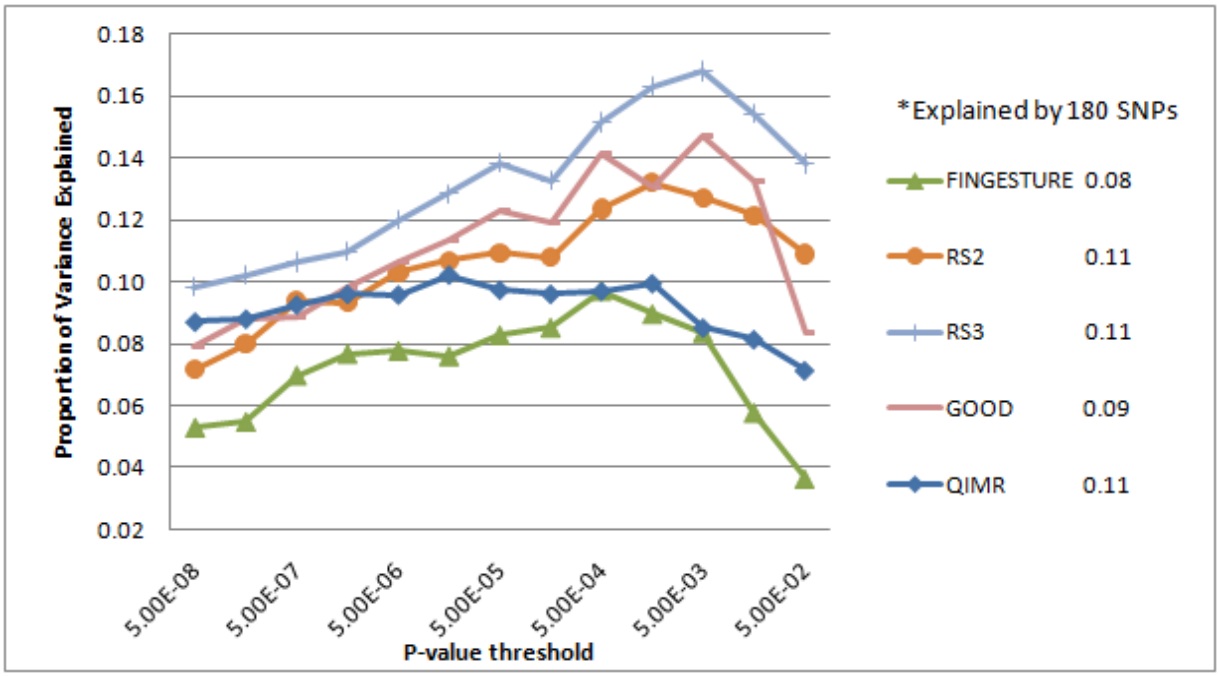


Figure 2B: Predicted number of height loci.

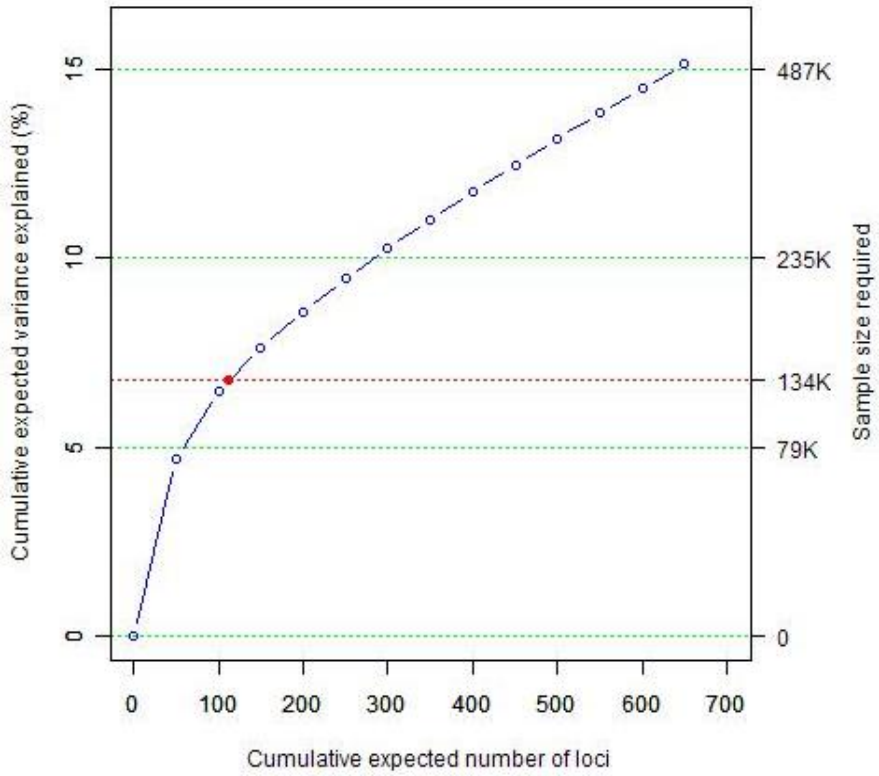
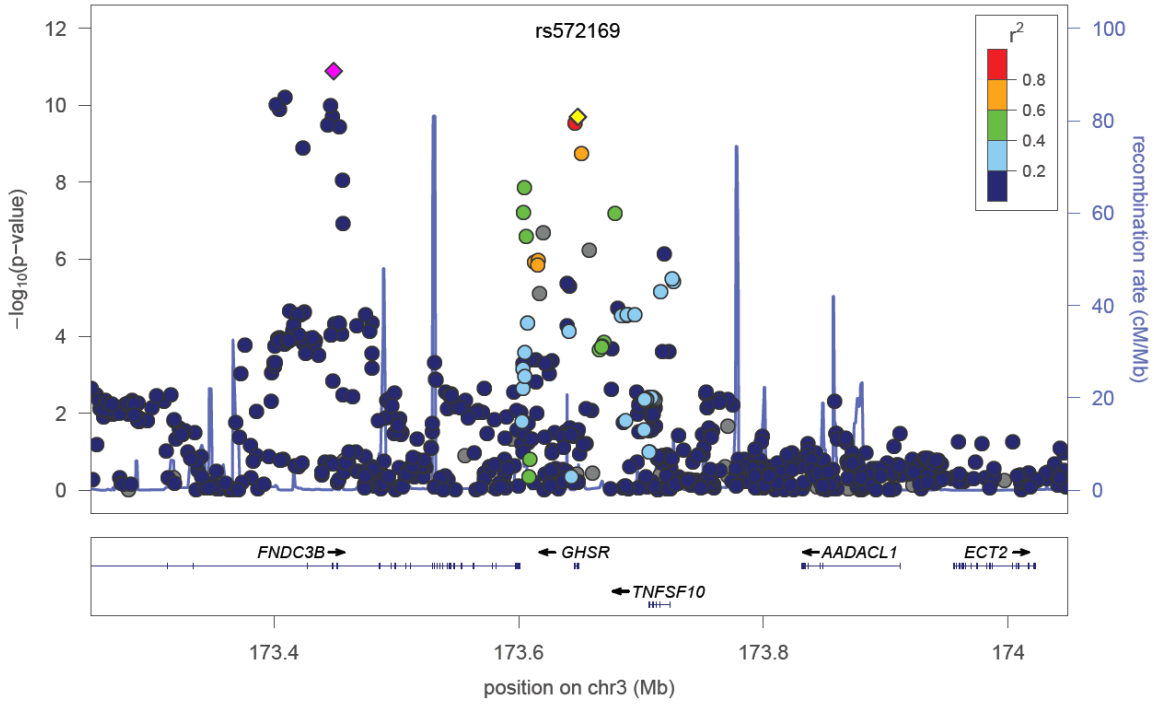
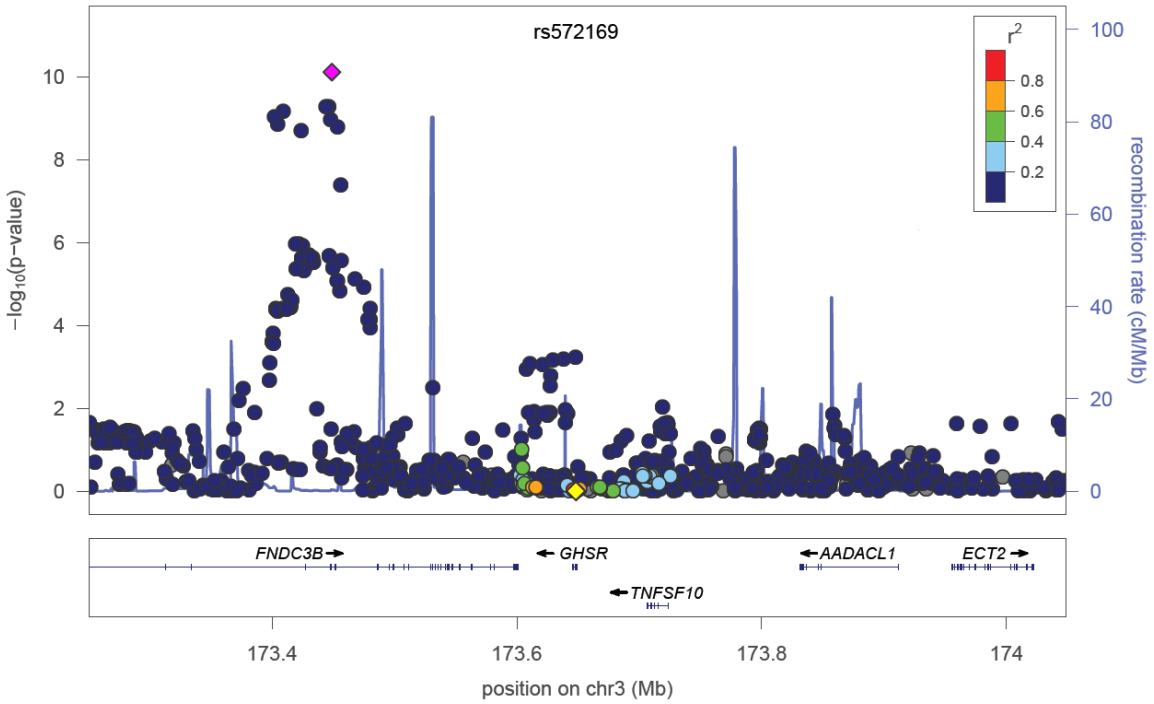


Figure 3. Examples of secondary signals before and after conditional analysis.

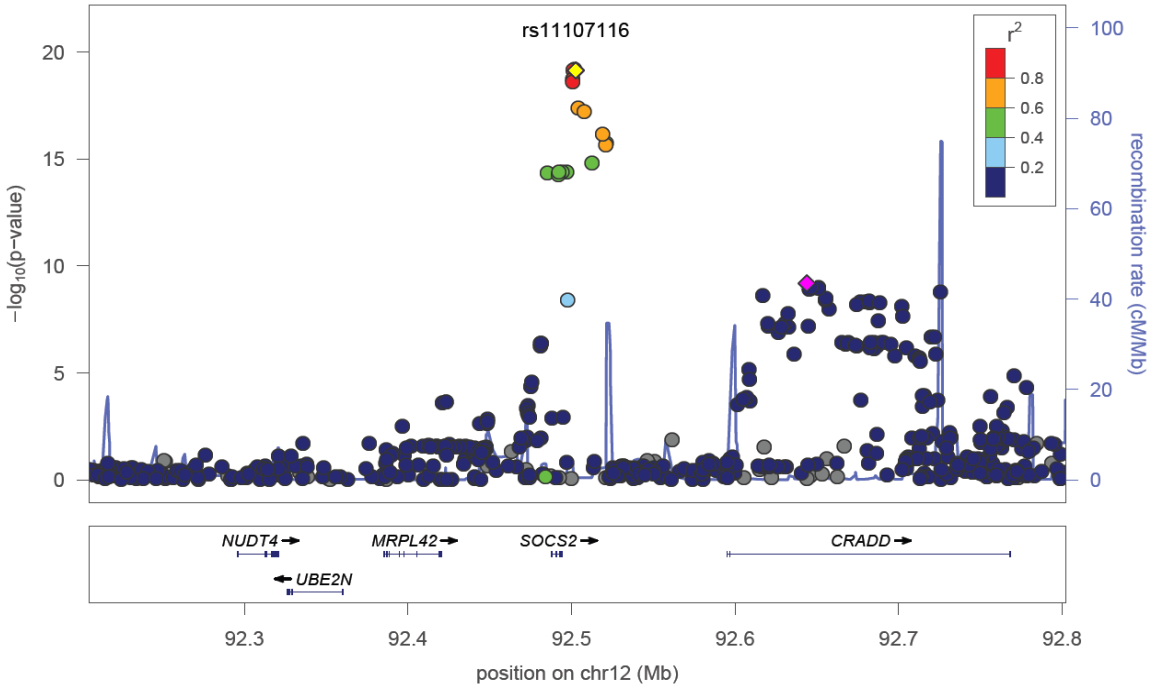
Unconditioned



Conditioned



Unconditioned



Conditioned

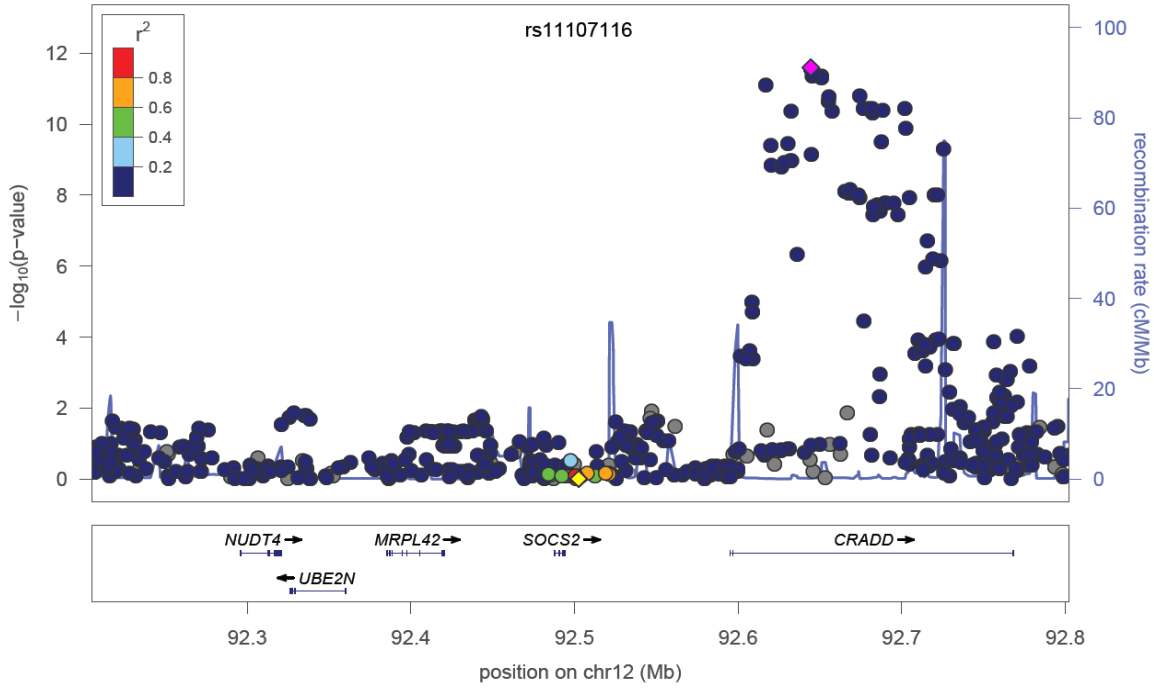
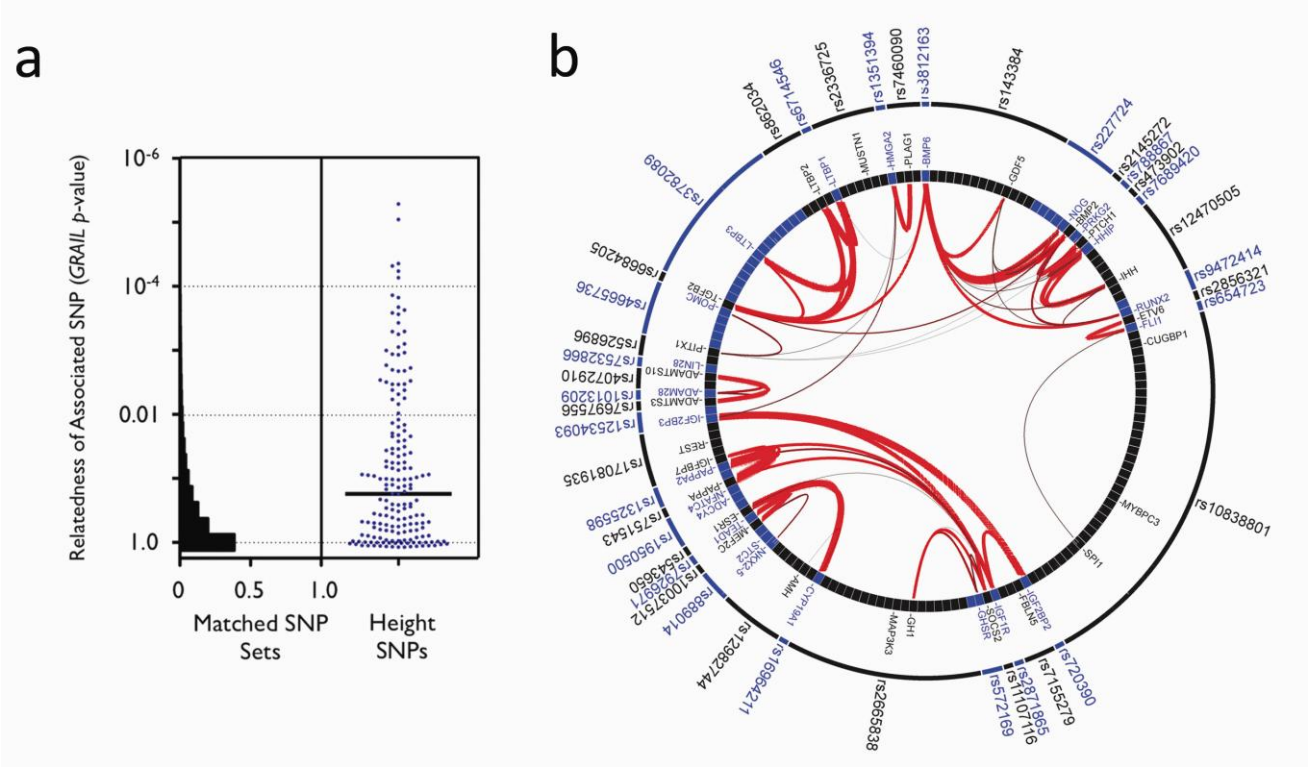


Figure 4: GRAIL analysis



References

1. Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA* 106, 9362-9367(2009).

2. Galton, F. Regression towards mediocrity in hereditary stature. *J. R. Anthropol. Inst.* 5, 329-348 (1885).

3. Fisher, R. A. The Correlation Between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh* 52, 399-433 (1918).

4. Frazer, K. A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449, 851-861 (2007).

5. Devlin, B. & Roeder, K. Genomic Control for Association Studies. *Biometrics* 55, 997-1004 (1999).

6. Reich, D.E. & Goldstein, D.B. Detecting association in a case-control study while correcting for population stratification. *Genetic Epidemiology* 20, 4-16 (2001).

7. Campbell, C. D. *et al.* Demonstrating stratification in a European-American population. *Nature Genet* 37, 868-872 (2005).

8. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* 461, 747-753 (2009).

9. Perola, M. *et al.* Combined genome scans for body stature in 6,602 European twins: evidence for common Caucasian loci. *PLoS Genet* 3, e97 (2007).

10. Silventoinen, K. *et al.* Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res* 6, 399-408 (2003).

11. Visscher, P. M. *et al.* Assumption-Free Estimation of Heritability from Genome-Wide Identity-by-Descent Sharing between Full Siblings. *PLoS Genet* 2, e41 (2006).

12. Weedon, M. N. *et al.* Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet* 40, 575-583 (2008).

13. Weedon, M. N. *et al.* A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nat Genet* 39, 1245-1250 (2007).

14. Lettre, G. *et al.* Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat Genet* 40, 584-591 (2008).

15. Gudbjartsson, D. F. *et al.* Many sequence variants affecting diversity of adult human height. *Nat Genet* 40, 609-615 (2008).

16. Sanna, S. *et al.* Common variants in the GDF5-UQCC region are associated with variation in human height. *Nat Genet* 40, 198-203 (2008).

17. Johansson, A. *et al.* Common variants in the JAZF1 gene associated with height identified by linkage and genome-wide association analysis. *Hum Mol Genet* 18, 373-380 (2009).

18. Estrada, K. *et al.* A genome-wide association study of northwestern Europeans involves the C-type natriuretic peptide signaling pathway in the etiology of human height variation. *Hum Mol Genet* 18, 3516-3524 (2009).

19. Soranzo, N. *et al.* Meta-analysis of genome-wide scans for human adult stature identifies novel Loci and associations with measures of skeletal frame size. *PLoS Genet* 5, e1000445 (2009).

20. Purcell, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460, 748-752 (2009).

21. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet* 5, e1000534 (2009).

22. Neptune, E. R. *et al.* Dysregulation of TGF-beta activation contributes to pathogenesis in Marfan syndrome. *Nat Genet* 33, 407-411 (2003).

23. Goldstein, D. B. Common genetic variation and human traits. *N Engl J Med* 360, 1696-1698 (2009).

24. Dickson, S. P., Wang, K., Krantz, I., Hakonarson, H. & Goldstein, D. B. Rare variants create synthetic genome-wide associations. *PLoS Biol* 8, e1000294, (2010).

25. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81, 559-575 (2007).

26. Nan, M. L., Steve, H. & Xin, X. Implementing a unified approach to family-based tests of association. *Genetic Epidemiology* 19, S36-S42 (2000).

SUPPLEMENTARY INFORMATION

Supplementary Methods

1. Primary genome-wide association meta-analysis (Stage 1)

In Stage 1, we combined the height summary statistics from 46 genome-wide association (GWA) studies in a meta-analysis of 133,653 individuals (60,587 males and 73,066 females).

1.1 Description of individual cohorts and genotyping methods

Descriptive characteristics, study design, sample size, sample quality control (QC) and anthropometric measurement technique for the studies included in Stage 1 are provided in **Supplementary Methods Table 1.** All individuals were Caucasians of European ancestry. Approximately 45% of the individuals were male, and the ages ranged from 14 to 103 years (99.7% of the samples were $\geq$18 years old). All participants provided written informed consent and the studies were approved by the respective Local Research Ethics committees or Institutional Review Boards.

Details on the genotyping platform used and genotype quality control procedures employed for each study are presented in **Supplementary Methods Table 2**.

1.2 Imputation

All cohorts were genotyped using commercially available Affymetrix (Affymetrix, Inc., Santa Clara, CA, USA), Illumina (Illumina, Inc., San Diego, CA, USA) genotyping arrays, or custom Perlegen (Perlegen Sciences, Inc., Mountain View, CA, USA) arrays. Quality control was performed independently for each study. To facilitate meta-analysis, each group performed genotype imputation using BIMBAM[1], IMPUTE[2], or MACH[3] and genotypes from the Phase II CEU HapMap[4]. Each imputation software estimates an overall imputation quality score for each SNP. For example, IMPUTE calculates the 'proper info' statistic which is a measure of the observed statistical information for the estimate of allele frequency

of the SNP, while MACH calculates the 'rsq_hat', which is the estimated r^2 between each imputed genotype and its true underlying genotype. Study-specific details are presented in **Supplementary Methods Table 2**.

1.3 GWA analyses in individual cohorts

Details on study-specific analysis software are summarized in **Supplementary Methods Table 2.** Each GWA study tested association between each imputed or genotyped SNP and sex-standardized height, assuming an additive inheritance model and adjusting for age and other appropriate covariates specific to the study (e.g. genotype-based principal components). Studies with unrelated individuals tested association under a linear regression framework. Studies with related samples used variance component or other linear mixed effects modeling to account for relatedness in the regression. The uncertainty of the imputed genotypes was taken into account in the association analysis using methods appropriate for the imputation software used.

The genomic control (GC) inflation factor was calculated for each of the GWA scans separately. The average GC inflation factor was 1.03 (**Supplementary Methods Table 2**). Genomic control correction was applied to results for each study prior to meta-analysis by multiplying SNP standard errors by the square root of the inflation factor.

1.4 Quality control checks of individual studies

Where applicable, the Stage 1 studies calculated separate summary GWA data in males and females and disease cases and controls. Except for studies with related individuals, we used the sex-specific summary results. Each file going into meta-analysis had the following information (columns): SNP, strand, N (sample size), effect allele (allele to which regression coefficient refers), other allele, EAF (effect allele frequency), imputation (posterior probability of imputed genotype, available from some programs), information type (imputation software used), information (imputation quality scores), *P*-value, beta (regression coefficient), standard error, and NxMAF (sample size multiplied by minor allele frequency).

Each file was processed through a cleaning script that performed several quality checks, including calculating the number of markers, ranges of test statistics, the genomic correction inflation factor, and NxMAF. From each study we excluded monomorphic SNPs and SNPs with poor imputation quality: rsq_hat < 0.3 (BIMBAM and MACH) or proper info < 0.4 (IMPUTE).

1.5 Meta-analysis of GWA studies

A total of 2,836,010 autosomal SNPs were meta-analyzed across 98 input files (many of the 50 cohorts had separate male-female and/or case-control files). We did not apply a minor allele frequency cut-off, but we did apply an arbitrary cut-off of NxMAF > 3 (equivalent to a minor allele count of 6) to guard against extremely rare variants present in only one or two samples (possible genotyping/imputation errors or private mutations), for which regression coefficients are not estimated well using the standard statistical methods employed in most GWA statistical programs.

We used the inverse-variance fixed effects meta-analysis method to combine the results from the individual studies. For comparison purposes, we also performed a sample size weighted *Z*-score-based fixed effects meta-analysis. The correlation coefficient between the $\log_{10}$ of the *P*-values of the inverse variance and sample size weighted meta-analysis was 0.99. SNP selection for follow-up was based on the meta-analysis of the inverse variance meta-analysis results. Meta-analyses were performed using the software program METAL (www.sph.umich.edu/csg/abecasis/metal).

1.6 Overall genomic control correction

After genomic control applied in each study the overall genomic control inflation factor (λ_{GC}) for the meta-analysis was 1.42. The possibility that such high inflation is due to effects of population stratification or genotyping biases alone is unlikely, considering the different results presented in Supplementary section 4 which argue against this. In an attempt to identify other sources for such inflation,

we removed all SNPs within 1Mb from the leading SNP in loci with SNPs reaching $P<5x10^{-8}$, which yielded a similarly high $\lambda_{GC}=1.33$. Next, in a simulated phenotype dataset we evaluated the potential role of multiple causal variants failing to reach genome-wide significance. Using a model comprising 120,000 subjects, 294,831 SNPs, and 1000 causal variants, the λ_{GC} increased in a near linear way from 1.15 to 1.32, as heritability (h^2) increased from 0.2 to 0.8. Alternatively, increasing the number of causal variants from 100 to 4000 while keeping heritability constant (at 0.52) increased the λ_{GC} from 1.1 to 1.6 [5]. The latter observed data are consistent with a model containing many causal variants that are in LD with multiple SNPs resulting in inflated test statistics. Although our data imply that a second GC correction on the meta-analysis statistics may be overly conservative, we decided to apply anyway a second genomic control correction to the meta-analysis standard errors and P-values.

1.7 Selection of SNPs for subsequent analyses

SNP selection criteria for validation by genotyping, *in silico* replication, and all additional analyses and simulations, based on the results of Stage 1 GWA meta-analysis, are described below.

2. *In silico* follow-up (Stage 2)

2.1 SNP selection

We took forward for replication 309 SNPs. These included the 207 index SNPs representing each of the 207 2Mb loci reaching $P<5x10^{-6}$ in Stage 1 and 102 SNPs that lie within the same 2Mb windows as the 207, but which were poorly correlated ($r^2<0.05$) with the index SNP in CEU HapMap II samples. No minimum sample size was used for SNPs taken forward for replication, although we note that the minimum N for the 207 variants taken forward from Stage 1 to Stage 2 was 78,550 (for SNP rs11714558 that reached Stage 1+2 *P*-value of $1.7x10^{-10}$). Subsequent analyses are based on the index SNPs from the 207 loci and the 19

SNPs within the 2 Mb windows that were confirmed to be independent by the conditional analysis described below.

2.2 Description of Stage 2 populations

Our *in silico* replication (Stage 2) included 50,074 individuals (12,651 males, 37,423 females) from 15 additional GWA studies. Approximately 26% of the subjects were male (one large study was entirely female), and ages ranged from 17 to 113 years (all but 5 individuals were $\geq$18 years old). Brief study descriptions, details on sample quality control, genotyping and imputation methods, and descriptive statistics, are provided in **Supplementary Methods Tables 1-3**.

2.3 Quality control checks of individual studies in Stage 2

The Stage 2 studies provided the same summary GWA statistics as Stage 1 studies, but only for the requested 309 SNPs. In addition to the QC checks performed in the stage 1 studies (section 1.4 above), we checked the direction of effects for the 309 SNPs in replication studies compared to the overall effects in the Stage 1 meta-analysis. In only one of the cohorts (Sorbs), fewer than 50% of the SNPs had effects in the same direction (47% for males, N=371; 50% for females, N=536). As expected, the largest study showed greatest consistency with the stage 1 meta-analysis results: 98% of SNPs in the same direction in the WGHS, N=32,099). We meta-analyzed these studies in METAL assuming a fixed effects model. When we examined the heterogeneity between Stage 2 studies, only one SNP (rs7567288) had a heterogeneity *P*-value smaller that that expected by chance ($P_{het} = 5.6 \times 10^{-6}$) (see **Supplementary Table 1**).

3. Meta-analysis of Stage 1 and Stage 2

The overall meta-analysis combined Stage 1 and Stage 2 results for the 309 SNPs using a fixed effects model. No SNP showed evidence of heterogeneity between Stage 1 and Stage 2 after accounting for the number of tests performed,

and only a single SNP had the opposite direction of effect in Stage 2 compared to Stage 1.

4. Validation analyses for genotyping and population stratification

4.1 Imputation validation

To validate genotype imputation, we directly genotyped 27 height associated SNPs from the 207 loci in 492 subjects from the WTCCC-T2D study. These 27 SNPs were tested because they were not present on any of the most commonly used arrays (used by >2 studies) and did not have any perfect directly-typed proxies (HapMap r^2=1). We also genotyped a random subset of 18 additional height associated SNPs from these 207 loci in the same samples. Genotyping was performed by Kbioscience (Herts, UK) using a KASPar-based singleplex assay (details of which are available on their website www.kbioscience.co.uk/chemistry/chemistry_Kasp_intro.htm). Forty-three SNPs passed genotyping quality control (HWE *P*>0.01; genotype success rate > 0.9; duplicate error rate <0.5%). We assessed imputation quality by determining the correlation between the directly ascertained genotypes and the genotype dosages produced by IMPUTE (the imputation program used in WTCCC-T2D). We then compared the observed R^2 to the proper_info statistic produced by IMPUTE (which is essentially a predicted R^2 between imputed genotype and actual genotype). The correlation between the predicted and the observed R^2 was high for both the random set of SNPs (*r*=0.92) and for SNPs that were not well captured (*r*=0.84). This suggests that imputation uncertainty has been appropriately accounted for in our analyses.

4.2 Direct genotyping in subjects from tails of height distribution

For additional validation, we genotyped randomly chosen SNPs representing 33 of the 207 associated loci in an independent samples of 2,181 European-American and 1,009 Polish subjects from the tails of the height distribution (5-10th and 90-95th percentile)[6]. These height case-control samples and the genotyping methods have been described previously[6]. For both panels, all

individuals were self-described "white" or "Caucasian." For the US panel, all subjects were born in the US, and all of their grandparents were born in either the US or Europe. All subjects in the Polish panel were born in Poland, and all grandparents were born in Europe or Russia. All subjects gave informed consent, and approval was obtained from the Institutional Review Board of Children's Hospital, Boston. Statistical analysis was performed using a Cochran-Mantel-Haenszel test, as implemented in PLINK[7]. The data set was stratified according to the country of origin of the grandparents to account for population stratification within the European American height panel[6].

Power to replicate the direction of effect of the top 180 height SNPs in the extreme height panel was calculated using the Genetic Power Calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/) based on the following assumptions: a sample size of 3,190 equally divided between individuals in the lower tail (5^{th}-10^{th} percentile) and the upper tail (90^{th}-95^{th} percentile) of the height distribution, variance explained between 0.005-0.3% of the height variation (consistent with our effect size estimates of Stage 2 data, using the equation from quantitative genetics $\sigma^2_g = 2*p*q*\alpha^2$, where σ^2_g is the additive genetic variance, p and q are the allele frequencies, and α is the effect size in SD units), and 3 different minor allele frequencies. Under these assumptions, power is minimally affected by minor allele frequency.

Variance explained	MAF=5%	MAF=25%	MAF=50%
0.005%	56%	56%	56%
0.01%	62%	62%	62%
0.05%	88%	88%	88%
0.1%	97%	97%	97%
0.2%	>99%	>99%	>99%
0.3%	>99%	>99%	>99%

4.3 Family-based association analyses

The family-based analysis performed to assess the influence of population stratification as a potential source of false positive associations in the discovered 180 loci, comprised the Framingham Heart (FramHS) and the Erasmus Rucphen Family-based (ERF) studies. The design of the studies has been described elsewhere[8,9]. The family-based analyses was performed in FramHS (n=5,510) using the QFAM --within procedure from PLINK, running 100,000 permutations to account for the dependence between related individuals. Effect sizes and directions in FramHS were the betas reported by PLINK from the within component but p values were empirical, based on the permutation testing. The extended pedigree of ERF was broken into nuclear families (totaling 1,826 individuals) and analyzed with FBAT[10] which uses a linear combination of offspring genotypes and traits to determine the test statistic. For imputed SNPs, only those with MACH rsq_hat>0.3 were analyzed, using the best guess genotypes from dosages (for FramHS, directly genotyped proxies were also analyzed for comparison and gave similar results). *P*-values were meta-analyzed using a weighted Z-score–based meta-analysis implemented in METAL; if data were only available from one study, the *P*-value from that study was used. Weights were defined based on effective sample size (actual sample size/lambda, where lambda is the genomic control inflation factor calculated from the GWA data of the family-based samples when ignoring relatedness). The direction of the effect allele in the FHS/ERF meta-analysis was compared to that observed in the GIANT meta-analysis using an exact sign-test statistic based on a binomial distribution. The average estimated effect sizes were essentially identical in the GIANT meta-analysis and the FramHS family-based sample (**Supplementary Table 3**), suggesting that there is minimal if any inflation of the GIANT effect sizes due to stratification.

4.4 Other population stratification analyses

We checked if the 180 height-associated variants included ancestry informative marker (AIM) SNPs previously identified as highly informative of the

sub-structure in European populations. We tested the correlations between height loci and 683 AIMs from 3 different sources[11,12,13]. These included AIMs from both the HLA and lactase loci. The largest HapMap CEU r^2 correlations between height SNPs and AIMs were observed for the *GDF5* and *EFEMP1* variants (r^2 = 0.3 and 0.35, respectively). All other pairwise correlations, including those at the HLA locus, had r^2<0.2.

We also assessed the absolute values of EIGENSTRAT[14] loading scores along the principal component of ancestry that corresponds roughly to the North-South intra-European axis that is correlated with height[6] (absolute values of loading scores are a measure of allele frequency differentiation along this axis). The absolute values of loading scores for the height-associated SNPs was not significantly greater than those of 1,000 sets of allele frequency matched SNPs (P=0.08). We also compared F_{st} values (a measure of the proportion of genetic diversity due to differences among populations) for the 180 SNPs with sets of matched SNPs and the F_{st} values of the height-associated SNPs were not different when calculated by cohort, although they were nominally significantly higher (P=0.04) when grouped by country. Together with the family-based analysis, these results strongly suggest that the observed associations with height are unlikely to be appreciably affected by population stratification.

5. Percentage variation explained and number of loci

5.1 Estimation of variance explained and polygene analysis

We used a method recently proposed by the International Schizophrenia Consortium[15] to evaluate the amount of phenotypic variance explained by our associated loci in an independent validation set including the Fingesture (Finland), RS-II & RS-III (Netherlands), GOOD (Sweden) and QIMR (Australia or UK origin) studies. To avoid the influence of potential cryptic relatedness between discovery and validation set, a “leave one out” analysis was performed, namely excluding in the discovery set in turn, all studies from one of the four European countries of the validation set (Finland, Netherlands, Sweden and UK).

The method followed three steps: 1) selection of markers to build a prediction model, 2) scoring each individual based on model and 3) estimation of variation explained using the scores as predictor.

First, we re-ran the meta-analysis using the "leave one out" approach and selected the SNPs that were genotyped in each validation study. For each of these four meta-analysis, a list of independent SNPs associated with height at various *P*-value thresholds (from $P<5\times10^{-8}$ to $P<0.05$) was computed (using the clumping procedure implemented in PLINK, with an LD-based threshold of $r^2\geq0.05$, and a physical distance of 1 Mb from the top hit).

Second, using the selected SNPs from the revised meta-analyses described above, we performed profile scoring for each individual of the five validation studies as implemented in PLINK, where:

$Score_i=\sum_{j=1 \text{ to } m} b_j x_{ij}$, where
m= number of SNPs
b_j =effect of allele at locus *j*
x_{ij}=number of reference alleles of individual *i* at locus *j*

Third, the measure of variance explained (adjusted R^2) is estimated from a linear regression model incorporating the score as the predictor and the age-adjusted standardized height residuals as outcome.

This approach was applied for the estimation of variance explained by the 43 previously published loci, the discovered 180 genome-wide significant loci and the polygene results incorporating different sets of markers at different significance thresholds.

5.2 Prediction of number of susceptibility loci

We utilized a new method (Park et *al*, submitted) to estimate the number of susceptibility loci that are likely to exist based on the distribution of effect sizes observed for established height loci and the power to detect those effects in the original scan. To be conservative and obtain unbiased estimates of the effect sizes, we only utilized the 118 loci that reach a significance threshold of $P<5x10^{-8}$

in the Stage 1 meta-analysis, and the Stage 2 replication data was used to estimate the effect sizes for these loci. Power was calculated based on the sample size for Stage 1 accounting for the number of SNPs that could be identified with the particular effect size. Only SNPs that had a power of at least 1% were used in the predication. The phenotypic variance explained was estimated by summing the product of each effect size and the number of loci predicted with that particular effect size. The genetic variance explained was estimated assuming heredity accounts for 80% of the variance in height. A parametric bootstrap method was used to obtain an estimate of the variability of the estimated number of loci.

6. Gene by gene (GxG) interaction, dominant and recessive analyses

6.1 Associated loci analyses

To perform the GxG, dominant and recessive analyses for just the associated loci, each individual study extracted genotype imputation dosages for each of the 207 lead SNPs from the Stage 1 meta-analysis (based on 2Mb distance pruning; $P<5x10^{-6}$). These dosages were also used for the conditional analysis described below.

An R-script (available on request) was provided to each individual study and was run using the extracted dosages. The allele coding was such that the height increasing allele (based on the Stage 1 meta-analysis) was always the dosage increasing allele (*i.e.* the height increasing allele was coded as allele 2). For the additive dosage and pairwise interaction (Y=b0+b1.A+b2.B+b3.AB+e; Test of b3 = 0) analyses, the dosages were then regressed against residuals of sex-standardized Z-score height, adjusted for age and appropriate covariates (e.g. principal components), as with the primary GWA study, under the appropriate models. For the additive best-guess (performed for quality control purposes), recessive, dominant and dominance deviation analyses “best guess” genotypes were assigned based on genotype dosage, and these genotypes were similarly regressed against Z-score height under the appropriate model.

We meta-analysed individual study results using METAL. We performed meta-analyses for the additive, dominant, recessive, dominant deviation and pairwise interaction terms. We excluded SNPs from individual studies where NxMAF < 10 and/or imputation quality was < 0.4. We also re-ran the meta-analyses excluding SNPs with a NxMAF < 30 and imputation quality < 0.9, because deviation from additivity is harder to detect if the genotype has not been accurately imputed. The results were essentially the same. As an additional quality check we compared the additive dosage and additive best guess results from this meta-analysis to that from obtained from the primary Stage 1 meta-analysis files, and the correlation were very high (r>0.99). Results for the single SNP models, and the top results from the GxG interaction analysis are presented in **Supplementary Table 5 and 6.**

6.2 Genome-wide joint effect analysis

For the genome-wide analysis we used 10,618 individuals from four WTCCC studies (T2D, CAD, HT, NBS) and the EPIC-obesity study where we had access to individual level genotype data and study and sex-standardised, age-adjusted height Z-scores. All the studies were genotyped using the Affymetrix 500K platform (Affymetrix, Inc., Santa Clara, CA, USA). After quality control (including genotype success rate >95%; MAF>1% and HWE *P*>0.0001), 343,249 autosomal SNPs were used in the analysis.

As a genome-wide pairwise interaction analysis was not computationally feasible we performed two separate analyses. First, we performed a pairwise analysis of all SNPs with individual SNP *P*<0.01 with each other (Y=b0+b1.A+b2.B+b3.AB+e; Test of b3 = 0). Second, we performed a genome-wide pairwise analysis testing the full model (an 8 d.f. model). SNP pairs generated here will include those driven by main effects as well as interaction. Therefore, we removed the 9 strongest single SNPs which accounted for a large fraction of the associated pairs, and assessed additive by additive interaction of the remaining pairs with a joint effects $P<1x10^{-8}$ using PLINK. A total of 371 pairs of SNPs with an additive by additive interaction $P<1x10^{-5}$ were taken forward into

replication in 16,100 samples from 4 cohorts, 3 of which (Rotterdam, CoLaus, DGI) were genotyped on the Affymetrix 500K platform (Affymetrix, Inc., Santa Clara, CA, USA). The fourth replication study, CGEMS, was genotyped on the Illumina platform (Illumina, Inc., San Diego, CA, USA), and where a SNP was not available an r^2 > 0.8 proxy was used. Of the 371 SNP pairs that were taken forward into replication, none showed strong evidence of replication (top $P_{Replication}$ = 0.01; top $P_{Overall} = 1x10^{-6}$; N~26,000).

7. Conditional analyses

To perform the conditional analysis, each individual study repeated the Stage 1 GWAS analysis, but included a set of 225 imputation dosages as covariates (those from the 180 SNPs representing the novel loci, plus 27 SNPs from the remaining loci reaching $P<5x10^{-6}$ in stage 1 and an additional 18 SNPs with $P<8x10^{-6}$). For quality control purposes, the files obtained from each of the individual studies were put through the same checks as for the Stage 1 analysis (described in section 1.4). Additional checks were performed to ensure that each of the 225 conditioned SNPs was no longer associated with height (all $P>0.2$) and that SNPs outside the 225 conditioned loci had similar *P*-values and effect sizes to the primary stage 1 analysis. Meta-analysis was performed in the same way as for the GWA studies in Stage 1 (including a NxMAF>3 cut-off and double GC correction).

8. Functional variant analyses

8.1 Non-synonymous enrichment analysis

For all 180 height SNPs, we retrieved all proxy SNPs in linkage disequilibrium ($r^2 \geq 0.8$ in HapMap phase II CEU) and annotated them according to whether they were missense, nonsense or neither. Annotation was based on the NCBI build 36.1. In total for the 180 height SNPs, we identified 2,550 proxies, including 0 nonsense and 31 missense SNPs. We repeated this analysis using 1,000 sets of 180 SNPs that were matched based on allele frequency (± 2.0%), nearby number of genes (± 10% of seed SNP count), and gene proximity (± 20kb).

Among these sets, the ranges for the number of proxies, nonsense SNPs, and missense SNPs were, respectively, 2566-4640, 0-1, and 8-49. After accounting for the number of proxy SNPs in each set, there were only four sets with a ratio (number of nonsynonymous SNPs / total number of proxies) equal or above the ratio observed for the 180 height SNPs (ratio: 0.0122, range of ratios observed in matched sets: 0.0024-0.0133). Similar results were obtained using a logistic regression framework, where control SNPs were matched only on allele frequency but the other matching parameters were used as covariates; here the "exposure" is being a height-associated SNP and the "outcome" is having a missense SNP as a proxy.

8.2 Association with CNVs

We used two different inventories of CNV that are in strong LD (HapMap CEU r^2>0.8) with a near-by non-overlapping HapMap SNP. There a total of 1330 unique CNV tags, 1138 from the Conrad *et al.* GSV data set[16] and 261 from analysis of the HapMap 2 data by McCarroll *et al.*[17]. We looked to identify those CNV-tagging SNPs that were highly correlated with one of the 199 index SNPs representing 180 confirmed height loci, but there was no such overlap between our and the CNV datasets. We then looked at the overlap between the CNV-tagging SNPs and our genome-wide Stage 1 results with $P<5x10^{-8}$. The only overlap occurred at *DLEU7* and chromosome 4 (near *HHIP*) loci; however, in both cases the height index SNP had much better *P*-value than the CNV-tagging SNP, and the index SNP was only either in weak LD (at *DLEU7* locus, r^2=0.56) or not at all correlated (at chromosome 4 locus) with the CNV tag.

8.3 Association with other traits

We downloaded from the NHGRI GWA study catalogue (http://www.genome.gov/26525384; accessed on 12th February 2010) all SNPs associated with diseases and traits other than height at genome-wide significance level of $P<5x10^{-8}$. We then identified all SNPs that mapped within 1Mb of at least one height SNP and had some correlation (HapMap CEU r^2>0.1) with the index

height SNP for each of the 180 associated loci. There were 22 such overlapping loci, some associated with multiple other traits and diseases (**Supplementary Table 8**). At 6 of the loci the height and 'other' trait SNP were either identical or strongly correlated (r^2>0.8). For one of these loci, *LIN28B,* the height effect is likely to be secondary to the large effect on pubertal timing, but the remaining five are likely to represent true pleiotropic effects.

9. Biological enrichment analyses

9.1 OMIM analysis

We searched the Online Mendelian Inheritance in Man (OMIM) database and identified 241 genes that underlie human syndromes characterized by abnormal skeletal growth (**Supplementary Table 9**). The gene list was initially obtained using search keywords 'short stature', 'overgrowth', 'skeletal dysplasia', and 'brachydactyly', and was manually curated blindly to our results. We then grouped the 180 height-associated SNPs into 175 non-overlapping gene regions (to avoid double counting), containing a total of 652 genes. For each region, we set the genomic boundaries using linkage disequilibrium cutoffs (r^2≤0.3 from the index height SNP) and then next recombination hotspots. Although these 175 regions contained only ~3.3% of all human genes, they included 21 genes from the curated OMIM height gene list (8.7%). We assessed the significance of this result by permutation: we generated 1,000 sets of 175 regions with similar gene content (±10%) and counted, in each set, the number of OMIM height genes within the regions. In these 1,000 permutations, the median number of OMIM height genes was 8 and the range was 1-19 (empirical *P*-value for an overlap of 21 OMIM genes is *P*<0.001).

9.2 Text-mining using GRAIL

The GRAIL algorithm was recently described[18]. As in the OMIM analysis, we used LD and recombination hotspots to define boundaries on the left and right of each height index SNP. This identified 652 genes in 175 regions (five regions

were overlapping when using our criteria to define genomic interval around height index SNP).

9.3 Pathway analysis

We applied an adaptation of the gene set enrichment analysis (GSEA) framework (Meta-Analysis Gene-set Enrichment of variaNT Associations, MAGENTA (Segrè et *al*., submitted)) to the height meta-analysis to determine whether the 180 height SNPs cluster near genes that belong to specific biological pathways and potentially to discover new pathways that may be enriched for modest height associations not yet identified. Specifically, for each gene in the genome we calculated a corrected gene association *P*-value based on the most significant SNP height association *P*-value of all SNPs in the gene region (110 kb upstream and 40 kb downstream to gene's most extreme transcript start and end sites, respectively), accounting for confounding effects such as gene size, number of SNPs per gene and linkage-related properties. Genes were grouped into pathways using annotations from the KEGG, PANTHER, and INGENUITY databases, downloaded from the Molecular Signatures Database (MsigDB, http://www.broad.mit.edu/gsea/msigdb/collections.jsp). Molecular function gene-sets were downloaded from the PANTHER website (http://www.pantherdb.org/). For each pathway, enrichment of highly ranked gene scores above the 95th percentile of all gene scores in the height meta-analysis, was evaluated compared to 10,000 randomly sampled gene sets of identical size from the genome. Results from this analysis show strong enrichment for genes that belong to the hedgehog signalling pathway (nominal GSEA *P*=0.0009, FDR=0.078) and the histone molecular function category (nominal GSEA *P*=0.0001, FDR=0.0028), many of which are near the top GIANT height SNPs. In total, there were 20 pathways, including the TGF-beta pathway, and 14 molecular function sets that were nominally significant (*P*=0.05) in our GSEA using MAGENTA (**Supplementary Table 11**).

10. URLs

Bayesian Imputation Based Association Mapping, BIMBAM, http://quartus.uchicago.edu/~yguan/bimbam/index.html;

population stratification detection software, EIGENSTRAT, http://genepath.med.harvard.edu/~reich/EIGENSTRAT.htm;

genotype imputation program, IMPUTE, http://www.stats.ox.ac.uk/~marchini/software/gwas/impute.html;

Markov chain haplotyping package, MACH, http://www.sph.umich.edu/csg/abecasis/MACH; MACH2QTL, http://www.sph.umich.edu/csg/abecasis/MACH/download;

pedigree analysis package, MERLIN, http://www.sph.umich.edu/csg/abecasis/Merlin;

meta-analysis tool for GWASs, METAL, http://www.sph.umich.edu/csg/abecasis/Metal/index.html;

whole-genome association analysis package, PLINK, http://pngu.mgh.harvard.edu/~purcell/plink;

whole-genome association analysis of imputed data, ProbABEL, http://mga.bionet.nsc.ru/~yurii/ABEL;

statistical computer software, R, http://www.r-project.org;

whole-genome association analysis package, SNPTEST, http://www.stats.ox.ac.uk/~marchini/software/gwas/snptest.html.

References

1. Scheet, P. & Stephens, M. A Fast and Flexible Statistical Model for Large-Scale Population Genotype Data: Applications to Inferring Missing Genotypes and Haplotypic Phase. *The American Journal of Human Genetics* 78, 629 (2006).

2. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 39, 906 (2007).

3. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype Imputation. *Annu. Rev. Genomics Hum. Genet.* 10, 387-406 (2009).

4. Frazer, K. A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449, 851-861 (2007).

5. Yang, J. *et al.* Genomic Control under Polygenic Inheritance. submitted.

6. Campbell, C. D. *et al.* Demonstrating stratification in a European-American population. *Nature Genet* 37, 868-872 (2005).

7. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81, 559-575 (2007).

8. Aulchenko, Y.S. *et al.* Linkage disequilibrium in young genetically isolated Dutch population. *Eur J Hum Genet* 12, 527-534 (2004).

9. Splansky, G.L. *et al*. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol*.165,1328-1335 (2007).

10. Nan, M. L., Steve, H. & Xin, X. Implementing a unified approach to family-based tests of association. *Genetic Epidemiology* 19, S36-S42 (2000).

11. Burton, P. R. *et al.* Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661-678 (2007).

12. Price, A. L. *et al.* Discerning the ancestry of European Americans in genetic association studies. *Plos Genetics* 4 (2008).

13. Tian, C. *et al.* Analysis and application of European genetic substructure using 300 KSNP information. *Plos Genetics* 4 (2008).

14. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genet* 38, 904-909 (2006).

15. Purcell, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460, 748-752 (2009).

16. Conrad, D. F. *et al.* Origins and functional impact of copy number variation in the human genome. *Nature* 464, 704-716 (2010).

17. McCarroll, S. A. *et al.* Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat Genet* 40, 1166-1174 (2008).

18. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet* 5, e1000534 (2009).

Supplementary Tables

Supplementary Table 1: Meta-analysis results

Supplementary Table 2: Extreme height association results

Supplementary Table 3: Family-based analyses results

Supplementary Table 4: Estimated total number of associated loci

Supplementary Table 5: Dominant/recessive results

Supplementary Table 6: Gene x Gene interaction results

Supplementary Table 7: Overlap with nsSNPs

Supplementary Table 8: Overlap with signals from GWAS of other traits and diseases

Supplementary Table 9: List of OMIM height genes

Supplementary Table 10: Overlap with OMIM height genes

Supplementary Table 11: Gene Set Enrichment Analysis results

Supplementary Methods Table 1: Study design, sample QC and references

Supplementary Methods Table 2: Genotyping, imputation and analytical details

Supplementary Methods Table 3: Descriptive statistics

Supplementary Table 1. Association results for Stage 1 (discovery GWAS), Stage 2 (in-silico replication), Stage 1+2 combined, and Stage 1+2 sex-specific meta-analyses, for the 180 independent signals that reached genome-wide significance ($P<5x10^{-8}$) in the combined Stage 1+2 analysis. I^2 represents the % heterogeneity of effect size between Stage 1 studies. P_{het} is the heterogeneity *P*-value.

SNP [a]	Chr	Position (bp)	Nearest/OMIM height gene [b]	Effect /other allele [c]	Frequency (effect allele)	STAGE 1 up to 133,653 samples				STAGE 2 up to 50,074 samples		STAGE 1 + STAGE 2 up to 183,727 samples		STAGE 1 + STAGE 2 SEX-SPECIFIC up to 73,238 males and 110,489 females				
						Beta	*P*-value [d]	I^2	P_{het}	Beta	P-value [d]	Beta	*P*-value [d]	Beta (M)	*P*-value [d] (M)	Beta (F)	*P*-value [d] (F)	P_{het} (MvsF)
rs425277	1	2059032	***PRKCZ***	T/C	0.28	0.024	1.70E-06	0	0.73	0.019	3.10E-03	0.022	2.10E-08	0.017	5.90E-03	0.027	6.70E-08	0.15
rs2284746	1	17179262	***MFAP2***	C/G	0.48	-0.035	5.60E-15	17.77	0.07	-0.049	2.50E-16	-0.04	3.90E-29	-0.041	1.60E-13	-0.039	1.80E-18	0.76
rs1738475	1	23409478	***HTR1D***	C/G	0.59	0.022	1.90E-06	0	0.69	0.031	1.60E-07	0.025	3.00E-12	0.02	2.80E-04	0.028	5.20E-10	0.25
rs4601530	1	24916698	***CLIC4***	T/C	0.26	-0.024	2.00E-06	15.60	0.10	-0.036	1.10E-07	-0.028	2.20E-12	-0.03	6.50E-07	-0.025	5.20E-07	0.47
rs7532866	1	26614131	***LIN28***	A/G	0.67	0.022	3.30E-06	0	0.54	0.02	2.60E-03	0.021	3.40E-08	0.017	4.30E-03	0.025	1.30E-07	0.23
rs2154319	1	41518357	***SCMH1***	T/C	0.75	-0.034	4.30E-10	0	0.86	-0.025	4.90E-04	-0.03	1.80E-12	-0.024	2.70E-04	-0.035	4.60E-11	0.13
rs17391694	1	78396214	***GIPC2***	T/C	0.12	0.04	5.90E-07	7.76	0.27	0.045	5.60E-06	0.042	1.70E-11	0.041	7.00E-06	0.042	5.90E-08	0.95
rs6699417	1	88896031	***PKN2***	T/C	0.61	0.022	1.70E-06	0	0.89	0.02	8.60E-04	0.021	5.00E-09	0.02	3.10E-04	0.02	6.40E-06	0.99
rs10874746	1	93096559	***RPL5***	T/C	0.37	-0.022	1.70E-06	0	0.55	-0.027	7.90E-06	-0.024	6.70E-11	-0.024	1.30E-05	-0.022	7.30E-07	0.78
rs9428104	1	118657110	***SPAG17***	A/G	0.24	-0.038	8.90E-13	0	0.98	-0.048	6.40E-12	-0.041	5.60E-23	-0.039	9.10E-10	-0.043	4.10E-17	0.55
rs11205277	1	148159496	**SF3B4**	A/G	0.58	-0.045	1.20E-18	0.02	0.48	-0.048	8.10E-15	-0.046	4.80E-32	-0.042	9.60E-12	-0.049	2.00E-24	0.36
rs17346452	1	170319910	***DNM3***	T/C	0.73	-0.038	3.30E-14	0	0.79	-0.045	4.00E-11	-0.04	1.40E-23	-0.037	1.10E-09	-0.042	6.60E-17	0.56
rs1325598	1	175058872	***PAPPA2***	A/G	0.43	-0.026	1.60E-08	0	0.88	-0.016	9.60E-03	-0.022	1.10E-09	-0.025	4.10E-06	-0.021	2.70E-06	0.52
rs1046934	1	182290152	***TSEN15***	A/C	0.64	-0.046	6.40E-22	0	0.80	-0.042	2.30E-11	-0.044	2.10E-31	-0.043	8.60E-14	-0.044	1.10E-20	0.94
rs10863936	1	210304421	***DTL***	A/G	0.53	-0.022	6.20E-07	3.05	0.40	-0.02	8.40E-04	-0.021	1.90E-09	-0.029	5.40E-08	-0.017	1.10E-04	0.06
rs6684205	1	216676325	***TGFB2***	A/G	0.71	-0.033	2.00E-11	0	0.61	-0.019	4.00E-03	-0.028	1.50E-12	-0.032	7.20E-08	-0.026	8.50E-08	0.41
rs11118346	1	217810342	***LYPLAL1***	T/C	0.47	-0.026	2.20E-09	9.57	0.22	-0.023	2.00E-04	-0.025	1.90E-12	-0.018	9.50E-04	-0.03	3.10E-11	0.05
rs10799445	1	225978506	***JMJD4***	A/C	0.77	0.031	1.20E-08	0	0.51	0.033	2.80E-06	0.032	2.40E-13	0.026	4.00E-05	0.036	7.10E-12	0.21
rs4665736	2	25041103	***DNAJC27***	T/C	0.54	0.034	1.40E-13	0	0.97	0.021	4.30E-04	0.029	7.30E-16	0.022	5.30E-05	0.034	3.40E-14	0.08
rs6714546	2	33214929	***LTBP1***	A/G	0.28	-0.025	2.20E-06	0	0.99	-0.026	1.70E-04	-0.026	1.60E-09	-0.019	3.40E-03	-0.029	2.70E-08	0.19
rs17511102	2	37814117	***CDC42EP3***	A/T	0.91	-0.06	1.30E-12	0	0.67	-0.061	1.70E-07	-0.06	1.60E-18	-0.061	1.80E-09	-0.06	1.20E-12	0.9
rs2341459	2	44621706	***C2orf34***	T/C	0.27	0.028	3.60E-08	0	0.75	0.02	4.40E-03	0.025	7.90E-10	0.031	2.40E-07	0.021	4.70E-05	0.14
rs12474201	2	46774789	***SOCS5***	A/G	0.35	0.023	1.00E-06	0	0.62	0.036	1.00E-08	0.028	2.60E-13	0.026	6.10E-06	0.028	2.90E-09	0.78
rs3791675	2	55964813	***EFEMP1***	T/C	0.23	-0.05	2.40E-20	22.09	0.03	-0.059	3.20E-17	-0.053	2.50E-35	-0.055	1.20E-17	-0.052	3.60E-23	0.71
rs11684404	2	88705737	***EIF2AK3***	T/C	0.67	-0.027	6.40E-09	14.78	0.12	-0.029	2.60E-06	-0.028	9.90E-14	-0.03	7.10E-08	-0.025	4.20E-08	0.46

SNP [a]	Chr	Position (bp)	Nearest/OMIM height gene [b]	Effect /other allele [c]	Frequency (effect allele)	STAGE 1 up to 133,653 samples				STAGE 2 up to 50,074 samples		STAGE 1 + STAGE 2 up to 183,727 samples		STAGE 1 + STAGE 2 SEX-SPECIFIC up to 73,238 males and 110,489 females				
						Beta	*P*-value [d]	I^2	P_{het}	Beta	P-value [d]	Beta	*P*-value [d]	Beta (M)	*P*-value [d] (M)	Beta (F)	*P*-value [d] (F)	P_{het} (MvsF)
rs7567288	2	134151294	***NCKAP5***	T/C	0.8	-0.031	6.70E-08	0	0.92	-0.033	8.40E-06	-0.032	2.10E-12	-0.029	2.10E-05	-0.033	4.10E-09	0.6
rs7567851	2	178392966	***PDE11A***	C/G	0.08	0.041	7.50E-07	25.21	0.01	0.028	9.50E-03	0.037	3.30E-08	0.033	8.20E-04	0.038	3.60E-06	0.7
rs1351164	2	217980143	***TNS1***	T/C	0.79	0.028	3.70E-07	0	0.87	0.044	2.70E-09	0.034	2.10E-14	0.033	4.30E-07	0.032	5.90E-09	0.83
rs12470505	2	219616613	***CCDC108/IHH***	T/G	0.9	0.048	1.30E-10	0	0.67	0.028	5.80E-03	0.041	8.90E-12	0.059	1.40E-10	0.032	2.50E-05	0.01
rs2629046	2	224755988	***SERPINE2***	T/C	0.55	0.025	2.20E-08	0	0.89	0.023	1.00E-04	0.024	7.90E-12	0.019	3.80E-04	0.027	7.20E-10	0.2
rs2580816	2	232506210	***NPPC***	T/C	0.19	-0.041	1.80E-12	0	0.80	-0.051	4.60E-11	-0.045	5.80E-22	-0.05	9.30E-13	-0.041	1.70E-12	0.23
rs12694997	2	241911659	***SEPT2***	A/G	0.24	-0.027	1.80E-07	3.06	0.40	-0.018	1.40E-02	-0.024	1.20E-08	-0.021	1.10E-03	-0.025	1.40E-06	0.61
rs2597513	3	13530836	***HDAC11***	T/C	0.9	-0.039	1.10E-07	9.85	0.22	-0.031	1.40E-03	-0.036	7.40E-10	-0.036	4.90E-05	-0.038	1.10E-07	0.83
rs13088462	3	51046753	***DOCK3***	T/C	0.94	-0.054	3.10E-07	0	0.80	-0.048	2.90E-04	-0.052	3.80E-10	-0.057	4.70E-06	-0.048	2.40E-06	0.56
rs2336725	3	53093779	***RTF1***	T/C	0.55	-0.026	3.50E-08	8.25	0.26	-0.028	5.20E-06	-0.027	9.70E-13	-0.028	1.00E-06	-0.026	1.30E-08	0.85
rs9835332	3	56642722	***C3orf63***	C/G	0.46	-0.022	8.70E-07	8.66	0.25	-0.032	5.70E-08	-0.026	5.30E-13	-0.026	2.10E-06	-0.025	2.10E-08	0.91
rs17806888	3	67499012	***SUCLG2***	T/C	0.88	0.04	1.10E-07	7.76	0.28	0.028	3.70E-03	0.036	2.10E-09	0.036	7.10E-05	0.035	1.20E-06	0.93
rs9863706	3	72520103	***RYBP***	T/C	0.22	-0.03	1.50E-08	0	0.69	-0.033	4.70E-06	-0.031	4.10E-13	-0.034	2.50E-07	-0.03	1.80E-08	0.6
rs6439167	3	130533446	***C3orf47***	T/C	0.21	-0.034	7.20E-10	0	0.89	-0.035	2.40E-06	-0.034	8.90E-15	-0.026	1.10E-04	-0.039	4.80E-13	0.09
rs9844666	3	137456906	***PCCB***	A/G	0.25	-0.028	3.10E-08	0	0.77	-0.017	1.70E-02	-0.024	3.50E-09	-0.016	8.60E-03	-0.029	1.80E-08	0.09
rs724016	3	142588260	***ZBTB38***	A/G	0.56	-0.067	4.50E-52	20.23	0.05	-0.075	2.90E-36	-0.07	3.10E-86	-0.066	8.80E-35	-0.071	5.70E-60	0.42
rs572169	3	173648421	***GHSR***	T/C	0.31	0.036	9.90E-14	3.61	0.38	0.03	3.40E-06	0.033	2.80E-18	0.03	2.80E-07	0.036	4.20E-14	0.4
rs720390	3	187031377	***IGF2BP2***	A/G	0.39	0.031	1.60E-10	19.54	0.05	0.026	1.80E-05	0.029	1.90E-14	0.036	4.40E-10	0.026	3.20E-08	0.14
rs2247341	4	1671115	***SLBP/FGFR3***	A/G	0.36	0.025	6.80E-08	17.58	0.08	0.026	3.80E-05	0.025	1.50E-11	0.027	1.60E-06	0.024	1.80E-07	0.67
rs6449353	4	17642586	***LCORL***	T/C	0.85	0.071	1.30E-27	0	0.69	0.081	2.60E-20	0.075	7.10E-46	0.074	2.10E-21	0.076	3.20E-32	0.88
rs17081935	4	57518233	***POLR2B***	T/C	0.2	0.031	4.80E-08	6.60	0.30	0.028	1.80E-04	0.03	3.70E-11	0.038	1.70E-08	0.025	6.60E-06	0.09
rs7697556	4	73734177	***ADAMTS3***	T/C	0.47	0.022	1.30E-06	0	0.71	0.038	2.90E-10	0.028	2.00E-14	0.03	4.80E-08	0.026	5.00E-09	0.56
rs788867	4	82369030	***PRKG2/BMP3***	T/G	0.68	-0.039	1.80E-15	0	0.52	-0.05	2.10E-14	-0.043	8.90E-28	-0.042	9.00E-13	-0.042	1.60E-18	0.95
rs10010325	4	106325802	***TET2***	A/C	0.49	0.021	2.30E-06	0	0.68	0.028	3.20E-06	0.024	3.90E-11	0.025	2.40E-06	0.022	3.40E-07	0.64
rs7689420	4	145787802	***HHIP***	T/C	0.16	-0.069	1.40E-29	10.51	0.20	-0.08	1.40E-23	-0.073	6.20E-51	-0.07	8.90E-22	-0.075	1.10E-35	0.61
rs955748	4	184452669	***WWC2***	A/G	0.24	-0.024	2.20E-06	0	0.52	-0.019	5.70E-03	-0.023	4.40E-08	-0.027	1.50E-05	-0.019	1.60E-04	0.29
rs1173727	5	32866278	***NPR3***	T/C	0.4	0.036	4.00E-15	1.45	0.44	0.032	1.10E-07	0.034	1.60E-21	0.038	4.60E-12	0.031	3.10E-12	0.27
rs11958779	5	55037656	***SLC38A9***	A/G	0.7	-0.028	8.00E-09	0	0.92	-0.026	4.90E-05	-0.027	1.80E-12	-0.028	1.20E-06	-0.027	2.30E-08	0.8
rs10037512	5	88390431	***MEF2C***	T/C	0.56	0.027	3.80E-09	22.57	0.03	0.04	2.20E-11	0.032	2.00E-18	0.035	1.70E-10	0.029	1.40E-10	0.3
rs13177718	5	108141243	***FER***	T/C	0.07	-0.041	4.10E-06	12.78	0.16	-0.037	2.20E-03	-0.04	3.00E-08	-0.051	2.30E-06	-0.034	1.30E-04	0.16

SNP [a]	Chr	Position (bp)	Nearest/OMIM height gene [b]	Effect /other allele [c]	Frequency (effect allele)	STAGE 1 up to 133,653 samples				STAGE 2 up to 50,074 samples		STAGE 1 + STAGE 2 up to 183,727 samples		STAGE 1 + STAGE 2 SEX-SPECIFIC up to 73,238 males and 110,489 females				
						Beta	*P*-value [d]	I^2	P_{het}	Beta	P-value [d]	Beta	*P*-value [d]	Beta (M)	*P*-value [d] (M)	Beta (F)	*P*-value [d] (F)	P_{het} (MvsF)
rs1582931	5	122685098	***CEP120***	A/G	0.47	-0.025	2.10E-08	0	0.98	-0.019	1.90E-03	-0.023	1.50E-10	-0.019	3.50E-04	-0.026	8.10E-09	0.31
rs274546	5	131727766	***SLC22A5***	A/G	0.4	-0.028	8.50E-10	0	0.92	-0.032	1.50E-07	-0.029	7.30E-16	-0.035	2.00E-10	-0.025	1.50E-08	0.13
rs526896	5	134384604	***PITX1***	T/G	0.73	0.032	1.90E-09	2.86	0.40	0.029	3.20E-05	0.03	2.30E-13	0.024	1.70E-04	0.035	1.70E-11	0.15
rs4282339	5	168188818	***SLIT3***	A/G	0.2	-0.035	3.40E-10	4.07	0.37	-0.038	3.10E-07	-0.036	6.60E-16	-0.034	4.40E-07	-0.037	1.70E-11	0.69
rs12153391	5	171136043	***FBXW11***	A/C	0.25	-0.033	8.70E-10	0	0.83	-0.024	5.20E-04	-0.03	3.60E-12	-0.027	2.10E-05	-0.032	2.00E-09	0.57
rs889014	5	172916720	***BOD1***	T/C	0.36	-0.029	4.50E-10	8.66	0.25	-0.032	2.10E-07	-0.03	9.40E-16	-0.032	9.70E-09	-0.028	8.30E-10	0.51
rs422421	5	176449932	***FGFR4/NSD1***	T/C	0.22	-0.033	1.40E-09	27.96	0.01	-0.028	1.40E-04	-0.031	1.10E-12	-0.03	7.10E-06	-0.034	5.20E-10	0.64
rs6879260	5	179663620	***GFPT2***	T/C	0.39	-0.028	5.60E-10	0	0.79	-0.01	9.70E-02	-0.022	1.60E-09	-0.02	3.40E-04	-0.025	3.00E-08	0.41
rs3812163	6	7670759	***BMP6***	A/T	0.54	-0.037	6.70E-16	23.10	0.03	-0.035	4.30E-09	-0.036	1.20E-23	-0.033	2.80E-09	-0.039	1.50E-18	0.36
rs1047014	6	19949472	***ID4***	T/C	0.75	-0.029	1.10E-07	0	0.55	-0.037	1.80E-07	-0.032	1.80E-13	-0.033	7.80E-07	-0.032	4.10E-09	0.9
rs806794	6	26308656	**Histone cluster**	A/G	0.7	0.053	5.50E-26	22.95	0.03	0.051	4.30E-15	0.052	1.20E-39	0.046	2.50E-14	0.057	5.30E-31	0.12
rs3129109	6	29192211	***OR2J3***	T/C	0.39	-0.026	3.30E-08	16.96	0.09	-0.041	1.60E-11	-0.032	2.40E-17	-0.029	2.60E-07	-0.032	3.30E-12	0.64
rs2256183	6	31488508	***MICA***	A/G	0.45	0.035	2.70E-14	0	0.54	0.051	8.30E-17	0.04	7.80E-29	0.043	4.40E-14	0.037	3.60E-17	0.43
rs6457620	6	32771977	**HLA locus**	C/G	0.51	-0.024	3.60E-08	0	0.98	-0.037	2.50E-10	-0.029	2.10E-16	-0.03	2.50E-08	-0.028	1.00E-10	0.81
rs2780226	6	34307070	***HMGA1***	T/C	0.92	-0.079	1.00E-18	20.61	0.05	-0.072	1.70E-10	-0.076	8.10E-28	-0.077	1.90E-12	-0.076	2.00E-19	0.96
rs6457821	6	35510783	***PPARD/FANCE***	A/C	0.02	-0.121	1.80E-11	3.24	0.40	-0.068	8.00E-03	-0.104	2.10E-12	-0.084	2.20E-04	-0.112	3.40E-10	0.29
rs9472414	6	45054484	***SUPT3H/RUNX2***	A/T	0.22	-0.031	2.40E-08	26.80	0.01	-0.019	8.70E-03	-0.026	1.80E-09	-0.029	6.90E-06	-0.026	1.70E-06	0.66
rs9360921	6	76322362	***SENP6***	T/G	0.89	-0.048	4.60E-11	17.19	0.08	-0.033	5.00E-04	-0.042	2.60E-13	-0.045	1.20E-07	-0.04	1.40E-08	0.62
rs310405	6	81857081	***FAM46A***	A/G	0.52	0.03	3.60E-11	0	0.89	0.02	8.10E-04	0.026	2.20E-13	0.023	2.60E-05	0.03	1.30E-11	0.25
rs7759938	6	105485647	***LIN28B***	T/C	0.68	-0.042	8.70E-18	6.39	0.30	-0.051	4.10E-15	-0.045	8.30E-31	-0.04	8.70E-12	-0.048	5.20E-23	0.26
rs1046943	6	109890634	***ZBTB24***	A/G	0.58	0.022	8.60E-07	0	0.67	0.016	7.20E-03	0.02	2.50E-08	0.024	1.20E-05	0.019	1.90E-05	0.46
rs961764	6	117628849	***VGLL2***	C/G	0.42	-0.023	2.40E-07	0	0.87	-0.026	1.20E-05	-0.024	1.30E-11	-0.024	1.20E-05	-0.025	8.90E-09	0.79
rs1490384	6	126892853	***C6orf173***	T/C	0.5	0.037	3.20E-16	15.83	0.10	0.028	1.80E-06	0.034	3.90E-21	0.037	5.30E-12	0.033	3.40E-14	0.55
rs6569648	6	130390812	***L3MBTL3***	T/C	0.76	-0.036	8.90E-12	16.88	0.08	-0.047	1.20E-11	-0.04	1.10E-21	-0.046	5.10E-13	-0.035	8.40E-12	0.14
rs7763064	6	142838982	***GPR126***	A/G	0.29	-0.045	6.40E-19	6.91	0.29	-0.055	7.20E-17	-0.048	1.10E-33	-0.044	2.10E-13	-0.051	5.30E-26	0.29
rs543650	6	152152636	***ESR1***	T/G	0.4	-0.032	1.40E-09	16.12	0.11	-0.037	2.10E-09	-0.034	1.20E-17	-0.029	3.30E-06	-0.036	1.30E-13	0.36
rs9456307	6	158849430	***TULP4***	A/T	0.06	-0.05	4.60E-07	0.20	0.47	-0.045	1.20E-03	-0.048	2.20E-09	-0.041	7.90E-04	-0.053	6.20E-08	0.38
rs798489	7	2768329	***GNA12***	T/C	0.3	-0.052	8.50E-25	0	0.55	-0.042	1.70E-10	-0.048	1.90E-33	-0.051	4.50E-17	-0.046	5.10E-21	0.53
rs4470914	7	19583047	***TWISTNB***	T/C	0.18	0.033	3.80E-08	5.76	0.32	0.023	3.40E-03	0.029	9.20E-10	0.03	4.80E-05	0.029	5.50E-07	0.93
rs12534093	7	23469499	***IGF2BP3***	A/T	0.22	-0.03	5.60E-08	3.09	0.39	-0.04	4.10E-08	-0.034	2.00E-14	-0.032	1.70E-06	-0.033	5.30E-10	0.84

SNP [a]	Chr	Position (bp)	Nearest/OMIM height gene [b]	Effect /other allele [c]	Frequency (effect allele)	STAGE 1 up to 133,653 samples				STAGE 2 up to 50,074 samples		STAGE 1 + STAGE 2 up to 183,727 samples		STAGE 1 + STAGE 2 SEX-SPECIFIC up to 73,238 males and 110,489 females				
						Beta	*P*-value [d]	I^2	P_{het}	Beta	P-value [d]	Beta	*P*-value [d]	Beta (M)	*P*-value [d] (M)	Beta (F)	*P*-value [d] (F)	P_{het} (MvsF)
rs1708299	7	28156471	***JAZF1***	A/G	0.3	0.042	1.50E-17	14.38	0.12	0.038	5.80E-09	0.04	5.80E-25	0.036	4.10E-10	0.044	3.30E-20	0.25
rs6959212	7	38094851	***STARD3NL***	T/C	0.32	-0.023	2.80E-06	0	0.52	-0.025	1.30E-04	-0.024	1.60E-09	-0.021	3.40E-04	-0.024	5.90E-07	0.66
rs42235	7	92086012	***CDK6***	T/C	0.31	0.055	7.30E-28	21.51	0.04	0.062	1.90E-20	0.057	7.70E-47	0.046	1.60E-14	0.063	3.70E-37	0.01
rs822552	7	148281567	***PDIA4***	C/G	0.74	-0.03	1.30E-07	0	0.48	-0.017	2.70E-02	-0.025	2.60E-08	-0.032	2.30E-06	-0.022	9.40E-05	0.24
rs2110001	7	150147955	***TMEM176A***	C/G	0.69	-0.033	9.80E-10	17.78	0.08	-0.028	4.40E-05	-0.031	3.30E-13	-0.029	3.40E-06	-0.032	4.40E-10	0.71
rs1013209	8	24172249	***ADAM28***	T/C	0.25	-0.029	4.50E-08	10.06	0.21	-0.019	7.30E-03	-0.025	1.60E-09	-0.026	4.80E-05	-0.026	8.90E-07	0.95
rs7460090	8	57356717	***SDR16C5***	T/C	0.87	0.055	9.60E-16	0	0.70	0.064	7.70E-13	0.058	8.20E-27	0.051	7.00E-10	0.064	1.10E-21	0.16
rs6473015	8	78341040	***PEX2***	A/C	0.72	-0.032	1.70E-10	12.48	0.16	-0.023	5.80E-04	-0.029	6.90E-13	-0.03	8.90E-07	-0.028	9.00E-09	0.84
rs6470764	8	130794847	***GSDMC***	T/C	0.2	-0.047	5.90E-17	17.64	0.07	-0.056	3.40E-13	-0.05	1.70E-28	-0.05	1.60E-13	-0.05	4.10E-18	0.95
rs12680655	8	135706519	***ZFAT***	C/G	0.6	0.03	4.80E-11	16.69	0.09	0.024	7.50E-05	0.028	1.60E-14	0.025	5.10E-06	0.029	3.70E-11	0.45
rs7864648	9	16358732	***BNC2***	T/G	0.32	0.025	4.90E-07	3.83	0.37	0.017	7.80E-03	0.022	2.10E-08	0.027	5.40E-06	0.019	9.70E-05	0.23
rs11144688	9	77732106	***PCSK5***	A/G	0.11	-0.055	1.50E-09	0	0.52	-0.04	9.10E-04	-0.049	9.60E-12	-0.044	3.30E-05	-0.057	3.90E-10	0.28
rs7853377	9	85742025	***C9orf64***	A/G	0.77	-0.026	3.10E-06	0	0.65	-0.021	3.50E-03	-0.024	4.50E-08	-0.018	6.10E-03	-0.027	5.00E-07	0.26
rs8181166	9	88306448	***ZCCHC6***	C/G	0.53	0.025	1.10E-07	26.48	0.01	0.028	3.90E-06	0.026	2.70E-12	0.019	8.30E-04	0.031	8.20E-12	0.07
rs2778031	9	90025546	***SPIN1***	T/C	0.24	0.027	3.60E-07	0	0.81	0.037	2.40E-07	0.031	9.00E-13	0.031	1.50E-06	0.029	2.20E-08	0.78
rs9969804	9	94468941	***IPPK***	A/C	0.44	0.028	5.60E-10	0	0.61	0.033	1.90E-08	0.03	7.70E-17	0.028	1.50E-07	0.029	7.30E-11	0.92
rs1257763	9	95933766	***PTPDC1***	A/G	0.04	0.069	2.50E-06	0	0.95	0.07	1.00E-04	0.069	9.90E-10	0.063	2.10E-04	0.075	1.00E-07	0.55
rs473902	9	97296056	***PTCH1/FANCC***	T/G	0.92	0.074	1.70E-14	0	0.61	0.05	6.80E-05	0.065	2.30E-17	0.061	6.70E-08	0.068	9.80E-13	0.62
rs7027110	9	108638867	***ZNF462***	A/G	0.23	0.034	1.30E-10	0	0.85	0.025	3.80E-04	0.031	2.30E-13	0.032	4.80E-07	0.03	8.90E-09	0.72
rs1468758	9	112846903	***LPAR1***	T/C	0.25	-0.026	1.50E-06	0	0.59	-0.026	1.90E-04	-0.026	1.40E-09	-0.031	1.20E-06	-0.022	2.50E-05	0.24
rs751543	9	118162163	***PAPPA***	T/C	0.72	0.029	4.50E-08	0	0.86	0.021	3.40E-03	0.026	6.50E-10	0.027	2.50E-05	0.026	6.70E-07	0.89
rs7466269	9	132453905	***FUBP3***	A/G	0.64	0.036	1.20E-14	37.95	0.00	0.024	7.50E-05	0.032	2.60E-17	0.032	2.70E-08	0.032	2.30E-12	0.92
rs7849585	9	138251691	***QSOX2***	T/G	0.33	0.032	3.40E-11	14.89	0.12	0.024	1.50E-04	0.029	4.70E-14	0.031	1.70E-07	0.028	3.30E-09	0.69
rs7909670	10	12958770	***CCDC3***	T/C	0.44	-0.022	1.30E-06	0	0.85	-0.02	7.30E-04	-0.021	3.20E-09	-0.028	3.60E-07	-0.016	2.60E-04	0.06
rs2145998	10	80791702	***PPIF***	A/T	0.49	-0.025	2.70E-08	2.75	0.40	-0.027	3.80E-06	-0.026	3.60E-13	-0.027	4.80E-07	-0.025	2.60E-08	0.68
rs11599750	10	101795432	***CPN1***	T/C	0.38	-0.023	7.60E-07	0	0.82	-0.036	6.90E-09	-0.028	1.60E-13	-0.023	3.40E-05	-0.03	9.00E-11	0.32
rs2237886	11	2767307	***KCNQ1***	T/C	0.11	0.043	3.10E-08	6.34	0.31	0.05	1.00E-06	0.046	2.20E-13	0.037	7.50E-05	0.05	4.30E-11	0.25
rs7926971	11	12654616	***TEAD1***	A/G	0.55	-0.024	7.30E-08	0	0.91	-0.019	1.40E-03	-0.023	4.40E-10	-0.025	3.50E-06	-0.02	8.30E-06	0.4
rs1330	11	17272605	***NUCB2***	T/C	0.35	0.024	4.40E-07	17.47	0.08	0.019	2.10E-03	0.022	4.90E-09	0.02	4.70E-04	0.024	3.10E-07	0.56
rs10838801	11	48054856	***PTPRJ/SLC39A13***	A/G	0.69	-0.031	1.80E-10	12.10	0.17	-0.02	1.90E-03	-0.027	3.50E-12	-0.024	5.40E-05	-0.031	7.70E-11	0.27

SNP [a]	Chr	Position (bp)	Nearest/OMIM height gene [b]	Effect /other allele [c]	Frequency (effect allele)	STAGE 1 up to 133,653 samples				STAGE 2 up to 50,074 samples		STAGE 1 + STAGE 2 up to 183,727 samples		STAGE 1 + STAGE 2 SEX-SPECIFIC up to 73,238 males and 110,489 females				
						Beta	*P*-value [d]	I^2	P_{het}	Beta	P-value [d]	Beta	*P*-value [d]	Beta (M)	*P*-value [d] (M)	Beta (F)	*P*-value [d] (F)	P_{het} (MvsF)
rs1814175	11	49515748	***FOLH1***	T/C	0.34	0.023	2.60E-06	0	0.62	0.02	1.60E-03	0.022	1.60E-08	0.016	5.60E-03	0.027	2.20E-08	0.13
rs5017948	11	51270794	***OR4A5***	A/T	0.18	0.027	4.70E-06	9.62	0.23	0.026	1.60E-03	0.027	3.10E-08	0.016	3.10E-02	0.036	1.60E-09	0.02
rs3782089	11	65093395	***SSSCA1***	T/C	0.06	-0.058	5.90E-09	0	0.63	-0.057	1.40E-05	-0.058	3.60E-13	-0.071	2.00E-09	-0.049	7.70E-07	0.13
rs7112925	11	66582736	***RHOD***	T/C	0.35	-0.023	8.50E-07	0	0.48	-0.023	2.00E-04	-0.023	9.00E-10	-0.026	5.90E-06	-0.022	2.30E-06	0.57
rs634552	11	74959700	***SERPINH1***	T/G	0.14	0.041	1.40E-09	2.32	0.42	0.035	4.40E-05	0.039	3.50E-13	0.036	7.00E-06	0.04	1.60E-09	0.69
rs494459	11	118079885	***TREH***	T/C	0.41	0.021	4.90E-06	19.42	0.05	0.02	1.10E-03	0.02	1.70E-08	0.023	1.90E-05	0.019	2.30E-05	0.5
rs654723	11	128091365	***FLI1***	A/C	0.62	0.024	6.70E-07	0	0.93	0.028	8.00E-06	0.025	3.60E-11	0.026	4.70E-06	0.025	1.30E-07	0.82
rs2856321	12	11747040	***ETV6***	A/G	0.64	-0.03	1.50E-10	0	0.99	-0.029	4.00E-06	-0.029	4.50E-15	-0.029	4.10E-07	-0.03	8.10E-11	0.83
rs10770705	12	20748734	***SLCO1C1***	A/C	0.33	0.031	4.60E-11	0	0.75	0.036	2.20E-08	0.033	8.00E-18	0.031	8.40E-08	0.033	3.80E-12	0.77
rs2638953	12	28425682	***CCDC91***	C/G	0.68	0.036	8.40E-14	2.95	0.40	0.026	5.40E-05	0.032	6.70E-17	0.024	3.10E-05	0.038	1.10E-15	0.04
rs2066807	12	55026949	***STAT2***	C/G	0.93	-0.052	9.60E-09	0	0.71	-0.058	1.90E-06	-0.054	1.00E-13	-0.047	2.20E-05	-0.056	1.30E-10	0.49
rs1351394	12	64638093	***HMGA2***	T/C	0.49	0.054	7.80E-34	24.54	0.02	0.073	1.60E-34	0.06	1.70E-65	0.054	1.40E-23	0.063	9.00E-48	0.14
rs10748128	12	68113925	***FRS2***	T/G	0.35	0.035	3.80E-11	20.87	0.04	0.042	1.20E-10	0.038	2.10E-20	0.043	1.10E-11	0.034	8.10E-12	0.23
rs11107116	12	92502635	***SOCS2***	T/G	0.22	0.052	1.70E-23	10.02	0.21	0.05	2.20E-12	0.052	1.40E-34	0.044	4.70E-12	0.057	1.90E-27	0.1
rs7971536	12	100897919	***CCDC53/GNPTAB***	A/T	0.46	-0.025	1.10E-07	0	0.64	-0.034	4.30E-08	-0.028	8.20E-14	-0.029	3.40E-07	-0.027	1.30E-08	0.75
rs11830103	12	122389499	***SBNO1***	A/G	0.78	-0.035	3.80E-10	0	0.76	-0.035	2.50E-06	-0.035	3.90E-15	-0.041	1.40E-09	-0.032	4.50E-09	0.27
rs7332115	13	32045548	***PDS5B/BRCA2***	T/G	0.62	-0.025	7.60E-08	0	0.86	-0.02	1.10E-03	-0.023	5.50E-10	-0.02	4.80E-04	-0.026	1.70E-08	0.37
rs3118905	13	50003335	***DLEU7***	A/G	0.29	-0.052	3.00E-25	0	0.58	-0.063	3.10E-22	-0.056	1.10E-45	-0.05	4.00E-17	-0.06	1.60E-34	0.15
rs7319045	13	90822575	***GPC5***	A/G	0.4	0.029	4.50E-10	0	0.89	0.019	1.80E-03	0.025	1.20E-11	0.027	8.40E-07	0.024	1.40E-07	0.6
rs1950500	14	23900690	***NFATC4***	T/C	0.29	0.032	3.90E-11	0	0.95	0.038	8.70E-09	0.034	2.20E-18	0.038	2.00E-10	0.031	1.60E-10	0.32
rs2093210	14	60027032	***SIX6***	T/C	0.58	-0.034	2.30E-12	0	0.56	-0.029	3.90E-06	-0.032	6.20E-17	-0.028	2.10E-06	-0.036	1.90E-14	0.23
rs1570106	14	67882868	***RAD51L1***	T/C	0.2	-0.026	4.90E-06	0.50	0.47	-0.026	4.70E-04	-0.026	8.10E-09	-0.023	5.40E-04	-0.027	1.70E-06	0.67
rs862034	14	74060499	***LTBP2***	A/G	0.36	-0.023	1.10E-06	12.90	0.15	-0.037	1.90E-09	-0.028	7.30E-14	-0.032	1.90E-08	-0.024	2.10E-07	0.24
rs7155279	14	91555634	***TRIP11***	T/G	0.36	-0.029	8.90E-10	21.48	0.04	-0.016	9.20E-03	-0.024	1.40E-10	-0.028	8.70E-07	-0.022	1.10E-06	0.38
rs16964211	15	49317787	***CYP19A1***	A/G	0.05	-0.051	2.50E-06	14.06	0.13	-0.049	1.60E-04	-0.05	1.70E-09	-0.067	8.10E-08	-0.036	5.30E-04	0.04
rs7178424	15	60167551	***C2CD4A***	T/C	0.47	-0.024	2.20E-07	0	0.62	-0.017	6.20E-03	-0.021	5.60E-09	-0.02	2.50E-04	-0.021	1.50E-06	0.88
rs10152591	15	67835211	***TLE3***	A/C	0.91	0.045	3.50E-08	0	0.50	0.034	1.50E-03	0.041	2.70E-10	0.033	8.60E-04	0.046	6.60E-09	0.28
rs12902421	15	69948457	***MYO9A***	T/C	0.97	-0.069	1.70E-06	0	0.51	-0.051	3.70E-03	-0.062	2.90E-08	-0.049	2.80E-03	-0.072	2.50E-07	0.25
rs5742915	15	72123686	***PML***	T/C	0.54	-0.031	3.00E-10	0	0.71	-0.031	5.30E-07	-0.031	1.00E-15	-0.039	3.90E-11	-0.027	1.10E-08	0.08
rs11259936	15	82371586	***ADAMTSL3***	A/C	0.48	-0.042	2.20E-21	3.92	0.37	-0.047	1.10E-15	-0.044	1.70E-35	-0.036	1.50E-11	-0.049	1.00E-29	0.03

SNP [a]	Chr	Position (bp)	Nearest/OMIM height gene [b]	Effect /other allele [c]	Frequency (effect allele)	STAGE 1 up to 133,653 samples				STAGE 2 up to 50,074 samples		STAGE 1 + STAGE 2 up to 183,727 samples		STAGE 1 + STAGE 2 SEX-SPECIFIC up to 73,238 males and 110,489 females				
						Beta	*P*-value [d]	I^2	P_{het}	Beta	P-value [d]	Beta	*P*-value [d]	Beta (M)	*P*-value [d] (M)	Beta (F)	*P*-value [d] (F)	P_{het} (MvsF)
rs16942341	15	87189909	***ACAN***	T/C	0.03	-0.134	1.30E-17	24.62	0.03	-0.124	4.50E-11	-0.13	3.80E-27	-0.139	1.60E-14	-0.122	1.40E-16	0.43
rs2871865	15	97012419	***IGF1R***	C/G	0.88	0.054	1.10E-12	32.60	0.002	0.062	3.50E-10	0.057	2.90E-21	0.052	1.80E-08	0.058	2.80E-15	0.54
rs4965598	15	98577137	***ADAMTS17***	T/C	0.68	-0.035	1.40E-13	0	0.81	-0.015	2.30E-02	-0.028	4.30E-13	-0.024	5.10E-05	-0.032	9.70E-12	0.21
rs11648796	16	732191	***NARFL***	A/G	0.74	-0.031	2.40E-07	0	0.87	-0.039	6.90E-08	-0.034	1.20E-13	-0.032	7.40E-06	-0.035	5.60E-10	0.71
rs26868	16	2189377	***CASKIN1***	A/T	0.46	0.03	3.50E-08	0	0.78	0.04	2.40E-10	0.034	9.00E-17	0.036	1.20E-08	0.034	9.80E-12	0.73
rs1659127	16	14295806	***MKL2***	A/G	0.34	0.024	2.90E-06	0	0.79	0.033	5.20E-07	0.027	1.10E-11	0.025	7.90E-05	0.027	2.90E-08	0.7
rs8052560	16	87304743	***CTU2/GALNS***	A/C	0.79	0.039	1.40E-08	0	0.63	0.015	7.40E-02	0.029	3.30E-08	0.025	2.20E-03	0.032	1.10E-06	0.47
rs4640244	17	21224816	***KCNJ12***	A/G	0.61	0.028	2.00E-07	13.00	0.15	0.017	1.20E-02	0.024	2.30E-08	0.023	2.80E-04	0.025	1.70E-06	0.78
rs3110496	17	24941897	***ANKRD13B***	A/G	0.33	-0.023	1.60E-06	0	0.69	-0.021	1.10E-03	-0.022	7.30E-09	-0.03	1.10E-07	-0.016	6.40E-04	0.04
rs3764419	17	26188149	***ATAD5/RNF135***	A/C	0.39	-0.037	8.90E-16	16.60	0.09	-0.032	1.50E-07	-0.035	1.80E-21	-0.034	1.30E-09	-0.036	7.80E-16	0.67
rs17780086	17	27367395	***LRRC37B***	A/G	0.15	0.035	4.40E-08	10.44	0.21	0.017	5.50E-02	0.028	2.60E-08	0.03	9.40E-05	0.028	5.30E-06	0.85
rs1043515	17	34175722	***PIP4K2B***	A/G	0.45	-0.022	1.30E-06	0	0.80	-0.024	6.60E-05	-0.023	2.90E-10	-0.028	2.00E-07	-0.019	2.20E-05	0.15
rs4986172	17	40571807	***ACBD4***	T/C	0.35	-0.028	7.10E-09	30.83	0.003	-0.037	2.50E-09	-0.032	2.30E-16	-0.035	1.70E-09	-0.03	3.10E-10	0.41
rs2072153	17	44745013	***ZNF652***	C/G	0.3	0.026	6.70E-08	0	0.86	0.013	4.30E-02	0.021	3.50E-08	0.031	1.60E-07	0.016	8.30E-04	0.03
rs4605213	17	46599746	***NME2***	C/G	0.34	0.023	9.30E-07	0	0.88	0.018	5.90E-03	0.021	2.70E-08	0.026	5.40E-06	0.018	2.10E-04	0.21
rs227724	17	52133816	***NOG***	A/T	0.65	-0.027	1.20E-08	0	0.92	-0.034	6.60E-08	-0.03	7.40E-15	-0.035	8.10E-10	-0.027	1.10E-08	0.2
rs2079795	17	56851431	***TBX2***	T/C	0.33	0.04	1.20E-16	0	0.81	0.04	1.50E-09	0.04	2.10E-24	0.033	7.80E-09	0.044	8.10E-20	0.12
rs2665838	17	59320197	***CSH1/GH1***	C/G	0.73	-0.037	2.00E-13	11.25	0.19	-0.052	7.00E-14	-0.042	5.10E-25	-0.042	2.20E-11	-0.042	3.00E-17	0.92
rs11867479	17	65601802	***KCNJ16/KCNJ2***	T/C	0.34	0.024	4.90E-07	0	0.87	0.026	5.40E-05	0.025	1.50E-10	0.023	7.00E-05	0.026	6.70E-08	0.68
rs4800452	18	18981609	***CABLES1***	T/C	0.79	0.048	2.40E-17	0	0.84	0.056	1.20E-14	0.051	4.20E-30	0.052	7.40E-15	0.05	8.40E-20	0.8
rs9967417	18	45213498	***DYM***	C/G	0.58	-0.038	2.60E-16	30.04	0.004	-0.039	3.20E-10	-0.038	9.30E-25	-0.041	3.40E-13	-0.036	1.30E-15	0.44
rs17782313	18	56002077	***MC4R***	T/C	0.76	-0.025	3.50E-06	13.42	0.14	-0.035	1.20E-06	-0.028	3.80E-11	-0.03	4.00E-06	-0.025	1.20E-06	0.55
rs12982744	19	2128193	***DOT1L***	C/G	0.6	-0.033	2.80E-12	0	0.97	-0.027	1.10E-05	-0.03	3.40E-16	-0.028	4.90E-07	-0.032	3.80E-12	0.6
rs7507204	19	3379834	***NFIC***	C/G	0.24	0.028	2.30E-07	0	0.88	0.049	2.10E-11	0.036	4.30E-16	0.025	1.70E-04	0.041	2.60E-14	0.05
rs891088	19	7135762	***INSR***	A/G	0.74	-0.025	1.70E-06	2.38	0.41	-0.035	1.80E-07	-0.029	2.40E-12	-0.025	6.10E-05	-0.031	1.10E-09	0.45
rs4072910	19	8550031	***ADAMTS10***	C/G	0.46	-0.029	2.50E-07	0	0.76	-0.034	2.20E-07	-0.031	3.60E-13	-0.025	1.30E-04	-0.033	3.10E-10	0.31
rs2279008	19	17144303	***MYO9B***	T/C	0.74	0.031	2.40E-07	0	0.63	0.018	9.50E-03	0.025	2.50E-08	0.022	2.00E-03	0.027	5.00E-07	0.48
rs17318596	19	46628935	***ATP5SL***	A/G	0.36	0.029	3.00E-09	0	0.79	0.037	2.10E-08	0.032	5.00E-16	0.043	1.30E-13	0.024	8.00E-07	0.01
rs1741344	20	4049800	***SMOX***	T/C	0.63	-0.026	3.50E-08	16.74	0.09	-0.016	1.00E-02	-0.023	3.30E-09	-0.02	4.10E-04	-0.024	2.60E-07	0.55
rs2145272	20	6574218	***BMP2***	A/G	0.65	-0.039	5.90E-16	19.29	0.06	-0.04	4.60E-10	-0.039	2.10E-24	-0.039	1.50E-11	-0.04	2.30E-17	0.85

SNP [a]	Chr	Position (bp)	Nearest/OMIM height gene [b]	Effect /other allele [c]	Frequency (effect allele)	STAGE 1 up to 133,653 samples				STAGE 2 up to 50,074 samples		STAGE 1 + STAGE 2 up to 183,727 samples		STAGE 1 + STAGE 2 SEX-SPECIFIC up to 73,238 males and 110,489 females				
						Beta	*P*-value [d]	I^2	P_{het}	Beta	P-value [d]	Beta	*P*-value [d]	Beta (M)	*P*-value [d] (M)	Beta (F)	*P*-value [d] (F)	P_{het} (MvsF)
rs7274811	20	31796842	***ZNF341***	T/G	0.23	-0.04	6.80E-14	7.93	0.26	-0.042	1.10E-09	-0.041	5.90E-22	-0.044	1.60E-11	-0.039	1.30E-13	0.52
rs143384	20	33489170	***GDF5***	A/G	0.58	-0.064	4.90E-39	21.58	0.04	-0.061	9.10E-22	-0.063	1.00E-58	-0.066	9.30E-30	-0.061	8.30E-38	0.47
rs237743	20	47336426	***ZNFX1***	A/G	0.21	0.034	7.20E-10	0	0.69	0.053	3.10E-13	0.041	1.30E-20	0.035	1.20E-07	0.043	6.80E-16	0.28
rs2834442	21	34612656	***KCNE2***	A/T	0.65	0.027	7.30E-09	0	0.80	0.024	9.70E-05	0.026	5.10E-12	0.025	9.10E-06	0.026	1.00E-08	0.9
rs4821083	22	31386341	***SYN3***	T/C	0.84	0.033	4.80E-08	0	0.70	0.027	1.40E-03	0.031	3.10E-10	0.036	1.40E-06	0.028	4.20E-06	0.41

[a] SNPs most likely to be representing a previously published height locus are highlighted in green.

[b] Gene regions are named after the gene nearest to the index SNP. A near-by (within 500kb from the index SNP) OMIM height gene (defined as a gene that when mutated results in a monogenic skeletal growth defect) is also included if it is not the nearest. All OMIM height genes are highlighted in blue.

[c] Alleles are indexed to the forward strand of NCBI Build 36.

[d] All p-values are based on the inverse-variance weighted meta-analysis model (fixed effects).

Supplementary Table 2. Association results for 33 SNPs selected randomly among the 180 SNPs that reached genome-wide significance (P=5x10^{-8}) in the Stage 1 meta-analysis and genotyped in European American (N=2,181) and Poland (N=1,009) extreme height panels. Results are combined using a Cochran-Mantel-Haenszel test.

			GIANT height meta-analysis				Results in extreme height panels				
GIANT height SNP	Chr	Position	Effect allele	Other allele	Effect size (Stage 1)	Stage 1+2 P-value	Effect allele	Other allele	OR [95% CI]	1-tailed P-value	Comment
rs143384	20	33489170	A	G	-0.0639	9.954E-59	G	A	1.2 [1.08-1.33]	0.0002	Same direction, 1-tailed P-value <0.05
rs2580816	2	232506210	T	C	-0.0412	5.837E-22	T	C	0.8 [0.7-0.91]	0.0002	Same direction, 1-tailed P-value <0.05
rs1738475	1	23409478	C	G	0.0216	2.952E-12	G	C	0.86 [0.78-0.96]	0.002	Same direction, 1-tailed P-value <0.05
rs12474201	2	46774789	A	G	0.0233	2.581E-13	A	G	1.16 [1.04-1.29]	0.003	Same direction, 1-tailed P-value <0.05
rs1351164	2	217980143	T	C	0.0279	2.081E-14	C	T	0.84 [0.74-0.96]	0.004	Same direction, 1-tailed P-value <0.05
rs822552	7	148281567	C	G	-0.0302	2.613E-08	G	C	1.15 [1.03-1.29]	0.007	Same direction, 1-tailed P-value <0.05
rs7849585	9	138251691	T	G	0.0324	4.724E-14	T	G	1.13 [1.02-1.26]	0.011	Same direction, 1-tailed P-value <0.05
rs1257763	9	95933766	A	G	0.0685	9.865E-10	A	G	1.33 [1.04-1.69]	0.012	Same direction, 1-tailed P-value <0.05
rs12534093	7	23469499	A	T	-0.0298	2.019E-14	A	T	0.87 [0.77-0.98]	0.012	Same direction, 1-tailed P-value <0.05
rs2871865	15	97012419	C	G	0.0535	2.862E-21	G	C	0.83 [0.71-0.98]	0.013	Same direction, 1-tailed P-value <0.05
rs310405	6	81857081	A	G	0.03	2.245E-13	G	A	0.89 [0.81-0.99]	0.016	Same direction, 1-tailed P-value <0.05
rs10037512	5	88390431	T	C	0.0267	2.011E-18	C	T	0.82 [0.69-0.99]	0.018	Same direction, 1-tailed P-value <0.05
rs1814175	11	49515748	T	C	0.023	1.645E-08	T	C	1.11 [1-1.24]	0.02	Same direction, 1-tailed P-value <0.05
rs16942341	15	87189909	T	C	-0.1335	3.807E-27	T	C	0.74 [0.55-1.01]	0.03	Same direction, 1-tailed P-value <0.05
rs4665736	2	25041103	T	C	0.0335	7.29E-16	C	T	0.92 [0.83-1.02]	0.05	Same direction, 1-tailed P-value <0.05
rs6684205	1	216676325	A	G	-0.0328	1.473E-12	G	A	1.09 [0.97-1.22]	0.07	Same direction
rs7567288	2	134151294	T	C	-0.0309	2.071E-12	C	T	1.11 [0.97-1.26]	0.07	Same direction
rs7697556	4	73734177	T	C	0.0219	1.958E-14	T	C	1.07 [0.96-1.18]	0.11	Same direction
rs11599750	10	101795432	T	C	-0.023	1.604E-13	T	C	0.94 [0.85-1.05]	0.13	Same direction
rs2066807	12	55026949	C	G	-0.052	1.025E-13	G	C	1.12 [0.92-1.35]	0.13	Same direction
rs751543	9	118162163	T	C	0.0287	6.537E-10	C	T	0.94 [0.84-1.05]	0.13	Same direction
rs7532866	1	26614131	A	G	0.0222	3.372E-08	G	A	0.94 [0.85-1.05]	0.14	Same direction
rs11118346	1	217810342	T	C	-0.0264	1.879E-12	T	C	0.96 [0.87-1.06]	0.20	Same direction
rs6439167	3	130533446	T	C	-0.0338	8.925E-15	T	C	0.93 [0.75-1.15]	0.24	Same direction
rs274546	5	131727766	A	G	-0.0278	7.254E-16	A	G	0.97 [0.87-1.07]	0.26	Same direction
rs10863936	1	210304421	A	G	-0.022	1.922E-09	G	A	1.03 [0.93-1.14]	0.27	Same direction

rs9360921	6	76322362	T	G	-0.0479	2.552E-13	G	T	1.05 [0.89-1.25]	0.28	Same direction
rs4986172	17	40571807	T	C	-0.0283	2.333E-16	T	C	0.97 [0.87-1.08]	0.29	Same direction
rs9456307	6	158849430	A	T	-0.0499	2.239E-09	A	T	0.98 [0.78-1.22]	0.42	Same direction
rs572169	3	173648421	T	C	0.0355	2.765E-18	T	C	1.02 [0.84-1.24]	0.42	Same direction
rs2110001	7	150147955	C	G	-0.0328	3.319E-13	G	C	0.99 [0.88-1.1]	0.61	Opposite direction (1-tailed P-value adjusted accordingly)
rs8052560	16	87304743	A	C	0.0392	3.324E-08	C	A	1.08 [0.95-1.23]	0.89	Opposite direction (1-tailed P-value adjusted accordingly)

Supplementary Table 3. Family-based association results for the 180 confirmed height SNPs in the Framingham Heart Study (FHS) and the Erasmus Rucphen Family (ERF) study. For each study, and the meta-analysis FHS+ERF, we compare the direction of effect observed with respect to the effect of the height-increasing allele in the GIANT meta-analysis.

SNP	GIANT meta-analysis			FHS (family-based test)			ERF (family-based test)		Meta-analysis FHS+ERF (family-based test)	
	Height Increasing allele	BETA	*P*-value	Direction of effect relative to GIANT	*P*-value	BETA relative to GIANT	Direction of effect relative to GIANT	*P*-value	Direction of effect relative to GIANT	*P*-value
rs724016	G	0.067	4.5E-52	Same	0.02	0.066	Same	0.61	Same	0.02
rs143384	G	0.064	4.9E-39	Same	3.5E-03	0.081	Same	0.65	Same	4.8E-03
rs1351394	T	0.054	7.8E-34	Same	1.3E-03	0.089	Same	0.73	Same	2.3E-03
rs7689420	C	0.069	1.4E-29	Same	0.06	0.068	Same	0.03	Same	0.01
rs42235	T	0.055	7.3E-28	Opposite	0.91	-0.003	Opposite	0.85	Opposite	0.85
rs6449353	T	0.071	1.3E-27	Same	0.04	0.070	Same	0.43	Same	0.03
rs806794	A	0.053	5.5E-26	Same	1.7E-03	0.106	Same	0.02	Same	1.2E-04
rs3118905	G	0.052	3.0E-25	Same	0.36	0.027	Same	0.57	Same	0.28
rs798489	C	0.052	8.5E-25	Same	0.09	0.052	Same	0.04	Same	0.02
rs11107116	T	0.052	1.7E-23	Same	0.05	0.065	Same	0.33	Same	0.03
rs1046934	C	0.046	6.4E-22	Same	0.03	0.062	Same	0.89	Same	0.05
rs11259936	C	0.042	2.2E-21	Same	0.06	0.053	Same	0.90	Same	0.09
rs3791675	C	0.050	2.4E-20	Same	0.67	0.014	Same	0.25	Same	0.38
rs7763064	G	0.045	6.4E-19	Same	0.43	0.023	Same	0.23	Same	0.22
rs2780226	C	0.079	1.0E-18	Same	6.4E-04	0.171	Same	0.02	Same	4.9E-05
rs11205277	G	0.045	1.2E-18	Same	0.12	0.043	Same	2.0E-03	Same	0.01
rs7759938	C	0.042	8.7E-18	Same	0.15	0.041	Same	0.65	Same	0.13
rs16942341	C	0.134	1.3E-17	Same	2.5E-03	0.236	Same	0.59	Same	3.1E-03
rs1708299	A	0.042	1.5E-17	Same	4.5E-03	0.090	Same	0.29	Same	2.6E-03
rs4800452	T	0.048	2.4E-17	Same	0.01	0.082	Same	0.49	Same	0.01
rs6470764	C	0.047	5.9E-17	Same	0.22	0.041	Same	0.87	Same	0.24
rs2079795	T	0.040	1.2E-16	Same	0.09	0.049	Same	0.69	Same	0.09
rs9967417	G	0.038	2.6E-16	Same	0.22	0.031	Same	0.67	Same	0.20
rs1490384	T	0.037	3.2E-16	Same	0.03	0.060	Opposite	0.45	Same	0.10

SNP	GIANT meta-analysis			FHS (family-based test)			ERF (family-based test)		Meta-analysis FHS+ERF (family-based test)	
	Height Increasing allele	BETA	*P*-value	Direction of effect relative to GIANT	*P*-value	BETA relative to GIANT	Direction of effect relative to GIANT	*P*-value	Direction of effect relative to GIANT	*P*-value
rs2145272	G	0.039	5.9E-16	Same	0.01	0.069	Same	0.88	Same	0.02
rs3812163	T	0.037	6.7E-16	Same	0.80	0.007	Same	0.18	Same	0.42
rs3764419	C	0.037	8.9E-16	Same	0.09	0.046	Same	0.49	Same	0.07
rs7460090	T	0.055	9.6E-16	Opposite	0.96	-0.002	Same	0.19	Same	0.60
rs788867	G	0.039	1.8E-15	Same	0.56	0.017	Same	0.36	Same	0.36
rs1173727	T	0.036	4.0E-15	Same	0.03	0.058	Same	0.40	Same	0.02
rs2284746	G	0.035	5.6E-15	Same	0.69	0.011	Same	0.09	Same	0.28
rs7466269	A	0.036	1.2E-14	Same	0.20	0.036	Same	0.05	Same	0.04
rs473902	T	0.074	1.7E-14	Same	0.26	0.056	Same	0.83	Same	0.26
rs2256183	A	0.035	2.7E-14	Same	0.79	0.008	Opposite	0.94	Same	0.84
rs17346452	C	0.038	3.3E-14	Same	0.92	0.003	Same	0.21	Same	0.52
rs7274811	G	0.040	6.8E-14	Same	0.01	0.080	Same	0.42	Same	0.01
rs2638953	C	0.036	8.4E-14	Same	0.07	0.056	Opposite	0.97	Same	0.11
rs572169	T	0.036	9.9E-14	Same	0.60	0.015	Same	0.03	Same	0.17
rs4665736	T	0.034	1.4E-13	Same	0.28	0.030	Same	0.11	Same	0.09
rs4965598	C	0.035	1.4E-13	Same	0.02	0.069	Same	0.52	Same	0.02
rs2665838	G	0.037	2.0E-13	Same	0.09	0.051	Same	0.40	Same	0.06
rs9428104	G	0.038	8.9E-13	Same	0.06	0.057	Opposite	0.19	Same	0.27
rs2871865	C	0.054	1.1E-12	Same	0.10	0.074	Same	0.08	Same	0.02
rs17511102	T	0.060	1.3E-12	Same	0.23	0.062	Same	0.50	Same	0.17
rs2580816	C	0.041	1.8E-12	Same	0.37	0.031	Opposite	0.67	Same	0.54
rs2093210	C	0.034	2.3E-12	Same	0.20	0.036	Same	0.07	Same	0.05
rs12982744	G	0.033	2.8E-12	Same	0.69	0.012	Opposite	0.46	Same	0.97
rs6569648	C	0.036	8.9E-12	Same	0.06	0.058	Same	0.46	Same	0.04
rs6457821	C	0.121	1.8E-11	Opposite	0.93	-0.009	Same	0.26	Same	0.69
rs6684205	G	0.033	2.0E-11	Same	0.47	0.021	Same	0.08	Same	0.16
rs7849585	T	0.032	3.4E-11	Opposite	0.91	-0.003	Same	0.19	Same	0.64
rs310405	A	0.030	3.6E-11	Opposite	0.81	-0.007	Same	0.16	Same	0.70

SNP	GIANT meta-analysis			FHS (family-based test)			ERF (family-based test)		Meta-analysis FHS+ERF (family-based test)	
	Height Increasing allele	BETA	*P*-value	Direction of effect relative to GIANT	*P*-value	BETA relative to GIANT	Direction of effect relative to GIANT	*P*-value	Direction of effect relative to GIANT	*P*-value
rs10748128	T	0.035	3.8E-11	Opposite	0.65	-0.014	Same	0.35	Same	0.99
rs1950500	T	0.032	3.9E-11	Same	0.28	0.033	Same	0.29	Same	0.15
rs10770705	A	0.031	4.6E-11	Opposite	0.29	-0.031	Opposite	0.60	Opposite	0.24
rs9360921	G	0.048	4.6E-11	Same	0.52	0.029	Opposite	0.65	Same	0.70
rs12680655	C	0.030	4.8E-11	Opposite	0.50	-0.019	Same	0.39	Opposite	0.82
rs12470505	T	0.048	1.3E-10	Same	0.22	0.053	Same	0.45	Same	0.15
rs7027110	A	0.034	1.3E-10	Same	0.16	0.046	Opposite	0.81	Same	0.24
rs2856321	G	0.030	1.5E-10	Same	0.32	0.028	Same	0.56	Same	0.25
rs720390	A	0.031	1.6E-10	Same	0.10	0.048	Same	0.19	Same	0.04
rs6473015	C	0.032	1.7E-10	Same	0.55	0.017	Opposite	0.62	Same	0.74
rs10838801	G	0.031	1.8E-10	Same	0.01	0.075	Same	0.04	Same	1.7E-03
rs5742915	C	0.031	3.0E-10	Same	0.05	0.059	Same	0.25	Same	0.02
rs4282339	G	0.035	3.4E-10	Same	0.34	0.032	Same	0.75	Same	0.32
rs2154319	C	0.034	4.3E-10	Same	0.27	0.035	Same	0.01	Same	0.03
rs7319045	A	0.029	4.5E-10	Same	0.01	0.078	Same	0.57	Same	0.01
rs889014	C	0.029	4.5E-10	Same	0.38	0.025	Same	0.70	Same	0.34
rs6879260	C	0.028	5.6E-10	Same	0.99	0.000	Same	0.13	Same	0.50
rs9969804	A	0.028	5.6E-10	Same	0.47	0.021	Opposite	0.71	Same	0.62
rs237743	A	0.034	7.2E-10	Same	0.24	0.041	Same	0.05	Same	0.06
rs6439167	C	0.034	7.2E-10	Same	0.12	0.049	Same	0.73	Same	0.12
rs274546	G	0.028	8.5E-10	Same	0.01	0.066	Opposite	0.41	Same	0.07
rs12153391	C	0.033	8.7E-10	Same	0.55	0.018	Same	0.15	Same	0.25
rs7155279	G	0.029	8.9E-10	Opposite	0.74	-0.009	Same	0.53	Opposite	0.98
rs2110001	G	0.033	9.8E-10	Same	0.15	0.042	Same	0.03	Same	0.03
rs422421	C	0.033	1.4E-09	Same	9.0E-05	0.126	Same	0.05	Same	1.2E-05
rs543650	G	0.032	1.4E-09	Same	0.27	0.037	Same	0.51	Same	0.20
rs634552	T	0.041	1.4E-09	Same	0.74	0.014	Opposite	0.10	Opposite	0.68
rs11144688	G	0.055	1.5E-09	Same	0.63	0.022	Same	0.34	Same	0.40

SNP	GIANT meta-analysis			FHS (family-based test)			ERF (family-based test)		Meta-analysis FHS+ERF (family-based test)	
	Height Increasing allele	BETA	*P*-value	Direction of effect relative to GIANT	*P*-value	BETA relative to GIANT	Direction of effect relative to GIANT	*P*-value	Direction of effect relative to GIANT	*P*-value
rs526896	T	0.032	1.9E-09	Same	0.97	0.001	Opposite	0.37	Opposite	0.72
rs1809889	T	0.032	1.9E-09	Opposite	0.48	-0.023	Same	0.10	Same	0.92
rs11118346	C	0.026	2.2E-09	Same	0.04	0.057	Same	0.68	Same	0.04
rs17318596	A	0.029	3.0E-09	Same	0.72	0.010	Opposite	0.62	Same	0.92
rs10037512	T	0.027	3.8E-09	Same	0.06	0.051	Same	0.73	Same	0.07
rs3782089	C	0.058	5.9E-09	Same	0.99	0.001	Same	0.57	Same	0.80
rs11684404	C	0.027	6.4E-09	Same	0.20	0.038	Same	0.36	Same	0.12
rs4986172	C	0.028	7.1E-09	Same	0.35	0.026	Opposite	0.12	Same	0.88
rs2834442	A	0.027	7.3E-09	Same	0.08	0.049	Opposite	0.82	Same	0.13
rs11958779	G	0.028	8.0E-09	Opposite	0.80	-0.008	Opposite	0.80	Opposite	0.73
rs2066807	G	0.052	9.6E-09	Same	0.03	0.129	Opposite	0.44	Same	0.10
rs10799445	A	0.031	1.2E-08	Same	0.51	0.021	Opposite	0.50	Same	0.77
rs227724	T	0.027	1.2E-08	Same	0.15	0.043	Opposite	0.87	Same	0.23
rs8052560	A	0.039	1.4E-08	Same	0.01	0.031	Opposite	0.34	Same	0.06
rs9863706	C	0.030	1.5E-08	Same	0.01	0.086	Opposite	0.57	Same	0.03
rs1325598	G	0.026	1.6E-08	Same	0.03	0.058	Same	0.62	Same	0.03
rs1582931	G	0.025	2.1E-08	Same	0.01	0.074	Same	0.38	Same	0.01
rs2629046	T	0.025	2.2E-08	Same	0.64	0.013	Same	0.62	Same	0.52
rs9472414	T	0.031	2.4E-08	Same	0.33	0.094	Same	0.48	Same	0.24
rs2145998	T	0.025	2.7E-08	Same	0.35	0.027	Opposite	0.80	Same	0.46
rs2237886	T	0.043	3.1E-08	Same	0.02	0.117	Same	0.52	Same	0.02
rs9844666	G	0.028	3.1E-08	Same	0.25	0.037	Opposite	0.25	Same	0.60
rs3129109	C	0.026	3.3E-08	Opposite	0.43	-0.022	Opposite	0.72	Opposite	0.38
rs10152591	A	0.045	3.5E-08	Same	0.20	0.063	Same	0.10	Same	0.06
rs1741344	C	0.026	3.5E-08	Same	0.05	0.057	Same	0.26	Same	0.02
rs2336725	C	0.026	3.5E-08	Same	0.05	0.051	Same	0.63	Same	0.05
rs26868	A	0.030	3.5E-08	Same	0.79	0.008	Same	0.58	Same	0.63
rs2341459	T	0.028	3.6E-08	Same	0.12	0.050	Opposite	0.82	Same	0.19

SNP	GIANT meta-analysis			FHS (family-based test)			ERF (family-based test)		Meta-analysis FHS+ERF (family-based test)	
	Height Increasing allele	BETA	*P*-value	Direction of effect relative to GIANT	*P*-value	BETA relative to GIANT	Direction of effect relative to GIANT	*P*-value	Direction of effect relative to GIANT	*P*-value
rs6457620	G	0.024	3.6E-08	Same	0.93	0.003	Opposite	0.69	Opposite	0.93
rs4470914	T	0.033	3.8E-08	Same	1.00	0.000	Same	1.00	Same	1.00
rs17780086	A	0.035	4.4E-08	Same	0.31	0.041	Opposite	0.97	Same	0.36
rs1013209	C	0.029	4.5E-08	Opposite	0.74	-0.011	Opposite	0.94	Opposite	0.74
rs751543	T	0.029	4.5E-08	Same	0.95	0.002	Opposite	0.02	Opposite	0.32
rs17081935	T	0.031	4.8E-08	Opposite	0.75	-0.011	Same	0.27	Same	0.85
rs4821083	T	0.033	4.8E-08	Same	0.07	0.063	Same	0.54	Same	0.06
rs12534093	T	0.030	5.6E-08	Same	0.59	0.018	Opposite	0.97	Same	0.64
rs2072153	C	0.026	6.7E-08	Same	0.41	0.026	Same	0.42	Same	0.27
rs7567288	C	0.031	6.7E-08	Opposite	0.93	-0.003	Same	0.90	Opposite	0.98
rs2247341	A	0.025	6.8E-08	Same	0.42	0.022	Same	0.12	Same	0.16
rs7926971	G	0.024	7.3E-08	Same	1.6E-03	0.082	Opposite	0.04	Same	0.05
rs7332115	G	0.025	7.6E-08	Same	0.24	0.033	Same	0.15	Same	0.09
rs1047014	C	0.029	1.1E-07	Same	0.41	0.028	Opposite	0.49	Same	0.66
rs17806888	T	0.040	1.1E-07	Opposite	0.04	0.077	Same	0.42	Same	0.03
rs2597513	C	0.039	1.1E-07	Same	0.08	0.078	Same	0.45	Same	0.06
rs7971536	T	0.025	1.1E-07	Same	0.17	0.036	Same	0.72	Same	0.17
rs8181166	C	0.025	1.1E-07	Opposite	0.35	-0.025	Opposite	0.82	Opposite	0.35
rs822552	G	0.030	1.3E-07	Opposite	0.44	-0.025	Opposite	0.95	Opposite	0.47
rs12694997	G	0.027	1.8E-07	Same	0.05	0.063	Opposite	0.20	Same	0.22
rs4640244	A	0.028	2.0E-07	Opposite	0.67	0.013	Same	0.05	Same	0.21
rs7178424	C	0.024	2.2E-07	Same	0.02	0.067	Opposite	0.29	Same	0.10
rs7507204	C	0.028	2.3E-07	Same	0.05	0.071	Same	0.51	Same	0.04
rs11648796	G	0.031	2.4E-07	NA	1.00	NA	Same	0.65	Same	0.65
rs2279008	T	0.031	2.4E-07	Same	0.44	0.027	Same	0.15	Same	0.18
rs961764	G	0.023	2.4E-07	Same	0.16	0.038	Same	0.48	Same	0.11
rs4072910	G	0.029	2.5E-07	NA	1.00	NA	Same	0.49	Same	0.49
rs13088462	C	0.054	3.1E-07	Same	0.96	0.003	Opposite	0.66	Opposite	0.88

SNP	GIANT meta-analysis			FHS (family-based test)			ERF (family-based test)		Meta-analysis FHS+ERF (family-based test)	
	Height Increasing allele	BETA	*P*-value	Direction of effect relative to GIANT	*P*-value	BETA relative to GIANT	Direction of effect relative to GIANT	*P*-value	Direction of effect relative to GIANT	*P*-value
rs2778031	T	0.027	3.6E-07	Same	0.04	0.066	Same	0.59	Same	0.04
rs1351164	T	0.028	3.7E-07	Same	0.78	0.010	Same	0.96	Same	0.78
rs1330	T	0.024	4.4E-07	Opposite	0.49	-0.018	Same	0.58	Opposite	0.70
rs9456307	T	0.050	4.6E-07	Opposite	0.56	-0.032	Same	0.94	Opposite	0.62
rs11867479	T	0.024	4.9E-07	Same	0.07	0.051	Same	0.95	Same	0.10
rs7864648	T	0.025	4.9E-07	Opposite	0.67	-0.013	Same	0.44	Opposite	0.97
rs17391694	T	0.040	5.9E-07	Opposite	0.43	-0.042	Same	0.51	Opposite	0.67
rs10863936	G	0.022	6.2E-07	Same	0.02	0.060	Opposite	0.90	Same	0.05
rs654723	A	0.024	6.7E-07	Same	0.38	0.024	Opposite	0.86	Same	0.48
rs7567851	C	0.041	7.5E-07	Same	0.32	0.046	Opposite	0.93	Same	0.40
rs11599750	C	0.023	7.6E-07	Same	0.38	0.025	Same	0.18	Same	0.17
rs7112925	C	0.023	8.5E-07	Same	0.23	0.034	Same	0.08	Same	0.06
rs1046943	A	0.022	8.6E-07	Same	0.42	0.023	Same	0.28	Same	0.23
rs9835332	G	0.022	8.7E-07	Same	0.13	0.038	Same	0.29	Same	0.07
rs4605213	C	0.023	9.3E-07	Opposite	0.73	-0.011	Same	0.49	Opposite	0.99
rs12474201	A	0.023	1.0E-06	Same	0.38	0.025	Opposite	0.37	Same	0.69
rs862034	G	0.023	1.1E-06	Same	0.05	0.059	Same	0.43	Same	0.03
rs1043515	G	0.022	1.3E-06	Same	0.15	0.039	Same	0.05	Same	0.03
rs7697556	T	0.022	1.3E-06	Same	0.02	0.063	Opposite	0.62	Same	0.07
rs7909670	C	0.022	1.3E-06	Opposite	0.77	-0.008	Same	0.71	Opposite	0.92
rs1468758	C	0.026	1.5E-06	Same	0.76	0.009	Opposite	0.88	Same	0.84
rs3110496	G	0.023	1.6E-06	Same	0.01	0.078	Same	0.12	Same	0.00
rs10874746	C	0.022	1.7E-06	Opposite	0.29	-0.030	Same	0.03	Opposite	0.98
rs12902421	C	0.069	1.7E-06	Same	0.11	0.145	Opposite	0.45	Same	0.27
rs425277	T	0.024	1.7E-06	Same	0.09	0.052	Same	0.99	Same	0.13
rs6699417	T	0.022	1.7E-06	Same	0.85	0.006	Opposite	0.51	Opposite	0.91
rs891088	G	0.025	1.7E-06	Opposite	0.70	-0.012	Same	0.48	Opposite	0.96
rs1738475	C	0.022	1.9E-06	Same	0.24	0.030	Opposite	0.90	Same	0.32

SNP	GIANT meta-analysis			FHS (family-based test)			ERF (family-based test)		Meta-analysis FHS+ERF (family-based test)	
	Height Increasing allele	BETA	*P*-value	Direction of effect relative to GIANT	*P*-value	BETA relative to GIANT	Direction of effect relative to GIANT	*P*-value	Direction of effect relative to GIANT	*P*-value
rs4601530	C	0.024	2.0E-06	Opposite	0.79	-0.008	Same	0.69	Opposite	0.95
rs6714546	G	0.025	2.2E-06	Same	0.16	0.045	Opposite	0.78	Same	0.25
rs955748	G	0.024	2.2E-06	Same	0.57	0.018	Opposite	0.59	Same	0.78
rs10010325	A	0.021	2.3E-06	Same	0.45	0.019	Same	0.33	Same	0.27
rs1257763	A	0.069	2.5E-06	Same	0.02	0.150	Opposite	0.71	Same	0.05
rs16964211	G	0.051	2.5E-06	Same	0.94	0.005	Opposite	0.14	Opposite	0.57
rs1814175	T	0.023	2.6E-06	Same	0.43	0.023	Same	0.01	Same	0.07
rs6959212	C	0.023	2.8E-06	Same	0.60	0.014	Same	0.19	Same	0.30
rs1659127	A	0.024	2.9E-06	Opposite	0.41	-0.023	Opposite	0.16	Opposite	0.17
rs7853377	G	0.026	3.1E-06	Same	0.90	0.004	Opposite	0.92	Same	0.94
rs7532866	A	0.022	3.3E-06	Same	0.24	0.033	Opposite	0.78	Same	0.35
rs17782313	C	0.025	3.5E-06	Opposite	0.60	-0.017	Same	0.28	Same	1.00
rs13177718	C	0.041	4.1E-06	Opposite	0.51	-0.033	Opposite	0.70	Opposite	0.45
rs5017948	A	0.027	4.7E-06	Same	8.9E-04	0.126	Same	0.07	Same	1.6E-04
rs1570106	C	0.026	4.9E-06	Same	0.59	0.019	Opposite	0.14	Opposite	0.87
rs494459	T	0.021	4.9E-06	Opposite	0.14	-0.040	Same	0.10	Opposite	0.56

Supplementary Table 4. Estimated number of height loci for each of the effect sizes observed in Stage 2 given the power to detect the association in Stage 1.

	SNP	MAF	Mean Difference	Standardized Effect size	Power	Estimated number of loci†
1	rs1325598	0.435	-0.016	1.18E-04	0.0151	74.5
2	rs9472414	0.217	-0.019	1.21E-04	0.0169	63.4
3	rs7155279	0.356	-0.016	1.22E-04	0.0171	62.3
4	rs1741344	0.375	0.016	1.23E-04	0.0178	59.3
5	rs1013209	0.252	-0.019	1.30E-04	0.0226	47.6
6	rs12470505	0.095	-0.028	1.35E-04	0.0263	41.5
7	rs6684205	0.294	0.019	1.45E-04	0.0350	30.9
8	rs2341459	0.279	0.020	1.53E-04	0.0434	24.6
9	rs4470914	0.178	0.023	1.53E-04	0.0436	24.5
10	rs6457821	0.018	-0.068	1.65E-04	0.0587	17.7
11	rs751543	0.275	-0.021	1.68E-04	0.0619	16.9
12	rs10838801	0.307	0.020	1.70E-04	0.0655	15.9
13	rs7319045	0.403	0.019	1.72E-04	0.0681	15.2
14	rs1582931	0.480	-0.019	1.82E-04	0.0848	12.2
15	rs4821083	0.151	-0.027	1.91E-04	0.1018	10.1
16	rs10152591	0.091	-0.034	1.93E-04	0.1044	9.8
17	rs310405	0.478	-0.020	2.00E-04	0.1185	8.7
18	rs6473015	0.274	0.023	2.10E-04	0.1426	7.3
19	rs9360921	0.113	0.033	2.15E-04	0.1529	6.7
20	rs12153391	0.250	-0.024	2.18E-04	0.1611	6.3
21	rs4665736	0.455	-0.021	2.21E-04	0.1681	6.0
22	rs7027110	0.230	0.025	2.21E-04	0.1692	6.0
23	rs2154319	0.244	0.025	2.27E-04	0.1839	5.5
24	rs17081935	0.203	0.028	2.48E-04	0.2444	4.2
25	rs11118346	0.477	-0.023	2.57E-04	0.2721	3.7
26	rs7849585	0.333	0.024	2.58E-04	0.2750	3.7
27	rs2629046	0.440	-0.023	2.63E-04	0.2906	3.5
28	rs422421	0.216	-0.028	2.64E-04	0.2927	3.5
29	rs2834442	0.343	-0.024	2.68E-04	0.3079	3.3
30	rs12680655	0.413	-0.024	2.75E-04	0.3286	3.1
31	rs7466269	0.370	-0.024	2.78E-04	0.3381	3.0
32	rs2638953	0.318	-0.026	2.82E-04	0.3529	2.9
33	rs11958779	0.306	0.026	2.90E-04	0.3783	2.7
34	rs634552	0.146	0.035	2.99E-04	0.4087	2.5
35	rs11144688	0.106	-0.040	3.03E-04	0.4230	2.4
36	rs720390	0.387	0.026	3.28E-04	0.5092	2.0
37	rs526896	0.279	-0.029	3.29E-04	0.5119	2.0
38	rs2110001	0.319	0.028	3.31E-04	0.5180	2.0
39	rs12982744	0.389	0.027	3.41E-04	0.5519	1.8
40	rs473902	0.079	-0.050	3.62E-04	0.6161	1.6
41	rs3782089	0.060	-0.057	3.67E-04	0.6317	1.6
42	rs9863706	0.216	-0.033	3.69E-04	0.6374	1.6
43	rs572169	0.311	0.030	3.73E-04	0.6494	1.5
44	rs2145998	0.499	0.027	3.75E-04	0.6564	1.5

	SNP	MAF	Mean Difference	Standardized Effect size	Power	Estimated number of loci†
45	rs2856321	0.358	0.029	3.76E-04	0.6587	1.5
46	rs2336725	0.450	0.028	3.85E-04	0.6842	1.5
47	rs10799445	0.235	-0.033	3.91E-04	0.6998	1.4
48	rs11684404	0.349	0.029	3.93E-04	0.7038	1.4
49	rs6439167	0.204	-0.035	3.97E-04	0.7163	1.4
50	rs1490384	0.499	0.028	3.98E-04	0.7169	1.4
51	rs11830103	0.213	0.035	4.01E-04	0.7253	1.4
52	rs2093210	0.405	0.029	4.08E-04	0.7431	1.3
53	rs2066807	0.070	0.058	4.43E-04	0.8183	1.2
54	rs2237886	0.099	0.050	4.53E-04	0.8359	1.2
55	rs4282339	0.203	-0.038	4.57E-04	0.8426	1.2
56	rs889014	0.360	-0.032	4.72E-04	0.8669	1.2
57	rs5742915	0.451	0.031	4.73E-04	0.8682	1.2
58	rs274546	0.406	-0.032	4.88E-04	0.8890	1.1
59	rs3764419	0.388	-0.032	4.89E-04	0.8910	1.1
60	rs1173727	0.407	0.032	4.94E-04	0.8974	1.1
61	rs227724	0.355	0.034	5.29E-04	0.9332	1.1
62	rs9969804	0.447	0.033	5.52E-04	0.9500	1.1
63	rs17511102	0.084	0.061	5.64E-04	0.9576	1.1
64	rs10770705	0.326	0.036	5.66E-04	0.9589	1.0
65	rs1950500	0.287	0.038	5.79E-04	0.9655	1.0
66	rs1708299	0.300	0.038	5.91E-04	0.9709	1.0
67	rs3812163	0.448	0.035	6.13E-04	0.9788	1.0
68	rs17318596	0.347	0.037	6.27E-04	0.9828	1.0
69	rs4986172	0.349	-0.037	6.35E-04	0.9848	1.0
70	rs7274811	0.240	-0.042	6.50E-04	0.9878	1.0
71	rs543650	0.415	-0.037	6.76E-04	0.9919	1.0
72	rs2079795	0.322	0.040	6.88E-04	0.9933	1.0
73	rs6457620	0.482	0.037	6.98E-04	0.9944	1.0
74	rs2145272	0.346	0.040	7.06E-04	0.9950	1.0
75	rs9967417	0.409	0.039	7.16E-04	0.9958	1.0
76	rs798489	0.301	-0.042	7.43E-04	0.9973	1.0
77	rs2780226	0.081	0.072	7.62E-04	0.9981	1.0
78	rs26868	0.455	0.040	7.74E-04	0.9984	1.0
79	rs2871865	0.115	-0.062	7.81E-04	0.9986	1.0
80	rs17346452	0.269	0.045	7.83E-04	0.9987	1.0
81	rs2580816	0.183	-0.051	7.91E-04	0.9988	1.0
82	rs10037512	0.449	-0.040	7.96E-04	0.9989	1.0
83	rs1046934	0.363	0.042	7.96E-04	0.9989	1.0
84	rs6569648	0.231	0.047	7.99E-04	0.9990	1.0
85	rs3129109	0.379	-0.041	8.06E-04	0.9991	1.0
86	rs10748128	0.351	0.042	8.07E-04	0.9991	1.0
87	rs9428104	0.241	-0.048	8.39E-04	0.9995	1.0
88	rs16942341	0.029	-0.124	8.52E-04	0.9996	1.0
89	rs11107116	0.228	0.050	8.68E-04	0.9997	1.0
90	rs7460090	0.127	-0.064	9.18E-04	0.9999	1.0
91	rs237743	0.214	0.053	9.26E-04	0.9999	1.0

	SNP	MAF	Mean Difference	Standardized Effect size	Power	Estimated number of loci†
92	rs6470764	0.189	-0.056	9.72E-04	1.0000	1.0
93	rs4800452	0.214	-0.056	1.05E-03	1.0000	1.0
94	rs788867	0.320	0.050	1.07E-03	1.0000	1.0
95	rs2665838	0.268	0.052	1.07E-03	1.0000	1.0
96	rs7759938	0.309	0.051	1.10E-03	1.0000	1.0
97	rs806794	0.315	-0.051	1.10E-03	1.0000	1.0
98	rs11259936	0.481	-0.047	1.12E-03	1.0000	1.0
99	rs11205277	0.416	0.048	1.13E-03	1.0000	1.0
100	rs2284746	0.482	-0.049	1.20E-03	1.0000	1.0
101	rs7763064	0.291	-0.055	1.23E-03	1.0000	1.0
102	rs3791675	0.236	-0.059	1.25E-03	1.0000	1.0
103	rs2256183	0.450	0.051	1.29E-03	1.0000	1.0
104	rs42235	0.308	0.062	1.62E-03	1.0000	1.0
105	rs6449353	0.149	-0.081	1.65E-03	1.0000	1.0
106	rs3118905	0.289	-0.063	1.65E-03	1.0000	1.0
107	rs7689420	0.165	-0.080	1.76E-03	1.0000	1.0
108	rs143384	0.435	0.061	1.83E-03	1.0000	1.0
109	rs1351394	0.496	0.073	2.63E-03	1.0000	1.0
110	rs724016	0.443	0.075	2.78E-03	1.0000	1.0
Estimated # of total loci				697.3		
Total phenotypic variance explained (%)						15.7
Total genotypic variance explained (%)						19.6

† Projections are made only for effect sizes for SNPs that reached $P<5x10^{-8}$ in Stage 1 and had at least 1% power.

Supplementary Table 5. Dominant, recessive and dominance deviation results for nominally significant (dominance deviation *P*<0.05) lead SNPs at the 207 loci with *P*<5x10^{-6} in Stage 1. The effect allele is the height increasing allele from Stage 1. Only SNPs with a dominance deviation *P*<0.05 are presented. The analysis is based on a subset of 103,034 individuals from Stage 1. None of the results remain significant at *P*<0.05 after correcting for the number of tests performed.

SNP	Effect allele	Other Allele	Additive beta (SE)	Additive P	Dominant beta (SE)	Dominant P	Recessive beta (SE)	Recessive P	Dom Dev beta (SE)	Dom Dev P
rs1047014	C	T	0.031 (0.006)	2.2x10^{-08}	0.041 (0.006)	9.3x10^{-11}	0.017 (0.013)	0.1991	0.026 (0.008)	0.002
rs17122670	A	G	0.032 (0.007)	8.8x10^{-06}	0.038 (0.008)	3.6x10^{-07}	-0.015 (0.027)	0.5645	0.044 (0.015)	0.003
rs425277	T	C	0.027 (0.005)	7.5x10^{-08}	0.023 (0.006)	2.0x10^{-04}	0.064 (0.012)	3.0x10^{-08}	-0.021 (0.008)	0.005
rs12982744	G	C	0.033 (0.005)	2.0x10^{-12}	0.049 (0.007)	9.1x10^{-14}	0.029 (0.009)	6.8x10^{-04}	0.018 (0.006)	0.006
rs1257763	A	G	0.077 (0.014)	4.0x10^{-08}	0.054 (0.012)	1.3x10^{-05}	0.358 (0.094)	1.3x10^{-04}	-0.132 (0.049)	0.007
rs2408058	G	A	0.035 (0.006)	8.4x10^{-09}	0.102 (0.02)	2.5x10^{-07}	0.033 (0.007)	2.5x10^{-06}	0.031 (0.012)	0.008
rs1708299	A	G	0.046 (0.005)	3.2x10^{-22}	0.048 (0.006)	2.0x10^{-14}	0.087 (0.01)	6.0x10^{-17}	-0.019 (0.007)	0.009
rs13177718	C	T	0.046 (0.009)	1.8x10^{-07}	0.164 (0.04)	4.3x10^{-05}	0.038 (0.009)	1.6x10^{-05}	0.055 (0.022)	0.012
rs9456307	T	A	0.056 (0.01)	1.7x10^{-08}	0.221 (0.053)	3.3x10^{-05}	0.05 (0.01)	8.5x10^{-07}	0.071 (0.028)	0.013
rs7601531	T	C	0.022 (0.005)	2.8x10^{-06}	0.041 (0.008)	6.5x10^{-07}	0.018 (0.007)	7.0x10^{-03}	0.015 (0.006)	0.019
rs2341459	T	C	0.031 (0.005)	7.1x10^{-10}	0.041 (0.006)	5.3x10^{-11}	0.028 (0.012)	1.9x10^{-02}	0.017 (0.008)	0.028
rs34651	C	T	0.057 (0.009)	4.0x10^{-11}	0.058 (0.009)	1.5x10^{-11}	0.017 (0.038)	0.65	0.044 (0.021)	0.032
rs1351394	T	C	0.059 (0.004)	3.1x10^{-40}	0.069 (0.007)	8.6x10^{-22}	0.085 (0.007)	1.7x10^{-32}	-0.013 (0.006)	0.032
rs17318596	A	G	0.032 (0.005)	3.5x10^{-11}	0.043 (0.006)	1.7x10^{-11}	0.029 (0.009)	8.4x10^{-04}	0.014 (0.007)	0.036
rs4072910	G	C	0.033 (0.006)	6.0x10^{-09}	0.051 (0.009)	1.2x10^{-08}	0.028 (0.008)	5.1x10^{-04}	0.015 (0.007)	0.036
rs42235	T	C	0.065 (0.005)	2.5x10^{-38}	0.067 (0.006)	2.5x10^{-26}	0.108 (0.011)	3.7x10^{-24}	-0.015 (0.007)	0.037
rs11648796	G	A	0.028 (0.006)	2.3x10^{-06}	0.034 (0.007)	3.6x10^{-07}	0.019 (0.014)	0.1882	0.019 (0.009)	0.040
rs10799445	A	C	0.028 (0.005)	1.0x10^{-07}	0.023 (0.014)	9.9x10^{-02}	0.037 (0.006)	7.8x10^{-09}	-0.018 (0.009)	0.043
rs822552	G	C	0.037 (0.006)	8.1x10^{-11}	0.043 (0.006)	1.6x10^{-11}	0.034 (0.013)	9.8x10^{-03}	0.017 (0.008)	0.044

Supplementary Table 6. Nominally significant (P<0.001) results for all pairwise tests between the lead SNPs at 207 loci with P<5x10^{-6} in Stage 1. The betas refer to the height increasing alleles from Stage 1. The additive effect results for each individual SNP is based on the Stage 1 meta-analysis. The results for the pairwise interaction analysis are based on a subset of 103,034 individuals from Stage 1.

Markers (SNP1/SNP2)	Additive effect SNP1	Additive effect SNP1 P	Additive effect SNP2	Additive effect SNP2 P	Pairwise interaction beta (SE)	Interaction P
rs2145998 rs6470764	0.025	2.7x10^{-08}	0.047	5.9x10^{-17}	0.036 (0.008)	7.7x10^{-06}
rs1741344 rs4800452	0.026	3.5x10^{-08}	0.048	2.4x10^{-17}	-0.035 (0.009)	3.8x10^{-05}
rs7853377 rs955748	0.026	3.1x10^{-06}	0.024	2.2x10^{-06}	0.035 (0.009)	6.1x10^{-05}
rs494459 rs6470764	0.021	4.9x10^{-06}	0.047	5.9x10^{-17}	-0.032 (0.008)	0.000104
rs3110496 rs7759938	0.023	1.6x10^{-06}	0.042	8.7x10^{-18}	-0.028 (0.007)	0.000146
rs1814175 rs9428104	0.023	2.6x10^{-06}	0.038	8.9x10^{-13}	-0.030 (0.008)	0.000163
rs3110496 rs7697556	0.023	1.6x10^{-06}	0.022	1.3x10^{-06}	0.025 (0.007)	0.000218
rs143384 rs17346452	0.064	4.9x10^{-39}	0.038	3.3x10^{-14}	-0.028 (0.008)	0.000231
rs1351164 rs6684205	0.028	3.7x10^{-07}	0.033	2.0x10^{-11}	-0.031 (0.009)	0.000356
rs1013209 rs16942341	0.029	4.5x10^{-08}	0.134	1.3x10^{-17}	0.095 (0.027)	0.000357
rs16942341 rs5017948	0.134	1.3x10^{-17}	0.027	4.7x10^{-06}	0.107 (0.030)	0.00042
rs2408058 rs6879260	0.035	2.2x10^{-08}	0.028	5.6x10^{-10}	-0.031 (0.009)	0.000507
rs806794 rs9428104	0.053	5.5x10^{-26}	0.038	8.9x10^{-13}	-0.029 (0.008)	0.000535
rs16892729 rs2154319	0.025	1.3x10^{-06}	0.034	4.3x10^{-10}	-0.031 (0.009)	0.000591
rs17081935 rs2110001	0.031	4.8x10^{-08}	0.033	9.8x10^{-10}	0.032 (0.009)	0.00062
rs4640244 rs9428104	0.028	2.0x10^{-07}	0.038	8.9x10^{-13}	-0.030 (0.009)	0.000658
rs1173727 rs2580816	0.036	4.0x10^{-15}	0.041	1.8x10^{-12}	0.029 (0.009)	0.000713
rs1046934 rs17017854	0.046	6.4x10^{-22}	0.028	4.0x10^{-06}	0.030 (0.009)	0.000817
rs2154319 rs6772112	0.034	4.3x10^{-10}	0.046	1.6x10^{-06}	0.054 (0.016)	0.00088
rs3791675 rs6684205	0.050	2.4x10^{-20}	0.033	2.0x10^{-11}	0.028 (0.008)	0.000935
rs3791675 rs7759938	0.050	2.4x10^{-20}	0.042	8.7x10^{-18}	0.027 (0.008)	0.000972

Supplementary Table 7. Height SNPs in linkage disequilibrium ($r^2 \geq 0.8$) with non-synonymous SNPs, using the HapMap phase II CEU data. For each gene, we annotated all reported isoforms.

Chr	Position	GIANT height SNP	Non-synonymous SNP	r^2	Amino acid change	Gene name	Gene isoform
1	148173037	rs11205277	rs11205303	0.89	ATG (Met) => GTG (Val) [exon6]	*MTMR11*	NM_001145862
1	148173037	rs11205277	rs11205303	0.89	ATG (Met) => GTG (Val) [exon5]	*MTMR11*	NM_181873
1	182287568	rs1046934	rs2274432	1	GGC (Gly) => GAC (Asp) [exon1]	*TSEN15*	NM_052965
1	182287568	rs1046934	rs2274432	1	GGC (Gly) => GAC (Asp) [exon1]	*TSEN15*	NM_001127394
1	182290152	rs1046934	rs1046934	1	CAA (Gln) => CAC (His) [exon2]	*TSEN15*	NM_052965
1	182290152	rs1046934	rs1046934	1	CAA (Gln) => CAC (His) [exon2]	*TSEN15*	NM_001127394
2	88656006	rs11684404	rs1805165	0.87	GCT (Ala) => TCT (Ser) [exon13]	*EIF2AK3*	NM_004836
2	88676238	rs11684404	rs13045	1	CAA (Gln) => CGA (Arg) [exon3]	*EIF2AK3*	NM_004836
2	88694388	rs11684404	rs867529	0.87	TCC (Ser) => TGC (Cys) [exon2]	*EIF2AK3*	NM_004836
2	241841521	rs12694997	rs7578199	0.88	AAT (Asn) => AGT (Ser) [exon10]	*HDLBP*	NM_005336
2	241841521	rs12694997	rs7578199	0.88	AAT (Asn) => AGT (Ser) [exon10]	*HDLBP*	NM_203346
3	56603071	rs9835332	rs7637449	0.92	CGA (Arg) => CAA (Gln) [exon10]	*CCDC66*	NM_001141947
3	56603071	rs9835332	rs7637449	0.92	CGA (Arg) => CAA (Gln) [exon10]	*CCDC66*	NM_001012506
3	56642722	rs9835332	rs9835332	1	ACA (Thr) => AGA (Arg) [exon11]	*C3orf63*	NM_015224
3	56642722	rs9835332	rs9835332	1	ACA (Thr) => AGA (Arg) [exon18]	*C3orf63*	NM_001112736
3	56691962	rs9835332	rs958755	1	CAA (Gln) => CCA (Pro) [exon1]	*C3orf63*	NM_001112736
4	57492171	rs17081935	rs3796529	1	CCA (Pro) => CTA (Leu) [exon4]	*REST*	NM_005612
5	131690961	rs274546	rs272893	1	ATA (Ile) => ACA (Thr) [exon5]	*SLC22A4*	NM_003059
5	176450403	rs422421	rs376618	0.87	CCC (Pro) => CTC (Leu) [exon3]	*FGFR4*	NM_022963
5	176450403	rs422421	rs376618	0.87	CCC (Pro) => CTC (Leu) [exon4]	*FGFR4*	NM_002011
5	176450403	rs422421	rs376618	0.87	CCC (Pro) => CTC (Leu) [exon4]	*FGFR4*	NM_213647
6	29071227	rs3129109	rs6456880	0.91	AAG (Lys) => CAG (Gln) [exon7]	*ZNF311*	NM_001010877
6	34322300	rs2780226	rs1150781	1	GGG (Gly) => GCG (Ala) [exon5]	*C6orf1*	NM_178508
6	34322300	rs2780226	rs1150781	1	GGG (Gly) => GCG (Ala) [exon5]	*C6orf1*	NM_001008704
6	34322300	rs2780226	rs1150781	1	GGG (Gly) => GCG (Ala) [exon5]	*C6orf1*	NM_001008703
6	35531864	rs6457821	rs7761870	1	TCA (Ser) => TTA (Leu) [exon2]	*FANCE*	NM_021922
6	35873021	rs6457821	rs2766597	1	CTG (Leu) => CCG (Pro) [exon1]	*CLPS*	NM_001832
6	109871228	rs1046943	rs1476387	0.96	AGG (Arg) => AGT (Ser) [exon9]	*SMPD2*	NM_003080
6	109934409	rs1046943	rs2277114	0.87	GTA (Val) => ATA (Ile) [exon35]	*AKD1*	NM_001145128
9	85807085	rs7853377	rs1982151	0.84	AAT (Asn) => AGT (Ser) [exon3]	*RMI1*	NM_024945
9	94324803	rs9969804	rs10120210	0.93	CAG (Gln) => CCG (Pro) [exon2]	*ECM2*	NM_001393
12	28303639	rs2638953	rs11049488	0.91	GCA (Ala) => ACA (Thr) [exon2]	*CCDC91*	NM_018318
12	55026949	rs2066807	rs2066807	1	ATG (Met) => ATC (Ile) [exon20]	*STAT2*	NM_005419
15	60046929	rs7178424	rs3784634	0.81	AGG (Arg) => AAG (Lys) [exon27]	*VPS13C*	NM_017684
15	60046929	rs7178424	rs3784634	0.81	AGG (Arg) => AAG (Lys) [exon29]	*VPS13C*	NM_020821
15	60046929	rs7178424	rs3784634	0.81	AGG (Arg) => AAG (Lys) [exon27]	*VPS13C*	NM_018080
15	60046929	rs7178424	rs3784634	0.81	AGG (Arg) => AAG (Lys) [exon29]	*VPS13C*	NM_001018088
15	72123686	rs5742915	rs5742915	1	TTC (Phe) => CTC (Leu) [exon9]	*PML*	NM_033238
15	82373128	rs11259936	rs4842838	1	GTG (Val) => TTG (Leu) [exon16]	*ADAMTSL3*	NM_207517
19	46595060	rs17318596	rs10853751	0.80	ACG (Thr) => ATG (Met) [exon1]	*EXOSC5*	NM_020158
19	46624115	rs17318596	rs284662	0.80	AGC (Ser) => GGC (Gly) [exon3]	*B3GNT8*	NM_198540
20	47275067	rs237743	rs11908296	0.97	GGA (Gly) => GTA (Val) [exon6]	*DDX27*	NM_017895
20	47299191	rs237743	rs6512577	1	ATG (Met) => ATA (Ile) [exon14]	*ZNFX1*	NM_021035

Supplementary Table 8. GIANT height variants associated with other traits and diseases reported in the NHGRI catalog of published GWAS at genome-wide level of significance ($P<5x10^{-8}$), based on a 1 megabase maximum distance and linkage disequilibrium ($r^2>0.1$) between the SNPs. Highlighted rows are those for which the GIANT height SNP and the NHGRI SNP showed a strong correlation ($r^2>0.8$).

GIANT height SNP	Chr	Position	Nearest or OMIM gene	GWAS SNP from NHGRI catalogue	GIANT height P-value for NHGRI SNP	Disease/Trait	r^2 between GIANT height and NHGRI SNP	D' between GIANT height and NHGRI SNP	Distance (kb) between GIANT height and NHGRI SNP	Height-increasing allele	Effect relative to height-increasing allele	Reference
rs11118346	1	217810342	***LYPLAL1***	rs2605100	1.42E-03	WHR (women)	0.17	0.69	99.495	A	lower WHR	Lindgren *et al.*, PLoS Genet 2009
rs720390	3	187031377	***IGF2BP2***	rs4402960	6.30E-01	Type 2 diabetes	0.48	0.86	36.996	T	higher T2D risk	Saxena *et al.*, Science 2007
				rs4402960	6.30E-01	Type 2 diabetes	0.48	0.86	36.996	T	higher T2D risk	Scott *et al.*, Science 2007
				rs6769511	6.45E-01	Type 2 diabetes	0.48	0.86	18.393	C	higher T2D risk	Unoki *et al.*, Nat Genet 2008
				rs4402960	6.30E-01	Type 2 diabetes	0.48	0.86	36.996	T	higher T2D risk	Zeggini *et al.*, Science 2007
rs10010325	4	106325802	***TET2***	rs7679673	2.12E-01	Prostate cancer	0.12	0.37	44.819	A	higher cancer risk	Eeles *et al.*, Nat Genet 2009
rs10037512	5	88390431	***MEF2C***	rs1366594	4.98E-09	BMD (hip)	0.97	1	21.386	A	lower BMD	Rivadeneira *et al.*, Nat Genet 2009
rs274546	5	131727766	***SLC22A5***	rs2188962	5.46E-07	Crohn's disease	0.42	1	70.938	T	higher Crohn's risk	Barrett *et al.*, Nat Genet 2008
				rs2522056	6.55E-04	Fibrinogen	0.21	0.81	101.859	G	lower fibrinogen	Dehghan *et al.*, Circ Cardiovasc Genet 2009
				rs1016988	1.34E-03	Fibrinogen	0.17	0.87	44.707	T	lower fibrinogen	Danik *et al.*, Circ Cardiovasc Genet 2009
				rs4143832	1.40E-01	Plasma eosinophil count	0.12	0.61	163.11	G	higher eosinophil count	Gudbjartsson *et al.*, Nat Genet 2009
rs2256183	6	31488508	***MICA***	rs2844479	9.64E-10	Weight	0.12	0.46	192.427	A	increased weight	Thorleifsson *et al.*, Nat Genet 2008
rs6457620	6	32771977	**HLA locus**	rs2187668	8.87E-03	Celiac disease	0.10	1	58.115	T	higher Celiac risk	van Heel *et al.*, Nat Genet 2007
				rs9271366	7.17E-06	Multiple sclerosis	0.35	1	77.145	G	higher MS risk	Bahlo *et al.*, Nat Genet 2009
				rs6457617	6.89E-08	Rheumatoid arthritis	1	1	0.148	C	lower RA risk	Julia *et al.*, Arthritis Rheum 2008
				rs6457617	6.89E-08	Rheumatoid arthritis	1	1	0.148	C	lower RA risk	WTCCC, Nature 2007
				same SNP	3.65E-08	Rheumatoid arthritis	-	-	-	G	n/a	Raychaudhuri *et al.*, Nat Genet 2008
				rs660895	4.65E-01	Rheumatoid arthritis	0.40	1	86.619	A	lower RA risk	Plenge *et al.*, N Engl J Med 2007
				rs2187668	8.87E-03	SLE	0.10	1	58.115	T	higher SLE risk	Hom *et al.*, N Engl J Med 2008
				rs9272346	6.12E-01	Type 1 diabetes	0.18	0.47	59.627	G	higher T1D risk	WTCCC, Nature 2007
				rs9272346	6.12E-01	Type 1 diabetes	0.18	0.47	59.627	G	higher T1D risk	Cooper *et al.*, Nat Genet 2008

GIANT height SNP	Chr	Position	Nearest or OMIM gene	GWAS SNP from NHGRI catalogue	GIANT height P-value for NHGRI SNP	Disease/Trait	r^2 between GIANT height and NHGRI SNP	D' between GIANT height and NHGRI SNP	Distance (kb) between GIANT height and NHGRI SNP	Height-increasing allele	Effect relative to height-increasing allele	Reference
				rs2395185	6.38E-01	Ulcerative colitis	0.22	0.59	230.832	T	lower UC risk	Silverberg *et al.*, Nat Genet 2009
				rs9268877	2.80E-01	Ulcerative colitis	0.12	0.39	232.852	G	lower UC risk	Franke *et al.*, Nat Genet 2008
				rs2395185	6.38E-01	Ulcerative colitis	0.22	0.59	230.832	T	lower UC risk	Asano *et al.*, Nat Genet 2009
				rs9268877	2.80E-01	Ulcerative colitis	0.12	0.39	232.852	G	n/a	Barrett *et al.*, Nat Genet 2009
rs7759938	6	105485647	***LIN28B***	rs314276	1.03E-16	Menarche (age at onset)	0.96	1	29.045	A	later menarche	Ong *et al.*, Nat Genet 2009
				rs314280	1.35E-10	Menarche (age at onset)	0.52	1	21.883	A	later menarche	Sulem *et al.*, Nat Genet 2009
				same SNP	8.69E-18	Menarche (age at onset)	-	-	-	C	later menarche	Perry *et al.*, Nat Genet 2009
				rs314277	3.65E-12	Menarche (age at onset)	0.25	1	28.708	A	later menarche	He *et al.*, Nat Genet 2009
rs1490384	6	126892853	***C6orf173***	rs9388489	1.03E-13	Type 1 diabetes	0.84	1	152.441	G	higher T1D risk	Barrett *et al.*, Nat Genet 2009
rs7763064	6	142838982	***GPR126***	rs3817928	1.97E-11	Pulmonary function	0.59	0.89	46.773	A	reduced pulmonary function	Hancock *et al.*, Nat Genet 2009
rs1708299	7	28156471	***JAZF1***	rs864745	1.31E-12	Type 2 diabetes	0.40	0.94	9.39	T	higher T2D risk	Zeggini *et al.*, Nat Genet 2008
rs6959212	7	38094851	***STARD3NL***	rs1524058	9.39E-05	BMD (spine)	0.73	1	7.951	C	lower BMD	Rivadeneira *et al.*, Nat Genet 2009
rs2110001	7	150147955	***KCNH2***	rs2968863	1.33E-04	QT interval	0.10	0.85	106.115	C	longer QT interval	Pfeufer *et al.*, Nat Genet 2009
				rs3807375	4.30E-05	QT interval	0.14	1.00	40.278	T	longer QT interval	Holm *et al.*, Nat Genet 2010
rs11599750	10	101795432	***CPN1***	rs11597390	1.54E-04	Liver enzymes levels	0.51	0.82	55.993	G	lower enzyme levels	Yuan *et al.*, Am J Hum Genet 2008
rs1330	11	17272605	***KCNJ11***	rs5215	6.70E-02	Type 2 diabetes	0.26	0.53	92.601	C	higher T2D risk	Zeggini *et al.*, Science 2007
rs494459	11	118079885	***DDX6***	rs4639966	2.77E-01	SLE	0.22	1	1.156	T	lower SLE risk	Han *et al.*, Nat Genet 2009
rs2066807	12	55026949	***STAT2***	rs2066808	2.75E-08	Psoriasis	1	1	2.709	G	lower psoriasis risk	Nair *et al.*, Nat Genet 2009
rs3110496	17	24941897	***ANKRD13B***	rs2138852	5.11E-01	Mean platelet volume	0.10	0.45	214.422	T	higher platelet volume	Soranzo *et al.*, Nat Genet 2009
				rs2138852	5.11E-01	Mean platelet volume	0.10	0.45	214.422	T	lower platelet volume	Meisinger *et al.*, Am J Hum Genet 2008
rs4986172	17	40571807	***ACBD4***	rs12946454	3.23E-07	Systolic blood pressure	0.76	0.94	8.16	A	lower systolic b.p.	Newton-Cheh *et al.*, Nat Genet 2009
rs2072153	17	44745013	***ZNF652***	rs16948048	4.22E-04	Diastolic blood pressure	0.32	1.00	50.452	A	lower diastolic b.p.	Newton-Cheh *et al.*, Nat Genet 2009
rs17782313	18	56002077	***MC4R***	rs12970134	5.52E-04	Body mass index	0.81	0.96	33.653	A	higher BMI	Thorleifsson *et al.*, Nat Genet 2008

GIANT height SNP	Chr	Position	Nearest or OMIM gene	GWAS SNP from NHGRI catalogue	GIANT height P-value for NHGRI SNP	Disease/Trait	r^2 between GIANT height and NHGRI SNP	D' between GIANT height and NHGRI SNP	Distance (kb) between GIANT height and NHGRI SNP	Height-increasing allele	Effect relative to height-increasing allele	Reference
				same SNP	3.48E-06	Body mass index	-	-	-	C	higher BMI	Willer *et al.*, Nat Genet 2008
				same SNP	3.48E-06	Body mass index	-	-	-	C	higher BMI	Loos *et al.*, Nat Genet 2008
				same SNP	3.48E-06	Obesity	-	-	-	C	higher obesity risk	Meyre *et al.*, Nat Genet 2009
				rs12970134	5.52E-04	Waist circumference	0.81	0.96	33.653	A	lower WC	Chambers *et al.*, Nat Genet 2008
				rs12970134	5.52E-04	Weight	0.81	0.96	33.653	A	increased weight	Thorleifsson *et al.*, Nat Genet 2008
rs2834442	21	34612656	***KCNE2***	rs9982601	4.66E-01	MI (early onset)	0.17	0.66	91.658	C	lower MI risk	Kathiresan *et al.*, Nat Genet 2009

Supplementary Table 9. List of 241 abnormal skeletal/growth genes identified in the OMIM database using the following keywords: short stature, overgrowth, skeletal dysplasia, brachydactyly, and manually curating the list blind to GIANT height results.

ACAN	*COL9A3*	*GJA1*	*NEU1*	*SIL1*
ADAMTS10	*COMP*	*GLB1*	*NF1*	*SLC26A2*
ADAMTS2	*CRTAP*	*GLI3*	*NIPBL*	*SLC29A3*
ADAMTSL2	*CTDP1*	*GNAS*	*NOG*	*SLC2A2*
AGPS	*CTSK*	*GNPAT*	*NPR2*	*SLC34A3*
ALG12	*CUL4B*	*GNPTAB*	*NSD1*	*SLC35C1*
ALMS1	*CUL7*	*GPC3*	*OCRL*	*SLC35D1*
ALPL	*CYP11B1*	*GUSB*	*OFD1*	*SLC37A4*
ANKH	*CYP19A1*	*HCCS*	*PAPSS2*	*SLC39A13*
ARL6	*CYP21A2*	*HESX1*	*PAX3*	*SLC4A4*
ARSB	*CYP27B1*	*HMGA2*	*PAX8*	*SLC6A8*
ARSE	*DHCR7*	*HOXD13*	*PCNT*	*SMARCAL1*
ATP6V0A2	*DYM*	*HPRT1*	*PEX7*	*SMC1A*
ATP7A	*EBP*	*HRAS*	*PHEX*	*SMC3*
ATP8B1	*EFNB1*	*HSPG2*	*PHF6*	*SMPD1*
ATR	*EIF2AK3*	*HYAL1*	*PITX2*	*SMS*
ATRX	*ERCC2*	*ICK*	*POU1F1*	*SOS1*
B3GALTL	*ERCC3*	*IDUA*	*PQBP1*	*SOST*
B4GALT7	*ESCO2*	*IFT80*	*PROP1*	*SOX3*
BBS1	*EVC*	*IGBP1*	*PTCH1*	*SPG20*
BBS10	*EVC2*	*IGF1*	*PTCH2*	*SRY*
BBS12	*EXT1*	*IGF1R*	*PTEN*	*STAT5B*
BBS2	*EXT2*	*IGF2*	*PTH1R*	*TAZ*
BBS4	*FANCA*	*IHH*	*PTPN11*	*TBCE*
BBS5	*FANCB*	*IKBKG*	*RAB23*	*TBX1*
BBS7	*FANCC*	*JAG1*	*RAB3GAP1*	*TBX15*
BBS9	*FANCD2*	*KCNJ2*	*RAB3GAP2*	*TCF4*
BMPR1B	*FANCE*	*KDM5C*	*RAF1*	*TGFBR1*
BRAF	*FANCF*	*KIAA1279*	*RAI1*	*TGFBR2*
BRCA2	*FANCG*	*KRAS*	*RBM28*	*THRB*
BTK	*FANCI*	*LBR*	*RECQL4*	*TNFRSF11B*
BUB1B	*FANCL*	*LEMD3*	*RMRP*	*TP63*
C7orf11	*FANCM*	*LEPRE1*	*RNF135*	*TRAPPC2*
CA2	*FBN1*	*LHX4*	*ROR2*	*TRIM32*
CCDC28B	*FBN2*	*LIFR*	*RPL11*	*TRIM37*
CEP290	*FGD1*	*LIG4*	*RPL35A*	*TRPS1*
CHD7	*FGF23*	*LMNA*	*RPL5*	*TRPV4*
CHRNG	*FGFR2*	*LRP5*	*RPS17*	*UBR1*
CHST3	*FGFR3*	*MAP2K1*	*RPS19*	*WNT7A*
CLCN5	*FLNA*	*MAP2K2*	*RPS24*	*WRN*
COL10A1	*FLNB*	*MATN3*	*RPS6KA3*	*ZBTB16*
COL11A1	*FOXC1*	*MC4R*	*RPS7*	
COL11A2	*FUCA1*	*MECP2*	*RUNX2*	
COL1A1	*G6PC*	*MGP*	*SBDS*	
COL1A2	*GALNS*	*MKKS*	*SDHA*	
COL2A1	*GDF5*	*MKS1*	*SECISBP2*	
COL5A1	*GH1*	*MMP13*	*SEMA3E*	
COL5A2	*GHR*	*MRPS16*	*SHH*	
COL9A1	*GHRHR*	*MYCN*	*SHOX*	
COL9A2	*GHSR*	*NBN*	*SHROOM4*	

Supplementary Table 10. Height SNPs found to be located near or in the abnormal skeletal/growth genes identified in the OMIM database.

SNP	Abnormal skeletal/growth gene (OMIM)	The closest gene to the height SNP is the abnormal skeletal/growth gene	The height SNP is in the abnormal skeletal/growth gene
rs16942341	*ACAN*	yes	yes
rs4072910	*ADAMTS10*	yes	no
rs16964211	*CYP19A1*	yes	yes
rs9967417	*DYM*	yes	yes
rs11684404	*EIF2AK3*	yes	yes
rs6457821	*FANCE*	no	no
rs143384	*GDF5*	yes	yes
rs2665838	*GH1*	no	no
rs572169	*GHSR*	yes	yes
rs7971536	*GNPTAB*	no	no
rs1351394	*HMGA2*	yes	yes
rs2871865	*IGF1R*	yes	yes
rs12470505	*IHH*	yes	no
rs17782313	*MC4R*	yes	no
rs227724	*NOG*	yes	no
rs422421	*NSD1*	no	no
rs473902	*PTCH1*	yes	yes
rs3764419	*RNF135*	no	no
rs10874746	*RPL5*	no	no
rs9472414	*RUNX2*	no	no
rs10838801	*SLC39A13*	no	no

Supplementary Table 11. Nominally significant biological pathways following gene set enrichment analysis of height meta-analysis.

Database	Biological pathway or gene-set	Original # genes in gene-set	# genes in gene-set analyzed by GSEA[§]	# genes in gene set ≤300kb from validated height SNPs	Nominal GSEA *P*-value	False discovery rate (FDR)	Genes 300 kb or less from validated height SNPs
KEGG	Hedgehog signaling pathway	54	50	9	0.0009	0.0777*	BMP6, IHH, PTCH1, WNT6, WNT9A, FBXW11, HHIP, WNT10A, WNT3A
KEGG	Gamma-hexachlorocyclohexane degradation	26	21	2	0.0028	0.0568*	DHRS1, LOC283871
KEGG	MAPK signaling pathway	269	243	23	0.0040	0.2796	ARRB1, CACNB1, CHUK, FGFR3, FGFR4, GNA12, MKNK2, MEF2C, MAP3K3, MOS, GADD45B, NF1, NFATC4, PPM1A, MAPK9, MAP2K3, RASA2, RPS6KA1, TGFB1, TGFB2, TNF, MAP3K14, RASGRP3
KEGG	Antigen processing and presentation	77	52	16	0.0132	0.3014	HLA-B, HLA-C, HLA-DMA, HLA-DMB, HLA-DOB, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB5, LTA, PSME1, PSME2, TAP1, TAP2
KEGG	TGF-beta signaling pathway	83	80	10	0.0167	0.3131	AMH, BMP6, ID4, LTBP1, TGFB1, TGFB2, TNF, GDF5, CUL1, NOG
KEGG	Type II diabetes mellitus	45	43	9	0.0172	0.2934	INSR, KCNJ11, PKM2, PRKCD, PRKCZ, MAPK9, ABCC8, TNF, SOCS2
KEGG	FC epsilon RI signaling pathway	79	73	8	0.0282	0.3237	CSF2, IL5, IL13, LYN, PRKCD, MAPK9, MAP2K3, TNF
KEGG	Folate biosynthesis	37	36	1	0.0305	0.3549	ATP13A2
KEGG	Citrate cycle TCA cycle	27	26	5	0.0417	0.3464	CS, PC, PCK2, SDHB, SUCLG2
Ingenuity	Hepatic Cholestasis	61	57	12	0.0237	1.5076	ABCC2, CYP27A1, ESR1, FGFR4, INSR, SLC4A2, TAP1, TAP2, TNF, MAP3K14, SLCO1B3, SLCO1C1
Ingenuity	VDR/RXR Activation	63	62	6	0.0341	0.6864	CSF2, GTF2B, PPARD, PSMC5, TGFB2, NCOA1
Ingenuity	Role of BRCA1 in DNA Damage Response	29	29	4	0.0507	0.4980	BRCA2, FANCC, FANCE, RAD50
Ingenuity	Fc Epsilon RI Signaling	20	17	5	0.0531	0.8701	CSF2, IL5, IL13, LYN, TNF

Database	Biological pathway or gene-set	Original # genes in gene-set	# genes in gene-set analyzed by GSEA[§]	# genes in gene set ≤300kb from validated height SNPs	Nominal GSEA *P*-value	False discovery rate (FDR)	Genes 300 kb or less from validated height SNPs
PANTHER	TGF-beta signaling pathway	64	59	8	0.0025	0.1844*	AMH, BMP3, BMP6, SKI, TGFB1, TGFB2, GDF5, DCP1A
PANTHER	Hedgehog signaling pathway	14	14	3	0.0042	0.2033*	IHH, PTCH1, FBXW11
PANTHER	Apoptosis signaling pathway	53	49	10	0.0109	0.2250*	BOK, CMA1, CTSG, GZMH, GZMB, LTA, LTB, TNFSF10, MAP3K14, RIPK3
PANTHER	Endothelin signaling pathway	19	19	4	0.0144	0.2319*	ADCY3, EDN2, PRKG2, ADCY4
PANTHER	Parkinson disease	43	41	5	0.0171	0.2348*	LYN, SEPT2, PSMB3, CUL1, STUB1
PANTHER	B cell activation	24	22	4	0.0223	0.2368*	CD79B, NFKBIL1, PRKCD, PRKCZ
PANTHER	Nicotinic acetylcholine receptor signaling pathway	42	39	3	0.0453	0.3888	MYO1F, MYO6, MYO9B
PANTHER, MF	Histone	86	31	29	0.0001†	0.0028*	HIST1H1C, HIST1H1D, HIST1H1E, HIST1H1T, HIST1H2AE, HIST1H2AD, HIST1H1A, HIST1H2AC, HIST1H2AB, HIST2H2AC, HIST1H3A, HIST1H3D, HIST1H3C, HIST1H3E, HIST1H3G, HIST1H3B, HIST1H4A, HIST1H4D, HIST1H4F, HIST1H4C, HIST1H4H, HIST1H4B, HIST1H4E, HIST1H4G, HIST1H3F, H1FX, H1FOO, HIST2H2AB, HIST2H3D
PANTHER, MF	Extracellular matrix glycoprotein	111	85	16	0.0015	0.1157*	ACAN, FBLN2, EFEMP1, GPC5, GP9, LTBP1, LTBP2, LTBP3, MFAP2, MSLN, FBLN5, EFEMP2, ADAMTSL3, HAPLN3, SCUBE3, MPFL
PANTHER, MF	Annexin	71	64	13	0.0038	0.1821*	AIF1, FBLN2, EFEMP1, LETM1, LTBP1, LTBP2, LTBP3, NUCB2, PRKCD, PKN2, PRKCZ, FBLN5, EFEMP2
PANTHER, MF	Transcription factor	198	127	13	0.0041	0.2089*	NR2F6, ESR1, NFIC, PPARD, BAT2, YEATS4, NCOA1, SCMH1, SFMBT1, MBTD1, GATAD1, L3MBTL3, VGLL2
PANTHER, MF	Exoribonuclease	35	25	7	0.0069	0.2009*	ISG20, PAN2, EXOSC2, EXOSC5, CNOT6, ISG20L1, PNPT1,

Database	Biological pathway or gene-set	Original # genes in gene-set	# genes in gene-set analyzed by GSEA[§]	# genes in gene set ≤300kb from validated height SNPs	Nominal GSEA *P*-value	False discovery rate (FDR)	Genes 300 kb or less from validated height SNPs
PANTHER, MF	Other transcription factor	349	298	30	0.0117	0.3260	RUNX3, E2F1, E2F2, ETS1, ETV5, ETV6, FLI1, ID4, IRF1, MEF2C, ATXN3, NFATC4, NRL, PA2G4, RELA, SKI, SNAPC4, STAT2, TEAD1, TEAD3, TBX4, CREB5, IRF9, FEV, UTP6, GNPTAB, LIN28, RFXDC1, FOXR1, LIN28B
PANTHER, MF	Metalloprotease	158	133	10	0.0158	0.4290	CPN1, PAPPA, ADAM7, ADAMTS3, ADAM28, MMP24, PMPCA, ADAMDEC1, PAPPA2, ADAMTS10
PANTHER, MF	Ligase	69	57	5	0.0206	0.3908	CTPS, DCI, SUCLG2, ZMIZ1, GDPD5
PANTHER, MF	ATP-binding cassette (ABC) transporter	46	34	7	0.0240	0.4536	ABCA3, ABCC2, TAP1, TAP2, ABCB6, RAD50, ABCB8
PANTHER, MF	Other phosphatase	82	71	8	0.0260	0.3693	PPAP2A, FIG4, MTMR11, NUDT4, NUDT3, INPP5E, ACPL2, LOC283871
PANTHER, MF	Damaged DNA-binding protein	27	25	3	0.0327	0.3545	BRCA2, RAD50, UTP6
PANTHER, MF	Other RNA-binding protein	192	151	14	0.0382	0.4413	CARS, STAU1, SLBP, FUBP1, FUBP3, IGF2BP3, IGF2BP2, CPSF6, HNRPUL1, ANKZF1, BRUNOL5, RBM45, ZFAND2B, C14orf21
PANTHER, MF	Major histocompatibility complex antigen	46	26	14	0.0386	0.3567	HFE, HLA-B, HLA-C, HLA-DMA, HLA-DMB, HLA-DOB,HLA-DQA1, HLA-DQA2,HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB5, MICA, MICB
PANTHER, MF	Kinase	30	29	2	0.0518	0.4439	DGKE, DCAKD
GO:0005694	Chromosome	147	111	29	5e-5†	0.0905*	HIST1H1C, HIST1H1D, HIST1H1E, HIST1H1T, HIST1H2AD, HIST1H2BD, HIST1H2BB, HIST1H1A, HMGA1, HMGA2, HIST1H2AC, HIST1H2AB, HIST2H2AC, HIST1H2BH, HIST1H2BC, HIST2H2BE, HIST1H4G, HIST1H3F, H1FX, RAD50, TINF2, CENPO, H1FOO, HIST2H2AB, C6orf173, SETD8, CENPP, HIST2H2BF, HIST2H3D
GO:0060389	Pathway-restricted SMAD protein phosphorylation	14	14	3	0.0001	0.0984*	BMP6, TGFB1, TGFB2

Database	Biological pathway or gene-set	Original # genes in gene-set	# genes in gene-set analyzed by GSEA[§]	# genes in gene set ≤300kb from validated height SNPs	Nominal GSEA *P*-value	False discovery rate (FDR)	Genes 300 kb or less from validated height SNPs
GO:0000786	Nucleosome	64	34	21	0.0002	0.0650*	HIST1H1C, HIST1H1D, HIST1H1E, HIST1H1T, HIST1H2AD, HIST1H2BD, HIST1H2BB, HIST1H1A, HIST1H2AC, HIST1H2AB, HIST2H2AC, HIST1H2BH, HIST1H2BC, HIST2H2BE, HIST1H4G, HIST1H3F, H1FX, H1FOO, HIST2H2AB, HIST2H2BF, HIST2H3D
GO:0006334	Nucleosome assembly	80	47	22	0.0003	0.0966*	HIST1H1C, HIST1H1D, HIST1H1E, HIST1H1T, HIST1H2AD, HIST1H2BD, HIST1H2BB, HIST1H1A, NAP1L4, HIST1H2AC, HIST1H2AB, HIST2H2AC, HIST1H2BH, HIST1H2BC, HIST2H2BE, HIST1H4G, HIST1H3F, H1FX, H1FOO, HIST2H2AB, HIST2H2BF, HIST2H3D
GO:0050680	Negative regulation of epithelial cell proliferation	22	21	9	0.0006	0.0705*	RUNX3, CDK6, CDKN1C, PPARD, PTCH1, TGFB1, TGFB2, TSC2, TINF2
GO:0009653	Anatomical structure morphogenesis	103	94	15	0.0007	0.1567*	CHUK, RPL10A, NEDD8, PITX1, PKD1, POU5F1, PTCH1, WHSC1, LST1, IGF2BP3, IGF2BP2, RCAN3, SCMH1, SIX4, WNT3A
GO:0032147	Activation of protein kinase activity	10	10	3	0.0008	0.0994*	INSR, TGFB2, LYK5
GO:0002474	Antigen processing and presentation of peptide antigen via MHC class I	11	10	3	0.0010	0.0764*	HFE, HLA-B, HLA-C
GO:0000175	3'-5'-exoribonuclease activity	10	10	4	0.0013	0.0679*	ISG20, EXOSC2, EXOSC5, PNPT1
GO:0007259	JAK-STAT cascade	27	26	7	0.0016	0.1300*	FGFR3, GH1, IL6ST, PKD1, STAT2, SOCS2, IL31RA
GO:0003007	Heart morphogenesis	29	28	4	0.0021	0.1520*	INSR, PTCH1, TGFB2, ZMIZ
GO:0030879	Mammary gland development	29	29	5	0.0023	0.1692*	BRCA2, IGF1R, PTCH1, TGFB1, WNT3A
GO:0043560	Insulin receptor substrate binding	12	12	4	0.0024	0.0888*	IGF1R, INSR, PRKCD, PRKCZ
GO:0000421	Autophagic vacuole membrane	13	12	3	0.0026	0.0907*	TM9SF1, ATG9A, ATG9B
GO:0005578	Proteinaceous extracellular matrix	225	198	25	0.0026	0.2502*	ACAN, ECM2, FBLN2, EFEMP1, GPC5, LOXL1, LTBP1, LTBP2, MFAP2, NTN2L, OMD, OGN, TGFB1, WNT6, WNT9A, ADAMTS3, FBLN5, MMP24, ANGPTL4, ASPN, ADAMTSL3, WNT10A, ADAMTS10, WNT3A, HAPLN3
GO:0007405	Neuroblast proliferation	14	14	3	0.0035	0.1436*	ID4, FRS2, HHIP

Database	Biological pathway or gene-set	Original # genes in gene-set	# genes in gene-set analyzed by GSEA[§]	# genes in gene set ≤300kb from validated height SNPs	Nominal GSEA *P*-value	False discovery rate (FDR)	Genes 300 kb or less from validated height SNPs
GO:0000080	G1 phase of mitotic cell cycle	10	7	4	0.0041	0.0909*	CDK6, CDKN1C, E2F1, MAP3K11
GO:0032355	Response to estradiol stimulus	53	53	7	0.0041	0.2674	GH1, IHH, INSR, NOS3, PTCH1, TGFB1, SOCS2
GO:0031965	Nuclear membrane	86	81	12	0.0058	0.3132	ABL1, MYO6, PML, TRIM27, NUPL2, NUP210, TMEM176B, DTL, INTS2, SENP2, QSOX2, LASS3
GO:0005743	Mitochondrial inner membrane	254	222	19	0.0059	0.3414	CYP27A1, DCI, LETM1, NDUFA7, NDUFB1, NDUFB10, PC, PHB, SDHB, SLC3A1, PPIF, ACAA2, ATP5L, ABCB8, PMPCA, C4orf14, COQ10A, DHRS1, SLC25A45
GO:0010628	Positive regulation of gene expression	27	25	3	0.0059	0.2293*	CSF2, MAPK9, TGFB1
GO:0000398	Nuclear mRNA splicing, via spliceosome	45	45	7	0.0061	0.2757	HNRPM, SFRS10, BAT1, SF3A2, SF3B4, TRA2A, LSM7
GO:0005242	Inward rectifier potassium channel activity	18	16	5	0.0061	0.1952*	KCNH2, KCNJ1, KCNJ2, KCNJ5, KCNJ12
GO:0017148	Negative regulation of translation	16	16	3	0.0063	0.1969*	EIF2AK3, IGF2BP3, IGF2BP2
GO:0016604	Nuclear body	16	16	3	0.0067	0.1936*	SKI, PCGF2, BTBD14A
GO:0042612	MHC class I protein complex	22	16	6	0.0076	0.2048*	HFE, HLA-B, HLA-C, MICA, MICB, PROCR
GO:0007067	Mitosis	192	180	16	0.0081	0.3591	CCNF, E4F1, SEPT2, YEATS4, HMGA2, TIMELESS, STAG1, SSSCA1, RGS14, PDS5B, FZR1, NCAPG, NUP37, FAM44B, NY-SAR-48, SETD8
GO:0007569	Cell aging	27	26	3	0.0083	0.2728	BRCA2, PML, ZMIZ1
GO:0060395	SMAD protein signal transduction	10	9	2	0.0085	0.1814*	BMP6, SKI
GO:0001763	Morphogenesis of a branching structure	10	9	2	0.0087	0.1781*	IHH, TGFB1
GO:0001501	Skeletal system development	126	113	11	0.0088	0.3536	PCSK5, NOG

Nominal gene set enrichment analysis (GSEA) *p*-values and false discovery rates are presented for the nominally significant pathways ($p<0.01$ for Gene Ontology (GO) and $p<0.05$ for the other databases), using MAGENTA (Segrè *et al*., in revision) applied to the height meta-analysis. The Bonferroni corrected cutoffs for the different databases are: KEGG (135 pathways): $p<0.0004$, Ingenuity (81 pathways): $p<0.0006$, PANTHER (94 pathways): $p<0.0005$, PANTHER, MF (Molecular Function classification; 217 gene sets): $p<0.0002$, and Gene Ontology (GO) biological process and molecular function terms (1,785 gene sets): $p<0.00003$.

† specifies a gene set that passes or is close to the Bonferroni cutoff. Since Bonferroni correction is stringent due to considerable gene overlap between pathways within each database, we further evaluated the statistical significance of each gene-set using a false discovery rate (FDR). FDR is defined for gene set *g* as the fraction of all randomized gene sets generated for all GSEA tests (10,000 permutations times the total number of gene sets tested) whose score is more significant than that of gene-set *g* divided by the fraction of tested gene-sets whose score is more significant than that of gene set *g*. A gene-set score refers here to the fraction of genes in a gene set whose gene *p*-value exceeds the 95th percentile of all gene *p*-values. For FDR calculation purposes, the differences in gene set size across the observed and randomized gene sets were accounted for by subtracting from each gene-set score the mean score of all randomized genes-sets of identical size, and dividing by the standard deviation of scores of all randomized genes-sets of identical size.

An asterisk (*) refers to pathways with an FDR<0.25 (i.e. one in four pathways more significant than the given gene-set is likely to be false).

§ The number of genes per gene-set analyzed by MAGENTA refers to the gene-set size after removal of genes with no SNPs in their gene region, and removal of all but one gene in each subset of genes in a given pathway that were assigned the same best local SNP due to physically proximity in the genome. Gene set size was restricted to between 10 and 1,000 genes. All genes within 300 kb of the validated height SNPs are listed, including those removed due to physical clustering adjustment.

Supplementary Methods Table 1. Study design, number of individuals and sample quality control for genome-wide association study cohorts

Study		Study design	Total sample size (N)	Sample QC		Samples in analyses (N)	Anthropo-metric assessment method	References
Short name	Full name			Call rate*	other exclusions			
Stage 1 (GWA studies)								
ADVANCE	Atherosclerotic Disease, VAscular FunctioN, and GenetiC Epidemiology	Population-based case-control (multi ethnic)	599 (Europeans)	>98.5%	1) duplicates 2) missing weight or height	584: 275 cases 309 ctrls	measured	Assimes T.L. et al. Susceptibility locus for clinical and subclinical coronary artery disease at chromosome 9p21 in the multi-ethnic ADVANCE study. Hum Mol Genet. (2008) 17(15):2320-8.
AGES	Age, Gene/Environment Susceptibility-Reykjavik Study	Population-based	3219	≥ 97%	1) mismatch with previous genotypes; 2) remove A/T & G/C SNPs; 3) remove SNPs not in HapMap	3219	measured	Harris T.B. et al. Age, Gene/Environment Susceptibility-Reykjavik Study: multidisciplinary applied phenomics. American Journal of Epidemiology (2007) 165 (9): 1076-87
Amish HAPI Heart Study	Amish Heredity and Phenotype Intervention Heart Study	Founder population	918	≥ 93%	1) Misidentified pedigree relationships 2) Misidentified sex	907	measured	Mitchell B.D. et al. The genetic response to short-term interventions affecting cardiovascular function: Rationale and design of the Heredity and Phenotype Intervention (HAPI) Heart Study. Am Heart J (2008) 823:828,
ARIC	Atherosclerosis Risk in Communities Study	Population-based	8861 (whites)	≥ 90%	1) True sex mismatch 2) Discordant genotype with earlier TaqMan genotyping. If >10/47 genotypes discordant -> exclude 3) First-degree relative 4) PC>8SD in Eigenstrat run (10 iterations with 10 PCs) 5) Outlier based on average IBS 6) missing height or other covariate	8110	measured	(1) The ARIC Investigators. Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. Am. J. Epidemiol. (1989) 129: 687-702. (2) Heard-Costa N.L. et al. NRXN3 is a novel locus for waist circumference: a genome-wide association study from the CHARGE Consortium. Plos Genet. (2009) 5(6): e1000539.
B58C-T1DGC	British 1958 birth cohort (Type 1 Diabetes Genetic Consortium controls)	Population-based	2592	≥ 98%	1) contamination; 2) non-European identity; 3) Missing body height.	2591	measured	(1) Strachan D.P. et al. Lifecourse influences on health among British adults: effects of region of residence in childhood and adulthood. Int J Epidemiol (2007) 36:522-531 (2) Barrett J.C. et al. The Type 1 Diabetes Genetics Consortium. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet (2009) 41:703-707

B58C-WTCCC	British 1958 birth cohort (Wellcome Trust Case Control Consortium controls)	Population-based birth cohort	1502	≥97%	1) contamination; 2) non-European identity and relatedness; 3) Missing body height.	1479	measured	The Wellcome Trust Case Control Consortium Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature (2007) 447: 661-678
BRIGHT	British Genetic of Hypertension (BRIGHT) study	Hypertension cases	2000	≥ 97%	1) heterozygosity <23% or >30%; 2) external discordance; 3) non-European ancestry; 4) duplicate/first/second degree relatives.	1806	measured	Caulfield M. et al. Genome-wide mapping of human loci for essential hypertension. Lancet.(2003) 361:2118-23.
CAPS1 cases	Cancer Prostate in Sweden 1	Case-control	505	> 95%	1) related individuals and duplicates; 2) ethnic outliers; 3) missing body weight and height.	489	self-reported	Duggan D. et al. Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP. J Natl Cancer Inst (2007) 99:1836-44
CAPS1 controls	Cancer Prostate in Sweden 1	Case-control	506	> 95%	1) related individuals and duplicates; 2) ethnic outliers; 3) missing body weight and height.	491	self-reported	Duggan D. et al. Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP. J Natl Cancer Inst (2007) 99:1836-44
CAPS2 cases	Cancer Prostate in Sweden 2	Case-control	1483	> 95%	1) related individuals and duplicates; 2) ethnic outliers; 3) missing body weight and height.	1483	self-reported	Duggan D. et al. Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP. J Natl Cancer Inst (2007) 99:1836-44.
CAPS2 controls	Cancer Prostate in Sweden 2	Case-control	519	> 95%	1) related individuals and duplicates; 2) ethnic outliers; 3) missing body weight and height.	519	self-reported	Duggan D. et al. Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP. J Natl Cancer Inst (2007) 99:1836-44
CAD-WTCCC	WTCCC Coronary Arteryt Disease cases	Case series	2000	≥ 97%	1) heterozygosity <23% or >30%; 2) discrepancy with external identifying information; 3) ethnic outliers; 4) related individuals and duplicates;	1879	self reported	The Wellcome Trust Case Control Consortium Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature (2007) 447: 661-678
CHS	Cardiovascular Health Study	Population-based	3232	>95%	1) Prevalent clinical CVD 2) African-americans 3) Sex discordant 4) Missing body weight and height	3228	measured	Fried L.P. et al. The Cardiovascular Health Study: design and rationale. Ann Epidemiol. (1991) 1: 263-276.
CoLaus	Cohorte Lausannoise	Population-based	6188	>90%	1) ethnic outliers; 2) related individuals and duplicates; 3) Missing height	5409	measured	Firmann M. et al. The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome BMC Cardiovascular Disorders (2008) 8:6

deCODE	deCODE genetics sample set	Population-based	38446	≥ 96%	Missing body weight and height.	26799	measured	Thorleifsson G. et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. Nat Genet. (2009) 41, 18-24.
DGI cases	Diabetes Genetics Initiative	Case-control	1464	≥ 95%	1) Related individuals and duplicates 2) Sex mismatch 3) Phenotype missing	1317	measured	Saxena R. et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science (2007) 316:1331-6.
DGI controls	Diabetes Genetics Initiative	Case-control	1467	≥ 95%	1) Related individuals and duplicates 2) Sex mismatch 3) Phenotype missing	1090	measured	Saxena R. et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science (2007) 316:1331-6..
EGCUT	Estonian Genome Center, University of Tartu	Population-based	1428	≥ 95%	1) Related individuals and duplicates 2) Sex mismatch 3) Phenotype missing	1417	measured	(1) Nelis M. et al. Genetic Structure of Europeans: A View from the North–East. PLoS ONE (2009) 4(5): e5472. (2) Metspalu A. et al. The Estonian Genome Project. Drug Development Research (2004) 62, 97-101.
EPIC-Obesity Study	European Prospective Investigation into Cancer and Nutrition - Obesity Study	Population-based	3821	≥ 94%	1) heterozygosity <23% or >30%; 2) >5.0% discordance in SNP pairs with r2= 1 in HapMap; 3) ethnic outliers; 4) related individuals and duplicates; 5) Missing body weight and height.	3552	measured	(1) Day N.E. et al. EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. British Journal of Cancer (1999) 80: 95-103. (2) Loos R.J. et al. Common variants near MC4R are associated with fat mass, weight and risk of obesity. Nat Genet (2008) 40: 768-775.
ERF (EUROSPAN)	Erasmus Rucphen Family	Family based	2300	> 95%	1)excess heterozygosity based on FDR 2)ehtnic outliers 3)sex mismatch 4)missing phenotype	2060	measured	(1) Aulchenko Y.S. et al. Linkage disequilibrium in young genetically isolated Dutch population. Eur J Hum Genet (2004) 12: 527-534 (2) Axenovich T.I. et al. Linkage analysis of adult height in a large pedigree from a Dutch genetically isolated population. Hum Genet. (2010) 126: 457-71.
Fenland	Fenland Study	Population-based	1500	≥ 95%	1) heterozygosity <27.3% or >28.8%; 2) duplicate check; 3) relatedness check	1402	measured	Willer C.J. et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. Nat Genet. (2009) 41:25-34
FHS controls	Family Heart Study	Case-control	434	≥ 98%	1) technical errors 2) discrepancies between reported sex and sex-diagnostic markers	415	measured	Higgins M. et al. NHLBI Family Heart Study: objectives and design, Am J Epidemiol (1996) 143, 1219–1228.
FHS cases	Family Heart Study	Case-control	463	≥ 98%	1) technical errors 2) discrepancies between reported sex and sex-diagnostic markers	441	measured	Higgins M. et al. NHLBI Family Heart Study: objectives and design, Am J Epidemiol (1996) 143, 1219–1228.

FRAM	Framingham Heart Study	Population-based, multi-generational	9274	≥ 97%	1) pHWE<1e-6call rate<97% 2) mishap p<1e-9 3) MAF<0.01 4) Mendelian errors>100 5) SNPs not in Hapmap or strandedness issues merging with Hapmap	8089	measured	(1) Dawber T.R. et al. An approach to longitudinal studies in a community: the Framingham Study. Ann N Y Acad Sci. (1963)107:539-556. (2) Feinleib M. et al. The Framingham Offspring Study. Design and preliminary data. Prev Med. (1975) 4:518-525. (3) Splansky G.L. et al. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. Am J Epidemiol. (2007) 165:1328-1335.
FTC	Finnish Twin Cohort	Monozygotic twins	152 pairs	≥ 95%	1) ethnic outliers; 2) related individuals and duplicates; 3) Missing body weight and body mass index.	125	measured	(1) Aulchenko Y.S. et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat Genet. (2009) 41:47-55.
FUSION controls	Finland-United States Investigation of NIDDM Genetics	Case-control	1174	> 97.5%	related individuals; missing BMI or height	1167	measured	Scott L.J. et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science (2007) 316:1341-1345.
FUSION cases	Finland-United States Investigation of NIDDM Genetics	Case-control	1161	> 97.5%	related individuals; missing BMI or height	1082	measured	Scott L.J. et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science (2007) 316:1341-1345.
GENMETS controls	Health 2000 / GENMETS substudy of Metabolic syndrome	Case-control	948	≥ 95%	1) ethnic outliers; 2) related individuals and duplicates; 3) Missing body weight and body mass index.	823	height calculated using BMI and weight	http://www.terveys2000.fi/indexe.html
GENMETS cases	Health 2000 / GENMETS substudy of Metabolic syndrome	Case-control	932	≥ 95%	1) ethnic outliers; 2) related individuals and duplicates; 3) Missing body weight and body mass index.	824	height calculated using BMI and weight	http://www.terveys2000.fi/indexe.html
GerMiFSI (cases only)	German Myocard Infarct Family Study I	Case-control	875	> 97%	1) related individuals and duplicates; 2)missin phenotypes 3) heterozygosity mean +- 3*sd outlier	600	measured	Samani N.J. et al. Genomewide association analysis of coronary artery disease. N Engl J Med.(2007) 357:443-453.
GerMiFSII (cases only)	German Myocard Infarct Family Study II	Case-control	1222	> 97%	1) related individuals and duplicates; 2)missin phenotypes 3) heterozygosity mean +- 3*sd outlier	1124	measured	Erdmann J. et al. New susceptibility locus for coronary artery disease on chromosome 3q22.3. Nat Genet. (2009) 41:280-282.

KORA S3	Cooperative Health Research in the Region of Augsburg, KOoperative Gesundheitsforschung in der Region Augsburg	Population-based	1644	≥ 93%	1) german passport; 2) missing height.	1643	measured	Wichmann H.E. et al. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen (2005) 67 Suppl 1, S26-30.
KORA S4	Cooperative Health Research in the Region of Augsburg, KOoperative Gesundheitsforschung in der Region Augsburg	Population-based	1814	≥ 93%	1) german passport; 2) missing height.	1811	measured	Wichmann H.E. et al. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen (2005) 67 Suppl 1, S26-30.
MICROS	MICROS (EUROSPAN)	Population-based	1098	≥ 97%	1) ethnic outliers; 2) duplicates; 3) Missing height.	1079	measured	Pattaro C. et al. The genetic study of three population microisolates in South Tyrol (MICROS): study design and epidemiological perspectives. BMC Med Genet (2007) 8:29
MIGEN	Myocardial Infarction Genetics Consortium	Case-control	6042	≥ 95%	1) Related individuals and duplicates 2) Sex mismatch 3) Phenotype missing	2652	measured	Kathiresan S. et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet. (2009) 41:334-41.
NBS-WTCCC	WTCCC National Blood Service donors	Population-based	1500	≥ 97%	1) heterozygosity <23% or >30%; 2) discrepancy with external identifying information; 3) ethnic outliers; 4) related individuals and duplicates;	1441	self reported	The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature (2007) 447, 661-678
NFBC1966	Northern Finland Birth Cohort 1966	Population-based	5654	≥ 95%	1) sex discrepancy with genetic data from X-linked markers; 2) withdrawn consent; 3) duplicates and first and second degree relatives; 4) contaminated samples	4499	measured	(1) Sabatti C. et al. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nat Genet (2008) 41: 35-46. (2) Sovio U. et al. Genetic determinants of height growth assessed longitudinally from infancy to adulthood in the northern Finland birth cohort 1966. PLoS Genet (2009) 5(3): e1000409
NHS	The Nurses' Health Study	Nested case-control	2368	>90%	1) Low genotying completion (<90%); 2) Unclear identity and admixed origin; 3) Missing height.	2265	self-reported	Hunter D. et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet. (2007) 39: 870-874.

NSPHS	Northern Sweden Population Health Study (EUROSPAN)	Population-based	720	≥ 97%	1) ethnic outliers; 2) duplicates; 3) Missing height.	652	measured	(1) Johansson A. et al. Common variants in the JAZF1 gene associated with height identified by linkage and genome-wide association analysis. Hum Mol Genet (2009) 18: 373-80. (2) Hicks A.A. et al. Genetic determinants of circulating sphingolipid concentrations in European populations. PLoS Genet. (2009) 5(10):e1000672
NTRNESDA	Netherlands Twin Register & the Netherlands Study of Depression and Anxiety	Case-control	3720	≥ 95%	1) evidence of sample contamination (heterozygosity); 2) ethnic outliers; 3) related individuals and duplicates; 5) missing body height.	3522	questionnaire and measured	(1) Boomsma D.I. et al. Netherlands Twin Register: from twins to twin families. Twin Res Hum Genet (2006) 9: 849–857. (2) Penninx B. et al. The Netherlands Study of Depression and Anxiety (NESDA): rationales, objectives and methods. Int J Methods Psychiatr Res (2008) 17: 121-140. (3) Boomsma D.I. et al. Genome-wide association of major depression: Description of samples for the GAIN major depressive disorder study: NTR and NESDA Biobank Projects. Eur J Hum Genet (2008) 16: 335–342.
ORCADES	Orkney Complex Disease Study (part of EUROSPAN)	Population-based	719	≥ 97%	1) ethnic outliers; 2) duplicates; 3) missing height.	695	measured	(1) Johansson A. et al. Common variants in the JAZF1 gene associated with height identified by linkage and genome-wide association analysis. Hum Mol Genet. (2009) 18: 373-380. (2) Hicks A.A. et al. Genetic determinants of circulating sphingolipid concentrations in European populations. PLoS Genet. (2009) 5(10):e1000672
PLCO	The Prostate, Lung Colorectal and Ovarian Cancer Screening Trial	Case-control	2298	≥ 94%	1) Sex discordance 2) Non-European ancestry 3) Related individuals and duplicates; 4) Missing height.	2244	self-reported	Yeager M. et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet (2007) 39: 645-649.
PROCARDIS	Precocious Coronary Artery Disease	Case-control	2573	> 95%	none	2312	measured	Broadbent H.M. et al. Susceptibility to coronary artery disease and diabetes is encoded by distinct, tightly linked SNPs in the ANRIL locus on chromosome 9p. Hum Mol Genet (2008) 17: 806-814.

RS-I	Rotterdam Study I	Population-based	7983	≥ 97.5%	1) sex mismatch with typed X-linked markers; 2) excess autosomal heterozygosity > 0.336~FDR>0.1%; 3) duplicates and/or 1st or 2nd degree relatives using IBS probabilities >97% from PLINK; 4) ethnic outliers using IBS distances > 3SD from PLINK; 5) Missing body weight and height.	5744	measured	(1) Estrada K. et al. A genome-wide association study of northwestern Europeans involves the C-type natriuretic peptide signaling pathway in the etiology of human height variation. Hum Mol Genet (2009) 18:3516-3524 (2) Estrada K. et al. GRIMP: a web- and grid-based tool for high-speed analysis of large-scale genome-wide association using imputed data. Bioinformatics (2009) 25:2750-2752 (3) Hofman A. et al. The Rotterdam Study: 2010 objectives and design update. Eur J Epidemiol (2009) 24: 553-572 (4) Hofman A. et al. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. Eur J Epidemiol (1991) 7: 403-422
RUNMC	Nijmegen Bladder Cancer Study (NBCS) & Nijmegen Biomedical Study (NBS), Radboud University Nijmegen Medical Centre	Population-based	3081	≥ 96%	Missing body weight and height.	2873	self-assessed and reported by questionnaire	(1) Wetzels J.F. et al. Age- and sex-specific reference values of estimated GFR in Caucasians: the Nijmegen Biomedical Study. Kidney Int (2007) 72, 632-637. (2) Kiemeney L.A. et al. Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. Nat Genet (2008) 40: 1307-1312.
SardiNIA	SARDINIA	Population-based	6148	≥ 90%	1) Morquio syndrome 2) Missing height	4298	measured	Pilia G. et al. Heritability of cardiovascular and personality traits in 6,148 Sardinians. PLoS Genet (2006) 2: e132
SASBAC cases	Swedish And Singapore Breast Association Consortium	Case-control	803	≥ 96%	1) related individuals and duplicates; 2) ethnic outliers; 3) missing body weight and height.	794	self-reported	(1) Magnusson C. et al. Breast-cancer risk following long-term oestrogen- and oestrogen-progestin-replacement therapy. Int J Cancer (1999) 81: 339-344. (2) Einarsdóttir K. et al. Comprehensive analysis of the ATM, CHEK2 and ERBB2 genes in relation to breast tumour characteristics and survival: a population-based case-control and follow-up study. Breast Cancer Res (2006) 8: R67.
SASBAC controls	Swedish And Singapore Breast Association Consortium	Case-control	764	≥ 96%	1) related individuals and duplicates; 2) ethnic outliers; 3) missing body weight and height.	758	self-reported	(1) Magnusson C. et al. Breast-cancer risk following long-term oestrogen- and oestrogen-progestin-replacement therapy. Int J Cancer (1999) 81: 339-344. (2) Einarsdóttir K. et al. Comprehensive analysis of the ATM, CHEK2 and ERBB2 genes in relation to breast tumour characteristics and survival: a population-based case-control and follow-up study. Breast Cancer Res (2006) 8: R67.

SEARCH / UKOPS	Studies of Epidemiology and Risk factors in Cancer Heredity / UK Ovarian Cancer Population Study	Population-based	1710	≥ 80%	1) ethnic outliers 2) duplicates 3) Missing height	1592	self-assessed	Song H. et al. A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. Nat Genet (2009) 41: 996-1000.
SHIP	Study of Health in Pomerania	Population-based	4310	≥ 92%	1) missing genotype or phenotype data	4092	measured	John U. et al. Study of health in Pomerania (SHIP): a health examination survey in an east German region: objectives and design. Soz-Präventivmed (2001) 46: 186-194.
T2D-WTCCC	WTCCC Type 2 Diabetes cases	case series	1999	≥ 97%	1) heterozygosity <23% or >30%; 2) discrepancy with external identifying information; 3) ethnic outliers; 4) related individuals and duplicates;	1903	measured	The Wellcome Trust Case Control Consortium Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature (2007) 447: 661-678
TwinsUK	TwinsUK	Twins pairs	2226	≥ 95%	1) heterozygosity <33% or >37%; 2) ethnic outliers; 3) related individuals and duplicates; 4) Missing body weight and height.	1479	measured	(1) Spector T.D., Williams F.M. The UK Adult Twin Registry (TwinsUK). Twin Res Hum Genet (2006) 9: 899-906. (2) Spector T.D., MacGregor A.J. The St. Thomas' UK Adult Twin Registry. Twin Res (2002) 5: 440-443.
VIS	VIS (EUROSPAN) and KORCULA	Population-based	795	≥ 97%	1) ethnic outliers; 2) duplicates; 3) Missing height.	784	measured	(1) Johansson A. et al. Common variants in the JAZF1 gene associated with height identified by linkage and genome-wide association analysis. Hum Mol Genet. (2009) 18: 373-380. (2) Hicks A.A. et al. Genetic determinants of circulating sphingolipid concentrations in European populations. PLoS Genet. (2009) 5(10):e1000672
Stage 2 (in-silico replication studies)								
BHS	Busselton Health Study	Population-based	1366	≥ 75%	1) ethnic outliers; 2) related individuals and duplicates; 3) Missing body waist and hip.	1328	measured	(1) James A.L. et al. Decline in lung function in the Busselton Health Study: the effects of asthma and cigarette smoking. Am J Respir Crit Care Med (2005) 171:109-114. (2) Hui J. et al. A genome-wide association scan for asthma in a general Australian population. Hum Genet (2008) 123:297-306

Corogene	Genetic Predisposition of Coronary Heart Disease in Patients Verified with Coronary Angiogram	Population-based	4130	≥ 95%	1) missing sex 2) related individuals and duplicates 3) (For this specific analysis) Missing body height	3758	measured	Soranzo, N. et al. A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. Nat. Genet (2009). 41: 1182-1190.
EGCUT	Estonian Genome Center, University of Tartu	Population-based	345	≥ 95%	1) Related individuals and duplicates 2) Sex mismatch 3) Phenotype missing	345	measured	(1) Nelis M. et al. Genetic Structure of Europeans: A View from the North–East. PLoS ONE (2009) 4(5): e5472. (2) Metspalu A. The Estonian Genome Project. Drug Development Research (2004) 62: 97-101.
FHS	Family Heart Study	Case-control	1808	≥ 98%	1) technical errors 2) discrepancies between reported sex and sex-diagnostic markers	1463	measured	Higgins M. et al. NHLBI Family Heart Study: objectives and design, Am J Epidemiol (1996) 143: 1219-1228.
FINGESTURE cases	Finnish Genetic Study of Arrhythmic Events	Disease cohort (MI cases only)	1103	≥ 97%	1) PLINK heterozygosity F-value <-0.05 or >0.05; 2) ethnic outliers; 3) related individuals and duplicates; 4) Missing body weight and height.	943	measured	Kaikkonen K.S. et al. Family history and the risk of sudden cardiac death as a manifestation of an acute coronary event. Circulation (2006) 114, 1462-7
GOOD	Gothenburg Osteoporosis and Obesity Determinants Study	Population-based	1056	≥ 97.5%	1) heterozygosity > 33%; 2) ethnic outliers; 3) related individuals and duplicates.	938	measured	Lorentzon M. et al. Free testosterone is a positive whereas free estradiol is a negative predictor of cortical bone size in young Swedish men-The GOOD Study. J Bone Miner Res (2005) 20: 1334-1341.
HBCS	Helsinki Birth Cohort Study	Birth cohort study	1872	≥ 95%	1) related individuals and duplicates 2) (From this specific analysis) Missing body height	1726	measured	Ylihärsilä H. et al. Body mass index during childhood and adult body composition in men and women aged 56-70 y. Am J Clin Nutr. (2008) 87:1769-1775. Kajantie E. et al. Size at birth as a predictor of mortality in adulthood: a follow-up of 350 000 person-years. Int J Epidemiol (2005) 34:655-663.
HYPERGENES controls	HYPERGENES	Case-control	1934	>90%	1) ethnic outliers 2) Missing body weight and height.	1838	measured	http://www.hypergenes.eu/
HYPERGENES cases	HYPERGENES	Case-control	2124	>90%	1) ethnic outliers 2) Missing body weight and height.	1787	measured	http://www.hypergenes.eu/

MGS	Molecular Genetics of Schizophrenia/NIMH Repository Control Sample	Population-based (survey research method)	2681	99.7%	1) call rate < 97% for samples,95% for SNPs 2) heterozygosity <26% or >28.5%; 3) excess duplicate discordancies or mendelian errors (SNPs); 4) ethnic outliers (principal component scores); 5) related individuals and duplicates; 6) Missing body weight or height.	2597	self-reported	(1) Shi J. et al. Common variants on chromosome 6p22.1 are associated with schizophrenia. Nature. (2009) 460: 753-757. (2) Sanders A.R. et al. No significant association of 14 candidate genes with schizophrenia in a large European ancestry sample: implications for psychiatric genetics. Am J Psychiatry. (2008) 165: 497-506.
NHS	The Nurses' Health Study	Nested case-control	3221	>98%	1) Low genotying completion (<98%); 2) Unclear identity and admixed origin; 3) related individuals and duplicates; 4) DNA contamination; 5) Missing height;	3217	self-reported	Qi L. et al. Genetic variants in ABO blood group region, plasma soluble E-selectin levels, and risk of type 2 diabetes. Hum Mol Genet. (2010) Feb 10, doi:10.1093/hmg/ddq057
RS-II	Rotterdam Study II	Population-based	3011	≥ 97.5%	1) sex mismatch with typed X-linked markers; 2) excess autosomal heterozygosity (F<-0.055); 3) duplicates and/or 1st degree relatives using IBD PiHAT >40% from PLINK; 4) ethnic outliers IBS distances > 4SD mean HaMAP CEU cluster from PLINK; 5) Missing body weight and height.	2124	measured	(1) Estrada K. et al. A genome-wide association study of northwestern Europeans involves the C-type natriuretic peptide signaling pathway in the etiology of human height variation. Hum Mol Genet (2009) 18:3516-3524 (2) Estrada K. et al. GRIMP: a web- and grid-based tool for high-speed analysis of large-scale genome-wide association using imputed data. Bioinformatics (2009) 25:2750-2752 (3) Hofman A. et al. The Rotterdam Study: 2010 objectives and design update. Eur J Epidemiol (2009) 24: 553-572 (4) Hofman A. et al. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. Eur J Epidemiol (1991) 7: 403-422
RS-III	Rotterdam Study III	Population-based	3932	≥ 97.5%	1) sex mismatch with typed X-linked markers; 2) excess autosomal heterozygosity (F<-0.055); 3) duplicates and/or 1st degree relatives using IBD PiHAT >40% from PLINK; 4) ethnic outliers IBS distances > 4SD mean HaMAP CEU cluster from PLINK; 5) Missing body weight and height.	2009	measured	(1) Estrada K. et al. A genome-wide association study of northwestern Europeans involves the C-type natriuretic peptide signaling pathway in the etiology of human height variation. Hum Mol Genet (2009) 18:3516-3524 (2) Estrada K. et al. GRIMP: a web- and grid-based tool for high-speed analysis of large-scale genome-wide association using imputed data. Bioinformatics (2009) 25:2750-2752 (3) Hofman A. et al. The Rotterdam Study: 2010 objectives and design update. Eur J Epidemiol (2009) 24: 553-572 (4) Hofman A. et al. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. Eur J Epidemiol (1991) 7: 403-422

Sorbs	Sorbs are self-contained population from Eastern Germany, European Descent	Population-based	1097	≥ 94%	1) sex mismatch; 2) ethnic outliers; 3) duplicates; 4) Missing body weight and height.	907	measured	Tönjes A. et al. Association of FTO variants with BMI and fat mass in the self-contained population of Sorbs in Germany. Eur J Hum Genet. (2010) 18:104-10.
WGHS	Women's Genome Health Study	Population-based	23,294	>98%	1) includes only WGHS participants with confirmed, self-reported European ancestry; 2) all SNPs have HWE p>10E-6; 3) all SNPs have genotype for >90% samples 4) only samples with biometric measures included in analysis	23099	self-report	Ridker P.M. et al. Rationale, design, and methodology of the Women's Genome Health Study: a genome-wide association study of more than 25,000 initially healthy American women. Clin Chem. (2008) 54:249-55. .
YFS	The Cardiovascular Risk in Young Finns Study	Population-based cohort	2,443	≥ 95%	1) missing sex 2) related individuals and duplicates 3) (From this specific analysis) Missing body height	1995	measured	Raitakari O.T. et al. Cohort profile: The cardiovascular risk in Young Finns Study. Int J Epidemiol. (2008) 37:1220-6
Polygene analysis study								
QIMR	Twin studies at the Queensland Instutite of Medical Research	Population-based	2,654	≥ 95%	1) close relatives based on pedigree information; 2) ethnic outliers; 3) Missing height.	1475	measured or self-report	Medland et al. Common Variants in the Trichohyalin Gene Are Associated with Straight Hair in Europeans. Amer J Hum Genet (2009) 85:750-5.
* Sample genotyping success rate; i.e. minimum percentage of successfully genotyped SNPs of GWAs per sample								

Supplementary Methods Table 2. Information on genotyping methods, quality control of SNPs, imputation, and statistical analysis.

	Genotyping						Imputation			Association analyses				
			Inclusion criteria					Inclusion criteria			λ_{GC}			
Cohort	Platform	Genotype calling algorithm	MAF	Call rate*	p for HWE	SNPs that met QC criteria	Imputation software	MAF	Imputation quality	SNPs in meta-analysis (after MAFxN>3 filter)	all	men	women	Analyses software
Stage 1 (GWA studies)														
ADVANCE cases	Illumina 550k	BeadStudio	none	≥98.5%	$>10^{-3}$	543,985	BIMBAM	>0%	none	2,193,902	NA	1.047	1.022	SNPTEST
ADVANCE controls	Illumina 550k	BeadStudio	none	≥98.5%	$>10^{-3}$	543,985	BIMBAM	>0%	none	2,206,332	NA	1.046	0.996	SNPTEST
AGES	Illumina Human370CNV	BeadStudio	≥1%	≥95%	$>10^{-6}$	308,340	MACH	>0%	r2-hat≥0.30	2,458,927	NA	1.075	1.082	ProbABEL
Amish HAPI Heart Study	Affymetrix GeneChip Human Mapping 500K	BRLMM	≥1%	≥95%	$>10^{-6}$	338,598	MACH	>0%	r2-hat≥0.30	2,291,092	1.057	0.938	1.045	MMAP
ARIC	Affymetrix Genome-Wide Human SNP Array 6.0	Birdseed	>1%	≥95%	$>10^{-5}$	685,812	MACH	>0%	r2-hat≥0.30	2,511,301	NA	1.021	1.039	ProbABEL
B58C-T1DGC	Illumina HumanHap 550 V.1	ILLUMINUS	>0%	none	none	539,458	MACH	>0%	r2-hat≥0.30	2,507,988	NA	1.024	1.014	ProbABEL
B58C-WTCCC	Affymetrix GeneChip Human Mapping 500K	CHIAMO	>5%	none	none	392,575	IMPUTE	>0%	proper-info≥0.40	2,448,428	NA	0.999	1.003	SNPTEST
BRIGHT	Affymetrix GeneChip Human Mapping 500K	CHIAMO	≥5%	≥95%	$>10^{-6}$	387,666	IMPUTE	>0%	proper-info≥0.40	2,429,136	NA	1.015	0.995	SNPTEST
CAPS1 cases	Affymetrix GeneChip Human Mapping 500K	BRLMM	≥1%	≥95%	$>10^{-7}$	330,124	IMPUTE	>0%	proper-info≥0.40	2,387,578	NA	0.993	NA	SNPTEST
CAPS1 controls	Affymetrix GeneChip Human Mapping 500K	BRLMM	≥1%	≥95%	$>10^{-7}$	330,124	IMPUTE	>0%	proper-info≥0.40	2,390,475	NA	0.995	NA	SNPTEST
CAPS2 cases	Affymetrix GeneChip Human Mapping 5.0K	BLRMM-P	≥1%	≥95%	$>10^{-7}$	348,163	IMPUTE	>0%	proper-info≥0.40	2,416,296	NA	1.044	NA	SNPTEST
CAPS2 controls	Affymetrix GeneChip Human Mapping 5.0K	BLRMM-P	≥1%	≥95%	$>10^{-7}$	348,163	IMPUTE	>0%	proper-info≥0.40	2,391,556	NA	1.041	NA	SNPTEST
CAD-WTCCC	Affymetrix GeneChip Human Mapping 500K	CHIAMO	>5%	≥95%	$>10^{-6}$	387,667	IMPUTE	>0%	proper-info≥0.40	2,430,482	NA	1.025	1.009	SNPTEST

CHS	Illumina 370-CNV	BeadStudio		>97%	$>10^{-5}$	306,655	BimBam	>0%	r2-hat ≥0.30	2,191,645	NA	1.11	1.15	R
CoLaus	Affymetrix GeneChip Human Mapping 500K	BRLMM	≥1%	≥70%	$>10^{-7}$	390,631	IMPUTE	>0%	proper-info ≥0.40	2,479,491	NA	1.013	1.034	QUICKTEST
deCODE	Illumina HumanHap300 or HumanHapCNV370	BeadStudio	≥1%	≥96%	$>10^{-6}$	290,447	IMPUTE	>0%	proper-info≥0.40	2,456,118	0.948	0.977	0.986	SNPTEST
DGI cases	Affymetrix 500K	BRLMM	≥1%	≥95%	$>10^{-6}$	386,731	MACH	>0%	r2-hat≥0.30	2,410,247	1.029	0.977	1.049	MACH2QTL
DGI controls	Affymetrix 500K	BRLMM	≥1%	≥95%	$>10^{-6}$	386,731	MACH	>0%	r2-hat≥0.30	2,408,993	1.029	1.045	0.995	MACH2QTL
EGCUT	Illumina Beadarray Human370CNV	BeadStudio	≥1%	≥98%	$>10^{-6}$	299,484	IMPUTE	>0%	proper-info≥0.40	2,429,620	NA	1.032	1.013	SNPTEST
EPIC-Obesity Study	Affymetrix GeneChip Human Mapping 500K	BRLMM	≥1%	≥90%	$>10^{-6}$	397,438	IMPUTE	>0%	proper-info≥0.40	2,420,624	NA	1.018	1.027	SNPTEST
ERF (EUROSPAN)	Illumina 318K, 370K, Affymetrix 250K	BRLMM, BeadStudio	>0.5%	>95%	$>10^{-6}$	NA	MACH	>0%	r2-hat≥0.30	2,463,846	1.031	1.012	1.019	ProbABEL
Fenland	Affymetrix SNP5.0	BRLMM	≥1%	≥90%	$>10^{-6}$	362,055	IMPUTE	>0%	proper-info≥0.40	2,406,753	NA	1.039	1.04	SNPTEST
FHS (cases + controls)	Illumina 1Million GeneChip	BeadStudio	≥1%	≥98%	$>10^{-6}$	874,830	MACH	>0%	r2-hat≥0.30	2,375,010	1.066	1.06	1.064	SAS
FRAM	Affymetrix 500K Affymetrix 50K supplemental	BRLMM	≥1%	≥97%	$>10^{-6}$	378,163	MACH	>0%	r2-hat≥0.30	2,455,455	1.071	1.027	1.062	R
FTC	Illumina HumanHap 318K	BeadStudio	≥1%	≥90%	$>10^{-6}$	304,582	MACH	>0%	r2-hat≥0.30	2,268,674	NA	NA	1.005	ProbABEL
FUSION controls	Illumina Infinium™ II HumanHap300 BeadChip	BeadStudio	>1%	≥90%	$\geq 10^{-6}$	315,635	MACH	>0%	r2-hat≥0.30	2,466,546	1.112	1.056	1.074	MACH2QTL
FUSION cases	Illumina Infinium™ II HumanHap300 BeadChip	BeadStudio	>1%	≥90%	$\geq 10^{-6}$	315,635	MACH	>0%	r2-hat≥0.30	2,466,546	1.08	1.077	1.027	MACH2QTL
GENMETS controls	Illumina HumanHap 610K	Illuminus	≥1%	≥95%	$>10^{-6}$	555,388	MACH	>0%	r2-hat≥0.30	2,345,066	NA	1.043	1.006	ProbABEL
GENMETS cases	Illumina HumanHap 610K	Illuminus	≥1%	≥95%	$>10^{-6}$	555,388	MACH	>0%	r2-hat≥0.30	2,343,751	NA	1.016	1.007	ProbABEL
GerMiFSI	Affymetrix NSP/STY	BRLMM	>1%	>97%	$>10^{-5}$	282,215	MACH	>0%	r2-hat≥0.30	2,333,219	NA	1.014	1.026	GenABEL
GerMiFSII	Affymetrix 6.0	Birdseed	>1%	>97%	$>10^{-5}$	653,149	MACH	>0%	r2-hat≥0.30	2,492,325	NA	1.07	1.015	GenABEL
KORA S3	Affymetrix 500K	BRLMM	none	none	none	490,032	MACH	>0%	r2-hat≥0.30	2,415,072	NA	1.018	1.016	MACH2QTL
KORA S4	Affymetrix 6.0	Birdseed	none	none	none	909,622	IMPUTE	>0%	proper-info≥0.40	2,109,266	NA	1.009	1.036	SNPTEST
MICROS	ILLUMINA318K	BeadStudio	≥1%	≥98%	$>10^{-6}$	318,237	MACH	>0%	r2-hat ≥0.30	2,435,539	1.004	1	0.994	ProbABEL

MIGEN	Affymetrix 6.0	Birdseed	≥1%	≥95%	$>10^{-6}$	727,496	MACH	>0%	r2-hat ≥0.30	2,288,269.4 (average)	NA	1.002 (average)	1.0015 (average)	MACH2QTL
NBS-WTCCC	Affymetrix GeneChip Human Mapping 500K	CHIAMO	>5%	≥95%	$>10^{-6}$	387,667	IMPUTE	>0%	proper-info ≥0.40	2,415,926	NA	1.002	1.008	SNPTEST
NFBC1966	Illumina HumanCNV-370DUO Analysis BeadChip	Standard Illumina BeadStudio	≥5%	≥95%	$>10^{-4}$	328,007	IMPUTE	>0%	proper-info ≥0.40	2,460,379	NA	1.037	1.053	SNPTEST
NHS	Illumina HumanHap550	Standard Illumina BeadStudio	≥1%	≥90%	none	510,073	MACH	>0%	r2-hat ≥0.30	2,520,546	NA	NA	1.005	MACH2QTL
NSPHS	ILLUMINA318K	BeadStudio	≥1%	≥98%	$>10^{-6}$	318,236	MACH	>0%	r2-hat≥0.30	2,382,373	1.023	1.03	1.015	ProbABEL
NTRNESDA	Perlegen - Affymetrix gene chip 600K	Proprietary Perlegen	>1%	≥95%	none	435,291	IMPUTE	>0%	proper-info≥0.40	2,493,317	NA	1.028	1.062	SNPTEST
ORCADES	ILLUMINA318K	BeadStudio	≥ 1%	≥98%	$>10^{-6}$	318,235	MACH	>0%	r2-hat≥0.30	2,433,999	1.004	0.966	1.042	ProbABEL
PLCO	Illumina HumanHap300 and Illumina HumanHap240	Illumina Bead Studio	none	≥90%	none	523,231	MACH	>0%	r2-hat≥0.30	2,527,780	NA	1.006	NA	MACH2QTL
PROCARDIS	HumanHap300 BeadChips	Illumina Beadstudio 2.0 software	>5%	≥95%	$>5x10^{-7}$	~820k	IMPUTE	>0%	proper-info≥0.40	2,580,770	NA	1.084	1.014	SNPTEST
RS-I	Illumina /HumanHap 550K V.3 ADHumanHap 550 V.3 DUO;	BeadStudio Genecall	≥1%	≥97.5%	$>10^{-6}$	512,349	MACH	>0%	(O/E)σ2 ratio≥0.1 r2-hat≥0.30	2,488,215	NA	1.045	1.064	MACH2QTL
RUNMC	Illumina HumanHapCNV370	BeadStudio	≥1%	≥96%	$>10^{-6}$	312,199	IMPUTE	>0%	proper-info≥0.40	2,465,662	0.996	0.996	0.996	SNPTEST
SardiNIA	Affymetrix 500K and Affymetrix 10K	BRLMM	≥5%	≥90%	$>10^{-6}$	356,359	MACH	>0%	r2-hat≥0.30	2,251,689	1.313	1.171	1.213	Merlin
SASBAC cases	Illumina HumanHap300+240S	Standard Illumina BeadStudio (GenCall)	≥3%	≥90%	$>10^{-7}$	510,578	IMPUTE	>0%	proper-info≥0.40	2,491,965	NA	NA	1.009	SNPTEST
SASBAC controls	Illumina HumanHap550	Standard Illumina BeadStudio (GenCall)	≥3%	≥90%	$>10^{-7}$	512,223	IMPUTE	>0%	proper-info≥0.40	2,474,508	NA	NA	1.012	SNPTEST

SEARCH / UKOPS	Illumina HumanHap 610 Quad	Illuminus	≥1%	≥95%	$>10^{-4}$	495,229	In-house method similar to IMPUTE	>0%	r2-hat≥0.30	2,486,650	NA	NA	1.02	Regression analysis on dosages
SHIP	Affymetrix Human SNP Array 6.0	Birdseed V2	≥0%	≥0%	≥0	869,224	IMPUTE	>0%	proper-info≥0.40	2,609,015	NA	1.034	1.046	SNPTEST v1.1.5 InforSense
T2D-WTCCC	Affymetrix GeneChip Human Mapping 500K	CHIAMO	>5%	≥95%	$>10^{-6}$	387,667	IMPUTE	>0%	proper-info≥0.40	2,425,374	NA	1.008	1.011	SNPTEST
TWINSUK	Illumina / HumanHap 300 & 550	Illuminus	≥1%	≥95%	$>10^{-6}$	295,702	IMPUTE	>0%	proper-info≥0.40	2,460,943	NA	NA	1.022	SNPTEST
VIS	Illumina HumanHap300v1	BeadStudio	≥1%	≥98%	$>10^{-6}$	317,465	MACH	>0%	r2-hat≥0.30	2,423,083	0.989	1.002	0.991	ProbABEL
Stage 2 (in-silico replication studies)														
BHS	Illumina Human 610-Quad	Illuminus	≥1%	≥95%	$>5.7x10^{-7}$	549,294	MACH	≥1%	r2-hat≥0.30	664	-	-	-	R
Corogene	Illumina BeadChip Human 610-Quad	Illuminus	≥1%	≥95%	$>10^{-6}$	554,988	MACH	≥1%	r2-hat≥0.30	663	-	1.079	1.084	PLINK
EGCUT	Illumina Beadarray Human370CNV	BeadStudio	≥1%	≥98%	$>10^{-6}$	316,924	IMPUTE	≥1%	proper-info≥0.30	662	-	1.034	1.025	SNPtest
FHS	Illumina 1Million GeneChip	BeadStudio	≥1%	≥98%	$>10^{-6}$	874,830	MACH	≥1%	r2-hat≥0.30	665	-	-	-	SAS
FINGESTURE cases	Affymetrix Genome-Wide Human SNP Array 6.0	Birdseed	≥5%	≥95%	$>10^{-6}$	606,717	MACH	>0%	r2-hat≥0.30	663	-	-	-	MACH2QTL
GOOD	Illumina Infinium HumanHap 610K	BeadStudio	≥1%	≥98%	$>10^{-6}$	521,160	MACH	>0%	r2-hat≥0.30	664	-	-	-	MACH2QTL
HBCS	Illumina custom made BeadChip Human 670-Quad	Illuminus	≥1%	≥95%	$>10^{-6}$	533491	MACH	≥1%	r2-hat≥0.30	663	-	1.000	1.002	PLINK
HYPERGENES controls	Illumina Human1M-Duov3_B	GenCall, BeadStudio	≥1%	≥90%	$>10^{-7}$	Center I: 861759, Center II: 872576	MACH	>0%	r2-hat≥0.30	642	-	-	-	Matlab
HYPERGENES cases	Illumina Human1M-Duov3_B	GenCall, BeadStudio	≥1%	≥90%	$>10^{-7}$	Center I: 861759, Center II: 872576	MACH	>0%	r2-hat≥0.30	642	-	-	-	Matlab

MGS	Affymetrix Genome-Wide Human SNP Array 6.0	Birdsuite 2.0	≥1%	≥95%	$>10^{-6}$	696,492	MACH	≥1%	r2-hat≥0.30	662	-	-	-	PLINK and local software
NHS	Affymetrix Genome-Wide Human 6.0 array	Birdseed algorithm v2	≥2%	≥98%	$>10^{-4}$	704,409	MACH	≥2%	r2-hat≥0.30	392	-	-	-	ProbABEL
RS-II	Illumina / HumanHap 550 V.3 DUO; Illumina / HumanHap 610 QUAD	Genomestudio Genecall	≥1%	≥97.5%	$>10^{-6}$	466,389	MACH	≥1%	(O/E)σ2 ratio≥0.1 r2-hat ≥0.30	664	-	1.004	1.012	MACH2QTL
RS-III	Illumina / HumanHap 610 QUAD	Genomestudio Genecall	≥1%	≥97.5%	$>10^{-6}$	514,073	MACH	≥1%	(O/E)σ2 ratio≥0.1 r2-hat ≥0.30	664	-	1.004	1.018	MACH2QTL
Sorbs	500K Affymetrix GeneChip (250K Sty and 250K Nsp arrays) and Affymetrix Genome-Wide Human SNP Array 6.0	BRLMM algorithm for 500K and Birdseed Algorithm for SNP Array 6.0	≥1%	≥95%	$>10^{-4}$	378,513	IMPUTE	>1%	proper-info>0.40	650	-	-	-	SNPTEST
WGHS	Illumina HumanHap300 Duo "+"	Beadstudio v 3.3	NA	≥90%	$>10^{-6}$	339,596	MACH	>0%	r2-hat≥0.30	663	-	-	-	R
YFS	Illumina custom made BeadChip Human 670-Quad	Illuminus	≥1%	≥95%	$>10^{-6}$	546,674	MACH	≥1%	r2-hat≥0.30	663	-	1.017	1.043	PLINK
Polygene analysis study														
QIMR	Illumina HumanHap 610 Quad	BeadStudio	≥1%	≥95%	$>10^{-6}$	493,578								
* SNP genotyping success rate; i.e. minimum percentage of successfully genotyped samples per SNP														

Supplementary Methods Table 3: Study-specific descriptive statistics

		Men								Women							
Study	**Trait**	**n**	**mean**	**SD**	**median**	**min**	**max**	**correlation with BMI**	**correlation with height**	**n**	**mean**	**SD**	**median**	**min**	**max**	**correlation with BMI**	**correlation with height**
Stage 1 (GWA studies)																	
ADVANCE cases	Age (yrs)	114	40.42	3.98	41.20	20.40	45.10	-0.10	-0.12	161	49.46	4.68	50.50	34.00	55.00	0.01	-0.15
	Height (m)	114	1.77	0.07	1.77	1.61	1.95	0.15	1.00	161	1.64	0.07	1.64	1.48	1.84	-0.14	1.00
	BMI (kg/m²)	114	31.39	5.77	30.89	19.48	54.32	1.00	0.15	161	31.40	8.17	30.65	17.30	61.08	1.00	-0.14
	Weight (kg)	114	99.03	21.16	97.59	64.05	181.44	0.92	0.49	161	83.98	21.78	81.74	48.58	153.00	0.95	0.16
ADVANCE controls	Age (yrs)	128	40.46	3.23	41.20	33.40	46.80	-0.03	0.15	183	48.69	4.45	49.80	34.80	55.40	0.09	-0.03
	Height (m)	128	1.79	0.07	1.78	1.58	1.96	0.02	1.00	181	1.66	0.06	1.66	1.45	1.80	-0.14	1.00
	BMI (kg/m²)	128	27.00	4.48	26.21	17.86	49.38	1.00	0.02	181	26.08	6.36	24.65	15.76	54.12	1.00	-0.14
	Weight (kg)	128	86.45	16.38	84.37	51.48	158.76	0.88	0.43	182	71.35	17.13	68.27	40.23	140.71	0.92	0.21
AGES Midlife	Age (yrs)	1352	49.69	5.87	50.00	34.00	75.00	0.05	-0.21	1867	52.00	6.54	52.00	34.00	77.00	0.15	-0.23
	Height (m)	1352	1.78	0.06	1.78	1.56	1.98	0.01	1.00	1867	1.64	0.05	1.64	1.45	1.83	-0.15	1.00
	BMI (kg/m²)	1351	25.62	3.09	25.48	16.94	38.61	1.00	0.01	1856	24.89	3.81	24.31	13.65	50.41	1.00	-0.15
	Weight (kg)	1351	81.32	11.41	80.40	51.00	139.00	0.87	0.51	1856	67.13	10.51	66.00	32.80	140.60	0.91	0.27
Amish HAPI Heart Study	Age (yrs)	471	46.2	16.9	43.0	20.0	99.0	0.25	-0.41	437	47.5	15.1	48.0	20.0	95.0	0.25	-0.40
	Height (m)	470	1.73	0.07	1.73	1.48	1.94	-0.05	1.00	437	1.61	0.06	1.61	1.39	1.75	-0.22	1.00
	BMI (kg/m²)	468	26.3	3.5	26.0	18.6	39.0	1.00	-0.05	437	28.5	5.7	28.3	16.9	47.1	1.00	-0.22
	Weight (kg)	468	78.6	11.7	77.0	49.4	112.8	0.86	0.45	437	73.5	14.4	71.9	37.8	114.3	0.93	0.16
ARIC	Age (yrs)	3823	54.69	5.70	55.00	44.00	66.00	-0.04	-0.16	4287	53.97	5.67	54.00	44.00	66.00	0.04	-0.15
	Height (m)	3823	1.76	0.06	1.76	1.49	1.99	-0.03	1.00	4287	1.62	0.06	1.62	1.37	1.87	-0.08	1.00
	BMI (kg/m²)	3822	27.48	4.01	26.97	17.21	56.26	1.00	-0.03	4286	26.63	5.52	25.45	14.38	55.20	1.00	-0.08
	Weight (kg)	3822	85.54	13.76	84.09	44.55	182.27	0.89	0.43	4286	70.00	14.99	66.82	36.36	141.82	0.94	0.26
B58C-T1DGC	Age (yrs)	1259	45.31	0.34	45.33	44.50	46.00	-0.02	-0.05	1328	45.27	0.34	45.25	44.50	46.00	0.00	-0.04
	Height (m)	1261	1.76	0.07	1.76	1.55	1.99	-0.03	1.00	1330	1.63	0.06	1.63	1.40	1.85	-0.07	1.00
	BMI (kg/m²)	1259	28.02	4.19	27.56	16.84	51.63	1.00	-0.03	1328	26.97	5.58	25.73	17.18	52.20	1.00	-0.07
	Weight (kg)	1259	87.05	14.41	86.00	50.80	177.10	0.89	0.43	1328	71.63	15.45	68.40	43.00	155.30	0.93	0.29
B58C-WTCCC	Age (yrs)	741	44.89	0.34	44.75	44.50	45.60	-0.01	-0.04	738	44.89	0.35	44.75	44.50	45.60	0.02	0.02
	Height (m)	741	1.76	0.07	1.76	1.52	2.02	-0.05	1.00	738	1.62	0.06	1.63	1.42	1.80	-0.10	1.00
	BMI (kg/m²)	741	27.84	4.29	27.23	15.93	48.41	1.00	-0.05	738	26.92	5.44	25.56	17.34	56.55	1.00	-0.10
	Weight (kg)	741	86.56	14.63	85.20	51.00	137.50	0.87	0.39	738	70.96	14.68	68.20	41.80	139.40	0.91	0.29

BRIGHT	Age (yrs)	719	56.29	11.15	57.00	21.00	84.00	-0.12	-0.24	1087	57.43	11.23	58.00	21.00	85.00	0.07	-0.24
	Height (m)	719	1.74	0.07	1.74	1.51	1.95	-0.06	1.00	1087	1.61	0.06	1.61	1.39	1.81	-0.08	1.00
	BMI (kg/m²)	719	27.74	3.28	27.68	17.20	38.26	1.00	-0.06	1087	27.36	4.04	27.03	16.85	41.66	1.00	-0.08
	Weight (kg)	719	84.22	11.90	83.45	51.00	121.00	0.80	0.54	1087	71.19	11.55	69.90	41.70	122.80	0.87	0.41
CAPS1 cases	Age (yrs)	505	68.15	7.38	67.90	49.50	81.10	-0.16	-0.19	NA	NA	NA	NA	NA	NA	NA	NA
	Height (m)	489	1.77	0.07	1.77	1.58	1.97	-0.04	1.00	NA	NA	NA	NA	NA	NA	NA	NA
	BMI (kg/m²)	484	26.42	3.48	26.01	18.36	41.77	1.00	-0.04	NA	NA	NA	NA	NA	NA	NA	NA
	Weight (kg)	485	82.50	12.26	82.00	47.00	135.00	0.86	0.47	NA	NA	NA	NA	NA	NA	NA	NA
CAPS1 controls	Age (yrs)	506	66.36	7.50	65.90	44.90	79.80	-0.17	-0.25	NA	NA	NA	NA	NA	NA	NA	NA
	Height (m)	491	1.77	0.07	1.76	1.58	2.01	0.04	1.00	NA	NA	NA	NA	NA	NA	NA	NA
	BMI (kg/m²)	483	26.49	3.58	26.25	16.60	58.36	1.00	0.04	NA	NA	NA	NA	NA	NA	NA	NA
	Weight (kg)	485	82.75	13.10	82.00	53.00	187.00	0.88	0.51	NA	NA	NA	NA	NA	NA	NA	NA
CAPS2 cases	Age (yrs)	1483	66.13	7.07	65.40	44.90	82.20	-0.09	-0.24	NA	NA	NA	NA	NA	NA	NA	NA
	Height (m)	1483	1.77	0.06	1.77	1.54	2.00	0.02	1.00	NA	NA	NA	NA	NA	NA	NA	NA
	BMI (kg/m²)	1423	26.34	3.37	25.95	15.74	55.24	1.00	0.02	NA	NA	NA	NA	NA	NA	NA	NA
	Weight (kg)	1424	82.53	12.24	82.00	47.00	185.00	0.88	0.49	NA	NA	NA	NA	NA	NA	NA	NA
CAPS2 controls	Age (yrs)	519	67.24	7.35	66.90	49.10	80.10	-0.05	-0.09	NA	NA	NA	NA	NA	NA	NA	NA
	Height (m)	519	1.76	0.06	1.76	1.59	1.98	-0.07	1.00	NA	NA	NA	NA	NA	NA	NA	NA
	BMI (kg/m²)	500	26.03	3.32	25.75	17.56	45.20	1.00	-0.07	NA	NA	NA	NA	NA	NA	NA	NA
	Weight (kg)	504	80.80	11.38	80.00	55.00	140.00	0.88	0.42	NA	NA	NA	NA	NA	NA	NA	NA
CAD-WTCCC	Age (yrs)	1491	59.96	7.98	61.00	35.00	82.00	-0.16	-0.10	388	60.28	8.47	61.00	36.00	81.00	-0.09	-0.12
	Height (m)	1491	1.74	0.07	1.74	1.40	1.98	-0.08	1.00	388	1.60	0.07	1.59	1.42	1.78	-0.09	1.00
	BMI (kg/m²)	1489	27.55	3.91	27.13	16.53	53.40	1.00	-0.08	387	27.84	5.23	27.18	12.81	51.73	1.00	-0.09
	Weight (kg)	1489	83.25	13.07	82.50	37.70	173.00	0.86	0.44	387	71.04	14.13	69.20	29.20	149.50	0.91	0.33
CHS	Age (yrs)	1281	73.00	5.66	72.00	65.00	95.00	-0.15	-0.22	1957	71.90	5.15	71.00	65.00	98.00	-0.14	-0.23
	Height (m)	1277	1.73	0.07	1.73	1.51	1.93	-0.04	1.00	1955	1.59	0.06	1.59	1.24	1.78	-0.04	1.00
	BMI (kg/m²)	1276	26.40	3.50	26.10	18.60	44.20	1.00	-0.04	1952	26.40	4.78	25.80	18.50	48.30	1.00	-0.05
	Weight (kg)	1276	79.70	11.90	79.00	50.00	145.00	0.86	0.46	1952	67.10	12.90	65.50	37.30	133.20	0.91	0.04
CoLaus	Age (yrs)	2547	52.92	10.77	52.20	34.90	75.10	0.18	-0.19	2862	53.88	10.72	53.70	35.00	75.40	0.16	-0.20
	Height (m)	2547	1.75	0.07	1.75	1.33	1.98	-0.15	1.00	2862	1.63	0.07	1.63	1.31	1.85	-0.21	1.00
	BMI (kg/m²)	2547	26.64	4.19	26.20	11.70	81.10	1.00	-0.15	2861	25.15	4.91	24.20	8.10	59.20	1.00	-0.21
	Weight (kg)	2547	81.54	13.41	79.90	36.50	175.40	0.85	0.38	2861	66.43	12.98	64.00	21.40	171.00	0.91	0.22
deCODE	Age (yrs)	9213	64.74	15.93	78.00	18.00	103.00	-0.15	-0.31	17586	57.94	18.46	43.00	11.50	108.00	0.08	-0.34
	Height (m)	9213	1.78	0.07	1.80	1.30	2.07	0.02	1.00	17586	1.65	0.06	1.69	1.34	1.99	-0.07	1.00
	BMI (kg/m²)	9213	27.71	4.70	45.99	14.52	72.14	1.00	0.02	17586	26.83	5.49	18.56	13.67	73.51	1.00	-0.07

	Weight (kg)	9213	87.89	16.57	149.00	40.00	216.00	0.88	0.44	17586	73.49	15.77	53.00	33.00	220.00	0.92	0.30
DGI cases	Age (yrs)	687	63.22	10.32	64.28	31.36	91.04	-0.19	-0.24	630	65.43	10.45	66.46	31.12	93.09	-0.26	-0.21
	Height (m)	687	1.74	0.06	1.74	1.43	2.00	-0.02	1.00	630	1.61	0.06	1.61	1.41	1.85	-0.08	1.00
	BMI (kg/m²)	687	28.15	3.87	27.97	18.05	46.71	1.00	-0.02	630	28.78	4.86	28.20	18.51	53.73	1.00	-0.08
	Weight (kg)	687	85.58	13.34	84.80	53.40	148.00	0.87	0.47	630	74.60	13.38	73.50	43.80	141.00	0.90	0.35
DGI controls	Age (yrs)	553	58.11	10.34	58.28	31.71	84.78	-0.01	-0.03	537	59.11	10.27	59.60	33.74	89.94	-0.02	-0.27
	Height (m)	553	1.76	0.06	1.76	1.57	2.00	-0.03	1.00	537	1.63	0.06	1.63	1.42	1.87	-0.10	1.00
	BMI (kg/m²)	553	26.62	3.20	26.37	16.95	43.89	1.00	-0.03	537	26.72	4.16	26.20	17.67	45.37	1.00	-0.10
	Weight (kg)	553	82.21	11.39	80.30	50.80	143.00	0.86	0.49	537	70.52	11.69	69.50	43.00	124.00	0.89	0.35
EGCUT	Age (yrs)	697	40.62	16.78	38.00	18.00	90.00	0.41	-0.39	720	42.88	15.93	42.00	18.00	92.00	0.35	-0.26
	Height (m)	697	1.79	0.07	1.79	1.58	2.03	-0.15	1.00	720	1.65	0.06	1.65	1.45	1.84	-0.14	1.00
	BMI (kg/m²)	697	26.05	4.61	25.39	15.82	54.00	1.00	-0.15	720	26.25	6.02	25.08	15.90	58.40	1.00	-0.14
	Weight (kg)	697	83.32	15.27	82.00	49.00	191.00	0.90	0.30	720	71.41	16.36	68.00	39.00	160.00	0.94	0.20
EPIC-Obesity Study	Age (yrs)	1621	59.8	9.0	60.0	39.0	77.0	0.03	-0.25	1931	58.8	8.9	59.0	39.0	77.0	0.08	-0.26
	Height (m)	1621	1.74	0.07	1.74	1.49	1.97	-0.05	1.00	1931	1.61	0.06	1.61	1.25	1.83	-0.13	1.00
	BMI (kg/m²)	1621	28.3	3.9	28.2	16.9	43.6	1.00	-0.05	1931	28.6	5.2	28.4	16.1	47.6	1.00	-0.13
	Weight (kg)	1621	85.5	13.3	85.0	42.8	137.6	0.87	0.45	1931	74.0	14.1	72.8	44.6	126.6	0.92	0.27
ERF (EUROSPAN)	Age (yrs)	890	50.14	14.98	50.67	18.00	88.60	0.14	-0.49	1170	49.30	15.34	49.52	18.03	92.10	0.27	-0.42
	Height (m)	890	1.75	0.07	1.75	1.52	1.96	-0.08	1.00	1170	1.61	0.07	1.62	1.41	1.83	-0.11	1.00
	BMI (kg/m²)	890	27.14	3.98	26.78	15.85	42.44	1.00	-0.08	1170	26.36	4.77	25.64	15.54	45.37	1.00	-0.11
	Weight (kg)	890	82.70	13.52	81.40	48.00	133.30	0.86	0.43	1170	68.96	13.14	67.00	42.10	133.90	0.90	0.32
Fenland	Age (yrs)	615	44.48	7.32	45.00	30.00	57.00	0.08	-0.09	787	45.34	7.18	46.00	30.00	57.00	0.09	-0.11
	Height (m)	615	1.77	0.07	1.77	1.59	2.01	-0.01	1.00	787	1.64	0.06	1.64	1.43	1.90	-0.07	1.00
	BMI (kg/m²)	615	27.62	4.07	27.27	18.62	56.66	1.00	-0.01	787	26.68	5.46	25.44	17.27	55.39	1.00	-0.07
	Weight (kg)	615	86.76	13.87	85.50	49.40	155.70	0.83	0.46	787	71.48	15.25	68.30	42.40	142.50	0.93	0.28
FHS controls	Age (yrs)	218	52.09	12.20	54.19	26.99	76.86	0.10	-0.18	216	58.25	8.57	59.10	27.33	81.09	-0.06	-0.11
	Height (m)	208	1.77	0.07	1.78	1.55	1.98	-0.07	1.00	207	1.62	0.06	1.63	1.46	1.81	0.04	1.00
	BMI (kg/m²)	208	27.74	3.59	27.10	19.56	42.51	1.00	-0.07	207	26.64	4.66	25.61	17.48	43.39	1.00	0.04
	Weight (kg)	208	87.22	13.09	84.80	57.61	131.09	0.83	0.50	207	70.34	13.59	67.59	43.09	122.05	0.93	0.41
FHS cases	Age (yrs)	220	54.20	11.87	55.75	26.38	74.14	0.02	-0.27	243	57.43	10.08	58.42	26.48	84.00	-0.04	-0.32
	Height (m)	208	1.77	0.07	1.77	1.58	1.96	-0.18	1.00	233	1.62	0.06	1.62	1.42	1.79	-0.02	1.00
	BMI (kg/m²)	208	28.51	4.68	28.15	15.96	45.72	1.00	-0.18	233	28.27	6.51	26.75	18.43	50.18	1.00	-0.02
	Weight (kg)	208	89.04	15.03	87.77	51.71	146.51	0.89	0.27	233	73.87	17.94	69.40	45.36	1.00	0.95	0.29
FRAM	Age (yrs)	3700	38.72	8.73	38.00	21.00	72.00	0.16	-0.05	4389	38.23	8.63	38.00	21.00	70.00	0.27	-0.08
	Height (m)	3700	1.77	0.07	1.77	1.52	2.00	-0.04	1.00	4389	1.63	0.06	1.63	1.40	1.85	-0.07	1.00

	BMI (kg/m²)	3700	27.07	4.18	26.61	16.91	56.54	1.00	-0.04	4384	24.88	5.25	23.57	14.96	60.58	1.00	-0.07
	Weight (kg)	3700	84.43	14.43	82.56	44.00	177.36	0.86	0.44	4384	65.84	14.64	62.60	38.10	170.10	0.89	0.35
FTC	Age (yrs)	NA	NA	NA	NA	NA	NA	NA	NA	126	63.49	12.08	66.28	26.52	75.94	0.27	-0.22
	Height (m)	NA	NA	NA	NA	NA	NA	NA	NA	125	1.61	0.06	1.61	1.47	1.78	-0.18	1.00
	BMI (kg/m²)	NA	NA	NA	NA	NA	NA	NA	NA	125	25.07	3.41	24.65	18.69	35.04	1.00	-0.18
	Weight (kg)	NA	NA	NA	NA	NA	NA	NA	NA	126	65.53	9.68	64.25	46.50	100.50	0.86	0.30
FUSION controls	Age (yrs)	572	63.41	7.62	64.00	46.00	90.91	-0.05	-0.24	599	63.71	7.27	64.75	42.60	89.15	0.05	-0.29
	Height (m)	569	1.74	0.06	1.74	1.56	1.91	-0.05	1.00	598	1.60	0.06	1.60	1.44	1.79	-0.12	1.00
	BMI (kg/m²)	572	27.02	3.53	26.78	19.22	51.07	1.00	-0.05	599	27.24	4.15	26.80	17.50	45.90	1.00	-0.12
	Weight (kg)	572	81.40	11.98	80.65	52.10	151.10	0.84	0.46	599	69.99	11.44	68.80	45.70	127.10	0.88	0.33
FUSION cases	Age (yrs)	623	62.06	7.33	62.41	40.77	77.81	-0.21	-0.18	469	63.66	7.75	64.01	45.00	83.19	-0.21	-0.22
	Height (m)	617	1.73	0.06	1.73	1.52	1.97	0.00	1.00	465	1.60	0.06	1.59	1.40	1.76	0.02	1.00
	BMI (kg/m²)	623	29.44	4.02	29.14	18.19	43.14	1.00	0.00	469	31.20	5.25	30.71	16.00	47.59	1.00	0.02
	Weight (kg)	623	88.43	13.58	88.00	50.90	144.00	0.89	0.42	469	79.51	14.70	76.90	35.00	125.50	0.91	0.39
GENMETS controls	Age (yrs)	401	48.91	10.15	49.00	30.00	74.00	0.03	-0.27	422	48.60	10.18	49.00	30.00	74.00	0.04	-0.26
	Height (m)	401	1.75	0.07	1.75	1.55	1.80	-0.16	1.00	422	1.75	0.07	1.75	1.55	1.96	-0.16	1.00
	BMI (kg/m²)	401	25.41	3.08	24.94	17.09	39.04	1.00	-0.16	422	25.34	3.15	24.92	17.09	39.04	1.00	-0.16
	Weight (kg)	401	78.03	10.33	77.00	54.00	116.00	0.82	0.43	422	77.62	10.60	77.00	51.00	113.00	0.81	0.45
GENMETS cases	Age (yrs)	410	49.11	10.55	49.00	30.00	75.00	-0.07	-0.24	414	52.25	11.62	51.00	30.00	75.00	-0.07	-0.27
	Height (m)	410	1.76	0.07	1.76	1.58	1.97	-0.13	1.00	414	1.61	0.07	1.61	1.35	1.82	-0.13	1.00
	BMI (kg/m²)	410	29.45	3.62	28.84	23.19	47.07	1.00	-0.13	414	29.62	4.88	28.68	20.58	45.78	1.00	-0.13
	Weight (kg)	410	91.16	12.56	89.00	65.00	151.00	0.81	0.47	414	76.60	13.40	75.00	49.00	123.00	0.87	0.36
GerMiFSI	Age (yrs)	394	57.27	8.57	59.00	32.00	82.00	-0.08	-0.17	206	60.39	8.67	61.00	36.00	82.00	0.04	-0.02
	Height (m)	394	1.75	0.06	1.75	1.59	1.97	-0.05	1.00	206	1.63	0.06	1.63	1.44	1.79	0.00	1.00
	BMI (kg/m²)	394	27.36	3.30	26.83	18.42	46.24	1.00	-0.05	206	27.17	4.17	26.91	19.05	40.75	1.00	0.00
	Weight (kg)	394	83.92	11.67	83.00	60.00	140.00	0.86	0.46	206	72.29	12.24	71.00	48.00	115.00	0.90	0.42
GerMiFSII	Age (yrs)	901	60.14	12.17	59.00	29.00	88.00	-0.01	-0.01	223	62.80	12.76	61.00	34.00	90.00	-0.17	0.17
	Height (m)	901	1.74	0.07	1.74	1.52	2.00	0.02	1.00	223	1.62	0.06	1.61	1.50	1.79	-0.20	1.00
	BMI (kg/m²)	901	27.82	3.54	27.41	18.44	54.08	1.00	-0.02	223	28.06	4.76	27.69	16.90	46.30	1.00	-0.20
	Weight (kg)	901	83.00	12.49	83.00	50.20	160.00	0.85	0.50	223	73.55	12.62	72.10	47.00	130.00	0.90	0.23
KORA S3	Age (yrs)	813	52.96	10.09	54.00	25.00	69.00	0.22	-0.33	831	52.09	10.08	53.00	25.00	69.00	0.33	-0.32
	Height (m)	813	1.74	0.07	1.74	1.51	1.96	-0.14	1.00	830	1.61	0.06	1.61	1.44	1.80	-0.25	1.00
	BMI (kg/m²)	813	27.69	3.45	27.29	18.73	40.67	1.00	-0.14	829	26.98	4.64	26.40	16.71	45.43	1.00	-0.25
	Weight (kg)	813	83.58	11.46	83.30	59.00	132.50	0.79	0.44	829	69.87	11.88	68.30	42.50	121.80	0.88	0.19
KORA S4	Age (yrs)	884	54.22	8.92	54.00	28.00	72.00	0.13	-0.31	930	53.62	8.80	53.00	25.00	74.00	0.32	-0.27

	Height (m)	883	1.74	0.07	1.74	1.56	1.95	-0.13	1.00	928	1.61	0.06	1.61	1.44	1.83	-0.18	1.00
	BMI (kg/m²)	883	27.99	3.91	27.59	18.31	55.11	1.00	-0.13	928	27.49	5.07	26.78	18.21	51.22	1.00	-0.18
	Weight (kg)	883	85.13	12.93	84.00	54.20	192.70	0.83	0.40	929	71.46	13.30	69.60	43.90	142.00	0.90	0.23
MICROS	Age (yrs)	475	45.09	15.67	41.97	18.19	87.85	0.28	-0.45	622	45.38	16.41	42.55	18.00	83.88	0.40	-0.52
	Height (m)	467	1.73	0.07	1.73	1.53	1.95	-0.07	1.00	612	1.61	0.07	1.61	1.40	1.79	-0.28	1.00
	BMI (kg/m²)	475	26.07	3.96	25.62	18.13	42.75	1.00	-0.07	622	25.28	5.32	24.27	14.03	71.26	1.00	-0.28
	Weight (kg)	468	78.38	13.32	76.90	47.00	127.50	0.86	0.43	612	65.16	13.19	63.00	36.60	169.00	0.91	0.13
MIGEN	Age (yrs)	1622	45.40	6.97	45.70	19.40	92.00	0.03	-0.08	1030	49.39	7.40	51.00	18.71	61.00	0.09	-0.14
	Height (m)	1622	1.76	0.08	1.75	1.53	2.08	0.02	1.00	1030	163.10	0.08	1.63	1.10	1.96	-0.06	1.00
	BMI (kg/m²)	1622	27.93	4.57	27.40	17.49	54.30	1.00	0.02	1030	27.96	7.03	26.35	14.78	78.41	1.00	-0.06
	Weight (kg)	1622	86.57	16.30	84.00	52.00	181.60	0.88	0.47	1030	74.42	19.71	70.00	43.09	205.02	0.94	0.27
NBS-WTCCC	Age (yrs)	696	45.41	11.77	47.00	17.00	69.00	0.06	-0.15	745	41.44	12.58	42.00	17.00	69.00	0.15	-0.19
	Height (m)	696	1.78	0.07	1.78	1.50	2.00	-0.07	1.00	745	1.65	0.07	1.65	1.48	1.83	-0.20	1.00
	BMI (kg/m²)	694	26.76	4.12	26.30	18.13	53.19	1.00	-0.07	743	25.75	4.46	24.86	18.08	47.22	1.00	-0.20
	Weight (kg)	694	85.03	14.35	82.73	54.09	173.00	0.88	0.41	743	69.74	12.21	66.82	50.00	127.27	0.89	0.25
NFBC1966	Age (yrs)	2250	31.00	0.00	31.00	31.00	31.00	NA	NA	2249	31.00	0.00	31.00	31.00	31.00	NA	NA
	Height (m)	2250	1.78	0.06	1.78	1.52	2.03	-0.04	1.00	2249	1.65	0.06	1.65	1.05	1.87	-0.10	1.00
	BMI (kg/m²)	2250	25.18	3.62	24.86	15.32	47.58	1.00	-0.04	2247	24.16	4.68	23.13	15.43	54.35	1.00	-0.10
	Weight (kg)	2250	80.15	12.72	78.70	49.40	150.40	0.89	0.42	2247	65.52	13.24	63.00	29.20	165.40	0.92	0.28
NHS	Age (yrs)	NA	NA	NA	NA	NA	NA	NA	NA	2265	54.32	6.67	55.00	21.00	66.00	0.05	-0.02
	Height (m)	NA	NA	NA	NA	NA	NA	NA	NA	2265	1.64	0.06	1.63	1.45	1.98	-0.10	1.00
	BMI (kg/m²)	NA	NA	NA	NA	NA	NA	NA	NA	2265	25.13	4.53	24.13	16.40	53.14	1.00	-0.10
	Weight (kg)	NA	NA	NA	NA	NA	NA	NA	NA	2265	149.16	27.78	144.00	84.00	310.00	0.92	0.29
NSPHS	Age (yrs)	309	47.56	20.83	48.00	15.00	87.00	0.31	-0.32	347	46.47	20.60	45.00	14.00	91.00	0.49	-0.40
	Height (m)	308	1.71	0.07	1.72	1.48	1.89	-0.04	1.00	344	1.58	0.07	1.59	1.40	1.75	-0.16	1.00
	BMI (kg/m²)	307	26.75	4.54	26.23	17.78	46.49	1.00	-0.04	340	25.97	5.07	24.98	16.44	46.68	1.00	-0.16
	Weight (kg)	307	78.42	14.66	77.00	51.00	138.00	0.88	0.42	342	64.99	13.11	63.00	38.00	121.00	0.89	0.29
NTRNESDA	Age (yrs)	1211	46.08	13.43	48.00	18.00	81.00	0.26	-0.29	2311	42.64	13.23	42.00	18.00	78.00	0.23	-0.23
	Height (m)	1211	1.82	0.07	1.82	1.59	2.07	-0.14	1.00	2311	1.69	0.06	1.69	1.50	1.96	-0.15	1.00
	BMI (kg/m²)	1210	26.05	3.92	25.62	15.95	50.21	1.00	-0.14	2306	25.15	4.82	24.19	14.61	53.27	1.00	-0.15
	Weight (kg)	1210	85.89	13.80	84.15	50.10	170.00	0.87	0.36	2306	71.83	14.06	69.20	44.00	167.00	0.92	0.24
ORCADES	Age (yrs)	332	54.27	15.73	54.66	17.29	93.75	0.29	-0.38	384	53.01	15.68	54.27	17.71	97.62	0.25	-0.38
	Height (m)	324	1.75	0.07	1.75	1.59	1.99	-0.22	1.00	371	1.61	0.06	1.61	1.38	1.78	-0.17	1.00
	BMI (kg/m²)	332	28.08	4.27	27.67	16.97	47.10	1.00	-0.22	384	27.48	5.18	26.60	18.47	47.63	1.00	-0.17
	Weight (kg)	324	85.76	13.21	84.25	44.40	148.40	0.87	0.28	371	71.06	13.69	69.10	45.60	123.10	0.92	0.22

PLCO	Age (yrs)	2244	64.2	5.1	64.0	55.0	74.0	-0.11	-0.11	NA	NA	NA	NA	NA	NA	NA	NA
	Height (m)	2244	1.78	0.07	1.78	1.55	2.03	-0.04	1.00	NA	NA	NA	NA	NA	NA	NA	NA
	BMI (kg/m²)	2236	27.5	3.8	27.1	13.3	48.2	1.00	-0.04	NA	NA	NA	NA	NA	NA	NA	NA
	Weight (kg)	2236	87.4	13.6	86.2	38.6	176.9	0.88	0.44	NA	NA	NA	NA	NA	NA	NA	NA
PROCARDIS	Age (yrs)	1700	59.29	7.08	60.00	34.00	82.00	-0.07	-0.14	612	61.21	6.72	62.00	33.00	81.00	0.03	-0.21
	Height (m)	1700	1.75	0.07	1.75	1.51	2.06	-0.10	1.00	612	1.63	0.07	1.64	1.44	1.85	-0.22	1.00
	BMI (kg/m²)	1700	27.60	3.80	27.14	18.34	48.23	1.00	-0.10	612	26.71	5.00	25.94	15.43	51.37	1.00	-0.22
	Weight (kg)	1700	84.51	12.91	83.50	51.00	159.00	0.84	0.44	612	71.21	13.30	69.00	42.00	145.00	0.89	0.23
RS-I	Age (yrs)	2427	68.13	8.16	67.05	55.01	97.81	-0.08	-0.31	3547	70.32	9.60	69.40	55.00	99.22	0.05	-0.38
	Height (m)	2372	1.75	0.07	1.75	1.51	1.98	-0.05	1.00	3375	1.61	0.07	1.62	1.01	1.92	-0.15	1.00
	BMI (kg/m²)	2372	25.68	2.99	25.61	14.19	38.19	1.00	-0.05	3372	26.74	4.10	26.31	15.43	59.50	1.00	-0.15
	Weight (kg)	2375	78.58	10.74	77.80	41.00	122.30	0.82	0.53	3383	69.59	11.29	68.70	40.10	146.50	0.85	0.37
RUNMC	Age (yrs)	1839	63.47	8.34	64.00	24.00	91.00	-0.02	-0.12	1132	55.41	11.14	64.00	25.00	91.00	0.17	-0.23
	Height (m)	1777	1.77	0.07	1.85	1.55	2.00	-0.10	1.00	1096	1.66	0.06	1.75	1.38	1.85	-0.15	1.00
	BMI (kg/m²)	1777	25.98	3.66	21.90	16.10	61.30	1.00	-0.10	1096	25.44	4.26	24.50	17.30	52.70	1.00	-0.15
	Weight (kg)	1777	81.49	12.33	75.00	46.00	185.00	0.87	0.40	1096	70.30	12.16	75.00	46.00	150.00	0.90	0.29
SardiNIA	Age (yrs)	1886	44.08	18.10	42.90	14.00	93.90	0.51	-0.46	2419	43.19	17.30	42.10	14.00	101.30	0.55	-0.50
	Height (m)	1883	1.66	0.07	1.66	1.44	1.96	-0.22	1.00	2415	1.55	0.06	1.55	1.31	1.78	-0.31	1.00
	BMI (kg/m²)	1885	26.15	4.11	25.90	14.90	42.90	1.00	-0.22	2416	24.75	5.03	23.80	13.90	53.30	1.00	-0.31
	Weight (kg)	1883	72.27	11.71	72.00	34.00	135.00	0.84	0.33	2415	59.17	11.40	57.00	32.00	145.00	0.90	0.11
SASBAC cases	Age (yrs)	NA	NA	NA	NA	NA	NA	NA	NA	795	62.64	6.26	63.00	50.00	75.00	0.11	-0.08
	Height (m)	NA	NA	NA	NA	NA	NA	NA	NA	794	1.64	0.06	1.65	1.47	1.82	-0.16	1.00
	BMI (kg/m²)	NA	NA	NA	NA	NA	NA	NA	NA	793	25.79	4.00	25.21	16.22	46.67	1.00	-0.16
	Weight (kg)	NA	NA	NA	NA	NA	NA	NA	NA	794	69.68	11.18	68.00	40.00	117.00	0.86	0.30
SASBAC controls	Age (yrs)	NA	NA	NA	NA	NA	NA	NA	NA	764	62.77	6.34	63.00	49.00	75.00	0.02	-0.05
	Height (m)	NA	NA	NA	NA	NA	NA	NA	NA	758	1.64	0.05	1.64	1.28	1.81	-0.06	1.00
	BMI (kg/m²)	NA	NA	NA	NA	NA	NA	NA	NA	755	25.52	4.10	25.22	16.94	59.52	1.00	-0.06
	Weight (kg)	NA	NA	NA	NA	NA	NA	NA	NA	760	68.67	11.69	67.00	42.00	168.00	0.89	0.33
SEARCH /UKOPS	Age (yrs)	NA	NA	NA	NA	NA	NA	NA	NA	1710	57.15	10.20	58.00	20.00	91.00	-0.09	-0.13
	Height (m)	NA	NA	NA	NA	NA	NA	NA	NA	1592	1.63	0.07	1.63	1.35	1.83	-0.14	1.00
	BMI (kg/m²)	NA	NA	NA	NA	NA	NA	NA	NA	1556	26.99	5.20	25.99	17.47	53.67	1.00	-0.14
	Weight (kg)	NA	NA	NA	NA	NA	NA	NA	NA	1581	71.32	13.99	69.00	44.00	135.17	0.91	0.27
SHIP	Age (yrs)	2019	50.88	16.43	52.00	20.00	80.00	0.25	-0.48	2073	48.58	16.02	48.00	20.00	81.00	0.41	-0.48
	Height (m)	2019	1.75	0.07	1.75	1.48	1.98	-0.12	1.00	2073	1.63	0.07	1.63	1.42	1.94	-0.26	1.00
	BMI (kg/m²)	2019	27.68	4.04	27.41	18.06	48.07	1.00	-0.12	2073	26.92	5.31	26.16	16.10	52.40	1.00	-0.26

	Weight (kg)	2019	85.06	13.56	83.80	49.90	156.40	0.83	0.40	2073	71.20	13.74	69.20	41.30	133.30	0.89	0.16
T2D-WTCCC	Age (yrs)	1105	58.95	9.91	59.00	29.00	96.00	-0.31	-0.17	798	57.94	10.45	59.00	27.00	85.00	-0.30	-0.16
	Height (m)	1105	1.75	0.07	1.75	1.50	1.98	-0.02	1.00	798	1.61	0.07	1.61	1.37	1.83	0.01	1.00
	BMI (kg/m²)	1105	30.29	5.36	29.71	18.02	55.91	1.00	-0.02	798	32.56	6.87	31.52	17.91	62.37	1.00	0.01
	Weight (kg)	1105	93.37	17.86	91.17	47.63	161.94	0.91	0.40	798	85.04	19.29	82.56	43.00	155.70	0.93	0.37
TwinsUK	Age (yrs)	NA	NA	NA	NA	NA	NA	NA	NA	1479	46.19	12.31	47.55	16.62	76.54	0.15	-0.20
	Height (m)	NA	NA	NA	NA	NA	NA	NA	NA	1479	1.62	0.06	1.63	1.42	1.80	-0.12	1.00
	BMI (kg/m²)	NA	NA	NA	NA	NA	NA	NA	NA	1477	25.02	4.80	24.06	13.22	52.71	1.00	-0.12
	Weight (kg)	NA	NA	NA	NA	NA	NA	NA	NA	1477	66.03	12.97	64.00	35.10	140.90	0.92	0.27
VIS	Age (yrs)	328	55.95	14.94	57.00	18.00	88.00	0.23	-0.40	467	56.97	15.64	57.00	18.00	93.00	0.30	-0.45
	Height (m)	325	1.76	0.07	1.76	1.58	2.04	-0.10	1.00	459	1.62	0.07	1.62	1.43	1.91	-0.20	1.00
	BMI (kg/m²)	328	27.55	3.69	27.49	18.36	40.69	1.00	-0.10	467	27.18	4.50	27.08	17.01	52.02	1.00	-0.20
	Weight (kg)	325	85.56	13.01	84.80	50.90	136.50	0.83	0.47	445	70.99	12.45	69.80	46.60	153.00	0.89	0.26
Stage 2 (in-silico replication studies)																	
BHS	Age (yrs)	558	53.47	17.15	53.65	17.60	91.40	0.15	-0.38	770	53.71	17.07	53.05	17.30	90.50	0.11	-0.43
	Height (m)	558	1.75	0.07	1.75	1.53	1.99	-0.09	1.00	770	1.62	0.06	1.62	1.35	1.90	-0.15	1.00
	BMI (kg/m²)	558	26.62	3.57	26.25	15.77	40.12	1.00	-0.09	769	25.49	4.42	24.66	16.82	40.77	1.00	-0.15
	Weight (kg)	558	81.80	12.31	80.25	46.40	127.00	0.83	0.47	769	67.06	12.02	65.00	34.80	109.00	0.90	0.29
Corogene	Age (yrs)	2266	59.66	12.83	61.00	25.00	92.00	-0.03	-0.26	1490	62.61	13.47	65.00	25.00	94.00	0.09	-0.30
	Height (m)	2267	1.76	0.07	1.76	1.34	2.03	-0.04	1.00	1491	1.62	0.07	1.62	1.05	1.85	-0.14	1.00
	BMI (kg/m²)	2265	27.39	4.23	26.79	15.95	54.88	1.00	-0.04	1491	26.87	5.21	26.07	13.63	57.68	1.00	-0.14
	Weight (kg)	2265	85.00	14.42	83.50	44.00	170.00	0.89	0.41	1491	70.14	13.88	68.30	36.00	144.00	0.90	0.28
EGCUT	Age (yrs)	135	40.93	17.81	36.50	18.00	80.00	0.33	-0.55	210	41.03	16.46	39.00	18.00	87.00	0.41	-0.37
	Height (m)	135	1.79	0.07	1.80	1.58	2.04	-0.14	1.00	210	1.66	0.07	1.66	1.44	1.84	-0.25	1.00
	BMI (kg/m²)	135	26.03	4.95	25.11	17.30	43.65	1.00	-0.14	210	25.63	6.09	24.02	17.00	48.24	1.00	-0.25
	Weight (kg)	135	83.68	16.41	80.50	50.00	143.00	0.91	0.27	210	70.46	16.22	66.50	40.00	136.00	0.93	0.10
FHS	Age (yrs)	662	48.20	13.70	46.30	25.60	85.70	0.15	-0.24	880	47.50	13.00	45.00	25.70	85.80	0.19	-0.26
	Height (m)	632	1.77	0.07	1.77	1.57	2.03	-0.09	1.00	831	1.63	0.06	1.63	1.41	1.96	-0.12	1.00
	BMI (kg/m²)	632	27.80	4.30	27.20	18.40	46.20	1.00	-0.09	831	27.10	6.10	26.10	16.50	55.00	1.00	-0.12
	Weight (kg)	632	87.10	14.60	85.30	55.30	140.60	0.88	0.39	831	72.30	16.60	68.90	41.70	144.20	0.94	0.22
FINGESTURE cases	Age (yrs)	745	61.19	10.58	62.00	34.00	85.00	-0.13	-0.33	198	67.44	10.33	68.00	31.00	85.00	-0.05	-0.28
	Height (m)	745	1.74	0.07	1.74	1.55	1.97	0.10	1.00	198	1.60	0.06	1.60	1.46	1.76	-0.02	1.00
	BMI (kg/m²)	739	27.22	3.93	27.02	16.20	44.80	1.00	0.10	196	28.14	5.17	27.98	16.67	46.09	1.00	-0.02
	Weight (kg)	743	82.32	14.09	81.00	42.00	150.00	0.89	0.53	197	71.91	14.06	71.60	37.50	112.00	0.92	0.38

GOOD	Age (yrs)	938	18.90	0.60	18.80	18.00	20.10	0.03	0.01	NA	NA	NA	NA	NA	NA	NA	NA
	Height (m)	938	1.82	0.07	1.82	1.61	2.03	-0.05	1.00	NA	NA	NA	NA	NA	NA	NA	NA
	BMI (kg/m²)	938	22.40	3.20	21.90	16.10	41.60	1.00	-0.05	NA	NA	NA	NA	NA	NA	NA	NA
	Weight (kg)	938	73.90	11.60	72.00	51.30	127.00	0.88	0.42	NA	NA	NA	NA	NA	NA	NA	NA
HBCS	Age (yrs)	737	61.41	2.75	60.80	57.00	69.30	-0.03	-0.15	991	61.55	3.05	60.90	56.70	69.80	-0.10	0.03
	Height (m)	736	1.77	0.06	1.77	1.59	1.97	-0.03	1.00	990	1.63	0.06	1.63	1.46	1.83	-0.09	1.00
	BMI (kg/m²)	736	27.56	4.30	27.01	18.75	68.39	1.00	-0.03	990	27.75	5.06	26.98	14.79	50.10	1.00	-0.09
	Weight (kg)	737	86.33	14.51	84.50	56.20	213.30	0.92	0.36	990	73.90	13.89	71.70	37.30	133.80	0.93	0.28
HYPERGENES - controls	Age (yrs)	1072	62.27	10.71	59.81	28.00	98.00	-0.09	-0.12	766	64.30	11.28	61.00	44.93	113.00	-0.15	-0.14
	Height (m)	1072	1.71	0.07	1.70	1.50	1.96	-0.14	1.00	766	1.60	0.06	1.60	1.40	1.81	-0.16	1.00
	BMI (kg/m²)	1072	25.95	3.27	25.59	10.15	40.77	1.00	-0.14	766	24.98	3.73	24.60	16.53	41.35	1.00	-0.16
	Weight (kg)	1072	76.10	10.59	75.00	29.00	118.00	0.81	0.46	766	64.25	10.13	63.00	41.00	110.00	0.87	0.34
HYPERGENES - cases	Age (yrs)	1189	49.41	10.42	50.00	17.63	84.00	0.04	-0.33	598	48.45	9.57	49.00	18.38	93.00	0.10	-0.19
	Height (m)	1189	1.72	0.07	1.72	1.48	1.96	-0.08	1.00	598	1.60	0.07	1.60	1.40	1.97	-0.10	1.00
	BMI (kg/m²)	1189	27.42	3.52	27.13	16.00	47.43	1.00	-0.08	598	26.88	4.96	26.21	17.45	52.35	1.00	-0.10
	Weight (kg)	1189	81.33	12.06	80.00	49.00	139.50	0.82	0.51	598	68.59	13.66	67.00	44.00	164.00	0.89	0.36
MGS	Age (yrs)	1247	52.67	16.01	52.00	18.00	90.00	0.02	-0.13	1350	48.48	16.29	48.00	18.00	90.00	0.03	-0.20
	Height (m)	1247	1.79	0.07	1.78	1.58	2.06	0.04	1.00	1350	1.64	0.07	1.65	1.35	2.01	-0.04	1.00
	BMI (kg/m²)	1247	30.85	6.45	29.84	15.83	72.56	1.00	0.04	1350	31.92	8.55	30.32	16.34	69.09	1.00	-0.04
	Weight (kg)	1247	98.77	22.67	95.25	53.98	249.48	0.93	0.38	1350	86.13	24.22	81.65	47.63	201.85	0.95	0.26
NHS	Age (yrs)	NA	NA	NA	NA	NA	NA	NA	NA	3217	53.22	6.96	54.00	22.00	65.00	-0.02	-0.07
	Height (m)	NA	NA	NA	NA	NA	NA	NA	NA	3217	1.64	0.08	1.63	1.35	1.83	-0.04	1.00
	BMI (kg/m²)	NA	NA	NA	NA	NA	NA	NA	NA	2988	27.13	5.63	26.00	17.01	54.87	1.00	-0.04
	Weight (kg)	NA	NA	NA	NA	NA	NA	NA	NA	2988	160.87	35.21	155.00	90.00	340.00	0.94	0.30
RS-II	Age (yrs)	973	64.48	7.59	61.89	55.14	93.95	-0.13	-0.22	1156	65.04	8.33	62.03	55.12	95.33	-0.03	-0.31
	Height (m)	971	1.76	0.06	1.76	1.57	2.03	-0.10	1.00	1153	1.63	0.06	1.63	1.42	1.90	-0.06	1.00
	BMI (kg/m²)	971	26.92	3.36	26.72	16.78	40.52	1.00	-0.10	1151	27.52	4.45	26.89	16.66	50.12	1.00	-0.06
	Weight (kg)	972	83.32	11.58	82.20	54.00	126.80	0.85	0.44	1151	72.77	12.74	71.10	36.20	150.00	0.90	0.38
RS-III	Age (yrs)	879	55.94	5.43	56.12	45.46	84.15	0.09	-0.24	1130	56.20	6.03	56.42	45.75	97.22	0.07	-0.23
	Height (m)	879	1.79	0.07	1.79	1.61	2.00	-0.07	1.00	1130	1.65	0.06	1.65	1.47	1.85	-0.10	1.00
	BMI (kg/m²)	879	28.03	4.07	27.31	18.42	46.68	1.00	-0.07	1130	27.48	5.06	26.55	14.02	56.87	1.00	-0.10
	Weight (kg)	879	89.75	14.32	87.70	58.30	153.50	0.88	0.41	1130	74.89	14.28	72.80	35.00	158.60	0.92	0.29
Sorbs	Age (yrs)	371	48.10	16.70	48.10	18.10	82.10	0.39	-0.43	536	48.00	15.90	48.60	18.00	88.40	0.49	-0.54
	Height (m)	371	1.77	0.07	1.77	1.58	1.95	-0.24	1.00	536	1.64	0.07	1.64	1.44	1.82	-0.32	1.00
	BMI (kg/m²)	371	27.20	4.00	26.80	19.00	43.90	1.00	-0.24	536	26.90	5.50	26.20	15.40	47.40	1.00	-0.32

	Weight (kg)	371	85.40	12.70	84.00	58.00	139.00	0.85	0.30	536	72.10	14.00	70.00	43.00	126.00	0.92	0.07
WGHS	Age (yrs)	NA	NA	NA	NA	NA	NA	NA	NA	23294	54.70	7.12	52.90	38.71	89.89	-0.02	-0.07
	Height (m)	NA	NA	NA	NA	NA	NA	NA	NA	23099	1.64	0.06	1.65	1.30	2.01	-0.06	1.00
	BMI (kg/m²)	NA	NA	NA	NA	NA	NA	NA	NA	22888	25.91	4.96	24.89	14.23	59.58	1.00	-0.06
	Weight (kg)	NA	NA	NA	NA	NA	NA	NA	NA	23046	70.00	14.18	68.04	38.56	175.09	0.92	0.32
YFS	Age (yrs)	1123	37.55	5.06	39.00	30.00	45.00	0.13	-0.12	1320	37.57	5.01	39.00	30.00	45.00	0.11	-0.06
	Height (m)	911	1.80	0.07	1.80	1.57	2.03	0.04	1.00	1084	1.66	0.06	1.66	1.45	1.89	-0.06	1.00
	BMI (kg/m²)	908	26.76	4.29	26.11	17.54	49.35	1.00	0.04	1081	25.32	5.03	24.34	16.56	58.82	1.00	-0.06
	Weight (kg)	908	86.56	15.65	85.00	54.00	166.00	0.91	0.45	1083	69.82	14.55	67.00	42.00	166.00	0.94	0.29
Polygene analysis study																	
QIMR	Age (yrs)	527	23.20	12.00	16.33	15.40	74.00	NA	0.15	948	29.86	14.95	26.00	15.70	84.00	NA	-0.15
	Height (m)	527	1.77	0.07	1.77	1.58	1.99	NA	1.00	948	1.64	0.07	1.64	1.44	1.93	NA	1.00

Supplementary Note

Author Contributions

GIANT Steering Committee

Goncalo R Abecasis, Themistocles L Assimes, Inês Barroso, Sonja I Berndt, Michael Boehnke, Ingrid B Borecki, Panos Deloukas, Caroline S Fox, Timothy M Frayling (lead), Leif C Groop, Talin Haritunians, Joel N Hirschhorn (chair), David J Hunter, Erik Ingelsson, Robert Kaplan, Ruth JF Loos, Mark I McCarthy, Karen L Mohlke, Kari E North, Jeffrey R O'Connell, Leena Peltonen, David Schlessinger, David P Strachan, Unnur Thorsteinsdottir, Cornelia M van Duijn, H.-Erich Wichmann

Discussion group (discussed and decided on analyses specific to the manuscript)

Goncalo R Abecasis, Sonja I Berndt, Michael Boehnke, Karol Estrada, Timothy M Frayling (chair), Daniel Gudbjartsson, Nancy L Heard-Costa, Joel N Hirschhorn (chair), Anne U Jackson, Hana Lango Allen, Guillaume Lettre, Mark I McCarthy, Fernando Rivadeneira, Albert Vernon Smith, Nicole Soranzo, André G Uitterlinden, Michael N Weedon

Writing group (drafted and edited manuscript)

Goncalo R Abecasis, Sonja I Berndt (lead), Michael Boehnke, Karol Estrada (lead), Timothy M Frayling (chair), Joel N Hirschhorn (chair), Hana Lango Allen (lead), Guillaume Lettre (lead), Mark I McCarthy, Fernando Rivadeneira (lead), Nicole Soranzo, André G Uitterlinden, Michael N Weedon (lead)

Height meta-analysis group (analyses specific to the manuscript)

Sonja I Berndt, Karol Estrada, Hana Lango Allen, Guillaume Lettre, Sailaja Vedantam, Michael N Weedon

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

Sonja I Berndt, Iris M Heid, Erik Ingelsson, Anne U Jackson, Hana Lango Allen, Cecilia M Lindgren, Ruth JF Loos (chair), Jian'an Luan, Reedik Mägi, Lu Qi, Joshua C Randall, Elizabeth K Speliotes, Gudmar Thorleifsson, Sailaja Vedantam, Michael N Weedon, Eleanor Wheeler, Cristen J Willer, Thomas Winkler, Tsegaselassie Workalemahu

Pathway analyses

Joel N Hirschhorn (lead), Guillaume Lettre (chair), Ken Sin Lo, Cameron Palmer, Soumya Raychaudhuri (lead), Ayellet V Segrè (lead)

Polygene analyses

Nilanjan Chatterjee (lead), Karol Estrada (lead), Timothy M Frayling, Michael E Goddard, Andrew C Heath, Joel N Hirschhorn, Erik Ingelsson, Guillaume Lettre, Reedik Mägi , Massimo Mangino, Nicholas G Martin, Mark I McCarthy (chair), Grant W Montgomery, Dale R Nyholt, Jeffrey R O'Connell, Ju-Hyun Park, Shaun Purcell, Albert Vernon Smith, Peter M Visscher (lead), Michael N Weedon, Cristen J Willer, Jian Yang

Secondary signals (conditional) analyses

Sonja I Berndt, Ingrid B Borecki, Mary F Feitosa, Teresa Ferreira (lead), Timothy M Frayling, Erik Ingelsson, Hana Lango Allen (lead), Cecilia M Lindgren (chair), Reedik Mägi, Karen L Mohlke, Joshua C Randall, Elizabeth K Speliotes, Michael N Weedon (chair), Robert Weyant (lead), Eleanor Wheeler, Cristen J Willer (chair)

Gene expression (eQTL) analyses

Sonja I Berndt (chair), Anna Dixon, Elin Grundberg, Ke Hao, Joel N Hirschhorn, Lee Kaplan, Tony Kwan, Liming Liang (lead), Tomi Pastinen (lead), Eric E Schadt (lead), Elizabeth K Speliotes, Dominique J Verlaan

Gene-by-gene interaction analyses

Tushar Bhangale, Timothy M Frayling, Cecilia M Lindgren, John RB Perry, Michael N Weedon (chair), Cristen J Willer, Andrew R Wood (lead)

Family-based and population stratification analyses

Elizabeth L Altmaier, Najaf Amin, Joel N Hirschhorn, Hana Lango Allen, Fernando Rivadeneira (chair), Cornelia M van Duijn, Sailaja Vedantam, Michael N Weedon

Imputation and signal validation by direct genotyping analyses

WTCCC-T2D cohort (imputation validation): Timothy M Frayling, Hana Lango Allen, Michael N Weedon (chair); GCI cohort (signal validation): Kristin G Ardlie, Joel N Hirschhorn (chair), Guillaume Lettre, Rany M Salem, Michael C Turchin

Cohort-specific contributions

Stage 1 – Genome-wide association cohorts

Cohort	Author	Overseeing (PI)	Geno-typing	Pheno-typing	Data analysis
ADVANCE	Devin Absher		X	X	X
	Themistocles L Assimes		X	X	X
	Carlos Iribarren	X			
	Joshua W Knowles		X	X	X
	Thomas Quertermous	X			
AGES	Thor Aspelund				X
	Gudny Eiriksdottir	X			
	Vilmundur Gudnason	X			
	Tamara B Harris	X			
	Lenore J Launer	X			
	Albert Vernon Smith				X
Amish	Quince Gibson				X
	Shen Haiqing		X	X	
	Jeffrey R O'Connell				X
	Alan R Shuldiner	X			
ARIC	Eric Boerwinkle	X	X	X	
	Keri L Monda				X
	Tom H Mosley, Jr	X			
	Kari E North	X			X
B58C-T1DGC and B58C-WTCCC	Wendy L McArdle		X		
	David P Strachan	X		X	X
BRIGHT	Mark J Caulfield	X			
	Anna Dominiczak			X	
	Martin Farrall			X	
	Toby Johnson				X
	Patricia B Munroe	X			
CAD-WTCCC	Anthony J Balmforth			X	
	Alistair S Hall	X			
	Suzanne Rafelt				X
	Nilesh J Samani	X			
	John R Thompson				X
CAPS	Henrik Grönberg	X		X	

Cohort	Author	Overseeing (PI)	Geno-typing	Pheno-typing	Data analysis
	Erik Ingelsson	X			X
	Fredrik Wiklund			X	X
	Jianfeng Xu			X	X
CHS	Alice Arnold	X		X	
	Nicole L Glazer				X
	Talin Haritunians		X		
	Robert Kaplan	X		X	
	Barbara McKnight				X
	Jerome I Rotter		X		
CoLaus	Jacques S Beckmann	X			
	Sven Bergmann	X			
	Toby Johnson				X
	Zoltán Kutalik				X
	Vincent Mooser	X	X		
	Dawn Waterworth	X	X		
deCODE	Daniel Gudbjartsson				X
	Kari Stefansson	X			
	Valgerdur Steinthorsdottir				X
	Gudmar Thorleifsson				X
	Unnur Thorsteinsdottir	X			
	G. Bragi Walters			X	
DGI	Peter Almgren			X	
	Leif C Groop	X		X	
	Joel N Hirschhorn				
	Guillaume Lettre				X
	Martin Ridderstråle			X	
	Elizabeth K Speliotes				X
	Sailaja Vedantam				X
EGCUT	Helene Alavere			X	
	Tõnu Esko				X
	Andres Metspalu	X			
	Mari Nelis		X		
	Mari-Liis Tammesoo				X
EPIC	Inês Barroso		X		
	Ruth JF Loos		X		X
	Nicholas J Wareham	X			
	Eleanor Wheeler		X		X
	Jing Hua Zhao				
ERF (EUROSPAN)	Najaf Amin				X
	Yurii S Aulchenko				X
	Ben Oostra	X			
	Cornelia M van Duijn	X		X	X
	M. Carola Zillikens				X
Family Heart Study	Ingrid B Borecki	X	X	X	
	Mary F Feitosa			X	X
	Shamika Ketkar			X	X

Cohort	Author	Overseeing (PI)	Geno-typing	Pheno-typing	Data analysis
	Michael A Province	X			
FENLAND	Ruth JF Loos	X		X	
	Jian'an Luan		X		X
	Nicholas J Wareham	X			
FRAM	Larry D Atwood	X	X		
	Adrienne L Cupples	X			X
	Nancy L Heard-Costa				X
	Julius Suh Ngwa				X
	Charles White				X
FTC	Jaakko Kaprio	X			
	Kirsi Pietiläinen			X	
	Samuli Ripatti				X
	Aila Rissanen	X			
	Ida Surakka				X
FUSION	Richard N Bergman	X			
	Michael Boehnke	X			
	Francis S Collins	X			
	Anne U Jackson				X
	Karen L Mohlke	X			
	Heather M Stringham			X	
	Jaakko Tuomilehto	X			
	Cristen J Willer				X
Genmets	Antti Jula			X	
	Seppo Koskinen			X	
	Leena Peltonen	X	X		
	Samuli Ripatti				X
	Veikko Salomaa	X		X	
	Ida Surakka				X
GerMIFSI and GerMIFSIII	Jeanette Erdmann	X			
	Christian Hengstenberg	X		X	
	Inke R König				X
	Michael Preuss				X
	Stefan Schreiber	X		X	
	Heribert Schunkert	X			
	H.-Erich Wichmann			X	
	Andreas Ziegler				X
KORA S3	Christian Gieger	X			X
	Iris M Heid				X
	Thomas Meitinger		X		
	Martina Müller				X
KORA S4	Eva Albrecht				X
	Thomas Illig		X		
	H.-Erich Wichmann	X		X	
	Thomas Winkler				X
MICROS (EUROSPAN)	Alessandro De Grandi		X	X	
	Andrew A Hicks		X		

Cohort	Author	Overseeing (PI)	Geno-typing	Pheno-typing	Data analysis
	Åsa Johansson				X
	Irene Pichler			X	
	Peter P Pramstaller	X			
MIGEN	Roberto Elosua				
	Aki S Havulinna			X	
	Sekar Kathiresan	X			
	Olle Melander	X			
	Christopher J O'Donnell	X			
	David S Siscovick	X			
	Elizabeth K Speliotes				X
	Benjamin F Voight	X			
NFBC1966	Lachlan Coin				X
	Paul Elliott		X	X	
	Nelson Freimer		X		
	Anna-Liisa Hartikainen		X	X	
	Marjo-Riitta Jarvelin	X	X	X	
	Markku Koiranen			X	
	Jaana Laitinen			X	
	Mark I McCarthy		X		
	Leena Peltonen		X		
	Anneli Pouta			X	
	Ulla Sovio				X
	Paavo Zitting		X	X	
NBS-WTCCC	Willem H Ouwehand	X			
	Jennifer G Sambrook			X	
NHS	Frank B Hu	X	X	X	
	David J Hunter	X	X	X	
	Peter Kraft		X		X
	Lu Qi		X	X	X
NSPHS (EUROSPAN)	Ulf Gyllensten	X			
	Wilmar Igl		X		
	Åsa Johansson		X	X	X
NTRNESDA	Dorret I Boomsma	X			
	Eco JC Geus		X		
	Jouke-Jan Hottenga		X		X
	Brenda W Penninx	X			
	Jan H Smit			X	
	Gonneke Willemsen			X	
ORCADES (EUROSPAN)	Harry Campbell		X		
	Åsa Johansson				X
	Veronique Vitart				X
	Sarah H Wild			X	
	James F Wilson	X			
	Alan F Wright	X		X	
PLCO	Sonja I Berndt	X		X	X
	Stephen J Chanock	X	X		

Cohort	Author	Overseeing (PI)	Geno-typing	Pheno-typing	Data analysis
	Richard B Hayes	X			
	Kevin B Jacobs		X		X
PROCARDIS	Martin Farrall				
	Anders Hamsten	X			
	Mark Lathrop	X	X		
	John F Peden			X	
	Hugh Watkins	X			
RS-I	Yurii S Aulchenko				X
	Karol Estrada		X		X
	Albert Hofman	X		X	
	Manfred Kayser	X			
	Marjolein J Peters		X		
	Fernando Rivadeneira	X	X	X	X
	André G Uitterlinden	X	X	X	X
	Cornelia M van Duijn	X		X	X
	Joyce B J van Meurs		X		
	M. Carola Zillikens				X
RUNMC	Katja K Aben	X			
	Martin den Heijer	X			
	Lambertus Kiemeney	X			
SardiNIA	Goncalo R Abecasis	X			X
	Andrea Maschio		X		
	Antonella Mulas		X		
	Serena Sanna				X
	David Schlessinger	X			
	Manuela Uda	X		X	
SASBAC	Per Hall	X		X	
	Erik Ingelsson	X			X
	Jianjun Liu			X	
SEARCH/UKOPS	Jonathan Patrick Tyrer				X
SHIP	Florian Ernst		X		X
	Wolfgang Hoffmann	X		X	
	Thomas Kocher	X			
	Astrid Petersmann		X		
	Carsten Oliver Schmidt			X	
	Henry Völzke	X			
T2D-WTCCC	Teresa Ferreira				X
	Timothy M Frayling	X		X	
	Andrew T Hattersley	X		X	
	Hana Lango Allen				X
	Cecilia M Lindgren	X	X		X
	Reedik Mägi				X
	Mark I McCarthy	X	X	X	
	Andrew P Morris				X
	John RB Perry				X
	Inga Prokopenko				X

Cohort	Author	Overseeing (PI)	Geno-typing	Pheno-typing	Data analysis
	Joshua C Randall				X
	Nigel W Rayner		X		X
	Neil R Robertson		X		X
	Michael N Weedon				X
	Andrew R Wood				X
TwinsUK	Massimo Mangino		X	X	X
	Nicole Soranzo	X	X		X
	Tim D Spector	X		X	
VIS (EUROSPAN) and KORCULA	Caroline Hayward		X		X
	Åsa Johansson				X
	Ivana Kolcic			X	
	Ana Marusic			X	
	Ozren Polasek			X	X
	Igor Rudan	X		X	
	Lina Zgaga			X	

Stage 2 – in silico replication cohorts

Cohort	Author	Overseeing (PI)	Geno-typing	Pheno-typing	Data analysis
BHS	John P Beilby	X		X	
	Matthew N Cooper				X
	Jennie Hui		X		
	Robert Lawrence				X
	Arthur W Musk	X		X	
	Lyle J Palmer	X			
Corogene	Marja-Liisa Lokki		X		
	Markku S Nieminen	X			
	Niina Pellikka				X
	Leena Peltonen	X	X		
	Markus Perola				X
	Juha Sinisalo			X	
EGCUT	Helene Alavere			X	
	Tõnu Esko				X
	Andres Metspalu	X			
	Mari Nelis		X		
	Mari-Liis Tammesoo				X
Family Heart Study	Ingrid B Borecki	X	X	X	
	Mary F Feitosa			X	X
	Shamika Ketkar			X	X
	Michael A Province	X			
FINGESTURE	Gabrielle Boucher				X
	Heikki V Huikuri	X		X	X
	Juhani Junttila			X	X
	John D Rioux	X			X
GOOD	Mattias Lorentzon		X	X	X

	Claes Ohlsson	X	X	X	X
	Liesbeth Vandenput			X	X
HBCS	Johan Eriksson	X		X	
	Eero Kajantie			X	
	Markus Perola		X		X
	Samuli Ripatti		X		X
	Elisabeth Widen		X		
HYPERGENES	Lorena Citterio			X	
	Daniele Cusi	X			
	Nicola Glorioso		X	X	
	Carlo Rivolta	X	X		
	Erika Salvi				X
	Laura Zagato			X	
MGS	Jubao Duan		X		
	Pablo V Gejman	X	X	X	
	Douglas F Levinson	X			X
	Alan R Sanders		X	X	
	Jianxin Shi				X
NHS	Frank B Hu	X	X	X	
	David J Hunter	X	X	X	
	Peter Kraft		X		X
	Lu Qi		X	X	X
RS-II and RS-III	Yurii S Aulchenko				X
	Karol Estrada		X		X
	Albert Hofman	X		X	
	Manfred Kayser	X			
	Marjolein J Peters		X		
	Fernando Rivadeneira	X	X	X	X
	André G Uitterlinden	X	X	X	X
	Cornelia M van Duijn	X		X	X
	Joyce B J van Meurs		X		
	M. Carola Zillikens				X
Sorbs	Peter Kovacs		X	X	
	Reedik Mägi				X
	Inga Prokopenko				X
	Michael Stumvoll	X			
	Anke Tönjes	X		X	
WGHS	Daniel I Chasman		X		X
	Guillaume Paré		X		
	Alex N Parker		X		
	Paul M Ridker	X			
YFS	Mika Kähönen	X		X	
	Johannes Kettunen		X		X
	Terho Lehtimäki	X		X	
	Niina Pellikka		X		X
	Olli Raitakari	X		X	
	Jorma Viikari	X		X	

Additional analyses cohorts

Cohort	Author	Overseeing (PI)	Geno-typing	Pheno-typing	Data analysis
GCI height extremes (additional genotyping)	Kristin G Ardlie	X			
	Joel N Hirschhorn	X			
	Guillaume Lettre			X	X
	Rany M Salem				X
	Michael C Turchin		X		X
QIMR (polygene analysis)	Andrew C Heath	X		X	
	Nick G Martin	X	X	X	
	Grant W Montgomery	X	X		
	Dale R Nyholt	X	X		X
	Peter M Visscher		X		X

Acknowledgments

ADVANCE (Atherosclerotic Disease, VAscular FunctioN, and GenetiC Epidemiology) - The ADVANCE study was supported by the Donald W. Reynolds Foundation, the NIH (R01-HL087647), and the NHLBI (R01-HL087647).

AGES (Age, Gene/Environment Susceptibility-Reykjavik Study) - The AGES Reykjavik Study has been funded by NIH contracts N01-AG-12100 and Z01-AG-007380, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). The study is approved by the Icelandic National Bioethics Committee, (VSN: 00-063) and the Data Protection Authority. The researchers are indebted to the participants for their willingness to participate in the study.

Amish HAPI Heart Study (Amish Heredity and Phenotype Intervention Heart Study) - This work was supported by NIH research grants U01 HL72515, U01 GM074518, R01 HL088119, and U01 HL084756. Partial funding was also provided by the Mid-Atlantic Nutrition and Obesity Research Center of Maryland (P30 DK072488).

ARIC (Atherosclerosis Risk in Communities Study) - The ARIC Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, N01-HC-55022, R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. The project described was supported by Grant Number UL1 RR 025005 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH) and NIH Roadmap for Medical Research, and its contents are solely the responsibility of the authors and do not necessarily represent the official view of NCRR or NIH. The authors thank the staff and participants of the ARIC Study for their important contributions.

B58C-T1DGC (British 1958 birth cohort, Type 1 Diabetes Genetic Consortium controls) - This research utilizes resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD), and Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062418. T1DGC GWAS data were deposited by the Diabetes & Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge (John Todd, Helen Stevens and Neil Walker) which is funded by the Juvenile Diabetes Research Foundation International, The National Institute for Health Research Cambridge Biomedical Research Centre and the Wellcome Trust (Strategic Award 079895). The British 1958 Birth Cohort collection is funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02.

B58C-WTCCC (British 1958 birth cohort, Wellcome Trust Case Control Consortium (WTCCC) controls) - We acknowledge use of genotypes and phenotypes from the British 1958 birth cohort collection, funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02.

BHS (Busselton Health Study) - The 1994-1995 Busselton Health Survey was funded by Healthway, Western Australia. The Busselton Health Studies are supported by the National Health and Medical Research Council of Australia and the Great Wine Estates Auctions. The BHS acknowledges the support of the Western Australian Genetic Epidemiology Resource and the Western Australian DNA Bank (both National Health and Medical Research Council of Australia Enabling Facilities).

BRIGHT (BTItish Genetic of HyperTension study) - The BRIGHT study is supported by the Medical Research Council of Great Britain (G9521010D) and the British Heart Foundation

(PG/02/128). Genotyping was funded by the Wellcome Trust (grant number; 076113/B/04/Z) as part of The Wellcome Trust Case Control Consortium. This work forms part of the research themes contributing to the translational research portfolio of Barts and the London Cardiovascular Biomedical Research Unit which is supported and funded by the National Institute of Health Research. Toby Johnson was supported through a VIP award from the Wellcome Trust to Queen Mary University of London in the 2009/2010 academic year. The BRIGHT study is extremely grateful to all the patients who participated in the study and the BRIGHT nursing team.

CAD-WTCCC (WTCCC Coronary Artery Disease cases) - Collection of the CAD-WTCCC cases (BHF Family Heart Study) was funded by the British Heart Foundation and the Medical Research Council and genotyping by the Wellcome Trust as part of the WTCCC. We thanks members of the BHF Family Heart Study Research Group for recruitment. NJS holds a personal chair supported by the BHF and also holds a UK NIHR Senior Investigator Award.

CAPS (Cancer Prostate in Sweden) - The CAPS study was supported by grants from the Swedish Research Council, the Swedish Cancer Society, and the National Cancer Institute. E.I. was supported by grants from the Swedish Research Council, the Swedish Heart-Lung Foundation, the Swedish Society of Medicine, the Swedish Foundation for Strategic Research, and the Royal Swedish Academy of Science while working with this article.

CHS (Cardiovascular Health Study) - The CHS research reported in this article was supported by contract numbers N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133, grant numbers U01 HL080295 and R01 HL087652 from the National Heart, Lung, and Blood Institute, R01AG031890, with additional contribution from the National Institute of Neurological Disorders and Stroke. A full list of principal CHS investigators and institutions can be found at http://www.chs-nhlbi.org/pi.htm. DNA handling and genotyping was supported in part by National Center for Research Resources grant M01RR00425 to the Cedars-Sinai General Clinical Research Center Genotyping core, National Institute of Diabetes and Digestive and Kidney Diseases grant DK063491 to the Southern California Diabetes Endocrinology Research Center, and the Cedars-Sinai Board of Governors' Chair in Medical Genetics (JIR).

CoLaus (Cohorte Lausannoise) - The CoLaus study received financial contributions from GlaxoSmithKline, the Faculty of Biology and Medicine of Lausanne, and the Swiss National Science Foundation (33CSCO-122661). The authors thank Peter Vollenweider, Vincent Mooser and Dawn Waterworth, Co-PIs of the CoLaus study. Special thanks to Murielle Bochud, Yolande Barreau, Mathieu Firmann, Vladimir Mayor, Anne-Lise Bastian, Binasa Ramic, Martine Moranville, Martine Baumer, Marcy Sagette, Jeanne Ecoffey and Sylvie Mermoud for data collection. JSB is supported by the Centre Hospitalier Universitaire Vaudois and the University of Lausanne, Switzerland, the Swiss National Science Foundation (grants nb 310000-112552) and the European Union HEALTH-F4-2007-201550 HYPERGENES grant. SB is supported by the Giorgi-Cavaglieri Foundation, the Swiss National Science Foundation (grant 3100AO-116323/1), the Leenaards Foundation, the European Union HEALTH-F4-2007-201550 HYPERGENES grant and the Swiss Institute of Bioinformatics.

COROGENE (study of Genetic Predisposition of Coronary Heart Disease in Patients Verified with Coronary Angiogram) - Finnish Foundation for Cardiovascular Research Finnish Academy SALVE-program PUBGENSENS grant no 10404."

deCODE (deCODE genetics sample set) - deCODE authors would like to thank participants in deCODE cardiovascular- and obesity studies and collaborators for their cooperation. We would also like to acknowledge the staff at the Clinical Research Centre (Iceland) and the deCODE Genetics biological materials and genotyping facilities for their work. The research performed at deCODE Genetics was part funded through the European Community's Seventh Framework Programme (FP7/2007-2013), ENGAGE project, grant agreement HEALTH-F4-2007- 201413.

DGI (Diabetes Genetics Initiative) - The Botnia (DGI) study have been supported by grants from Folkhälsan Research Foundation, Sigrid Juselius Foundation, Ministry of Education, Nordic Center of Excellence in Disease Genetics, Gyllenberg Foundation, Swedish Cultural Foundation in Finland, Finnish Diabetes Research Foundation, Foundation for Life and Health in Finland, Finnish Medical Society, Paavo Nurmi Foundation, Perklén Foundation, Ollqvist Foundation, Närpes Health

Care Foundation, the Municipal Health Care Center and Hospital in Jakobstad, Health Care Centers in Vasa, Närpes and Korsholm. This work was also partially supported by NIH grant R01-DK075787 to JNH and by K23-DK080145 and F32-DK079466 to EKS.

EGCUT (Estonian Genome Center, University of Tartu) - EGCUT received support from FP7 grants ((201413 ENGAGE, 212111 BBMRI, 205419 ECOGENE, 245536 OPENGENE). EGCUT also received targeted financing from Estonian Government SF0180142s08 and from the European Union through the European Regional Development Fund, in the frame of Centre of Excellence in Genomics. EGC authors want to acknowledge EGCUT personnel, especially Ms. Merli Hass. The genotyping of the EGCUT samples were performed in Estonian Biocentre Genotyping Core Facility, EGC authors want to acknowledge Mr. Viljo Soo for their contribution in that. EGCUT data analyzes were carried out in part in the High Performance Computing Center of University of Tartu.

EPIC-Obesity (obesity cases and controls from the European Prospective Investigation into Cancer and Nutrition - Norfolk) - The EPIC Norfolk Study is funded by program grants from the Medical Research Council UK and Cancer Research UK. Genotyping of samples for the obesity cases and controls was carried out at the WT Sanger centre and was supported by the Wellcome Trust funding to Inês Barroso (grant 077016/Z/05/Z).

ERF (Erasmus Rucphen Family) (EUROSPAN) - The study was supported by grants from The Netherlands Organisation for Scientific Research (NWO), Erasmus MC, the Centre for Medical Systems Biology (CMSB) and EUROSPAN (European Special Populations Research Network) FP6 STRP grant number 018947 (LSHG-CT-2006-01947). We are grateful to all general practitioners for their contributions, to Petra Veraart for her help in genealogy, Jeannette Vergeer for the supervision of the laboratory work and Peter Snijders for his help in data collection.

Fenland (Fenland Study) - The Fenland Study is funded by the Wellcome Trust and the Medical Research Council, as well as by the Support for Science Funding programme and CamStrad. We are grateful to all the volunteers for their time and help, and to the General Practitioners and practice staff for help with recruitment. We thank the Fenland Study co-ordination team and the Field Epidemiology team of the MRC Epidemiology Unit for recruitment and clinical testing.

FHS (Family Heart Study) - The FHS is funded by a NHLBI grant R01-HL087700, and NIDDK grants R01-DK068336 and R01-DK075681.

FinGesture (Finnish Genetic Study of Arrhythmic Events) - We thank the study participants. We also thank Juhani Junttila, Kari Kaikkonen, and Marja-Leena Kortelainen for study concept and design, and data acquisition and interpretation. The FinGesture cohort is supported by the Juselius Foundation (Helsinki, Finland) and the Council of Health of the Academy of Finland (Helsinki, Finland). Authors would like to thank John D. Rioux, Sylvain Foisy, and Gabrielle Boucher for their contributions to the design, implementation and analysis of the GWA study of the FinGesture cohort. In addition, we would like to acknowledge the support of the Montreal Heart Institute Foundation.

FRAM (Framingham Heart Study) - This research was conducted in part using data and resources from the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine. The analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. This work was partially supported by the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract No. N01-HC-25195) and its contract with Affymetrix, Inc for genotyping services (Contract No. N02-HL-6-4278). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center.

FTC (Finnish Twin Cohort) - Medical Research Fund of the Helsinki University Hospital, this project is supported by the European Commission under the programme 'Quality of Life and Management of the Living Resources' of 5th Framework Programme (no. QLG2-CT-2002-01254). This research was supported through funds from The European Community's Seventh Framework Programme (FP7/2007-2013), ENGAGE Consortium, grant agreement HEALTH-F4-2007-

201413, Nordic Center of Excellence in Disease Genetics, Center of Excellence in Complex Disease Genetics of the Academy of Finland (grants 213506 and 129680). K.H. Pietiläinen was supported by grants from Yrjö Jahnsson, Jalmari and Rauha Ahokas, Biomedicum Helsinki and Novo Nordisk Foundations.

FUSION (Finland-United States Investigation of NIDDM Genetics) - Support for FUSION was provided by NIH grants DK062370 (to M.B.), DK072193 (to K.L.M.), T32 HG00040 (to T.M.T.), and K99-HL094535 (to C.J.W.), and intramural project number Z01-HG000024 (to F.S.C.), and an ADA Mentor-Based Postdoctoral Fellowship grant to M.B. K.L.M. is a Pew Scholar for the Biomedical Sciences. Genome-wide genotyping was conducted by the Johns Hopkins University Genetic Resources Core Facility SNP Center at the Center for Inherited Disease Research (CIDR), with support from CIDR NIH contract no. N01-HG65403.

GCI height extremes (Genomics Collaborative, Incorporated) - Genotyping and analysis were supported by March of Dimes grant 6-FY-09-507

Genmets (Health 2000 / GENMETS substudy) - This research was supported through funds from The European Community's Seventh Framework Programme (FP7/2007-2013), ENGAGE Consortium, grant agreement HEALTH-F4-2007- 201413, Nordic Center of Excellence in Disease Genetics, Center of Excellence in Complex Disease Genetics of the Academy of Finland (grants 213506, 129494 and 129680).

GerMIFS I and GerMIFS II (German Myocardial Infarction Family Study I & II) - BMBF funded project Atherogenomics (FKZ: 01GS0831) and Cardiogenics integrated project (LSH-2006-037593) of the European Union.

GOOD (Gothenburg Osteoporosis and Obesity Determinants Study) - Financial support was received from the Swedish Research Council, the Swedish Foundation for Strategic Research, the ALF/LUA research grant in Gothenburg, the Lundberg Foundation, the Emil and Vera Cornell Foundation, the Torsten and Ragnar Söderberg's Foundation, Petrus and Augusta Hedlunds Foundation, the Västra Götaland Foundation, the Göteborg Medical Society, and the Sahlgrenska Center for Cardiovascular and Metabolic Research (CMR, no. A305:188), which is supported by the Swedish Foundation for Strategic Research.

HBCS (Helsinki Birth Cohort Study) - Helsinki Birth Cohort Study has been supported by grants from the Academy of Finland, the Finnish Diabetes Research Society, Folkhälsan Research Foundation, Novo Nordisk Foundation, Finska Läkaresällskapet, Signe and Ane Gyllenberg Foundation.

HYPERGENES - HYPERGENES (European Network for Genetic-Epidemiological Studies: building a method to dissect complex genetic traits, using essential hypertension as a disease model) is a Large Cooperative Project funded by EU within the FP7 (HEALTH-F4-2007-201550). To HYPERGENES consortium took part: (1) University of Milano and Fondazione Filarete with Daniele Cusi, Project Coordinator, Fabio Macciardi co-PI, Cristina Barlassina, Erika Salvi, Cristian Cosentino, Sara Lupoli, Federica Torri, Maurizio Marconi, Gianna Petrini, Vincenzo Toschi, Simona Sitia, Giancarlo Mariotti, Maurizio Turiel; (2) University of Leuven, Division of Hypertension and Cardiovascular Rehabilitation, Department of Cardiovascular Diseases, with Robert Fagard, Yu Jin, Tatiana Kuznetsova, Tom Richart, Jan A. Staessen, and Lutgarde Thijs; (3) Jagiellonian University Medical College, Krakow, with Kalina Kawecka-Jaszcz, Katarzyna Stolarz-Skrzypek, Agnieszka Olszanecka, Wiktoria Wojciechowska, Małgorzata Kloch-Badełek; (4) IBM Israel – Science and Technology LTD, with Amnon Shabo, Ariel Frakash, Simona Cohen, Boaz Carmeli, Dan Pelleg, Michal Rosen-Zvi, Hani Neuvrith-Telem; (5) I.M.S. – Istituto di Management Sanitario S.r.l., Milan, with Pietro Conti, Costanza Conti, Mariella D'Alessio; (6) Institute of Internal Medicine, Siberian Branch of Russian Academy of Medical Science, Novosibirsk, with Yuri Nikitin, Galina Simonova, Sofia Malyutina, Elena Pello; (7) Imperial College of Science, Technology and Medicine, with Paolo Vineis and Clive J Hoggart; (8) INSERM – Institut National de la Santé et de la Recherche Médicale U772, with Xavier Jeunemaitre, Pierre-François Plouin, Michel Azizi (9) University of Warwick. Cardiovascular Medicine & Epidemiology Group, Clinical Sciences Research Institute, with Francesco P Cappuccio, Michelle A Miller, Chen Ji; (10) Università degli Studi di Sassari-AOU. Hypertension and Cardiovascular Prevention Centre, with Nicola Glorioso, Giuseppe Argiolas,

Francesca Fau, Silvia Pitzoi, Emanuela Bulla, Roberta Zaninello, Patrizia Bulla, Simone Fadda, Gianclaudia Cappai, Siria Motroni, Chiara Maria Troffa; (11) STMICROELECTRONICS SRL, with Tony Barbuzzi; (12) University of Lausanne. Department of Medical Genetics, with Carlo Rivolta, Jacques S. Beckmann, Zoltan Kutalik, Paola Benaglio, Sven Bergmann, Murielle Bochud, Diana Marek, and Peter Bastian; (13) Pharnext S.A.S., Paris, with Daniel Cohen and Ilya Chumakov; (14) Softeco Sismat Spa, Genova, with Enrico Morten; (15) Shanghai Institute of Hypertension, with Jiguang Wang and Li Yan; (16) Charles University in Prague. Department of Internal Medicine II, Pilsen, with Jan Filipovsky, Jitka Seidlerova, Otto Mayer Jr., Milena Dolejsova, Jana Hirmerova, Jana Strizova; (17) University of Padova, Department of Clinical and Experimental Medicine, with Edoardo Casiglia and Valérie Tikhonoff; (18) Medical University of Gdansk. Hypertension Unit, Department of Hypertension and Diabetology, with Krzysztof Narkiewicz, Marzena Chrostowska, Wojciech Sakiewicz, Michal Wojtowicz, Michal Hoffmann; (19) University Vita-Salute San Raffaele, with Paolo Manunta, Chiara Lanzani, Maria Teresa Sciarrone, Lorena Citterio, Laura Zagato, Giuseppe Bianchi. Regarding the present work, cases and controls were recruited within specific cohorts/networks: FLEMENGHO/EPOGH cohort (Coordinator J Staessen, contributors Units 2,3,6,16,17,18); Wandsworth Heart & Stroke Study (WHSS, Coordinator F Cappuccio, contributing Unit 9); Milano-Sassari cohort (coordinator D Cusi, contributors Units 1,10,19); SOPHIA cohort (coordinator N Glorioso, contributing Unit 10); (20) National Centre of Competence in Research "Frontiers in Genetics", University of Geneva, with Patrick Descombes.

KORA studies (Cooperative Health Research in the Region of Augsburg, KOoperative Gesundheitsforschung in der Region Augsburg) - We thank all study participants and would like to acknowledge the KORA (Cooperative Health Research in the Region of Augsburg) group. The KORA research platform was initiated and financed by the Helmholtz Center Munich, German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Part of this work was financed by the NIH subcontract from the Children's Hospital, Boston, US, (H.E.W., I.M.H, prime grant 1 R01 DK075787-01A1 to J.N.H.), the German National Genome Research Net NGFN2 and NGFNplus (01GS0823 to H.E.W), by MC Health - the Munich Center of Health Sciences as part of LMUinnovativ (H.E.W.), and by the Austrian GEN-AU-Programme "GOLD" to the Division of Genetic Epidemiology, Innsbruck Medical University. The LINUX platform for meta-analyses computations was funded by the Department of Epidemiology and Preventive Medicine at Regensburg University Medical Center.

MGS (Molecular Genetics of Schizophrenia/NIMH Repository Control Sample) - The Molecular Genetics of Schizophrenia project was carried out by 10 research sites and PIs: Pablo V. Gejman, Study Coordinator (NorthShore University Healthcare Research Institute, Evanston, IL), Douglas F. Levinson (Stanford University), Bryan J. Mowry (University of Queensland), Donald Black (University of Iowa), Robert Freedman (University of Colorado), C. Robert Cloninger (Washington University), Jeremy Silverman (Mt. Sinai Medical School), Nancy Buccola (Louisiana State University - New Orleans), William Byerley (University of California at San Francisco), and Farooq Amin (Emory University). This study was supported by NIH R01 grants (MH67257 to N.G.B., MH59588 to B.J.M., MH59571 to P.V.G., MH59565 to R.F., MH59587 to F.A., MH60870 to W.F.B., MH59566 to D.W.B., MH59586 to J.M.S., MH61675 to D.F.L., MH60879 to C.R.C., MH81800 to P.V.G., MH79469 to P.V.G., and MH79470 to D.F.L.), NARSAD (National Alliance for Research on Schizophrenia and Depression) Young Investigator Awards (to J.D. and A.R.S.), the Genetic Association Information Network (GAIN), the Walter E. Nichols, M.D., and Eleanor Nichols endowments, at Stanford University, and by The Paul Michael Donovan Charitable Foundation. Genotyping was carried out by the Genotyping and Analysis at the Broad Institute of Harvard and MIT (S. Gabriel and D.B.M.), which is supported by grant U54 RR020278 from the National Center for Research Resources. Genotyping of half of the control sample presented here was carried out with support from GAIN. The GAIN quality control team (G.R. Abecasis and J. Paschall) made important contributions to the project.

MICROS (EUROSPAN) - For the MICROS study, we thank the primary care practitioners Raffaela Stocker, Stefan Waldner, Toni Pizzecco, Josef Plangger, Ugo Marcadent and the personnel of the Hospital of Silandro (Department of Laboratory Medicine) for their participation and collaboration in the research project. In South Tyrol, the study was supported by the Ministry of Health and

Department of Educational Assistance, University and Research of the Autonomous Province of Bolzano and the South Tyrolean Sparkasse Foundation.

MIGEN (Myocardial Infarction Genetics Consortium) - National Heart, Lung, and Blood Institute's STAMPEED genomics research program (R01 HL087676) and the National Center for Research Resources (U54 RR020278). This work was also partially supported by NIH grants K23-DK080145 and F32-DK079466 to EKS

NBS-WTCCC (WTCCC National Blood Service donors) - We acknowledge use of DNA from The UK Blood Services/Wellcome Trust collection of Common Controls (UKBS/WT collection), funded by the Wellcome Trust (grants 076113/C/04/Z and 084183/Z/07/Z) and by the National Institute for Health research. The collection was established as part of the Wellcome Trust Case Control Consortium.

NFBC1966 (Northern Finland 1966 Birth Cohort) - We thank Professor Paula Rantakallio (launch of NFBC1966 and initial data collection), Ms Sarianna Vaara (data collection), Ms Tuula Ylitalo (administration), Mr Markku Koiranen (data management), Ms Outi Tornwall and Ms Minttu Jussila (DNA biobanking). Financial support was received from the Academy of Finland (project grants 104781, 120315, 129269, and Center of Excellence in Complex Disease Genetics), University Hospital Oulu, Biocenter, University of Oulu, Finland (75617), the European Commission (EURO-BLCS, Framework 5 award QLG1-CT-2000-01643), NHLBI grant 5R01HL087679-02 through the STAMPEED program (1RL1MH083268-01), NIH/NIMH (5R01MH63706:02), ENGAGE project and grant agreement HEALTH-F4-2007-201413, and the Medical Research Council, UK (Grants G0500539, G0600331, PrevMetSyn). The DNA extractions, sample quality controls, biobank up-keeping and aliquotting were performed in the National Public Health Institute, Biomedicum Helsinki, Finland and supported financially by the Academy of Finland and Biocentrum Helsinki. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

NHS (The Nurses' Health Study) - The NHS is funded by the NIH (grants U01HG004399, DK58845, CA65725, CA87969, CA49449, CA67262, CA50385, U01CA098233, R01HL71981), and the Boston Obesity Nutrition Research Center (DK46200). We acknowledge the NHS participants for their contributions in making this study possible.

NSPHS (Northern Sweden Population Health Study) (EUROSPAN) - The Northern Swedish Population Health Study (NSPHS) was funded by the Swedish Medical Research Council (Project Number K2007-66X-20270-01-3) and the Foundation for Strategic Research (SSF). NSPHS as part of EUROSPAN (European Special Populations Research Network) was also supported by European Commission FP6 STRP grant number 01947 (LSHG-CT-2006-01947). We are grateful for the contribution of district nurse Svea Hennix for data collection and Inger Jonasson for logistics and coordination of the health survey. Finally, we would like to thank all the participants from the community for their interest and willingness to contribute to this study.

NTRNESDA (Netherlands Twin Register & the Netherlands Study of Depression and Anxiety) - Funding support was provided by the Netherlands Scientific Organization (904-61-090, 904-61-193, 480-04-004, 400-05-717, Center for Medical Systems Biology (NWO Genomics), SPI 56-464-1419); Neuroscience Campus Amsterdam and the EMGO+ institute; the European Union (EU/WLRT-2001-01254), ZonMW (Geestkracht program, 10-000-1002), NIMH (RO1 MH059160) and matching funds from participating institutes in NESDA and NTR. The genotyping of samples was provided through the Genetic Association Information Network (GAIN). The dataset(s) used for the analyses described in this manuscript were obtained from the database of Genotype and Phenotype (dbGaP) found at http://www.ncbi.nlm.nih.gov/gap through dbGaP accession number phs000020.v2.p1.

ORCADES (Orkney Complex Disease Study) (EUROSPAN) - The Orkney Complex Disease Study (ORCADES) was supported by the Chief Scientist Office of the Scottish Government (grant no. CZB/4/276), the Royal Society and the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947). DNA extractions were performed at the Wellcome Trust Clinical Research Facility in Edinburgh. We would like to acknowledge the invaluable

contributions of Lorraine Anderson and the research nurses in Orkney, the administrative team in Edinburgh and the people of Orkney.

PLCO (Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial) - This research was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics and by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS. The authors thank Drs. Christine Berg and Philip Prorok of the Division of Cancer Prevention, National Cancer Institute, the screening center investigators and staff of the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial, Mr. Tom Riley and staff at Information Management Services, Inc., Ms. Barbara O'Brien and staff at Westat, Inc., and the study participants for making this study possible.

Procardis (Precocious Coronary Artery Disease) The Procardis consortium was funded by EC Sixth Framework Programme (Procardis:LSHM-CT- 2007- 037273) and AstraZeneca AB. RC, MF and HW are supported by the British Heart Foundation Centre for Research Excellence; MF and HW acknowledge support from the Wellcome Trust; RC acknowledge support from the MRC; AH obtained support for this project from the Swedish Heart-Lung Foundation, the Swedish Medical Research Council (8691), the Knut and Alice Wallenberg Foundation, the Karolinska Institute and the Stockholm County Council (560183).

QIMR Polygene (Twins studies at the Queensland Instutite of Medical Research) We are grateful to the twins and their families for their generous participation in these studies. We would like to thank staff at the Queensland Institute of Medical Research: Dixie Statham, Ann Eldridge and Marlene Grace for sample collection, Anjali Henders, Megan Campbell, Lisa Bowdler, Steven Crooks and staff of the Molecular Epidemiology Laboratory for sample processing and preparation, Scott Gordon, Brian McEvoy, Belinda Cornes and Beben Benyamin for data QC and preparation, and David Smyth and Harry Beeby for IT support. We acknowledge funding from the Australian National Health and Medical Research Council (NHMRC grants 241944, 389875, 389891, 389892, 389938, 442915, 442981, 496739, 496688, 552485 and 613672), the U.S. National Institute of Health (grants AA07535, AA10248, AA014041, AA13320, AA13321, AA13326 and DA12854) and the Australian Research Council (ARC grant DP0770096).

RS (The Rotterdam Study) (GWA and replication studies) - The GWA study was funded by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Consortium for Healthy Aging (NCHA) project nr. 050-060-810. We thank Pascal Arp, Mila Jhamai, Dr Michael Moorhouse, Marijn Verkerk, and Sander Bervoets for their help in creating the GWAS database. The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are very grateful to the participants and staff from the Rotterdam Study, the participating general practioners and the pharmacists. We would like to thank Dr. Tobias A. Knoch, Luc V. de Zeeuw, Anis Abuseiris, and Rob de Graaf as well as their institutions the Erasmus Computing Grid, Rotterdam, The Netherlands, and especially the national German MediGRID and Services@MediGRID part of the German D-Grid, both funded by the German Bundesministerium fuer Forschung und Technology under grants #01 AK 803 A-H and # 01 IG 07015 G for access to their grid resources.

RUNMC (The Nijmegen Bladder Cancer Study & The Nijmegen Biomedical Study) - Collection of samples and data in the RUNMC study was funded in part by the European Commission (POLYGENE: LSHC-CT-2005) and Radboud University Nijmegen Medical Centre.

SardiNIA - We thank all of the volunteers who participated in the study, Monsignore Piseddu, Bishop of Ogliastra, the mayors and citizens of the Sardinian towns (Lanusei, Ilbono, Arzana, and Elini), the head of the Public Health Unit ASL4 for their volunteerism and cooperation, and team of physicians, nurses and the recruitment personnel. Funding was provided by the National Institute on Aging, NIH (contracts NO1-AG12109 to the SardiNIA/ProgeNIA team; contract AG00675-03 to

D.S, and contracts HG002651; HG005214, HG005581;HL084729; MH084698 and 263-MA-410953 to the University of Michigan (G.R.A)). G.R.A. is a Pew Scholar for the Biomedical Sciences.

SASBAC (Swedish And Singapore Breast Association Consortium) - The SASBAC study was supported by funding from the Agency for Science, Technology and Research of Singapore (A*STAR), the United States National Institute of Health (NIH) and the Susan G. Komen Breast Cancer Foundation.

SEARCH/UKOPS (Studies of Epidemiology and Risk factors in Cancer Heredity / UK Ovarian Cancer Population Study) - SEARCH is funded by a programme grant from Cancer Research UK. We thank Jonathan Morrison, Honglin Song, Paul Pharoah and the SEARCH team. UKOPS is funded by a grant from the Eve Appeal. We thank Simon Gayther, Susan Ramus and the UKOPS team.

SHIP (Study of Health in Pomerania) - SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania. Genome-wide data have been supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg-West Pomerania. The University of Greifswald is a member of the 'Center of Knowledge Interchange' program of the Siemens AG. Data analyses were further supported by the DIAB Core project of the German Network of Diabetes.

Sorbs (Sorbs are self-contained population from Eastern Germany, European Descent) - Dr Knut Krohn, Microarray Core Facility of the Interdisciplinary Centre for Clinical Research (IZKF), University of Leipzig, Germany; grants from the German Research Council (KFO-152 to Michael Stumvoll) and from IZKF (B27 to Michael Stumvoll, Peter Kovacs and Anke Tönjes). The research of Dr. Inga Prokopenko is funded in part through the European Community's Seventh Framework Programme (FP7/2007-2013), ENGAGE project, grant agreement HEALTH-F4-2007- 201413. Dr. Reedik Mägi from the Wellcome Trust Centre for Human Genetics, University of Oxford, UK for data handling for this study.

T2D-WTCCC (WTCCC Type 2 Diabetes cases) - Research funding for sample collection, genotyping and data analysis for the T2D-WTCCC and other cohorts for which the Oxford group had responsibility came from the British Diabetes Association, BDA Research, Diabetes UK, Oxford NIHR Biomedical Research Centre, European Commission (ENGAGE: HEALTH-F4-2007-201413; EURODIA: LSHG-CT-2004-518153, Wellcome Trust (072960, 076113, 083270, 085301, 079557, 081682, 075491) UK Medical Research Council (G0000649, G0601261) and NIDDK (R01-DK-073490). In addition, Cecilia Lindgren is funded by WT086596/Z/08/Z (Wellcome Trust Research Career Development Fellowship); Reedik Mägi is funded by European Commission under the Marie Curie Intra-European Fellowship; Andrew Morris is funded by WT081682/Z/06/Z (Wellcome Trust Senior Research Fellowship); Mark McCarthy receives personal funding from the Oxford NIHR Biomedical Research Centre.

TwinsUK - The study was funded by the Wellcome Trust (Grant ref. 079771); European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F2-2008-ENGAGE and the European Union FP-5 GenomEUtwin Project (QLG2-CT-2002-01254) and Framework 6 Project EUroClot. The study also receives support from the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London. We thank the staff from the TwinsUK, the DNA Collections and Genotyping Facilities at the Wellcome Trust Sanger Institute for sample preparation; Quality Control of the Twins UK cohort for genotyping (in particular Amy Chaney, Radhi Ravindrarajah, Douglas Simpkin, Cliff Hinds, and Thomas Dibling); Paul Martin and Simon Potter of the DNA and Genotyping Informatics teams for data handling; Le Centre National de Génotypage, France, led by Mark Lathrop, for genotyping; Duke University, North Carolina, USA, led by David Goldstein, for genotyping; and the Finnish Institute of Molecular Medicine, Finnish Genome Center, University of Helsinki, led by Aarno Palotie. Nicole Soranzo acknowledges financial support from the Wellcome Trust (Grant 091746/Z/10/Z).

VIS (EUROSPAN) and KORCULA - The studies in the Croatian islands were supported through the grants from the Medical Research Council UK to H.C., A.F.W. and I.R.; and Ministry of Science, Education and Sport of the Republic of Croatia to I.R. (number 216-1080315-0302). The authors collectively thank a large number of individuals for their individual help in organizing, planning and carrying out the field work related to the project and data management: Professor Pavao Rudan and the staff of the Institute for Anthropological Research in Zagreb, Croatia (organization of the field work, anthropometric and physiological measurements, and DNA extraction in Vis); Professor Ariana Vorko-Jovic and the staff and medical students of the Andrija Stampar School of Public Health of the Faculty of Medicine, University of Zagreb, Croatia (questionnaires, genealogical reconstruction and data entry); Dr Branka Salzer from the biochemistry lab “Salzer”, Croatia (measurements of biochemical traits); local general practitioners and nurses (recruitment and communication with the study population); and the employees of several other Croatian institutions who participated in the field work, including but not limited to the University of Rijeka, Croatia; Croatian Institute of Public Health; Institutes of Public Health in Split and Dubrovnik, Croatia. SNP Genotyping of the Vis samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, WGH, Edinburgh, Scotland.

WGHS (Women's Genome Health Study) - The WGHS is funded by the Donald W. Reynolds Foundation (Las Vegas, NV), the Fondation LeDucq (Paris, France), the National Heart, Lung and Blood Institute (NHLBI; HL043851 and HL69757) and the National Cancer Institute (NCI; CA047988). Funding for genotyping and collaborative scientific support was provided by Amgen.

YFS (The Cardiovascular Risk in Young Finns Study) - Academy of Finland (grant no. 117797, 121584 and 126925), the Social Insurance Institution of Finland, University Hospital Medical funds to Tampere, and Turku University Hospitals, the Finnish Foundation of Cardiovascular Research.

CHAPTER 4: ASSESSING THE COMBINED IMPACT OF 18 COMMON GENETIC VARIANTS OF MODEST EFFECT SIZES ON TYPE 2 DIABETES RISK

Hana Lango, The UK Type 2 Diabetes Genetics Consortium, Colin NA Palmer, Andrew D Morris, Eleftheria Zeggini, Andrew T Hattersley, Mark I McCarthy, Timothy M Frayling, Michael N Weedon

Published: Diabetes (2008) 57: 3129

Abstract

Objectives Genome-wide association studies have dramatically increased the number of common genetic variants that are robustly associated with type 2 diabetes (T2D). A possible clinical use of this information is to identify individuals at high risk of developing the disease, so that preventative measures may be more effectively targeted. Here we assess the ability of 18 confirmed T2D variants to differentiate between T2D cases and controls.

Research design and methods We assessed index SNPs for the 18 independent loci in 2598 controls and 2309 cases from the GoDARTS study. The discriminatory ability of the combined SNP information was assessed by grouping individuals based on number of risk alleles carried and determining relative odds of T2D, and by calculating the area-under the receiver-operator characteristic curve (AUC).

Results Individuals carrying more risk alleles had higher risk of T2D. For example, 1.2% of individuals with 25-28 risk alleles had an odds ratio of 4.2 (95% CI: 2.11, 8.56) against the 1.8% with 10-12 risk alleles. The AUC (a measure of discriminative accuracy) for these variants was 0.60. The AUC for age, BMI and sex was 0.78, and adding the genetic risk variants only marginally increased this to 0.80.

Conclusions Currently, common risk variants for T2D do not provide strong predictive value at a population level. However, the joint effect of risk variants identified sub-groups of the population at substantially different risk of disease. Further studies are needed to assess whether individuals with extreme numbers of risk alleles may benefit from genetic testing.

Introduction

Recent genome-wide association (GWA) studies, which assay >300,000 single nucleotide polymorphisms (SNPs) across many thousands of individuals, have led to the discoveries of variants predisposing to many common complex diseases, including: type 2 diabetes [1-6], coronary artery disease [7-9], prostate cancer [10, 11], Crohn's disease [12-14] and many others (see http://www.genome.gov/26525384 for an up to date list of all GWA studies). The variants identified by these GWA studies are common in the general population (minor allele frequency [MAF] > 1%), but most have, individually, only small effects on disease risk, with odds ratios (ORs) typically <1.3.

Despite the relatively small predisposing effects conferred, these variants provide important, novel, insights into disease biology. For example, variants of a number of genes, such as *HHEX*, *CDKN2A/B* and *CDKAL1* implicate defects in pancreatic beta-cell development and function as important in type 2 diabetes etiology [4, 15, 16], whereas the discovery that variants in *FTO* are associated with BMI, opened up novel areas of investigation for obesity biology [17-19]. By gaining further knowledge of the underlying biology, and promoting potential therapeutic and preventative approaches, these insights are likely to be the most important outcome from these GWA studies.

A more immediate clinical utility may be to use the identified risk variants to aid the determination of an individual's risk of developing a particular disease. Several companies such as deCODE genetics and 23andme have begun to use SNPs identified from these GWA studies, offering up to 1 million SNP GWA scans (http://www.decodeme.com, https://www.23andme.com) or individual disease-associated SNP tests (http://www.decodediagnostics.com). It is, however, unclear how useful the currently identified variants will be in predicting disease.

One of the disease traits for which the GWA approach has been most successful is type 2 diabetes. Together with candidate gene approaches, eighteen common variants, including *FTO* and two independent signals in the *CDKN2A/B*

region, have now been convincingly shown to associate with the disease [1-6, 20-26]. In this study, we aimed to assess the combined discriminatory power of these common, modest effect variants, using over 4900 individuals from the GoDARTS (Genetics of Diabetes Audit and Research Tayside) study, which was also used as for replication of associated signals in the WTCCC-T2D GWA study[5].

Methods

SNP selection and genotyping

We only included variants that have been convincingly shown to associate with type 2 diabetes. We used variants reviewed in [27] and those described in [5] and [6], except for: the E23K (rs5219; r^2 with GWA-SNP rs5215 = 0.89) variant of *KCNJ11* [22] and rs7903146 (r^2 with GWA-SNP rs7901695 = 0.80) of *TCF7L2* [23, 28] where we genotyped a SNP shown to have stronger association with type 2 diabetes, but which were not genotyped on the genome-wide association chips; the *TCF2* locus where we used rs757210 [26], instead of rs4430796 [24] (r^2 = 0.61); and the *ADAM30/NOTCH2* locus where we used rs2641348 in *ADAM30* as a proxy for rs2934381 (r^2 = 0.92).

Genotyping was performed by KBioscience (Herts., UK) who designed and used assays based on either their proprietary competitive allele-specific PCR (KASPar) method or a modified TaqMan-based assay, details of which are available on their website (www.kbioscience.co.uk/chemistry/index.htm). Genotyping quality control measures for the SNPs are described in [5, 6, 25].

GoDARTS study and participants

The GoDARTS study is a sub-study of the Diabetes Audit and Research Tayside (DARTS) study [29], which aims to identify all known diabetes patients in the Tayside region of Scotland using electronic database retrieval. The samples used in this study are a sub-sample of the type 2 diabetes patients identified and are described previously [6]. Briefly, the GoDARTS study includes individuals of white European descent, living in the Tayside region when recruited. The diagnosis of diabetes in cases was based on either current treatment with diabetes-specific medication, or laboratory evidence of hyperglycaemia if treated with diet alone. Patients with confirmed diagnosis of monogenic diabetes and those treated with regular insulin therapy within 1 year of diagnosis were excluded. Cases in this study had an age at diagnosis between 35 and 70, inclusive. Controls had not

been diagnosed with diabetes at the time of recruitment or subsequently, and were excluded if there was evidence of hyperglycaemia during recruitment (fasting glucose >7.0 mmol/l, HbA1c >6.4%) or were > 80 years old. The study was approved by the Tayside Medical Ethics Committee. Informed consent was obtained from all study participants. **Table 1** presents the clinical characteristics of subjects used in this study.

Statistical analysis

All statistical analyses were performed in StataSE v10.0 for Windows (StataCorp LP, Texas, USA). We used logistic regression for all individual SNP analyses. To test for deviation from a within-loci additive model, we performed likelihood ratio test of an additive model against a general 2df model. To test for gene-gene interaction across all pairs of loci, we used likelihood ratio tests to compare an additive model to a model with an interaction term. We combined information from multiple SNPs by using an allele count model, where we summed the number of risk alleles carried by each individual. This assumes that each of the alleles has an equal and additive effect on type 2 diabetes risk.

We used logistic regression on the general model (i.e. individual SNP genotypes as indicator variables) to construct the ROC curves and calculate the AUCs. We also performed these ROC analyses on the allele count model for comparison to the general model.

Results

Genotyping data on all of the variants were available for 2309 type 2 diabetes cases and 2598 controls. Characteristics of these participants are shown in **Table 1**. **Supplementary Table 1** presents a comparison of clinical characteristics for these subjects against the 1739 who were not successfully genotyped across all SNPs. Individually, the variants have similar effect sizes in this study compared to those reported in other large studies (**Table 2**; [1-6, 20-26]) and the range of odds ratios from 1.00 to 1.36 most likely reflects stochastic variation. Several variants are not associated at $P<0.05$ in the sample used here, but are still included in the analyses as they are confirmed type 2 diabetes risk variants, and the lack of significance is the result of relatively low power in this number of subjects.

Based on these and larger datasets, all the variants appear to have an additive mode of inheritance [1-6, 20-26]. The *CDKAL1* locus was reported by Steinthorsdottir *et al.* to fit a recessive model [4], but other large studies do not support this. There is no evidence of interaction between any of the SNPs based on these data (**Supplementary Table 2**), or on the larger analyses previously published. Therefore, we assumed an additive genetic model. We found no evidence of any interaction between the individual variants and BMI or age (lowest interaction P values = 0.14 and 0.02, respectively). We performed the analysis with and without the *FTO* variant, the one variant shown to predispose to type 2 diabetes through a primary effect on BMI [18].

Comparing extremes

The proportion of case subjects and control subjects grouped according to the number of risk alleles they carry is shown in **Figure 1**. The distribution of risk alleles follows a normal distribution in both cases and controls, with a shift towards a higher number of risk alleles in the case subjects. There is an increase in odds ratios for type 2 diabetes with the increasing number of risk alleles, against the baseline group of 1.8% of individuals carrying 10-12 risk alleles. 1.2% of

individuals with 25 or more risk alleles have an odds ratio of 4.2 (95% CI: 2.11, 8.56) against the baseline reference group. Similarly, 11.5% of this study population carrying 22 or more risk alleles had an OR of 2.3 (95% CI: 1.73, 2.93) for type 2 diabetes compared to the 8.2% of individuals with <15 risk alleles.

Figure 2 plots the odds ratios relative to the median number of 18 risk alleles. Those with 25 or more risk alleles were over twice as likely to have type 2 diabetes (OR: 2.18, 95% CI: 1.24, 3.81) compared to those with the median number of risk alleles. The *TCF7L2* variant had a stronger effect than the other variants (OR = 1.36, compared to 1.00-1.25 for the rest), so these results may be slight underestimates, since the additive model used for the allele counting assumes equal effects across all SNPs.

We performed the same analyses for two sub-groups of the cohort, one including only obese individuals (with BMI of 30 kg/m^2 or greater, n=1803), the other non-obese individuals (BMI less than 30 kg/m^2, n=3083). The results were similar across these sub-groups. For example, the 1.4% of obese individuals with >24 risk alleles had an OR = 5.5 (95% CI: 2.11, 14.36) compared to the 1.9% of obese individuals with <13 risk alleles. The corresponding odds ratio for the non-obese subjects was 3.31 (95% CI: 1.34, 8.16), for the 1.8% and 1.1% of individuals with <13 and >24 risk alleles, respectively.

ROC curve

We evaluated the discriminatory power of a genetic test based on the 18 type 2 diabetes variants by calculating the area under the receiver-operating characteristic (ROC) curve. Using the general model (as opposed to the additive model which assumes equal and additive effects), the ROC curve for the 18 type 2 diabetes variants studied here is 0.60 (**Figure 3**). We performed the same analysis for the obese and non-obese sub-groups of the cohort. The AUCs for the obese and non-obese groups were 0.58 and 0.60, respectively. Similar result was obtained when we removed the *FTO* variant (obese 0.58, non-obese 0.59). We also tested whether the risk variants would add to the discriminatory power of BMI,

age and sex alone (AUC = 0.78 in our study). A model that includes BMI, age, sex and the 18 variants has an AUC of 0.80 (**Figure 3**); although marginal, the increase in the AUC was statistically significant (P=2.88x10^{-12}). The AUC remained virtually the same (AUC=0.80) when the *FTO* variant was removed from the model.

The effect of BMI and age

Supplementary Table 3 presents the individual SNP type 2 diabetes associations adjusted for BMI. As expected, the *FTO* association is weakened on adjusting for BMI (OR = 1.00 [0.92, 1.10]), and the *TCF7L2* strengthened (OR = 1.46 [1.32, 1.61]). Testing the combined effect of the risk variants on clinical features of the type 2 diabetes patients, we found that the number of risk alleles was associated with an earlier age at diagnosis of 0.15 years per risk allele (95% CI: -0.29, -0.01; P=0.038). We also observed an overall modifying effect on BMI (-0.14 BMI units per risk allele, 95% CI: -0.23, -0.05, P=3.41x10^{-3}), but this finding is mainly explained by the known association of the *TCF7L2* variant alone with BMI, in type 2 diabetes cases [30, 31]. Here, each *TCF7L2* risk allele was associated with a difference in BMI of -0.69 kg/m^2 (95% CI: -1.06, -0.31; P=3.18x10^{-4}), while the combined effect of all other variants without *TCF7L2* could just be detected (-0.10 kg/m^2 per risk allele; 95% CI: -0.20, 0.01; P=0.036). The difference in BMI and age at diagnosis was more noticeable when we compared individuals with low and high number of risk alleles. For example, carriers of 23 or more risk alleles (11.8%) were, on average, diagnosed 4.2 years earlier (95% CI: -6.45, -1.87; P=4.21x10^{-4}) and had 1.60 kg/m^2 lower BMI (95% CI: -3.35, 0.08; P=0.062) than those carrying fewer than 15 (8.6%) risk alleles.

Discussion

Recent success in identifying common variants predisposing to type 2 diabetes has led to suggestions that they may be useful in predicting an individual's risk of the disease. In this study we evaluated the ability of 18 confirmed predisposing variants to discriminate between individuals with and without type 2 diabetes, using the GoDARTS study. The samples used in this study were not enriched for family history or low BMI, factors that may inflate effect sizes. Although the GoDARTS cohort was a part of the WTCCC-T2D GWA study [5, 6], it was only used as a stage 2 replication set for the follow-up of the initial hits. This means that there should be a minimal effect of the "winner's curse" [32], the upward bias of the effect size in the discovery samples compared to subsequent replication studies.

The combined information identifies individuals at different risks of disease

By comparing individuals with fewest type 2 diabetes risk alleles with those carrying the most risk alleles, combining genetic information allowed us to identify subgroups of the population at a distinctly differing risk of disease. For example, we were able to distinguish about 1% of the population carrying more than 25 risk alleles that had over four times increased risk of diabetes compared to the 2% with 10-12 risk alleles. The high-risk group also had over twice the odds for type 2 diabetes, than those with the median number of risk alleles. These figures were similar in individuals who were obese and not obese, a major risk factor for type 2 diabetes and easily measurable. Obese individuals carrying large numbers of type 2 diabetes risk alleles may therefore be a particular group worth studying to test potential intervention strategies. This may be important given that the escalating rates of obesity and type 2 diabetes suggest that efforts aimed at the whole population are not effective, and that intensive, but expensive, lifestyle interventions aimed at increasing exercise and improving diet can result in weight loss and a reduced risk of type 2 diabetes [33-36].

The current variants are not particularly discriminative, but explain only a small amount of the heritability of type 2 diabetes

Rather than focusing on individuals with "extreme" numbers of risk alleles, at a population level the utility of genetic tests may be better classified by receiver-operating characteristic (ROC) curves. One of the most important factors in the validity of a genetic test in clinical practice is its ability to discriminate between individuals who will and will not develop the disease. A clinically relevant AUC threshold clearly depends on a whole range of factors (for example, the cost of the test, and the availability of preventative measures), but as an example from current clinical practice, oxidized-LDL cholesterol has an AUC of ~0.80 for coronary artery disease [37], making it a good discriminator between patients and healthy controls. The 18 type 2 diabetes variants had an inadequate discriminatory ability with an AUC of 0.60, a slight improvement on the AUC of 0.55 based on *TCF7L2* alone. These data imply that genetic tests for type 2 diabetes (and many other complex diseases) that are offered by several commercial companies currently have limited predictive value. However, there are many more variants to be identified, since these 18 variants only explain a small amount of the heritability of type 2 diabetes: the sibling relative risk for type 2 diabetes is ~3 [38], and the combination of these variants would only account for a sibling relative risk of ~1.07. As more susceptibility variants are found for type 2 diabetes, genetic testing that utilizes the inexpensive and rapid genotyping technologies may eventually become more clinically useful.

The use of genetic information in addition to age, sex and BMI

For many complex diseases, there are already well-established risk factors that can be used to predict someone's chances of developing the disease. Incorporating genetic information may be justified on the basis that current preventative measures are expensive, and that prevention at a population level is not effective, so the more selective we can be the better. In type 2 diabetes, family history, age, BMI, ethnicity and lifestyle all contribute to an individual's risk of the disease. In our study the AUC for BMI, age and sex (we did not have family history

data) combined was 0.78, a moderate diagnostic value. The genetic risk variants had a poor discriminatory ability alone (AUC=0.60), and only marginally increased the discriminatory power of the test when combined with BMI, age and sex (AUC=0.80), suggesting that they add little to the already known predictive factors.

Risk variants modify clinical characteristics of individuals with type 2 diabetes

Type 2 diabetes often occurs in individuals who are not overweight or obese, and can be diagnosed at a relatively young age. This may be because these individuals have a stronger genetic risk component than more "typical" type 2 diabetes patients. Therefore, we tested the extent to which patients with the stronger genetic predisposition tended to be leaner, and how much younger they were at diagnosis. There were notable differences between the 11.8% and 8.6% of the population carrying either high or low number of disease predisposing alleles, respectively. Patients with high genetic risk had an average BMI of 30.3 kg/m^2 compared with 31.9 kg/m^2 in those with low genetic risk, and were diagnosed at an average age of 55.2 years, compared to 59.3 years for patients with relatively low genetic risk. These results support an important role for genetic predisposition to type 2 diabetes in non-obese, young-onset cases.

Weighting variants and the optimal ROC curve

The simple allele count model we used for some of our analyses of "extremes" assumes that each risk allele has the same effect size, and the effects are additive both within and between loci. While we found no strong evidence for deviation from additivity, clearly some SNPs have stronger effects than others. This is most evident for *TCF7L2*, where the allelic odd ratio is 1.37, significantly larger than any of the other variants. One way to overcome this is to weigh SNPs differently; however, we decided not to do this in this study for a number of reasons. First, all our AUC analyses are based on a general model, where the assumption of equal effects is not made. Second, as Janssens *et al.* [39] previously showed, when the odds ratios of the individual variants are relatively low (as here)

there is little difference in the discriminative accuracy of the test based on the simple allele count model and a model which allows each variant to have a different effect size (the AUCs here are 0.583 and 0.603, respectively, although this was statistically significant (*P*=0.001)). Third, it is unclear what the most appropriate weights to use would be. Fourth, an allele count model provides important advantages for simplicity and visualization of the results.

Recently, Lu and Elston [40] proposed using an optimal ROC analysis approach rather than the standard approach we have used. While the authors proved theoretically that their method is more powerful, the results presented by Lu and Elston [40] showed that the two methods produce the same results when there are few loci, and no interactive effects. As we still have only a relatively few loci, there is no evidence of any non-additive effects within or between loci, and the ROC curve is concave [40] the two methods should produce the same results. We tested this using the 10 SNPS that were significant (at *P*<0.05) in our study. Using these variants the results were the same for both methods (AUC for the Lu and Elston method = 0.596; AUC for the standard method = 0.596).

Strengths and limitations of our study

Our study was relatively large in terms of the number of samples, and number of common variants used. We had over 2000 cases and 2000 controls, after excluding individuals who were not successfully genotyped for all of the variants included in the study. The 18 variants we used had all been convincingly shown in previous studies to associate with type 2 diabetes.

One of the main limitations of our study is that it was not prospective and, therefore, we are unable to truly determine the predictive power of these variants. Although the results of this study only apply to the Tayside population, it is likely, based on previous data [41-43], that our prediction estimates are reasonably accurate, and that the effect sizes observed are likely to be representative of those in similar populations. A second limitation is that, although the results are applicable to the Tayside and similar populations, they may not apply to

populations of substantially different ethnic origin or those exposed to different social and environmental circumstances. A third limitation concerns the caveat that the majority of the type 2 diabetes associated SNPs identified to date and used in this study are not the causal variants. This means that the predictive power of these susceptibility loci is likely to be an underestimate. Fine mapping and sequencing approaches are needed to identify the variants causal to these associations, which often have stronger effects than the currently identified variants. These follow-up studies may also reveal additional causal variants at these loci that cannot be detected by GWA methods because of, for example, low frequency, but that may have higher penetrance and, therefore, would be much more powerful predictors.

In conclusion, the combined information from the currently known susceptibility variants allows us to identify subgroups of the population at substantially increased odds of getting type 2 diabetes. These individuals could be targeted with more effective preventative measures. On a population level, these variants appear to be of limited use in discriminating between individuals who will and will not develop type 2 diabetes. As more variants are identified, tests with better predictive performance should become available, and could eventually become a valuable addition to clinical practice.

Acknowledgements

We thank all study participants. M.N.W. is a Vandervell Foundation Research Fellow. E.Z. is a Wellcome Trust Research Career Development Fellow. We thank the Wellcome Trust for funding.

Table 1. Characteristics of study participants.

Variable	Cases	Controls
Number	2309	2598
Males, %	56	51
Age at diagnosis, years (SD)	55.7 (9.0)	NA
Body mass index, kg/m^2 (SD)	31.5 (6.1)	26.9 (4.5)
HbA1c, % (SD)	7.8 (1.5)	5.5 (0.3)

Table 2. Summary of type 2 diabetes variants in 2598 controls and 2309 cases from the Dundee cohort.

SNP	Gene / Region	Risk allele frequency	Additive model test *P*	Odds ratio (95% CI)	*P* value
rs7903146	*TCF7L2*	0.30	0.70	1.36 (1.24, 1.48)	3.97×10^{-12}
rs5219	*KCNJ11*	0.36	0.058	1.25 (1.15, 1.36)	8.54×10^{-8}
rs10811661	*CDKN2A/2B*	0.85	0.24	1.21 (1.08, 1.35)	8.82×10^{-4}
rs1801282	*PPARG*	0.87	0.46	1.21 (1.07, 1.36)	2.18×10^{-3}
rs2641348*	*ADAM30 / NOTCH2*	0.11	0.68	1.15 (1.01, 1.30)	3.20×10^{-2}
rs564398	*CDKN2A/2B*	0.59	0.95	1.13 (1.04, 1.22)	3.61×10^{-3}
rs4402960	*IGF2BP2*	0.33	0.76	1.12 (1.03, 1.22)	7.62×10^{-3}
rs8050136	*FTO*	0.41	0.32	1.11 (1.02, 1.20)	1.43×10^{-2}
rs10946398	*CDKAL1*	0.34	0.19	1.11 (1.02, 1.21)	1.47×10^{-2}
rs13266634	*SLC30A8*	0.70	0.60	1.10 (1.01, 1.20)	2.57×10^{-2}
rs7961581	*TSPAN8 / LGR5*	0.29	0.87	1.09 (1.00, 1.19)	5.56×10^{-2}
rs12779790	*CDC123*	0.20	0.15	1.10 (0.99, 1.21)	7.58×10^{-2}
rs10010131	*WFS1*	0.60	0.54	1.07 (0.99, 1.16)	9.19×10^{-2}
rs757210	*TCF2*	0.37	0.18	1.07 (0.99, 1.16)	1.09×10^{-1}
rs4607103	*ADAMTS9*	0.77	0.60	1.05 (0.96, 1.16)	2.89×10^{-1}
rs1111875	*HHEX-IDE*	0.62	0.19	1.02 (0.94, 1.11)	5.98×10^{-1}
rs7578597	*THADA*	0.91	0.33	1.04 (0.90, 1.19)	6.07×10^{-1}
rs864745	*JAZF1*	0.50	0.50	1.00 (0.93, 1.09)	9.70×10^{-1}

Only samples that were successfully genotyped for all 18 variants are included. Additive model test P refers to a test of deviation from additivity of alleles at each SNP. *This SNP falls within the *ADAM30* gene and is a proxy (r^2=0.92 in HapMap CEU) for rs2934381 in the *NOTCH2* gene, which showed stronger association in [5].

Figure 1. Distribution of risk alleles in type 2 diabetes cases (black bars) and controls (grey bars).

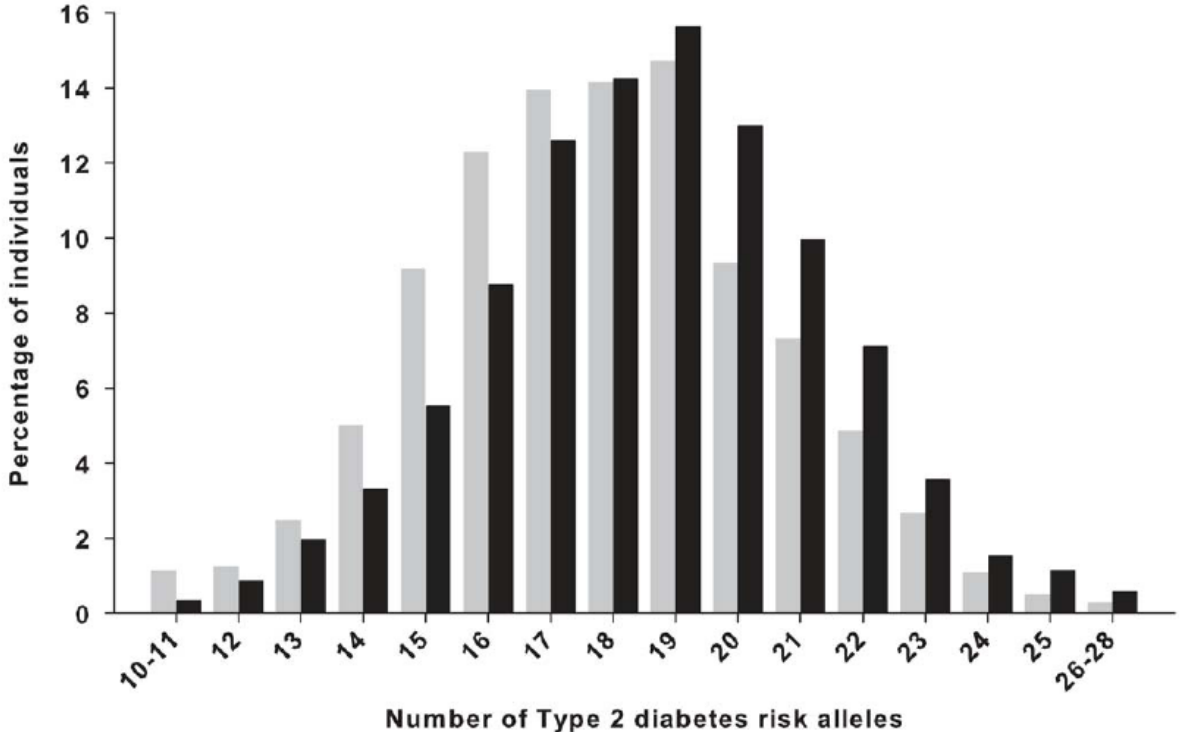


Figure 2. A plot showing odds ratios (OR) by number of type 2 diabetes risk alleles. The odds ratios are given relative to the median number of 18 risk alleles (black circle). The vertical bars represent 95% confidence intervals.

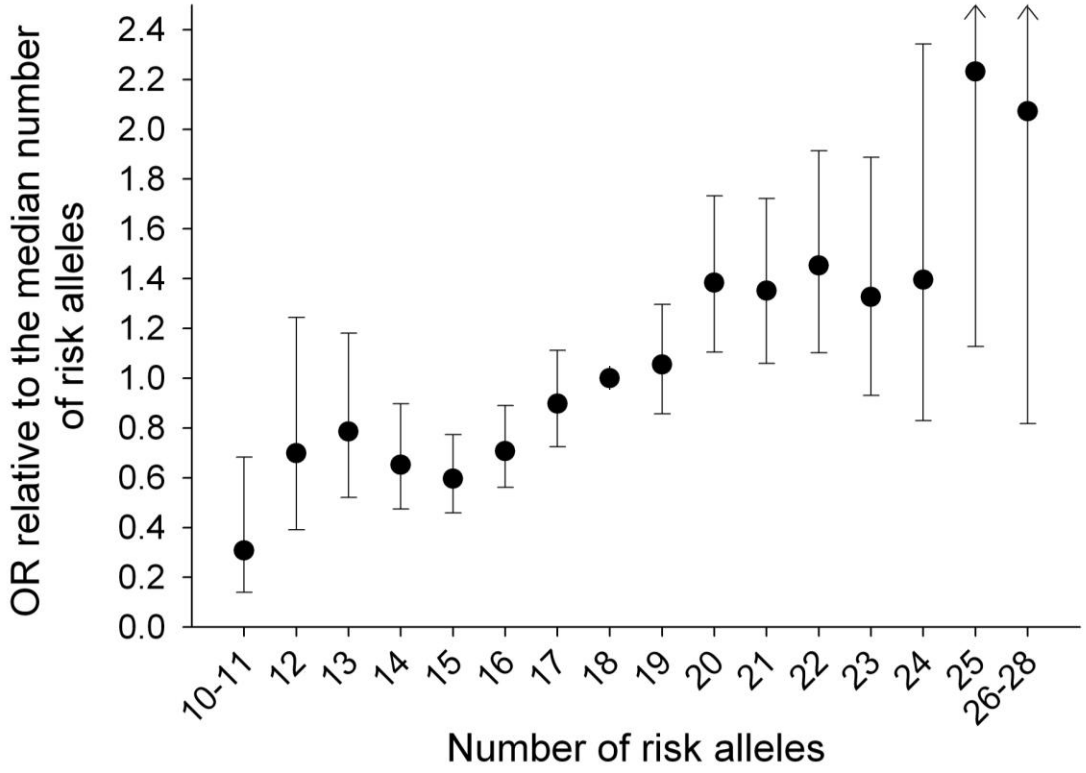
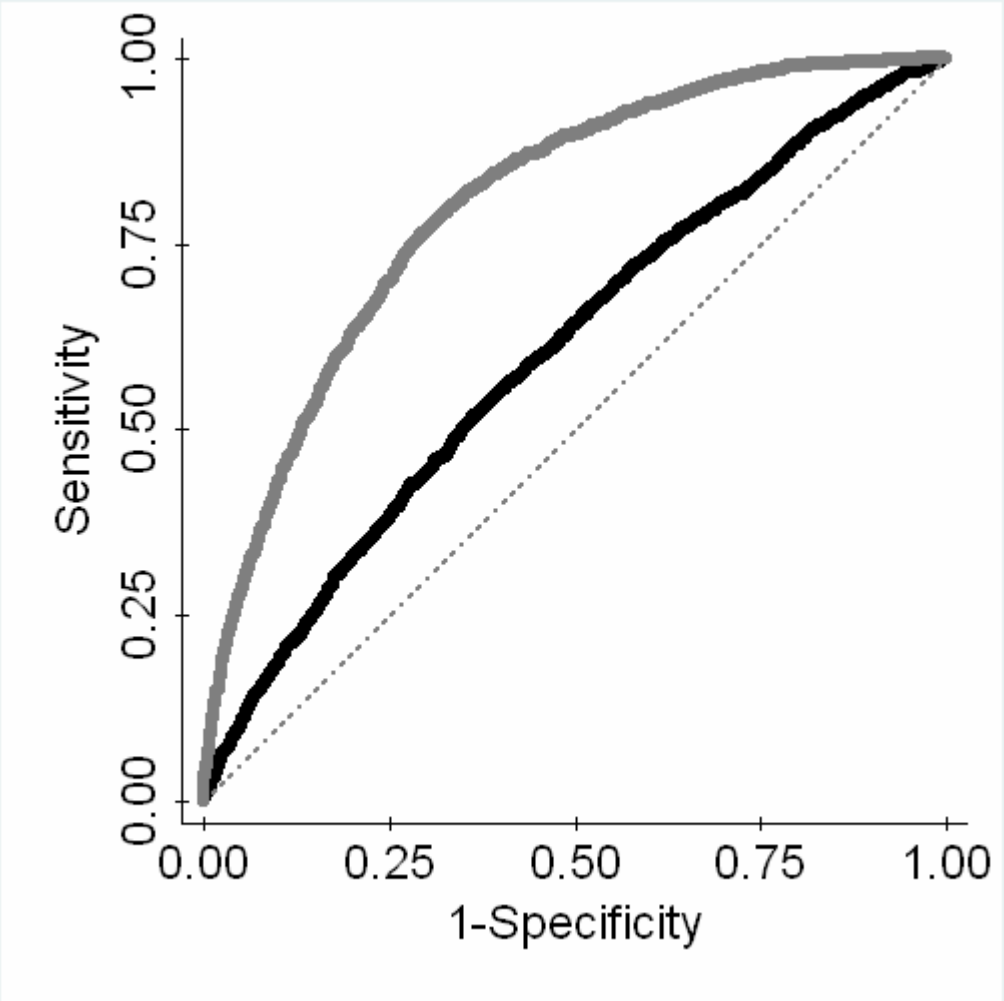


Figure 3. ROC plot for a model containing all type 2 diabetes variants, BMI, age and sex (gray line; AUC = 0.80), and the 18 variants alone (black line, AUC = 0.60).



References

1. Saxena, R. et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 316, 1331-1336 (2007).

2. Scott, L. J. et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 316, 1341-1345 (2007).

3. Sladek, R. et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 445, 881-885 (2007).

4. Steinthorsdottir, V. et al. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. Nature Genetics 39, 770-775 (2007).

5. Zeggini, E. et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet advanced online publication (2008).

6. Zeggini, E. et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 316, 1336-1341 (2007).

7. Helgadottir, A. et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. Science 316, 1491-1493 (2007).

8. McPherson, R. et al. A common allele on chromosome 9 associated with coronary heart disease. Science 316, 1488-1491 (2007).

9. Samani, N. J. et al. Genomewide association analysis of coronary artery disease. New England Journal Of Medicine 357, 443-453 (2007).

10. Gudmundsson, J. et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nature Genetics 39, 631-637 (2007).

11. Yeager, M. et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nature Genetics 39, 645-649 (2007).

12. Hampe, J. et al. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. Nature Genetics 39, 207-211 (2007).

13. Parkes, M. et al. Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. Nature Genetics 39, 830-832 (2007).

14. Rioux, J. D. et al. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. Nature Genetics 39, 596-604 (2007).

15. Grarup, N. et al. Studies of association of variants near the HHEX, CDKN2A/B, and IGF2BP2 genes with type 2 diabetes and impaired insulin release in 10,705 Danish subjects - Validation and extension of genome-wide association studies. Diabetes 56, 3105-3111 (2007).

16. Pascoe, L. et al. Common variants of the novel type 2 diabetes genes CDKAL1 and HHEX/IDE are associated with decreased pancreatic beta-cell function. Diabetes 56, 3101-3104 (2007).

17. Dina, C. et al. Variation in FTO contributes to childhood obesity and severe adult obesity. Nature Genetics 39, 724-726 (2007).

18. Frayling, T. M. et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science 316, 889-894 (2007).

19. Scuteri, A. et al. Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. Plos Genetics 3, 1200-1210 (2007).

20. Altshuler, D. et al. The common PPARγ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nature Genetics 26, 76-80 (2000).

21. Florez, J. C. et al. Haplotype structure and genotype-phenotype correlations of the sulfonylurea receptor and the islet ATP-sensitive potassium channel gene region. Diabetes 53, 1360-8 (2004).

22. Gloyn, A. L. et al. Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. Diabetes 52, 568-72 (2003).

23. Grant, S. F. A. et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nature Genetics 38, 320-323 (2006).

24. Gudmundsson, J. et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. Nature Genetics 39, 977-983 (2007).

25. Sandhu, M. S. et al. Common variants in WFS1 confer risk of type 2 diabetes. Nature Genetics 39, 951-953 (2007).

26. Winckler, W. et al. Evaluation of common variants in the six known maturity-onset diabetes of the young (MODY) genes for association with type 2 diabetes. Diabetes 56, 685-693 (2007).

27. Frayling, T. M. Genome-wide association studies provide new insights into type 2 diabetes aetiology. Nature Reviews Genetics 8, 657-662 (2007).

28. Helgason, A. et al. Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution. Nat Genet 39, 218-25 (2007).

29. Morris, A. D. et al. The diabetes audit and research in Tayside Scotland (DARTS) study: electronic record linkage to create a diabetes register. DARTS/MEMO Collaboration. Bmj 315, 524-8 (1997).

30. Florez, J. C. et al. TCF7L2 polymorphisms and progression to diabetes in the Diabetes Prevention Program. New England Journal Of Medicine 355, 241-250 (2006).

31. Weedon, M. N. The importance of TCF7L2. Diabetic Medicine 24, 1062-1066 (2007).

32. Lohmueller, K. E., Pearce, C. L., Pike, M., Lander, E. S. & Hirschhorn, J. N. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nature Genetics 33, 177-182 (2003).

33. Lindstrom, J. et al. Sustained reduction in the incidence of type 2 diabetes by lifestyle intervention: follow-up of the Finnish Diabetes Prevention Study. Lancet 368, 1673-1679 (2006).

34. Knowler, W. C. et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. New England Journal Of Medicine 346, 393-403 (2002).

35. Tuomilehto, J. et al. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. New England Journal Of Medicine 344, 1343-1350 (2001).

36. Pan, X. R. et al. Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance - The Da Qing IGT and diabetes study. Diabetes Care 20, 537-544 (1997).

37. Johnston, N., Jernberg, T., Lagerqvist, B., Siegbahn, A. & Wallentin, L. Improved identification of patients with coronary artery disease by the use of new lipid and lipoprotein biomarkers. American Journal Of Cardiology 97, 640-645 (2006).

38. Kobberling, J. & Tillil, H. in The genetics of diabetes mellitus (ed. Kobberling J Tattersall, R.) 201-210 (Academic Press, London, 1982).

39. Janssens, A. C. et al. The impact of genotype frequencies on the clinical validity of genomic profiling for predicting common chronic diseases. Genet Med 9, 528-35 (2007).

40. Lu, Q. & Elston, R. C. Using the optimal receiver operating characteristic curve to design a predictive genetic test, exemplified with type 2 diabetes. Am J Hum Genet 82, 641-51 (2008).

41. Lyssenko, V. et al. Genetic prediction of future type 2 diabetes. Plos Medicine 2, 1299-1308 (2005).

42. Lyssenko, V. et al. Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes. Journal Of Clinical Investigation 117, 2155-2163 (2007).

43. Weedon, M. N. et al. Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. Plos Medicine 3, 1877-1882 (2006).

SUPPLEMENTARY INFORMATION

Supplementary Table 1. Clinical characteristics of individuals from the original sample who were excluded from the analyses owing to incomplete genotype data, compared to the individuals successfully genotyped for all 18 SNPs and included in the analyses.

Variable		Individuals excluded from the analyses	Individuals used in the analyses	*P*-value for difference
Total number		1739	4907	NA
Percentage of cases		45.0	47.1	0.14
Percentage of males		55.7	53.5	0.12
Age at diagnosis (cases), years (SD)		55.2 (8.8)	55.7 (9.0)	0.20
Body mass index, kg/m^2 (SD)	cases	31.9 (6.3)	31.5 (6.1)	0.07
	controls	26.6 (4.6)	26.9 (4.5)	0.12

Supplementary Table 2. Interaction test *P* values for all combinations of 18 type 2 diabetes SNPs.

SNP1	SNP2	Interaction *P*
rs7903146	rs12779790	0.0038
rs4402960	rs13266634	0.012
rs13266634	rs1801282	0.019
rs7903146	rs7961581	0.043
rs10010131	rs12779790	0.052
rs5219	rs7578597	0.056
rs564398	rs757210	0.056
rs4402960	rs757210	0.057
rs12779790	rs864745	0.085
rs10811661	rs757210	0.090
rs1111875	rs13266634	0.092
rs12779790	rs7961581	0.093
rs13266634	rs5219	0.11
rs757210	rs12779790	0.15
rs1111875	rs5219	0.15
rs10946398	rs5219	0.16
rs5219	rs2641348	0.16
rs564398	rs12779790	0.17
rs5219	rs1801282	0.17
rs7903146	rs5219	0.19
rs10811661	rs10946398	0.21
rs10946398	rs864745	0.21
rs5219	rs12779790	0.22
rs10946398	rs12779790	0.22
rs10811661	rs5219	0.22
rs8050136	rs5219	0.23
rs1111875	rs2641348	0.24
rs8050136	rs1801282	0.24
rs10946398	rs10010131	0.25
rs564398	rs10010131	0.25
rs10811661	rs7961581	0.25
rs4402960	rs12779790	0.25
rs8050136	rs4607103	0.26
rs757210	rs2641348	0.27
rs1111875	rs757210	0.28
rs2641348	rs7578597	0.28
rs10811661	rs4402960	0.29
rs7903146	rs10010131	0.29
rs8050136	rs10946398	0.30
rs7903146	rs7578597	0.30
rs4402960	rs4607103	0.31
rs10946398	rs564398	0.31
rs8050136	rs1111875	0.32

rs5219	rs10010131	0.32
rs5219	rs4607103	0.32
rs8050136	rs7903146	0.32
rs2641348	rs864745	0.32
rs10811661	rs7903146	0.34
rs5219	rs757210	0.36
rs1111875	rs12779790	0.37
rs5219	rs7961581	0.37
rs10946398	rs1801282	0.38
rs1111875	rs10010131	0.38
rs12779790	rs7578597	0.38
rs10946398	rs1111875	0.38
rs4402960	rs5219	0.39
rs1111875	rs7578597	0.40
rs564398	rs7961581	0.40
rs10811661	rs7578597	0.41
rs12779790	rs4607103	0.41
rs8050136	rs564398	0.41
rs5219	rs864745	0.41
rs1111875	rs864745	0.42
rs10010131	rs4607103	0.42
rs10946398	rs4402960	0.43
rs564398	rs4607103	0.44
rs10811661	rs8050136	0.45
rs10811661	rs864745	0.46
rs13266634	rs2641348	0.46
rs10811661	rs1111875	0.46
rs10946398	rs13266634	0.46
rs564398	rs5219	0.46
rs7903146	rs4607103	0.47
rs757210	rs4607103	0.47
rs4607103	rs7961581	0.47
rs8050136	rs7961581	0.48
rs7903146	rs1801282	0.48
rs10946398	rs7903146	0.49
rs4607103	rs7578597	0.49
rs10946398	rs7578597	0.50
rs13266634	rs10010131	0.51
rs8050136	rs7578597	0.51
rs10811661	rs12779790	0.52
rs2641348	rs7961581	0.52
rs4402960	rs7961581	0.52
rs564398	rs2641348	0.54
rs10946398	rs757210	0.55
rs4402960	rs7578597	0.57
rs10811661	rs10010131	0.57
rs564398	rs13266634	0.57

rs8050136	rs757210	0.58
rs7903146	rs2641348	0.58
rs1111875	rs7903146	0.58
rs8050136	rs864745	0.59
rs1801282	rs864745	0.61
rs13266634	rs4607103	0.61
rs1111875	rs1801282	0.61
rs10811661	rs13266634	0.62
rs10946398	rs7961581	0.62
rs1111875	rs4402960	0.63
rs10946398	rs2641348	0.63
rs7578597	rs864745	0.64
rs1801282	rs4607103	0.65
rs757210	rs7578597	0.65
rs10811661	rs2641348	0.65
rs7903146	rs757210	0.65
rs4402960	rs864745	0.67
rs2641348	rs4607103	0.67
rs1801282	rs7578597	0.67
rs1801282	rs7961581	0.68
rs1111875	rs7961581	0.69
rs757210	rs7961581	0.70
rs12779790	rs2641348	0.70
rs10010131	rs757210	0.72
rs8050136	rs12779790	0.72
rs10811661	rs1801282	0.72
rs8050136	rs4402960	0.73
rs13266634	rs757210	0.73
rs1801282	rs12779790	0.74
rs7961581	rs864745	0.74
rs10811661	rs564398	0.76
rs10010131	rs7961581	0.76
rs10811661	rs4607103	0.77
rs13266634	rs864745	0.77
rs13266634	rs7578597	0.78
rs13266634	rs7961581	0.78
rs564398	rs7578597	0.80
rs8050136	rs2641348	0.80
rs13266634	rs12779790	0.80
rs1111875	rs4607103	0.81
rs757210	rs864745	0.81
rs1801282	rs2641348	0.81
rs1801282	rs757210	0.83
rs4402960	rs1801282	0.83
rs4402960	rs7903146	0.84
rs7578597	rs7961581	0.85
rs13266634	rs7903146	0.88

rs1111875	rs564398	0.89
rs10010131	rs864745	0.90
rs4402960	rs10010131	0.90
rs10946398	rs4607103	0.92
rs8050136	rs13266634	0.92
rs10010131	rs2641348	0.92
rs8050136	rs10010131	0.93
rs4607103	rs864745	0.93
rs1801282	rs10010131	0.96
rs10010131	rs7578597	0.97
rs564398	rs7903146	0.97
rs564398	rs864745	0.98
rs7903146	rs864745	0.98
rs564398	rs1801282	0.98
rs564398	rs4402960	0.99
rs4402960	rs2641348	1.00

Supplementary Table 3 BMI-adjusted odds ratios (OR) for 4886 individuals with non-missing values for both 18 SNPs and BMI. The SNPs are in the same order as in Table 2.

SNP	Gene / Region	OR adjusted for BMI (95% CI)	Adjusted OR *P* value
rs7903146	*TCF7L2*	1.46 (1.32, 1.61)	5.33×10^{-15}
rs5219	*KCNJ11*	1.28 (1.17, 1.40)	1.10×10^{-7}
rs10811661	*CDKN2A/2B*	1.27 (1.13, 1.44)	1.09×10^{-4}
rs1801282	*PPARG*	1.24 (1.09, 1.41)	1.47×10^{-3}
rs2641348	*ADAM30 / NOTCH2*	1.20 (1.05, 1.38)	8.05×10^{-3}
rs564398	*CDKN2A/2B*	1.11 (1.02, 1.21)	2.20×10^{-2}
rs4402960	*IGF2BP2*	1.11 (1.01, 1.22)	2.44×10^{-2}
rs8050136	*FTO*	1.00 (0.92, 1.10)	9.14×10^{-1}
rs10946398	*CDKAL1*	1.10 (1.01, 1.21)	3.40×10^{-2}
rs13266634	*SLC30A8*	1.15 (1.05, 1.27)	3.85×10^{-3}
rs7961581	*TSPAN8 / LGR5*	1.09 (0.99, 1.20)	6.88×10^{-2}
rs12779790	*CDC123*	1.10 (0.98, 1.22)	1.01×10^{-1}
rs10010131	*WFS1*	1.09 (1.00, 1.18)	6.29×10^{-2}
rs757210	*TCF2*	1.13 (1.03, 1.23)	8.29×10^{-3}
rs4607103	*ADAMTS9*	1.07 (0.97, 1.19)	1.74×10^{-1}
rs1111875	*HHEX-IDE*	1.05 (0.96, 1.15)	2.85×10^{-1}
rs7578597	*THADA*	1.07 (0.92, 1.24)	3.86×10^{-1}
rs864745	*JAZF1*	1.01 (0.93, 1.11)	7.51×10^{-1}

CHAPTER 5: POLYGENIC RISK VARIANTS FOR TYPE 2 DIABETES SUSCEPTIBILITY MODIFY AGE AT DIAGNOSIS IN MONOGENIC HNF1A DIABETES

Hana Lango Allen, Stefan Johansson, Sian Ellard, Beverley Shields, Jens K Hertel, Helge Ræder, Kevin Colclough, Anders Molven, Timothy M Frayling, Pål R Njølstad, Andrew T Hattersley, Michael N Weedon

Published: Diabetes (2009) 59: 266

Abstract

Objective: Mutations in the *HNF1A* gene are the most common cause of MODY. There is a substantial variation in the age at diabetes diagnosis, even within families where diabetes is caused by the same mutation. We investigated the hypothesis that common polygenic variants that predispose to type 2 diabetes might account for the difference in age at diagnosis.

Research Design and Methods: Fifteen robustly associated T2D variants were successfully genotyped in 410 individuals from 203 *HNF1A*-MODY families, from two study centers in the UK and Norway. We assessed their effect on the age at diagnosis both individually and in a combined genetic score by summing the number of T2D risk alleles carried by each patient.

Results: We confirmed the effects of environmental and genetic factors known to modify the age at *HNF1A*-MODY diagnosis, namely intrauterine hyperglycemia (-5.1 years if present, $P=1.6\text{x}10^{-10}$), and *HNF1A* mutation position (-5.2 years if at least two isoforms affected, $P=1.8\text{x}10^{-2}$). Additionally, our data showed strong effects of sex (females diagnosed 3.0 years earlier, $P=6.0\text{x}10^{-4}$), and age at study (0.3 years later diagnosis per year increase in age, $P=4.7\text{x}10^{-38}$). There were no strong individual SNP effects; however, in the combined genetic score model, each additional risk allele was associated with 0.35 years earlier diabetes diagnosis ($P=5.1\text{x}10^{-3}$).

Conclusions: We show that T2D risk variants of modest effect sizes reduce the age at diagnosis in *HNF1A*-MODY. This is one of the first studies to demonstrate that clinical characteristics of a monogenic disease can be modified by common polygenic variants.

Introduction

Maturity onset diabetes of the young (MODY) is a young-onset, dominantly inherited non-insulin dependent diabetes mellitus resulting from β-cell dysfunction[1]. There are at least eight genetic subgroups of MODY [1, 2], with most patients having mutations in transcription factor genes. Hepatocyte nuclear factor 1 alpha (*HNF1A*) mutations are the commonest cause of MODY in many series [3, 4]. *HNF1A* diabetes is characterized by progressive failure of β-cell function, resulting in increasing hyperglycemia throughout life [1]. Initially, basal insulin secretion is maintained but it cannot be increased in the presence of hyperglycaemia [5].

The severity and clinical presentation of MODY varies according to MODY genetic subtype [6]. In addition, there can be considerable variation both between and within families where diabetes is caused by mutations in the same gene. In *HNF1A* diabetes the age of diagnosis is widely variable (4 - 74 years [7]), and, although the mutations are highly penetrant, only 63% of mutation carriers develop diabetes by the age of 25 [8]. The variation in diagnosis is influenced by social and environmental factors. Within families early age at diagnosis tends to fall in the younger generations, in part owing to increased awareness of the familial nature of the condition [9, 10]. If the mother had diabetes during pregnancy, intrauterine exposure to hyperglycemia of maternal diabetes is associated with diabetes being diagnosed on average 12 years earlier compared to subjects not exposed to maternal hyperglycaemia [9, 10].

It is likely that there are genetic modifiers of the age of onset of *HNF1A* diabetes, namely the position of the *HNF1A* mutation [11, 12]. However, much of the variation in age at diagnosis within families, where diabetes is caused by the same mutation, cannot be explained by social or environmental factors and this supports the notion that there are likely to be genetic modifiers independent of the *HNF1A* mutation. A genome-wide search for genetic modifiers of diagnosis age found no single large linkage peak [13], suggesting that the age of onset is a complex genetic

trait. A previous study of one large pedigree has shown that severity of the *HNF1A* diabetes phenotype was increased (earlier age of diagnosis and more severe hyperglycaemia) when type 2 diabetes was present in the non-carrier parent [14]. We, therefore, hypothesised that common genetic variants that predispose to type 2 diabetes might modify the severity of the disease and explain some of the variation in the age at *HNF1A* diabetes diagnosis.

Methods

HNF1A mutation patients

The subjects were 410 *HNF1A* mutation carriers with diabetes from two sources: The Department of Molecular Genetics, Royal Devon and Exeter NHS Foundation Trust, Exeter, UK (N=298 from 140 families) and the Center for Diabetes Genetics, Department of Pediatrics, Haukeland University Hospital, Bergen, Norway (N=112 from 63 families). They were all established MODY patients previously recruited for *HNF1A* sequencing on the basis of clinical criteria, such as family history or first degree relative with diabetes, onset of diabetes typically before age 25, or low dose insulin requirement (full details are available at http://www.diabetesgenes.org and http://www.mody.no).

All patients gave consent for genetic testing and had *HNF1A* mutations identified by direct sequencing. To avoid population heterogeneity, individuals who were not Caucasian Europeans were excluded. Clinical characteristics of the patients are shown in Table 1. Two thirds were females, and the majority of the probands had an *HNF1A* mutation in exons 1-6. Such mutations are regarded as clinically most severe as they affect all 3 isoforms of the gene product. The most common *HNF1A* mutation, P291fsinsC in exon 4, accounted for 32% of all cases in this study.

Classification and assessment of non-polygenic modifiers

We assessed the association between age at diagnosis and the following non-polygenic factors: sex, age at study, BMI, exposure to *in-utero* hyperglycemia, and the position of *HNF1A* mutation. Mother's diabetic status at pregnancy was calculated from the age of diagnosis and the date of birth of the mother and child. Where this information was incomplete (usually owing to mother's information not being available as deceased), we assumed that the patient was not exposed to *in-utero* hyperglycaemia. Previous studies showed that mutation position impacts the severity of the disease by determining the number of the affected *HNF1A*

isoforms: patients with mutations in exons 1-7, affecting two or all three isoforms, were diagnosed earlier than patients with mutations in exons 8-10, affecting only one isoform [11, 12]. To account for this effect, we classified mutations of patients in this study according to those two groups. Intronic mutations were assigned to a group according to the mutation position and hence *HNF1A* isoform impacted on.

SNP Selection and Genotyping

We decided to include in this study only those variants, or their proxies, robustly shown to predispose to type 2 diabetes in Caucasians. Seventeen common susceptibility variants had been identified and robustly replicated at the time of our study [15-20], recently reviewed in [21, 22]. These include SNPs in or near *PPARG*, *KCNJ11*, *TCF7L2*, *IGF2BP2*, *CDKN2A/2B*, *CDKAL1*, *SLC30A8*, *HHEX/IDE*, *FTO*, *WFS1*, *HNF1B* (*TCF2*), *MC4R*, *NOTCH2*, *ADAMTS9*, *THADA*, *TSPAN8/LGR5*, *CDC123/CAMK1D* and *JAZF1* genes. **Table 3** lists all 17 SNPs assessed in our study, of which we were able to combine 15 for the joint analysis, because *JAZF1* and *NOTCH2* loci failed genotyping in the Norwegian samples. At four of the 15 loci different SNPs were genotyped by the two study centres: rs757210 (UK) and rs4430796 (Norway) at the *HNF1B* locus (HapMap CEU r^2=0.61, D'=0.96); rs10946398 and rs7754840 at *CDKAL1* (r^2 and D'=1); rs8050136 and rs9939609 at *FTO* (r^2 and D'=1); and rs1111875 and rs5015480 at *HHEX/IDE* (r^2 and D'=1). We combined genotypes for each of the four proxy pairs, coded with respect to the type 2 diabetes risk allele.

In the UK samples, genotyping of *TCF7L2*, *KCNJ11* and *PPARG* SNPs was performed in-house, using a TaqMan-based assay. The probes were supplied by AppliedBiosystems (Foster City, CA). Genotyping of the remaining 14 variants was performed by KBiosciences (Herts, UK), who designed and used assays based on either their proprietary competitive allele-specific PCR (KASPar) method or a modified TaqMan assay, details of which are available on their website (www.kbioscience.co.uk/chemistry/index.htm). Genotyping success rate was >96% for each SNP, overall duplicate concordance rate was 99.9% (1 discrepancy from 1153 comparisons), and in the Hardy-Weinberg equilibrium test, used as an

additional genotyping quality check, all *P*-values were >0.01 for the full dataset and >0.05 for 140 unrelated probands.

In the Norwegian samples, genotyping was carried out by the multiplex MassARRAY® *iPLEX*™ System (SEQUENOM Inc., San Diego, CA, USA) at the Norwegian national technology platform CIGENE. *NOTCH2* SNP (rs2934381) failed the assay design, while *JAZF1* SNP (rs864745) had a poor genotype call rate. For the remaining 15 SNPs, which we were able to combine with the UK data, genotype concordance rate was 100% for internal controls (n=108 genotypes). Final genotyping call-rate was 99.2% after exclusion of samples with bad quality or lacking DNA. All tests for Hardy-Weinberg equilibrium had *P*-values >0.05.

Statistical methods

We performed family-based association analyses using ASSOC program from S.A.G.E. (Statistical Analysis for Genetic Epidemiology) software package, version 5.4.2 for Linux [23]. Assuming randomly sampled independent pedigrees, ASSOC simultaneously tests for associations between a quantitative trait and one or more covariates of interest, and estimates familial variance components from the given familial correlations. In our study, the trait of interest was age at diabetes diagnosis, while the main covariate of interest was the number of type 2 diabetes risk alleles. As one of the parameters for the ASSOC program we set the family effect option to “true”, thus including the random nuclear family effect as an additional term in the regression model. Relationships between family members were fully established for most pedigrees. In some of the large pedigrees we included parents and relatives that had no data for the analyses, but were used by the program to accurately connect all related individuals. Singletons, and 5 UK family members for whom we could not establish how they were related to other members of their pedigrees, were automatically treated as one-person pedigrees and required no special handling in the model.

We assessed the effect of each risk variant on the age at diagnosis individually, jointly and by using an allele counting method to assign a genetic risk score to each patient (the sum of the number of risk alleles a person carries). The allele counting method assumed equal and additive effects of the individual variants. We repeated this analysis using a weighted allele approach, where the genetic score was based on the previously reported odds ratios for type 2 diabetes (obtained from a recent review [22]). For each patient we first calculated the sum across SNPs of the number of risk alleles at each SNP multiplied by the log of the odds ratio (OR) for that SNP (i.e. genotypes were coded as 0, log(OR), 2xlog(OR), rather than 0, 1, 2 in the allele count model). To obtain a rescaled "weighted allele count" score, we multiplied each log(OR) score by 30 (maximum number of risk alleles), and divided the product by 1.81, the sum of the 15 risk homozygote log(OR) weights .

Although family relationships were fully accounted for in the analyses, it is possible that the results could have been affected by the skewed allele distributions. Therefore, we analysed the effects of both individual SNPs and the combined genetic score on age at diagnosis in 203 unrelated probands, using the youngest individual from each pedigree (**Supplementary Table 1**).

We used StataSE v10.0 for Windows (StataCorp LP, Brownsville, TX, USA) to generate adjusted age at diagnosis, using the 'predict' function after running linear regression that fitted family id, study, sex, age at study, presence of intrauterine hyperglycemia and mutation position in the same regression model. This enabled us to use the full dataset, with adjusted ages at diagnosis, for linear regression and cumulative diabetes incidence analyses (**Figures 1 and 2**), rather than a much smaller sample of unrelated singletons and phenotypically homogeneous individuals. All figures were generated using SigmaPlot (Systat Software Inc., CA, USA). Power calculations were performed using QUANTO power calculator, version 1.2.4 [24].

Results

The analyses included 410 diabetic *HNF1A* mutation carriers from 203 families who were successfully genotyped for all 15 type 2 diabetes risk variants.

We confirmed the strong effects of age at study, mutation position and intrauterine hyperglycemia on the severity of *HNF1A* diabetes clinical presentation (**Table 2**). These associations were independent of the polygenic risk factors (**Table 2**). On average, patients were diagnosed 5.1 years earlier if the mother was diabetic during pregnancy ($P = 1.6x10^{-10}$), 5.2 years earlier if the mutation affected at least two *HNF1A* isoforms ($P = 1.8x10^{-2}$), and 0.3 years later for every additional year of their age at study ($P = 4.7x10^{-38}$). In addition we observed a strong effect of sex in our data, where females were diagnosed 3.0 years earlier than males ($P = 6.0x10^{-4}$), but there was no association with BMI (available for 305 subjects; $P = 0.99$).

We included those variables that had individual effect on age at diagnosis (i.e. all of the above apart from BMI) as covariates in the individual and joint SNP models, to reduce the remaining variance in the age at diagnosis and, therefore, increase our power to detect the effect of polygenic modifiers. We repeated these analyses excluding age at study, to make sure that its strong association with age at diagnosis did not drive the SNP association (**Supplementary table 2**). As expected, the results were not statistically significantly different to the fully adjusted model (all t-test $P > 0.32$). Although for some of the SNPs the effects on age at diagnosis were slightly stronger when age at study was excluded, the standard errors were larger, resulting in similar overall *P*-values.

Individual type 2 diabetes risk variants were not strongly associated with the age at diagnosis, as shown in **Table 3**. However, of the 15 variants, 11 risk alleles for type 2 diabetes in the unadjusted analyses, and 10 in the adjusted analyses, were associated with reduced age at diagnosis, in a direction consistent with polygenic studies. When we included all 15 variants in the regression model, there was borderline evidence of an overall joint effect on the age at diagnosis (*P* =

0.062). The 15 variants explain 6.4% of the total proportion of diagnosis age variance, whilst the non-polygenic factors (sex, age at study, mutation position, and presence of intrauterine hyperglycemia) explain 37.9%; combining these together, they explain 42.1% of the total variance in the *HNF1A*-MODY age at diagnosis in these families.

We then generated a single genetic risk score representing the combined genetic susceptibility for type 2 diabetes (**Table 3**). In the allele count model, each additional risk allele was associated with 0.35 years reduction in age at diagnosis (P = 0.005). The association strength was weaker when we used unrelated probands (0.28 years earlier age at diagnosis per one additional risk allele; P = 0.094; **Supplementary Table 1**), which most probably reflects reduced power. The correlation between the decreasing age at diagnosis and the increasing number of risk alleles appears to be linear for the full dataset of 410 patients (**Figure 1A**). **Figure 1B** presents the results for 203 unrelated probands only. Looking at the impact of risk alleles on the cumulative incidence of diabetes, the effect was most noticeable around age 30, where diabetes developed in 80% of *HNF1A* mutation carriers with 9-14 polygenic risk alleles, compared to 93% with 17-22 risk alleles (**Figure 2**).

The weighted allele score yielded similar results to the allele count model (P = 0.005). Stratified analysis showed that the impact of the allele count score was of similar magnitude in the two cohorts individually, with all t-test P-values > 0.1 (**Supplementary table 3**).

Discussion

We have shown that type 2 diabetes risk variants of modest effect sizes when combined are associated with a reduced age at diagnosis in monogenic *HNF1A* diabetes. This association is independent of other genetic and environmental modifiers, namely the *HNF1A* mutation position, age at study, sex, and mother's diabetes status during pregnancy. Thus, this is one of the first studies to demonstrate that clinical characteristics of a monogenic disease can be influenced by common variants that predispose to the polygenic form of that disease. To our knowledge, only two other studies, of breast cancer [25] and Alzheimer's disease [26], have identified polygenic variants that act as modifiers of disease onset age.

In support of previous findings, an increase in the age of patients at the time of genetic testing is strongly associated with an older age at diabetes diagnosis. It is not known if this represents an earlier diagnosis as a result of the increasing awareness of diabetes in the family by their physicians, or a genuine decrease in age of onset in succeeding generations. The former is likely to be a large contributor. In addition, there is strong evidence that the age at diagnosis is affected by genetic factors. We confirm previous findings by Harries *et al.* [11] and Bellanné-Chantelot [12] that patients with mutations affecting at least two of the three known *HNF1A* isoforms were diagnosed earlier than patients with mutations affecting only one *HNF1A* isoform.

In our study we provide evidence for additional genetic modifiers, the robustly replicated type 2 diabetes risk variants. Combining the effect of the variants by adding up the total number of risk alleles carried, each additional risk allele was associated with 0.49 and 0.35 year earlier age at diagnosis, in the unadjusted and adjusted models, respectively. Most of the genetic variants predisposing to type 2 diabetes act through reducing β-cell function, rather than increasing insulin resistance. This is true of the three risk variants with strongest effects observed in this study, the SNPs in the *HNF1B*, *SLC30A8* and *CDKAL1*

genes. It is possible that they interact with the β-cell dysfunction resulting from the *HNF1A* mutation, leading to an increased rate of β-cell destruction and, therefore, earlier onset of diabetes. Furthermore, mutations of *HNF1B*, also known as *TCF2*, are another known cause of MODY, accounting for about 2% of cases [27].

This study does have limitations. Although we included 410 subjects, one of the largest cohorts of *HNF1A* patients ever reported, some simple power calculations suggest that we were still under-powered to detect the impact of individual loci. For example, we had only 29% power to detect an individual SNP explaining 1% of the variation in age at diagnosis (and this is assuming independence of the individuals in the study) at $P<0.01$; in singleton-only analysis the power was 13%. These patients were not studied prospectively and, therefore, the age at diagnosis does not accurately reflect the age at onset of diabetes. We would anticipate that if age of onset was studied using prospective data, the impact of these type 2 diabetes loci would be greater.

In conclusion, we show that type 2 diabetes risk variants of modest effect sizes act as an additional modifier of age at diagnosis in *HNF1A*-MODY. This is one of the first studies to demonstrate that common variants associated with a polygenic disease can also influence clinical characteristics of a monogenic form of the disease.

Acknowledgements

MNW is a Vandervell Foundation Research Fellow. BS and SE are employed as core staff within the NIHR-funded Peninsula Clinical Research Facility. Parts of this study were supported by the University of Bergen, Haukeland University Hospital, Helse Vest, Innovest, and the Research Council of Norway (to PRN). We thank all MODY patients and nurses who took part and contributed to this study.

Table 1. Characteristics of the 410 *HNF1A*-MODY patients included in the analyses.

	UK	Norway
Families / singletons, N	140 / 87	63 / 41
Examined individuals, N	298	112
Males, N (%)	100 (33.6)	41 (36.6)
Average number of individuals in non-singleton families (range)	3.0 (2-9)	3.1 (2-8)
Number of individuals with *HNF1A* mutations affecting		
Isoform A only (exons 8-10)	30 (10.1%)	3 (2.7%)
Isoforms A and B only (exon 7)	22 (7.4%)	4 (3.6%)
Isoforms A, B and C (exons 1-6)	246 (82.5%)	105 (93.7%)
Age at study, years *	37.1 $\pm$ 17.0 (8-87)	33.4 + 17.6 (6-73)
Age at diabetes diagnosis, years *	21.9 $\pm$ 11.2 (4-70)	20.3 + 10.0 (6-60)
BMI, kg/m^2 *†	24.1 $\pm$ 4.1 (15.9-50.7)	23.9 + 3.7 (15.8-33.6)

* Data are presented as means $\pm$ SD (range).

† BMI was only available for 224 and 81 individuals in UK and Norway studies, respectively.

Table 2. Results of regression analyses of non-polygenic factors on the age at diabetes diagnosis in 410 *HNF1A*-MODY patients, with and without the inclusion of the polygenic risk score in the regression model. All effect sizes are in years. Genetic score is the number of risk alleles, carried by each patient, from the 15 type 2 diabetes susceptibility variants.

	Without genetic score			**With genetic score**		
	Effect size	**Std Error**	***P*-value**	**Effect size**	**Std Error**	***P*-value**
Study (Norway=0, UK=1)	1.31	0.99	0.18	1.44	0.97	0.14
Sex (effect w.r.t. females)	-2.97	0.86	6.0×10^{-4}	-2.93	0.85	5.7×10^{-4}
BMI, kg/m^2 (per unit increase) *	0.0009	0.13	0.99	0.033	0.13	0.80
Presence of intrauterine hyperglycaemia	-5.06	0.79	1.6×10^{-10}	-4.86	0.79	6.5×10^{-10}
Position of *HNF1A* mutation †	-5.22	2.21	1.8×10^{-2}	-5.67	2.18	9.4×10^{-3}
Age at study (per year increase)	0.29	0.02	4.7×10^{-38}	0.29	0.02	1.5×10^{-37}

w.r.t. = with respect to.

* BMI was only available for 305 individuals.

†The position of the mutation has been dichotomised into those affecting exons 8-10 (isomer A only; N=33) versus those affecting exons 1-7 (N=377). The age at diagnosis is lower for patients with mutations affecting exons 1-7.

Table 3. Effects of individual type 2 diabetes risk variants and the combined genetic scores on the age at diabetes diagnosis in 410 *HNF1A*-MODY patients. All effect sizes are in years change of age at diagnosis per risk allele. The 410 patients were successfully genotyped for all 15 SNPs that were included in the combined genetic scores. All analyses took into account family relationships and included a random family effect in the regression model. Individual SNP effects are based on risk allele count method. *P* values are unadjusted for multiple testing. Results are presented in order of the adjusted effect sizes.

		Unadjusted results			Adjusted results ‡		
		Effect Size	Std Error	*P*-value	Effect Size	Std Error	*P*-value
Individual SNP effects							
Gene region	SNP						
HNF1B (TCF2) *	rs757210 / rs4430796	-1.85	0.58	0.0014	-1.07	0.43	0.014
SLC30A8	rs13266634	-1.07	0.64	0.095	-0.90	0.50	0.070
CDKAL1 *	rs10946398 / rs7754840	-1.22	0.59	0.038	-0.87	0.46	0.059
TCF7L2	rs7903146	-0.55	0.63	0.39	-0.65	0.46	0.16
ADAMTS9	rs4607103	0.27	0.69	0.70	-0.59	0.51	0.25
TSPAN8	rs7961581	-0.97	0.63	0.13	-0.53	0.44	0.22
JAZF1 †	rs864745	-0.45	0.72	0.53	-0.46	0.53	0.38
FTO *	rs8050136 / rs9939609	-0.30	0.63	0.63	-0.42	0.47	0.37
KCNJ11	rs5219	-0.15	0.63	0.82	-0.34	0.50	0.50
CDKN2A/2B	rs10811661	-0.89	0.87	0.31	-0.25	0.65	0.70
WFS1	rs10010131	-0.08	0.61	0.89	-0.21	0.46	0.65
CDC123	rs12779790	0.83	0.75	0.27	0.07	0.55	0.91
HHEX/IDE *	rs1111875 / rs5015480	-0.27	0.61	0.66	0.19	0.44	0.66
PPARG	rs1801282	-1.46	0.99	0.14	0.36	0.76	0.64
IGF2BP2	rs4402960	0.45	0.64	0.48	0.43	0.47	0.36
THADA	rs7578597	0.50	1.00	0.62	0.55	0.78	0.48
NOTCH2 †	rs2934381	1.31	1.32	0.32	0.82	1.00	0.41
Combined SNP effect							
Allele count score		-0.49	0.17	0.0043	-0.35	0.13	0.0051
Weighted score (log odds)		-0.49	0.15	0.0013	-0.33	0.12	0.0046

*At 4 loci different SNPs, representing the same signal, were genotyped by the two study centres, in which case they are shown as UK / Norway SNPs.

† Results for *JAZF1* and *NOTCH2* SNPs were available only for UK samples (N=296 and 297, respectively).

‡ Adjusted results include study, sex, age at study, presence of intrauterine hyperglycaemia, and mutation position (2 groups, according to exon affected, 1-7 or 8-10) as covariates in the regression model.

Figure 1. Mean age at diabetes diagnosis (black triangles) and frequency (bars) of *HNF1A*-MODY patients at each number of the type 2 diabetes risk alleles carried. Only individuals genotyped for all 15 variants are included. **A =** full dataset of 410 patients; **B =** 203 unrelated probands (youngest family members). Ages at diagnosis were adjusted for family (**A** only), study, sex, age at study, exposure to mother's hyperglycemia *in utero*, and position of *HNF1A* mutation. Black lines are fitted age at diagnosis linear regression lines. Both y-axis are on the same scale in panels **A** and **B**.

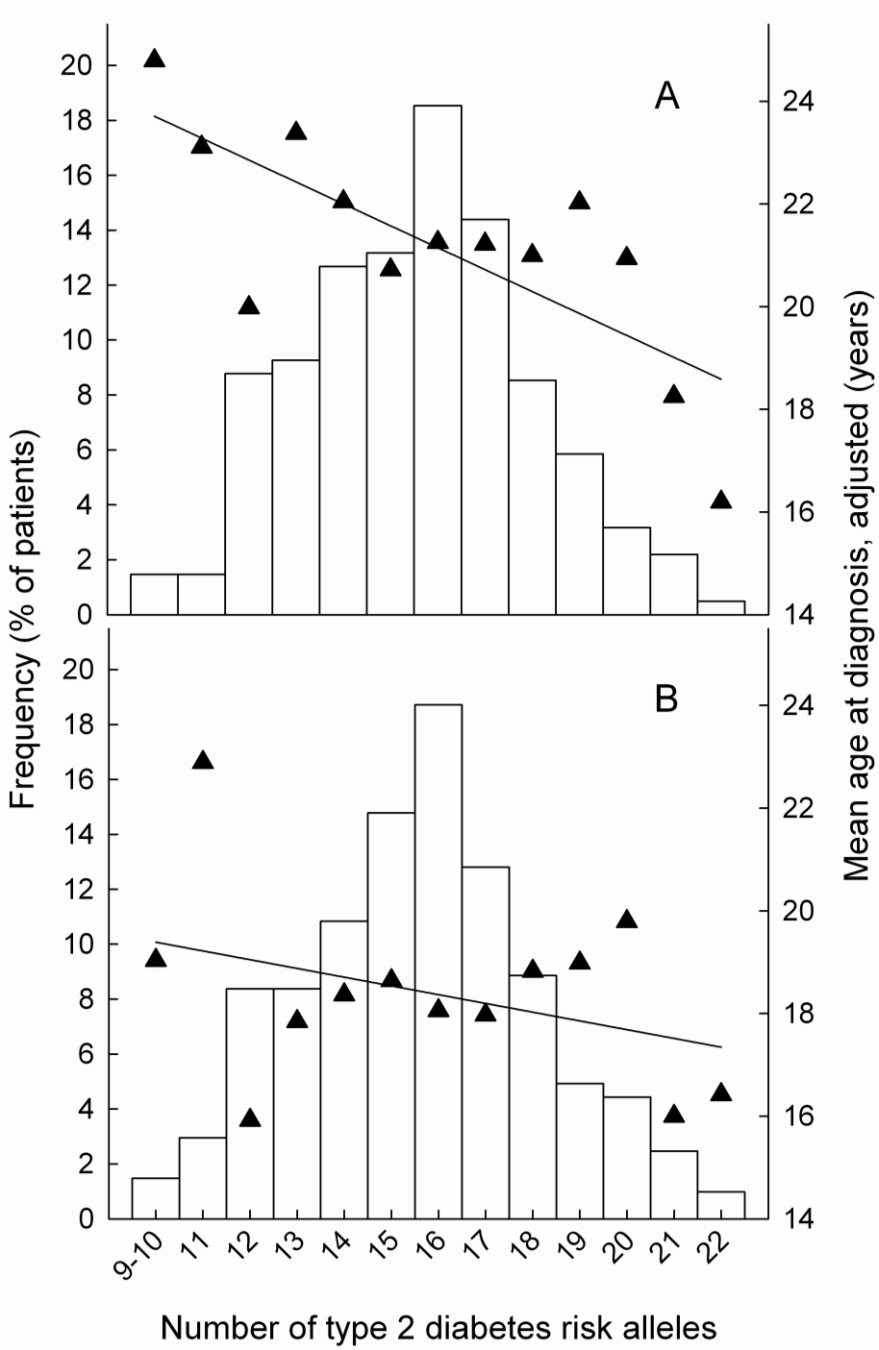
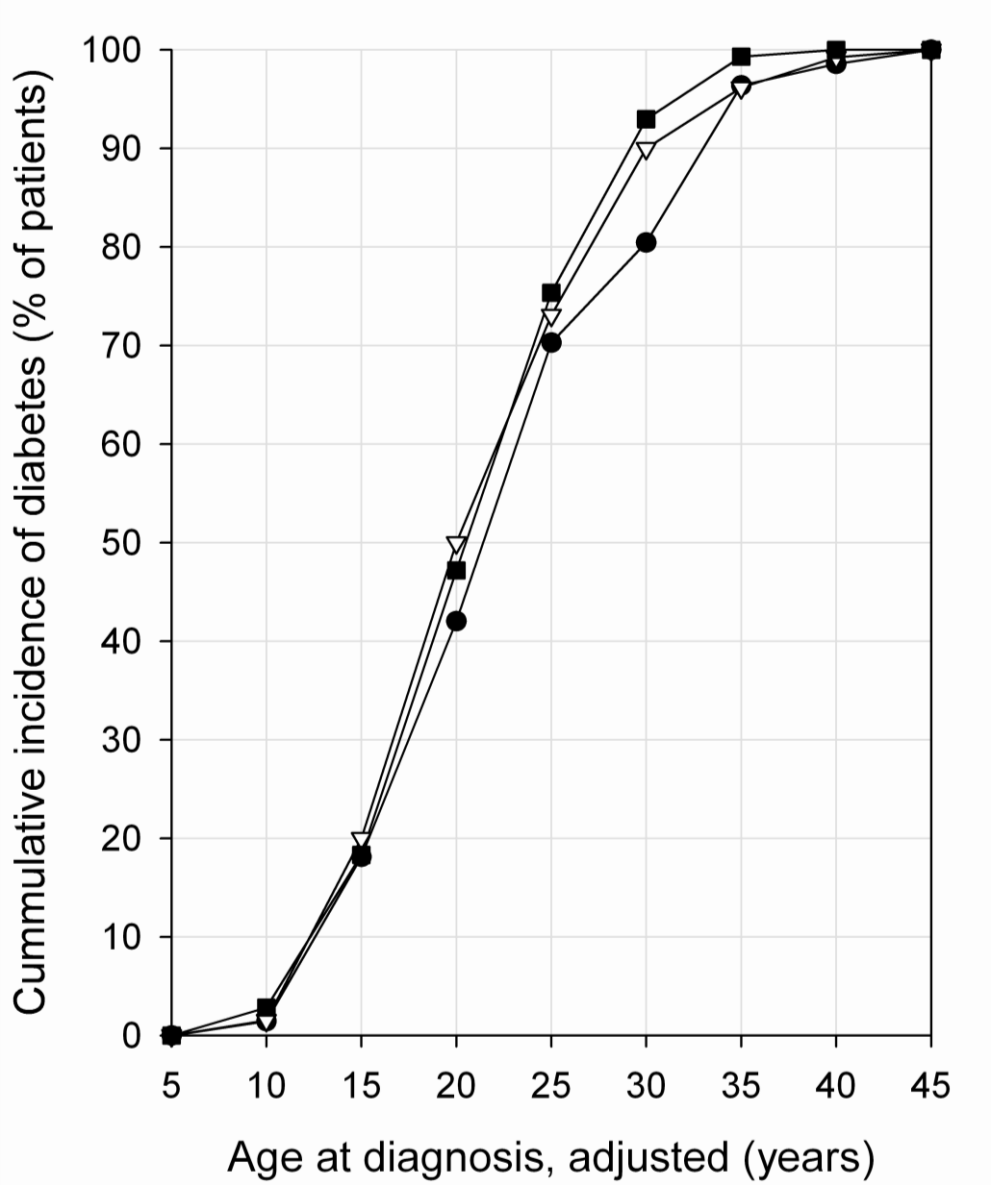


Figure 2. Cumulative incidence of diabetes in 410 *HNF1A*-MODY patients, by type 2 diabetes risk allele count category. Black circles = 9-14 risk alleles, N=138; white triangles = 15-16 risk alleles, N=130; black squares = 17-22 risk alleles, N=142. Only individuals genotyped for all 15 variants are included. The ages at diabetes diagnosis ware adjusted for family, study, sex, age at study, exposure to mother's hyperglycemia *in utero*, and position of *HNF1A* mutation.



References

1. Murphy, R., Ellard, S. & Hattersley, A. T. Clinical implications of a molecular genetic classification of monogenic beta-cell diabetes. Nature Clinical Practice Endocrinology & Metabolism 4, 200-213 (2008).

2. Hattersley, A., Bruining, J., Shield, J., Njolstad, P. & Donaghue, K. ISPAD Clinical Practice Consensus Guidelines 2006-2007 - The diagnosis and management of monogenic diabetes in children. Pediatric Diabetes 7, 352-360 (2006).

3. Frayling, T. et al. Mutations in the Hepatocyte Nuclear Factor 1 Alpha gene are a common cause of maturity-onset diabetes of the young in the United Kingdom. Diabetes 46, 720-725 (1997).

4. Bjorkhaug, L. et al. Hepatocyte nuclear factor-1 alpha gene mutations and diabetes in Norway. Journal Of Clinical Endocrinology And Metabolism 88, 920-931 (2003).

5. Byrne, M. M. et al. Altered insulin secretory responses to glucose in diabetic and nondiabetic subjects with mutations in the diabetes susceptibility gene MODY3 on Chromosome 12. Diabetes 45, 1503-1510 (1996).

6. Hattersley, A. T. Maturity-onset diabetes of the young: clinical heterogeneity explained by genetic heterogeneity. Diabetic Medicine 15, 15-24 (1998).

7. Frayling, T. M. et al. Beta-cell genes and diabetes: molecular and clinical characterization of mutations in transcription factors. Diabetes 50, S94-100. (2001).

8. Shepherd, M., Hattersley, A. T. & Sparkes, A. Genetic testing in maturity onset diabetes of the young (MODY): a new challenge for the diabetic clinic. Practical Diabetes 18, 16-21 (2001).

9. Stride, A. et al. Intrauterine Hyperglycemia Is Associated With an Earlier Diagnosis of Diabetes in HNF-1alpha Gene Mutation Carriers. Diabetes Care 25, 2287-91 (2002).

10. Klupa, T. et al. Determinants of the development of diabetes (maturity-onset diabetes of the young-3) in carriers of HNF-1 alpha mutations - Evidence for parent-of-origin effect. Diabetes Care 25, 2292-2301 (2002).

11. Harries, L. W. et al. Isomers of the TCF1 gene encoding hepatocyte nuclear factor-1 alpha show differential expression in the pancreas and define the relationship between mutation position and clinical phenotype in monogenic diabetes. Hum. Mol. Genet. 15, 2216-2224 (2006).

12. Bellanne-Chantelot, C. et al. The type and the position of HNF1A mutation modulate age at diagnosis of diabetes in patients with maturity-onset diabetes of the young (MODY)-3. Diabetes 57, 503-508 (2008).

13. Kim, S. H. et al. Genetic modifiers of the age at diagnosis of diabetes (MODY3) in carriers of hepatocyte nuclear factor-1 alpha mutations map to chromosomes 5p15, 9q22, and 14q24. Diabetes 52, 2182-2186 (2003).

14. Tack, C. J. J., Ellard, S. & Hattersley, A. T. A severe clinical phenotype results from the co-inheritance of type 2 susceptibility genes and a hepatocyte nuclear factor-1 alpha mutation. Diabetes Care 23, 424-425 (2000).

15. Saxena, R. et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 316, 1331-1336 (2007).

16. Scott, L. J. et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 316, 1341-1345 (2007).

17. Sladek, R. et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 445, 881 (2007).

18. Zeggini, E. et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nature Genetics 40, 638-645 (2008).

19. Zeggini, E. et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 316, 1336-1341 (2007).

20. Hertel, J. K. et al. Genetic analysis of recently identified type 2 diabetes loci in 1,638 unselected patients with type 2 diabetes and 1,858 control participants from a Norwegian population-based cohort (the HUNT study). Diabetologia 51, 971-977 (2008).

21. Frayling, T. M. Genome-wide association studies provide new insights into type 2 diabetes aetiology. Nature Reviews Genetics 8, 657-662 (2007).

22. Prokopenko, I., McCarthy, M. I. & Lindgren, C. M. Type 2 diabetes: new genes, new understanding. Trends in Genetics 24, 613 (2008).

23. S.A.G.E. (2008).

24. Gauderman, W. & Morrison, J. (2006).

25. Antoniou, A. C. et al. RADS1 135G -> C modifies breast cancer risk among BRCA2 mutation carriers: Results from a combined analysis of 19 studies. American Journal Of Human Genetics 81, 1186-1200 (2007).

26. Bertram, L. et al. Genome-wide Association Analysis Reveals Putative Alzheimer's Disease Susceptibility Loci in Addition to APOE. American Journal Of Human Genetics 83, 623-632 (2008).

27. McCarthy, M. I. & Hattersley, A. T. Learning From Molecular Genetics: Novel Insights Arising From the Definition of Genes for Monogenic and Type 2 Diabetes. Diabetes 57, 2889-2898 (2008).

SUPPLEMENTARY INFORMATION

Supplementary table 1. Effects of individual type 2 diabetes risk variants and the combined genetic scores on the age at diabetes diagnosis in 203 *HNF1A*-MODY unrelated probands (youngest family members). All effect sizes are in years change of age at diagnosis per risk allele. The 203 patients were successfully genotyped for all 15 SNPs that were included in the combined genetic score. Individual SNP effects are based on risk allele count method. *P* values are unadjusted for multiple testing. Results are presented in order of Table 3 adjusted results.

		Unadjusted analysis			Adjusted analysis‡		
		Effect Size	Std Error	*P*-value	Effect Size	Std Error	*P*-value
Individual SNP effects							
Gene region	SNP						
HNF1B (TCF2) *	rs757210 / rs4430796	-0.64	0.74	0.39	-0.37	0.58	0.52
SLC30A8	rs13266634	-2.09	0.82	0.011	-1.29	0.66	0.052
CDKAL1 *	rs10946398 / rs7754840	-0.93	0.79	0.24	-0.64	0.62	0.30
TCF7L2	rs7903146	0.76	0.85	0.37	-0.46	0.69	0.51
ADAMTS9	rs4607103	0.13	0.89	0.89	-0.89	0.70	0.20
TSPAN8	rs7961581	-1.15	0.82	0.16	-0.79	0.64	0.22
JAZF1 †	rs864745	-1.23	0.93	0.19	-1.31	0.74	0.078
FTO *	rs8050136 / rs9939609	0.05	0.81	0.95	-0.38	0.64	0.55
KCNJ11	rs5219	-0.33	0.87	0.71	0.17	0.68	0.81
CDKN2A/2B	rs10811661	-0.54	1.13	0.63	-0.42	0.88	0.64
WFS1	rs10010131	0.35	0.79	0.66	0.05	0.62	0.94
CDC123	rs12779790	0.73	0.96	0.45	0.42	0.76	0.58
HHEX/IDE *	rs1111875 / rs5015480	-0.04	0.78	0.96	-0.33	0.61	0.59
PPARG	rs1801282	-0.99	1.37	0.47	0.46	1.08	0.67
IGF2BP2	rs4402960	0.17	0.83	0.84	0.28	0.66	0.67
THADA	rs7578597	0.65	1.44	0.65	0.67	1.13	0.55
NOTCH2 †	rs2934381	2.41	1.70	0.16	1.70	1.33	0.20
Combined SNP effect							
Allele count score		-0.26	0.21	0.23	-0.28	0.17	0.094

*At 4 loci different SNPs, representing the same signal, were genotyped by the two study centres, in which case they are shown as UK / Norway SNPs.

† Results for *JAZF1* and *NOTCH2* variants were available only for UK samples (N=140 and 139, respectively).

‡ Adjusted results include study, sex, age at study, presence of intrauterine hyperglycaemia, and mutation position (2 groups, according to exon affected, 1-7 or 8-10) as covariates in the regression model.

Supplementary table 2. Effects of individual type 2 diabetes risk variants and the combined genetic scores on the age at diabetes diagnosis in 410 *HNF1A*-MODY patients, in an adjusted analysis excluding age at study. All effect sizes are in years change of age at diagnosis per risk allele. The 410 patients were successfully genotyped for all 15 SNPs that were included in the combined genetic score. All analyses took into account full family relationships and included a random family effect in the regression model. Individual SNP effects are based on risk allele count method. *P* values are unadjusted for multiple testing. Results are presented in order of the Table 3 adjusted effect sizes.

		Adjusted results excluding age at study			T-test *P* for difference between results with age at study
		Effect Size	Std Error	*P*-value	
Individual SNP effects					
Gene region	SNP				
HNF1B (TCF2) *	rs757210 / rs4430796	-1.19	0.52	0.023	0.86
SLC30A8	rs13266634	-1.00	0.59	0.092	0.90
CDKAL1 *	rs10946398 / rs7754840	-1.36	0.52	0.0095	0.48
TCF7L2	rs7903146	-0.34	0.57	0.54	0.67
ADAMTS9	rs4607103	0.07	0.62	0.91	0.41
TSPAN8	rs7961581	-0.79	0.56	0.16	0.72
JAZF1 †	rs864745 (n=296)	-0.47	0.64	0.46	0.99
FTO *	rs8050136 / rs9939609	-0.60	0.57	0.29	0.81
KCNJ11	rs5219	0.20	0.56	0.72	0.47
CDKN2A/2B	rs10811661	-0.79	0.78	0.31	0.59
WFS1	rs10010131	-0.36	0.55	0.51	0.83
CDC123	rs12779790	0.22	0.67	0.74	0.86
HHEX/IDE *	rs1111875 / rs5015480	-0.36	0.53	0.50	0.42
PPARG	rs1801282	-0.81	0.90	0.37	0.32
IGF2BP2	rs4402960	0.72	0.58	0.22	0.70
THADA	rs7578597	0.70	0.93	0.45	0.90
NOTCH2 †	rs2934381 (n=297)	1.31	1.15	0.26	0.75
Combined SNP effect					
Allele count score, adjusted		-0.40	0.15	0.0072	0.80

*At 4 loci different SNPs, representing the same signal, were genotyped by the two study centres, in which case they are shown as UK / Norway SNPs.

† Results for *JAZF1* and *NOTCH2* SNPs were available only for UK samples (N=296 and 297, respectively).

Supplementary table 3. Stratified-by-study analysis of non-polygenic factors, individual type 2 diabetes risk variants, and the combined genetic scores on the age at diabetes diagnosis in 410 *HNF1A*-MODY patients. All analyses took into account full family relationships and, except for those marked ‡, included a random family effect in the regression model All effect sizes are in years change of age at diagnosis per risk allele. Individual SNP effects are based on risk allele count method adjusted for sex, age at study, presence of intrauterine hyperglycaemia, and mutation position. All *P* values are unadjusted for multiple testing. Individual SNP results are in the same order as in Table 3.

		UK (N=298)			Norway (N=112)			T-test *P* for difference
		Effect size	Std Error	*P*-value	Effect size	Std Error	*P*-value	
Non-polygenic factors (without genetic score)								
Sex (effect w.r.t. females)		-3.63	1.06	$6.3x10^{-4}$	-1.48	1.49	0.32	0.27
Presence of intrauterine hyperglycaemia		-5.27	0.96	$4.1x10^{-8}$	-4.58	1.29	$4.0x10^{-4}$	0.69
Position of *HNF1A* mutation		-5.27	2.34	0.024	-3.11	5.57	0.58	0.67
Age at study (per year increase)		0.29	0.03	$3.3x10^{-26}$	0.30	0.04	$1.1x10^{-13}$	0.86
Individual SNP effects								
Gene region	SNP							
HNF1B *	rs757210 / rs4430796	-0.76	0.57	0.18	-1.53	0.65	0.018	0.45
SLC30A8	rs13266634	-0.74	0.60	0.22	-0.83	0.28	0.003	0.93
CDKAL1 *	rs10946398 / rs7754840	-1.18	0.53	0.026	-0.08	0.92	0.93	0.29
TCF7L2	rs7903146	-0.39	0.54	0.48	-1.19	0.55	0.031	0.40
ADAMTS9 ‡	rs4607103	-0.65	0.61	0.29	-0.70	0.86	0.42	0.96
JAZF1 †	rs864745	-0.46	0.53	0.38	*NA*	*NA*	*NA*	*NA*
FTO *	rs8050136 / rs9939609	-0.47	0.57	0.41	-0.24	0.63	0.70	0.81
TSPAN8 ‡	rs7961581	-0.47	0.52	0.37	-0.56	0.82	0.49	0.93
CDKN2A/2B	rs10811661	0.38	0.74	0.61	-1.16	0.40	0.004	0.21
KCNJ11	rs5219	-0.43	0.59	0.46	0.22	0.94	0.82	0.56
WFS1	rs10010131	-0.20	0.55	0.71	-0.56	0.80	0.48	0.72
CDC123 ‡	rs12779790	-0.09	0.64	0.89	0.13	0.96	0.89	0.85
HHEX-IDE *	rs1111875 / rs5015480	-0.002	0.54	1.00	0.34	0.23	0.15	0.70
THADA	rs7578597	0.69	0.87	0.43	-0.45	1.97	0.82	0.54
IGF2BP2	rs4402960	-0.09	0.55	0.88	1.46	0.90	0.10	0.14
PPARG	rs1801282	0.08	0.92	0.93	1.04	1.25	0.41	0.57
NOTCH2 †	rs2934381	0.82	1.00	0.41	*NA*	*NA*	*NA*	*NA*
Combined SNP effects								
Allele count score, unadjusted		-0.43	0.20	0.036	-0.73	0.32	0.021	0.43
Allele count score, adjusted		-0.35	0.15	0.020	-0.43	0.24	0.069	0.78

*At 4 loci different SNPs, representing the same signal, were genotyped by the two study centres, in which case they are shown as UK / Norway SNPs.

† Results for *JAZF1* and *NOTCH2* SNPs were available only for UK samples (N=296 and 297, respectively).

‡ Because of the small sample size relative to the number of covariates, the random family effect could not be fully fitted in the regression model for the Norwegian sample. Therefore, for these 3 variants, this term was excluded from the model in both studies (in the Exeter sample the results with and without the random family term were nearly identical).

CHAPTER 6: DISCUSSION AND CONCLUSIONS

The field of complex trait genetics has progressed substantially throughout the duration of research that is included in this thesis. Owing to the success of genome-wide association (GWA) studies, the number of common variants robustly associated with common human traits and diseases has substantially increased (listed in the NHGRI GWAS catalogue, http://www.genome.gov/26525384). Genome-wide association studies are based on two major advancements in the field of genetics: the dissection of human genome into distinct linkage disequilibrium (LD) blocks, catalogued by the HapMap project [1, 2], and the technological advances in high-throughput genotyping. Rather than testing one SNP at a time, it is possible to test 300,000 – 1 million SNPs simultaneously, and capture up to ~80% of the common (>5% MAF) genetic variation owing to the underlying LD structure. Following some essential GWAS procedures such as selection of subjects with well characterized phenotypes, as well as stringent quality checks of samples and genotypes for artifacts of DNA preparation and genotyping process, has become standard practice among many research groups. Importantly, though, as demonstrated in Chapter 2, there has been a realization that, rather than trying to replicate GWA findings in additional independent cohorts, a more powerful approach for discovering the contributing common variants is to combine efforts and, since it is often not possible to share raw genotype data, perform a meta-analysis of a much larger set of samples. The necessity for organized data sharing and agreements has led to the formation of many trait-specific consortia, discussed below. In such meta-analyses there are usually no additional, independent, suitably sized cohorts left to confirm positive associations, so only those that reach a conservative statistical significance level, usually $P<5x10^{-8}$, are reported.

Because different genotyping platforms include different subsets of HapMap SNPs, initially it was difficult for many collaborating studies to exchange or combine their data, unless they used the same platform. However, it is now possible to use genotype imputation programs that combine the information from the known genotypes and LD structure around them to infer missing genotypes of nearly 2.5 million HapMap SNPs [3]. Although sharing of individual level data is

often not possible because of ethical restrictions, Chapter 3 describes a combined analysis of association statistics (effect sizes and p-values) of these imputed SNPs, a meta-analysis approach that is now being adopted for many other traits and diseases.

Common variants have small effects and have explained small proportion of the heritable component

As Chapters 2 and 3 demonstrate, many of the associated variants have modest effect sizes and can only be identified once a GWA study has large enough sample size and is sufficiently powered to detect them. In many cases, including height, the model polygenic trait used here, only a small fraction of the heritable component is accounted for. This may be because GWA studies are not designed to detect rare and structural variants, gene-by-gene and gene-by-environment interactions [4], or epigenetic effects [5], all of which have been suggested to contribute to the heritability of the trait (although epigenetic effects by definition do not, as explained below). Unsurprisingly, this perceived lack of success has caused some to start questioning the usefulness of performing ever-larger GWA studies and investing resources into what appears to be diminishing genetic returns, since one will either find associations with random variants across the genome, or the effect sizes will be so small that they become irrelevant [6].

This thesis tackles some of these questions by performing the largest GWA study to date, as part of the Genetic Investigation of Anthropometric Traits (GIANT) consortium. The Chapter 2 study [7] was published in parallel with another three large height GWA studies [8-10]. It was apparent that each study was in fact underpowered to detect many of the associations it reported (at the genome-wide significance level), suggesting that more could be discovered if the sample sizes are increased. This led to the formation of the GIANT consortium and data sharing between many research groups. Crucial to this was the availability of imputation methods, which allowed for meta-analysis of summary statistics of 2.5 million HapMap SNPs polymorphic in populations of European origin, regardless of which

one of the many available genotyping platforms was used by each individual group.

In addition to GIANT, several other international consortia have been established over the past couple of years to investigate common diseases and related intermediate traits. These include the Diabetes Genetics Replication and Meta-analysis (DIAGRAM) consortium, a collaboration between DGI, FUSION and WTCCC-T2D groups that were part of the initial wave of successful GWA studies [11-15]. The first meta-analysis by DIAGRAM provided evidence for 6 additional type 2 diabetes loci [16], and since then the consortium has been expanded to include additional cohorts. Other consortia have been established to look at intermediate, often quantitative, disease traits. A good example is the Meta-Analyses of Glucose and Insulin Related Traits Consortium (MAGIC), which has investigated diabetes-related traits in non-diabetic individuals. The initial analysis identified *MTNR1B* as a novel fasting glucose and type 2 diabetes locus [17], and was followed by an even larger, imputation facilitated meta-analysis that identified nine novel loci associated ($P<5x10^{-8}$) with fasting glucose and one with fasting insulin, five of which were also associated, in the expected direction, with type 2 diabetes [18]. Interestingly, one of the five signals, in *ADCY5* gene, was also associated with birth weight in the Early Growth Genetics (EGG) Consortium meta-analysis of six GWA studies with imputed genotypes [19].

Despite the increasing number of variants identified, the percentage of the explained genetic variance still remains relatively small for most common traits studied so far. The GIANT consortium height study, presented in Chapter 3, with a sample size of over 130,000 individuals showed that some of the genetic component of the 'missing heritability' can indeed be found among additional common variants of small effect sizes identified by expanding the sample size and thus increasing power to detect them. However, the proportion of the genetic variance explained for height is still only ~20%, and up to 50% if all variants of similar effect sizes are identified. The study found no evidence that epistasis or non-additive effects are associated with height. Although it is often suggested that

these and epigenetic effects might explain additional heritable variance, these factors do not contribute to genetic variance because they are not part of the heritability calculations (ratio of additive genetic factors to total phenotypic variation) [20]. The question then remains, where is the 'missing' genotypic heritability, and what can our knowledge of the associated variants be used for?

Insights into genetic architecture and biological mechanisms

Although it is possible that the associated variants are tagging rare, as yet unidentified, causal variants of larger effects, the GIANT height study does not rule out the common variant / common disease hypothesis [21]. A recent study suggested that most of the heritability is not missing, but has not yet been detected because most of it can be explained by large number of common variants with effect sizes too small to reach statistical significance in the GWA studies [22]. It is likely that insertions/deletion or large structural variants contribute to common traits. Although these are not represented directly on the genotyping platforms, they can be assessed indirectly because many of the SNPs on the chips are tags for a large number of known copy-number variants (CNVs). However, the GIANT height study did not identify any strong associations, and this has mainly been the case for common complex diseases. Rare deletions and/or duplications have, so far, only been shown to associate with schizophrenia [23, 24] and autism [25], while some of the signals associated with autoimmune diseases and type 2 diabetes do tag common CNVs [26].

Chapters 2 and 3 show that, even though many common variants for height have been identified, they are not randomly distributed across the genome. Rather, they implicate functionally relevant genes and pathways, and are themselves functional (amino acid changing) polymorphisms more often than expected by chance. Therefore, the identified common variants are likely to implicate genes relevant to trait or disease under study, thus providing a number of novel candidate regions for drug and therapeutic targeting, or for mutations in biologically related monogenic diseases with unknown causes. Even though the proportion of variance explained by the common variants is relatively small, the

increase in our understanding of the biological mechanisms disease aetiology may be substantial.

Several interesting features of the common trait architecture have emerged through the height studies: there are multiple variants with independent effects at individual gene loci (allelic heterogeneity), and the same variants can affect multiple traits (pleiotropy). Thus, they may represent the genetic links between traits and diseases that are suspected to share some of their aetiology, and provide cause/effect distinctions between those that the epidemiological studies have shown to be correlated. For example, a variant in *LIN28B* gene that is associated with reduced adult height [10] is also associated with earlier onset of puberty [27, 28], which is consistent with epidemiological observations.

The GIANT height study has provided several good candidate genes that might have mutations responsible for as-yet unexplained skeletal and developmental disorders. For example, the study identified associations in *FGFR4* and *STAT2* genes, which have similar functions to already known human growth-related genes, *FGFR3* and *STAT5B*, respectively. In type 2 diabetes, common variants in the *KCNJ11* gene region have only small effect on the disease, yet the membrane protein coded by this gene forms a potassium channel that is the target of sulphonylureas, major anti-diabetic drugs that act by increasing insulin from the pancreatic beta cells. It is worth noting that *KCNJ11* was already known to the diabetes community before the GWA studies, because its association with diabetes was discovered through candidate gene studies [29, 30]. Among the type 2 diabetes genes identified through the GWA approach, a zinc transporter gene *SLC30A8* seems particularly interesting for drug targeting, since it has been shown that down-regulation of the gene in beta cells leads to reduced insulin secretion in response to hyperglycemia [31].

Disease prediction

The immediate utility of disease associated variants discovered through GWA studies appears to be the identification of molecular pathways involved in

the disease and intermediate quantitative traits, which should eventually lead to new therapeutic targets. In the long-term, there is the prospect of disease susceptibility prediction, which would be useful for diseases where preventative measures are effective. Chapter 4 assesses how well the variants robustly associated with type 2 diabetes predict disease status. The study shows that they would not be useful for disease prediction, and add little to the discriminatory power of other well-established diabetes risk factors. However, a combined genetic score could be used to identify individuals with high genetic predisposition to disease.

One of the limitations of this study was that it was not prospective population-based study, and consequently, the predictive power of the variants could not be accurately determined. Since then, several prospective studies have assessed the combined effect of the known type 2 diabetes variants [32-34]. They have all reached similar conclusions – the combined genetic score had only modest ability to predict the future development of diabetes (all had AUC of around 0.6), and provided only slightly improved prediction when added to the other known risk factors.

Combined with an individual's lifestyle and environmental exposures, the genetic information can be then used to guide decisions about disease prevention, monitoring and management. For example, those at higher genetic risk of breast and bowel cancer could be offered more regular screening. Similar discriminatory ability of the combined genetic score was observed for the 20 height associated variants described in Chapter 2. This information may be used in medicine, for example to determine if a child's growth is reaching its genetic potential, as well as in forensics [35].

Another clinical application of common variants may be in better clinical characterisation of monogenic diseases, which are often heterogeneous in terms of disease onset, severity, progression and other clinical characteristics. A good starting point is to investigate variants that have been shown to associate with biologically related common diseases and traits. This is the case in Chapter 5,

which showed that *HNF1A*-MODY patients with higher load of common type 2 diabetes predisposing variants had earlier age at diabetes diagnosis, and required smaller load of other risk factors, namely their BMI, to get the disease. Another disease where common variants have been shown to have a strong modifying effect includes sickle cell anaemia, where foetal haemoglobin (HbF) expression is an established and heritable disease modifier, such that high HbF levels lead to slower disease progression and fewer complications in patients with the sickle cell disease. Several common variants have now been shown to affect HbF persistence in adults [36, 37], so an immediate clinical application would be the ability to better predict disease severity and, therefore, improve disease management. In the long term, these variants are potential targets of new therapies based on increasing HbF expression.

Future directions

Most of the associated signals detected through GWA studies are not the causal variants themselves, but are detected because they are correlated with rarer, more penetrant variants. A recent study suggested that many of the current GWA signals could reflect effects of several rare, deleterious variants that have emerged more recently, on the same haplotypes as the common SNP for which the 'synthetic' association is observed [38]. However, rare variants cannot be directly detected through the current GWA approaches, and imputation of deeper sets of SNPs can only help to narrow down the region containing the causal variant(s).

There are several emerging tools and technologies that can be used in future studies to get closer to the causal variants. These include deeper imputation with 1000 Genomes data, fine mapping of the associated loci with custom-designed genotyping chips, and studying more non-European populations that have different LD structure (especially African populations, who have shorter LD blocks and thus narrower regions of associated loci). One fine-mapping project currently underway is the genotyping of samples with metabolic trait phenotypes on the metabochip, an Illumina custom-made chip based on GWA results and designed by several collaborating consortia including DIAGRAM, MAGIC and

GIANT. The most direct way of detecting rare functional variants, however, is sequencing of the associated regions, possibly in samples enriched for genetic predisposition, for example those with familial background, or at the extremes of the trait distribution or young-onset disease cases.

This was the approach I took to search for rare variants in height genes selected from the 20 presented in Chapter 2: *HMGA2* and *ZBTB38* that contained the two most-associated signals, *JAZF1* that is also a type 2 diabetes locus; and two hedgehog signaling genes *IHH* and *HHIP*. After designing primers to cover all exons and flanking regulatory and intronic regions, these genes were sequenced in a panel of 48 tall and 48 short individuals at the extremes of the height distribution in the Exeter Family Study (EFS) cohort. Several novel missense variants were identified, and so far those in *JAZF1* have been genotyped in 1700 unrelated individuals from the EFS cohort. These included 229Phe>Leu variant initially seen in two tall individuals, and an intron4 variant identified in four short individuals. Genotyping results, shown in the table below, show a directionally consistent trend for association, but are not statistically significant and the new variants probably do not confer a major effect on height. Sequencing of *JAZF1* in 300 type 2 diabetes young-onset cases and 300 controls is currently under way.

Variant	Genotype	Genotype count	Mean Z-height	P-value
F229L	AA AG GG	1681 6 0	-0.007 0.418 NA	0.296
Intron4 (ex3+68nt)	GG GA AA	1591 92 1	0.006 -0.125 -0.931	0.164

Owing to the emergence to the next generation sequencing technologies and the substantial reduction in costs, many researchers are now embarking on whole-exome, whole-genome and sequence-capture sequencing projects. Whole genome sequencing will clearly be needed to identify rare variants where

associated loci lie outside the known gene and regulatory regions, although proving causality in such cases will be more difficult. Examples of common disease rare variant detection by whole genome/exome sequencing are still rare. The most exciting success story is perhaps in type 1 diabetes, where resequencing of exomes and splice sites identified four rare protective variants in *IFIH1* gene [39], a locus already implicated through type 1 diabetes GWA studies.

To follow up the height GWA study, a Nimblegen Capture Array has been designed to cover 3.7 megabases around the height associated loci, including entire smaller regions between recombination hotspots, and exons of all genes in larger regions. Four individuals have now been 'captured' on this array and sequenced on one of the next-generation sequencers, the Illumina Genome Analyzer II. Furthermore, for quality control purposes the same individuals have been genotyped on the Affymetrix 6.0 platform, and the sequencing and genotyping data will be analysed shortly. In a separate multi-centre project, we are currently sequencing 1500 type 2 diabetes patients and 1500 controls.

Undoubtedly, as the targeted sequencing within my own research group has hinted, a large number of novel non-synonymous variants will be identified, and deciding which are functionally relevant will be the next major challenge. Furthermore, as the preliminary metabochip analyses are already demonstrating, genotyping and correctly calling rare variants is tricky because the approach is based on genotype clustering. In my project I knew which samples had the rare variants and was, therefore, able to include positive controls that served as rare genotype reference during clustering process. However, many of the next-generation sequencing projects use pooled samples to reduce costs, in which case sample tagging approaches can be used to identifying which rare variant comes from which sample. New developments in genotyping methodology and clustering algorithms are needed to facilitate the search for rare variants in common traits and diseases.

Conclusions

Using human height, type 2 diabetes and *HNF1A*-MODY as examples of a quantitative trait, polygenic disease and monogenic disease, respectively, this thesis has explored the role of common genetic variation identified through genome-wide association approaches in these types of human traits. It has demonstrated that common variants are not only statistically associated with common traits and diseases, but can reveal novel biology, disease aetiology, and genetic architecture; explain epidemiological observations; help better characterise both complex and single-gene diseases; and have potential to be used in disease prediction, therapeutic targeting, and personalised disease management. It is clear that much of the near future work in complex traits genetics will focus on the search for rare variants of larger effects, which may underlie many of the current associations with the common variants of modest effects. The already established collaborations within the common disease genetics community, and the constantly reducing costs of whole genome sequencing, promise to yield many exciting findings in near future.

References

1. The International HapMap Consortium. A haplotype map of the human genome. Nature 437, 1299-1320 (2005).

2. Frazer, K. A. et al. A second generation human haplotype map of over 3.1 million SNPs. Nature 449, 851-861 (2007).

3. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype Imputation. Annual Review Of Genomics And Human Genetics 10, 387-406 (2009).

4. Frazer, K. A., Murray, S. S., Schork, N. J. & Topol, E. J. Human genetic variation and its contribution to complex traits. Nature Reviews Genetics 10, 241-251 (2009).

5. Kong, A. et al. Parental origin of sequence variants associated with complex diseases. Nature 462, 868-U59 (2009).

6. Goldstein, D. B. Common Genetic Variation and Human Traits. N Engl J Med 360, 1696-1698 (2009).

7. Weedon, M. N. et al. Genome-wide association analysis identifies 20 loci that influence adult height. Nat Genet 40, 575 (2008).

8. Sanna, S. et al. Common variants in the GDF5-UQCC region are associated with variation in human height. Nat Genet 40, 198 (2008).

9. Gudbjartsson, D. F. et al. Many sequence variants affecting diversity of adult human height. Nat Genet 40, 609 (2008).

10. Lettre, G. et al. Identification of ten loci associated with height highlights new biological pathways in human growth. Nat Genet 40, 584 (2008).

11. Saxena, R. et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 316, 1331-1336 (2007).

12. Scott, L. J. et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 316, 1341-1345 (2007).

13. Sladek, R. et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 445, 881 (2007).

14. Zeggini, E. et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 316, 1336-1341 (2007).

15. Steinthorsdottir, V. et al. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. Nature Genetics 39, 770-775 (2007).

16. Zeggini, E. et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nature Genetics 40, 638-645 (2008).

17. Prokopenko, I. et al. Variants in MTNR1B influence fasting glucose levels. Nat Genet 41, 77 (2009).

18. Dupuis, J. et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet 42, 105 (2010).

19. Freathy, R. M. et al. Variants in ADCY5 and near CCNL1 are associated with fetal growth and birth weight. Nat Genet 42, 430 (2010).

20. Visscher, P. M., Hill, W. G. & Wray, N. R. Heritability in the genomics era [mdash] concepts and misconceptions. Nat Rev Genet 9, 255 (2008).

21. Reich, D. E. & Lander, E. S. On the allelic spectrum of human disease. Trends Genet 17, 502-10 (2001).

22. Yang, J. et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42, 565 (2010).

23. Stefansson, H. et al. Large recurrent microdeletions associated with schizophrenia. Nature 455, 232 (2008).

24. Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature 455, 237 (2008).

25. Pinto, D. et al. Functional impact of global rare copy number variation in autism spectrum disorders. Nature advance online publication (2010).

26. Craddock, N. et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Nature 464, 713-U86 (2010).

27. Ong, K. K. et al. Genetic variation in LIN28B is associated with the timing of puberty. Nat Genet 41, 729 (2009).

28. Perry, J. R. B. et al. Meta-analysis of genome-wide association data identifies two loci influencing age at menarche. Nat Genet 41, 648 (2009).

29. Gloyn, A. L. et al. Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. Diabetes 52, 568-72 (2003).

30. Nielsen, E. M. et al. The E23K variant of Kir6.2 associates with impaired post-OGTT serum insulin response and increased risk of type 2 diabetes. Diabetes 52, 573-7 (2003).

31. Fu, Y. et al. Down-Regulation of ZnT8 Expression in INS-1 Rat Pancreatic Beta Cells Reduces Insulin Content and Glucose-Inducible Insulin Secretion. PLoS ONE 4, e5679 (2009).

32. van Hoek, M. et al. Predicting Type 2 Diabetes Based on Polymorphisms From Genome-Wide Association Studies: A Population-Based Study. Diabetes 57, 3122-3128 (2008).

33. Lyssenko, V. et al. Clinical Risk Factors, DNA Variants, and the Development of Type 2 Diabetes. New England Journal of Medicine 359, 2220-2232 (2008).

34. Meigs, J. B. et al. Genotype Score in Addition to Common Risk Factors for Prediction of Type 2 Diabetes. N Engl J Med 359, 2208-2219 (2008).

35. Visscher, P. M. Sizing up human height variation. Nat Genet 40, 489 (2008).

36. Thein, S. L. & Menzel, S. Discovering the genetics underlying foetal haemoglobin production in adults. British Journal Of Haematology 145, 455-467 (2009).

37. Steinberg, M. H. Genetic Etiologies for Phenotypic Diversity in Sickle Cell Anemia. TheScientificWorldJOURNAL 9, 46 (2009).

38. Dickson, S. P., Wang, K., Krantz, I., Hakonarson, H. & Goldstein, D. B. Rare Variants Create Synthetic Genome-Wide Associations. PLoS Biol 8, e1000294 (2010).

39. Nejentsev, S., Walker, N., Riches, D., Egholm, M. & Todd, J. A. Rare Variants of IFIH1, a Gene Implicated in Antiviral Responses, Protect Against Type 1 Diabetes. Science 324, 387-389 (2009).

APPENDIX I: LIST OF PUBLICATIONS

1. Frayling TM*, Timpson NJ*, Weedon MN*, Zeggini E*, Freathy RM, Lindgren CM, Perry JR, Elliott KS, **Lango H**, Rayner NW, Shields B, Harries LW, Barrett JC, Ellard S, Groves CJ, Knight B, Patch AM, Ness AR, Ebrahim S, Lawlor DA, Ring SM, Ben-Shlomo Y, Jarvelin MR, Sovio U, Bennett AJ, Melzer D, Ferrucci L, Loos RJ, Barroso I, Wareham NJ, Karpe F, Owen KR, Cardon LR, Walker M, Hitman GA, Palmer CN, Doney AS, Morris AD, Davey-Smith G, Hattersley AT and McCarthy MI: *A Common Variant in the FTO Gene Is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity.* ***Science*** **(2007) 316:889-94.**

Personal contribution: Prepared DNA samples, performed data management, QC and analysis of WTCCC-T2D dataset. Commented on the manuscript.

2. Zeggini E*, Weedon MN*, Lindgren CM*, Frayling TM*, Elliott KS, **Lango H**, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney AS, McCarthy MI and Hattersley AT: *Replication of Genome-Wide Association Signals in U.K. Samples Reveals Risk Loci for Type 2 Diabetes.* ***Science*** **(2007) 317:1035-1036.**

Personal contribution: Prepared DNA samples, performed data management, QC and analysis of WTCCC-T2D dataset. Commented on the manuscript.

3. Wellcome Trust Case Control Consortium: *Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.* ***Nature*** **(2007) 447:661-78.**

Personal contribution: Prepared DNA samples, performed data management, QC and analysis of WTCCC-T2D dataset.

4. Manjinder S Sandhu*, Michael N Weedon*, Katherine A Fawcett, Jon Wasson, Sally L Debenham, Allan Daly, **Hana Lango**, Timothy M Frayling, Rosalind J Neumann, Richard Sherva, Ilana Blech, Paul D Pharoah, Colin N A Palmer, Charlotte Kimber, Roger Tavendale, Andrew D Morris, Mark I McCarthy, Mark Walker, Graham Hitman, Benjamin Glaser, M Alan Permutt, Andrew T Hattersley, Nicholas J Wareham & Inês Barroso: *Common variants in WFS1 confer risk of type 2 diabetes.* ***Nature Genetics*** **(2007) 39:951-53.**

Personal contribution: Assisted statistical analyses. Commented on the manuscript.

5. Michael N Weedon*, Guillaume Lettre*, Rachel M Freathy*, Cecilia M Lindgren*, Benjamin F Voight, John R B Perry, Katherine S Elliott, Rachel Hackett, Candace Guiducci, Beverley Shields, Eleftheria Zeggini, **Hana Lango**, Valeriya Lyssenko, Nicholas J Timpson, Noel P Burtt, Nigel W Rayner, Richa Saxena, Kristin Ardlie, Jonathan H Tobias, Andrew R Ness, Susan M

Ring, Colin N A Palmer, Andrew D Morris, Leena Peltonen, Veikko Salomaa, The Diabetes Genetics Initiative, The Wellcome Trust Case Control Consortium, George Davey Smith, Leif C Groop, Andrew T Hattersley, Mark I McCarthy, Joel N Hirschhorn & Timothy M Frayling: *A common variant of HMGA2 is associated with adult and childhood height in the general population.* ***Nature Genetics*** **(2007) 39:1245-50.**

Personal contribution: Prepared DNA samples, performed data management, QC and analysis of WTCCC-T2D dataset. Commented on the manuscript.

6. **Hana Lango** & Michael N Weedon: *What will whole genome searches for susceptibility genes for common complex disease offer to clinical practice?* ***Journal Of Internal Medicine*** **(2008) 263: 16-27.**

Personal contribution: co-wrote the manuscript.

7. Rachel M Freathy, Nicholas J Timpson, Debbie A Lawlor, Anneli Pouta, Yoav Ben-Shlomo, Aimo Ruokonen, Shah Ebrahim, Beverley Shields, Eleftheria Zeggini, Michael N Weedon, Cecilia M Lindgren, **Hana Lango**, David Melzer, Luigi Ferrucci, Giuseppe Paolisso, Matthew J Neville, Fredrik Karpe, Colin N A Palmer, Andrew D Morris, Paul Elliott, Marjo-Riitta Jarvelin, George Davey Smith, Mark I McCarthy, Andrew T Hattersley & Timothy M Frayling: *Common variation in the FTO gene alters diabetes-related metabolic traits to the extent expected, given its effect on BMI.* ***Diabetes*** **(2008) 57:1419-26.**

Personal contribution: Assisted with statistical analysis. Commented on the manuscript.

8. Michael N Weedon*, **Hana Lango***, Cecilia M Lindgren, Chris Wallace, David M Evans, Massimo Mangino, Rachel M Freathy, John RB Perry, Suzanne Stevens, Alistair S Hall, Nilesh J Samani, Beverly Shields, Inga Prokopenko, Martin Farrall, Anna Dominiczak, Diabetes Genetics Initiative, The Wellcome Trust Case Control Consortium, Toby Johnson, Sven Bergmann, Jacques S Beckmann, Peter Vollenweider, Dawn M Waterworth, Vincent Mooser, Colin NA Palmer, Andrew D Morris, Willem H Ouwehand, Cambridge GEM Consortium, Mark Caulfield, Patricia B Munroe, Andrew T Hattersley, Mark I McCarthy, Timothy M Frayling: *Genome-wide association analysis identifies 20 loci that influence adult height.* ***Nature Genetics*** **(2008) 40: 575-583.**

Personal contribution: Prepared DNA samples, performed data management, QC and analysis of WTCCC-T2D dataset. Performed meta-analysis. Co-wrote the manuscript.

9. David Melzer, John R. B. Perry, Dena Hernandez, Anna-Maria Corsi, Kara Stevens, Ian Rafferty, Fulvio Lauretani, Anna Murray, J. Raphael Gibbs, Giuseppe Paolisso, Sajjad Rafiq, Javier Simon-Sanchez, **Hana Lango**, Sonja Scholz, Michael N. Weedon, Sampath Arepalli, Neil Rice, Nicole Washecka,

Alison Hurst, Angela Britton, William Henley, Joyce van de Leemput, Rongling Li, Anne B. Newman, Greg Tranah, Tamara Harris, Vijay Panicker , Colin Dayan, Amanda Bennett, Mark I. McCarthy, Aimo Ruokonen, Marjo-Riitta Jarvelin, Jack Guralnik, Stefania Bandinelli, Timothy M. Frayling, Andrew Singleton, Luigi Ferrucci: *A Genome-Wide Association Study Identifies Protein Quantitative Trait Loci (pQTLs)*. ***PLoS Genetics*** **(2008) 4(5):e1000072.**

Personal contribution: Involved in data management, initial QC and analysis of the InCHIANTI dataset. Performed stratification analyses. Commented on the manuscript.

10. **Hana Lango**, The UK Type 2 Diabetes Genetics Consortium, Colin NA Palmer, Andrew D Morris, Eleftheria Zeggini, Andrew T Hattersley, Mark I McCarthy, Timothy M Frayling, Michael N Weedon: *Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk.* ***Diabetes*** **(2008) 57:3129-35.**

Personal contribution: Designed and led the project, performed all statistical analyses and wrote the manuscript.

11. Rafiq S, Melzer D, Weedon MN, **Lango H**, Saxena R, Scott LJ; DIAGRAM Consortium, Palmer CN, Morris AD, McCarthy MI, Ferrucci L, Hattersley AT, Zeggini E, Frayling TM: *Gene variants influencing measures of inflammation or predisposing to autoimmune and inflammatory diseases are not associated with the risk of type 2 diabetes.* ***Diabetologia*** **(2008) 51:2205-13.**

Personal contribution: Involved in data management, initial QC and analysis of the WTCCC-T2D and InCHIANTI datasets. Performed InCHIANTI imputation. Assisted statistical analyses.

12. Freathy RM, Bennett AJ, Ring SM, Shields B, Groves CJ, Timpson NJ, Weedon MN, Zeggini E, Lindgren CM, **Lango H**, Perry JR, Pouta A, Ruokonen A, Hyppönen E, Power C, Elliott P, Strachan DP, Järvelin MR, Smith GD, McCarthy MI, Frayling TM, Hattersley AT. *Type 2 diabetes risk alleles are associated with reduced size at birth.* ***Diabetes*** **(2009) 58:1428-33.**

Personal contribution: Assisted data management and wrote scripts for file/data formatting; helped with statistical analysis. Commented on the manuscript.

13. **Hana Lango Allen**, Stefan Johansson, Sian Ellard, Jens K Hertel, Beverley Shields, Helge Ræder, Kevin Colclough, Anders Molven, Timothy M Frayling, Pål R Njølstad, Andrew T Hattersley, Michael N Weedon: *Polygenic Risk Variants for Type 2 Diabetes Susceptibility Modify Age at Diagnosis in Monogenic HNF1A Diabetes.* ***Diabetes*** **(2010) 59:266-71.**

Personal contribution: Designed and led the project, performed all statistical analyses and wrote the manuscript.

14. Wellcome Trust Case Control Consortium, Craddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, Robson S, Vukcevic D, Barnes C, Conrad DF, Giannoulatou E, Holmes C, Marchini JL, Stirrups K, Tobin MD, Wain LV, Yau C, Aerts J, Ahmad T, Andrews TD, Arbury H, Attwood A, Auton A, Ball SG, Balmforth AJ, Barrett JC, Barroso I, Barton A, Bennett AJ, Bhaskar S, Blaszczyk K, Bowes J, Brand OJ, Braund PS, Bredin F, Breen G, Brown MJ, Bruce IN, Bull J, Burren OS, Burton J, Byrnes J, Caesar S, Clee CM, Coffey AJ, Connell JM, Cooper JD, Dominiczak AF, Downes K, Drummond HE, Dudakia D, Dunham A, Ebbs B, Eccles D, Edkins S, Edwards C, Elliot A, Emery P, Evans DM, Evans G, Eyre S, Farmer A, Ferrier IN, Feuk L, Fitzgerald T, Flynn E, Forbes A, Forty L, Franklyn JA, Freathy RM, Gibbs P, Gilbert P, Gokumen O, Gordon-Smith K, Gray E, Green E, Groves CJ, Grozeva D, Gwilliam R, Hall A, Hammond N, Hardy M, Harrison P, Hassanali N, Hebaishi H, Hines S, Hinks A, Hitman GA, Hocking L, Howard E, Howard P, Howson JM, Hughes D, Hunt S, Isaacs JD, Jain M, Jewell DP, Johnson T, Jolley JD, Jones IR, Jones LA, Kirov G, Langford CF, **Lango-Allen H**, Lathrop GM, Lee J, Lee KL, Lees C, Lewis K, Lindgren CM, Maisuria-Armer M, Maller J, Mansfield J, Martin P, Massey DC, McArdle WL, McGuffin P, McLay KE, Mentzer A, Mimmack ML, Morgan AE, Morris AP, Mowat C, Myers S, Newman W, Nimmo ER, O'Donovan MC, Onipinla A, Onyiah I, Ovington NR, Owen MJ, Palin K, Parnell K, Pernet D, Perry JR, Phillips A, Pinto D, Prescott NJ, Prokopenko I, Quail MA, Rafelt S, Rayner NW, Redon R, Reid DM, Renwick, Ring SM, Robertson N, Russell E, St Clair D, Sambrook JG, Sanderson JD, Schuilenburg H, Scott CE, Scott R, Seal S, Shaw-Hawkins S, Shields BM, Simmonds MJ, Smyth DJ, Somaskantharajah E, Spanova K, Steer S, Stephens J, Stevens HE, Stone MA, Su Z, Symmons DP, Thompson JR, Thomson W, Travers ME, Turnbull C, Valsesia A, Walker M, Walker NM, Wallace C, Warren-Perry M, Watkins NA, Webster J, Weedon MN, Wilson AG, Woodburn M, Wordsworth BP, Young AH, Zeggini E, Carter NP, Frayling TM, Lee C, McVean G, Munroe PB, Palotie A, Sawcer SJ, Scherer SW, Strachan DP, Tyler-Smith C, Brown MA, Burton PR, Caulfield MJ, Compston A, Farrall M, Gough SC, Hall AS, Hattersley AT, Hill AV, Mathew CG, Pembrey M, Satsangi J, Stratton MR, Worthington J, Deloukas P, Duncanson A, Kwiatkowski DP, McCarthy MI, Ouwehand W, Parkes M, Rahman N, Todd JA, Samani NJ, Donnelly P. *Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls.* ***Nature*** **(2010) 464:713-20.**

Personal contribution: Prepared WTCCC-T2D DNA samples, performed initial data management, QC and association analysis.

15. Hana Lango Allen*, Karol Estrada*, Guillaume Lettre*, Sonja Berndt*, Michael N Weedon*, Fernando Rivadeneira*, *et al*., for the GIANT Consortium. *Hundreds of variants influence human height and cluster within genomic loci and biological pathways.* ***Nature*** **(2010) 467:832-38**

Personal contribution: WTCCC-T2D dataset analyst. Performed cleaning and formatting of files as part of the central GIANT analysis group. Co-lead of the conditional analyses working group. Performed height-specific analyses and co-wrote the manuscript.

16. Elizabeth K. Speliotes*, Cristen J. Willer*, Sonja I. Berndt*, Keri L. Monda*, Gudmar Thorleifsson*, Anne U. Jackson, **Hana Lango Allen**, *et al.*, for the GIANT Consortium. A*ssociation analyses of 249,796 individuals reveal 18 new loci associated with body mass index.* ***Nature Genetics*** **(2010) 42:937-48**

Personal contribution: WTCCC-T2D dataset analyst. Performed cleaning and formatting of files as part of the central GIANT analysis group. Co-lead of the conditional analyses working group.

17. Iris M. Heid*, Anne U. Jackson*, Joshua C. Randall*, Thomas W. Winkler*, Lu Qi*, Valgerdur Steinthorsdottir*, Gudmar Thorleifsson*, M. Carola Zillikens, Elizabeth K. Speliotes, Reedik Mägi, Tsegaselassie Workalemahu, Charles C. White, Nabila Bouatia-Naji, Tamara B. Harris, Sonja I. Berndt, Erik Ingelsson, Cristen J. Willer, Michael N. Weedon, Jian'an Luan, Sailaja Vedantam, Tõnu Esko, Tuomas O. Kilpeläinen, Zoltán Kutalik, Shengxu Li, Keri L. Monda, Anna L. Dixon, Christopher C. Holmes, Lee M. Kaplan, Liming Liang, Josine L. Min, Miriam F. Moffatt, Cliona Molony, George Nicholson, Eric E. Schadt, Krina T. Zondervan, Mary F. Feitosa, Teresa Ferreira, **Hana Lango Allen**, *et al.*, for the GIANT Consortium. *Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution.* ***Nature Genetics*** **(2010) 42:949-60**

Personal contribution: WTCCC-T2D dataset analyst. Performed cleaning and formatting of files as part of the central GIANT analysis group. Co-lead of the conditional analyses working group.