Growth Disorders: Gene Discovery

HORMONE RESEARCH

Horm Res 2009;71(suppl 2):5–13
DOI: 10.1159/000192430

Published online: April 29, 2009

Progress in Genome-Wide Association Studies of Human Height

Joel N. Hirschhorn[a–c] Guillaume Lettre[a, b]

[a]Divisions of Endocrinology and Genetics and Program in Genomics, Children's Hospital, Boston, Mass., [b]Broad Institute of Harvard and MIT, Cambridge, Mass., and [c]Department of Genetics, Harvard Medical School, Boston, Mass., USA

Key Words
Genome-wide association studies · Height · Polygenic traits · Skeletal growth · Stature

Abstract
Human height (stature) is a strongly genetic trait, with up to 90% of the variation in height within a population determined by a combination of multiple inherited factors. Recent advances in genetics and genomics now permit comprehensive genome-wide surveys of common genetic variations in those variants that are associated with stature. The first such studies have borne fruit, identifying over 40 genetic loci that can be reproducibly shown to have an influence on adult height. These unbiased searches throughout the genome identified several loci that also harbour rare mutations responsible for more severe alterations in height or skeletal growth. Although the predictive value of the common variants thus far discovered remains low, the identification of these loci has led to new insights into the biology of human growth, and may help identify genes that underlie previously uncharacterized syndromes of abnormal skeletal growth.

Genetics of Height

Scientists have studied the inherited basis of height since the beginning of human genetics. Galton and Pearson used height to illustrate and quantify the strong resemblance of offspring to their parents [1, 2]. The concept of regression, a foundation of statistical analysis, was developed to describe the relationship between offspring height and midparental height (offspring height is strongly correlated with midparental height but shows a slight 'regression' towards the mean [1]). In 1918, Fisher used height as an example in reconciling the apparent conflict between Mendelian theories of inheritance of discrete genetic factors and the continuous nature of quantitative traits [3]. Specifically, Fisher proposed that the continuous variation in height observed in the general population was the result of the combined effects of numerous independent genetic factors. Although this idea was generally accepted, as of early 2007 none of the common genetic variants influencing height had been identified.

The strong influence of genetics deduced from the familial resemblances for height is also apparent from studies of monozygotic and dizygotic twins. Monozygotic twins are much more similar in height than dizygotic twins, and estimates from twin studies indicate that up to 90% of the variation in adult height within a population is due to inherited DNA variation [4, 5]. Average

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0301–0163/09/0718–0005$26.00/0

Accessible online at:
www.karger.com/hre

Dr. Joel N. Hirschhorn
Divisions of Endocrinology and Genetics and Program in Genomics
Children's Hospital
Boston, MA 02115 (USA)
Tel. +1 617 919 2129, Fax +1 617 730 0253, E-Mail joelh@broad.mit.edu

height has also increased substantially over the last few generations, likely due to changing environmental factors such as improved nutrition; this changing impact of the environment over time is a secular trend. Nonetheless, within a particular population at any particular time, genetics plays the predominant role in determining height.

Despite the strong role of genetics, efforts to identify the genes that influence height had, until recently, met with little success, with the exception of genes underlying rare, monogenic syndromes with altered growth [6]. Patients with these syndromes are usually identified because they have growth hormone (GH) deficiency or resistance, overgrowth or skeletal dysplasias. The genes underlying these syndromes have highlighted fundamental molecular determinants of growth, such as GH and its receptor [7–9], pituitary transcription factors, members of the GH signalling pathway [10–12] and key factors in the biology of growth plate, cartilage and bone [13, 14]. However, genes related to these syndromes do not explain much of the overall variation in height in the general population. Furthermore, many patients with syndromic short stature likely have mutations in as yet unidentified genes and biological pathways. Thus, much of the genetic variation in human growth remains unexplained, presenting an opportunity for new insights into the biology of human growth.

Many initial efforts to identify the genes underlying stature in the general population took the form of linkage studies. In linkage studies, height data from families was combined with genetic data from hundreds of genetic markers spread across the genome to identify chromosomal regions that were more often co-inherited in pairs of relatives with similar heights than in pairs of relatives with divergent heights. A number of studies identified several regions where the statistical evidence of linkage surpassed genome-wide thresholds of significance [15–19], defined as a statistical result strong enough that it is unlikely to be observed by chance even when examining the entire genome. However, no one region was clearly reproducible; even a large meta-analysis that combined data from multiple linkage studies did not point overwhelmingly to a single region [4]. Thus, consistent with Fisher's proposal, there are likely no single genetic loci with major effects on height in the general population (if there were, the linkage evidence would be overwhelming). It remains unclear why several studies observed regions that had genome-wide significant linkage, unless these regions are in fact enriched with genetic variants that play a role in height.

Genetic Association Studies of Height

The other major approach to finding genes for stature has been the association study. In this type of analysis, individual genetic variants are tested in individuals with different heights. If a genotype or allele at a particular variant shows a significant statistical association with increased or decreased height, then that variant (or perhaps a close-by variant that is correlated with the tested variant) is proposed to play a causal role in influencing height. In the past, association studies were limited to studies of particular genes and variants suggested by prior hypotheses ('candidate genes'). Recently, advances in technology, knowledge of patterns of human variation, and data management and analysis tools have permitted genome-wide association studies [20], in which hundreds of thousands of variants (usually single nucleotide polymorphisms [SNPs]) are tested in patient samples. This new approach can provide an unbiased and fairly comprehensive survey of common genetic variants (usually considered to be those in which a minor allele has a frequency in the population of at least 5%).

Although simple in practice, association studies have a few important potential pitfalls [21]. First, rigorous statistical thresholds must be achieved to have a high degree of certainty that an apparent association is not merely due to statistical fluctuation. Although there is no precise agreed-upon threshold for significance, for genome-wide association studies, in which hundreds of thousands of markers are tested, p values in the range of 10^{-7}–10^{-8} are probably required [21, 22]. Similar thresholds are probably required even in studies where fewer markers are tested [22–24].

Technical artefacts can lead to false-positive associations and can arise either from incomplete or inaccurate genotyping or phenotyping, or from population stratification [25, 26]. Avoidance of genotyping-related artefacts requires careful attention to quality control and the generation of high-quality genotype data. Stratification, which is essentially confounding of an association by ancestry, stems from the sometimes unavoidable inclusion of individuals with different genetic ancestries in a single analysis. This problem can be particularly troublesome for height, which can vary substantially even among populations with fairly similar ancestries. Indeed, we used height to provide the first demonstration of a false-positive association in a population of European ancestry [25]. Fortunately, methods to detect and correct for stratification exist where genome-wide genotype data are available [27] and resources are being developed even

for studies where dense genotype data are unavailable [28].

Finally, adequate power is essential to avoid false negatives. A review of association studies and subsequent experience with genome-wide association studies suggest that most common genetic variants that influence polygenic traits will have modest effects. Thus, large sample sizes and combined consideration of multiple studies will often be beneficial in testing genetic variants for association with polygenic phenotypes such as height.

Genome-Wide Association Studies of Height

In 2007, the first large genome-wide studies were published [22, 29–41], and opened the way to a flood of new discoveries of genetic variants that affected the risk of common disease or quantitative traits. The first publicly available genome-wide association data for height were produced by the Diabetes Genetics Initiative of Broad, Lund and Novartis [31]. The initial results of this study did not reveal any associations that reached genome-wide levels of statistical significance. However, joint consideration of these data with similar data from the UKT2D portion of the Wellcome Trust Case Control Consortium pointed to a SNP in the 3′ untranslated region of the *HMGA2* gene, which showed strong evidence of association with height (combined $p = 4 \times 10^{-8}$). Subsequent replication in over 22,000 adults provided overwhelming evidence of replication, demonstrating that this association was correct [42]. Effects on height were also seen in children as young as 7 years of age, although there was no detectable effect on birth length.

HMGA2 was an excellent candidate gene for height: not only was its orthologue mutated in the Pygmy mouse [43], but a patient with a disruption in the *HMGA2* gene (probably leading to overexpression) showed severe overgrowth as well as multiple benign tumours [44]. More recently, a microdeletion encompassing *HMGA2* was described that caused a syndrome with multiple features, including short stature [45]. Intriguingly, *HMGA2*, a chromatin protein of unknown function, is largely not expressed after birth except in a variety of tumours, including pituitary adenomas [46]. One possibility is that *HMGA2* regulates cell number in the pituitary or in other tissues relevant to growth (such as the growth plate), with an eventual outcome on adult height.

Data drawn primarily from two other genome-wide association studies, FUSION and SardiNIA, provided strong evidence for an association between height and a second genetic variant, near the *GDF5* gene [47]. *GDF5* was also a reasonable candidate gene for influencing height, as mutations in this gene cause brachydactyly and a form of skeletal dysplasia [48–50]. Interestingly, the same SNP also influences the susceptibility to osteoarthritis [51], indicating that height SNPs can have multiple (pleiotropic) effects, and suggesting that bone/cartilage is a relevant site of action of the *GDF5* variant.

Bolstered by these initial discoveries, a number of groups with genome-wide association data decided to collaborate and combine their results, eventually coming together under the rubric of a consortium called GIANT (Genetic Investigation of ANthropometric Traits). The first results from this collaborative effort represented efforts to combine data from each of two halves of the GIANT consortium, groups from the UK and the remainder of the consortium. Although none of the best results from the individual genome-wide studies had association p values that were lower than expected by chance, combining the results from six studies from the USA, Scandinavia, Sardinia and Germany yielded a large excess of highly significant results (fig. 1), providing indisputable evidence of the power of increasing sample size (and of collaboration). Replication of these potential associations yielded strong evidence for 10 new height loci, in addition to strong confirmation of the associations at *HMGA2* and *GDF5* [52]. Similar efforts from the UK groups provided evidence for 20 new loci [53], including several of the loci from Lettre et al. [52] (table 1). Simultaneously, a group from deCODE Genetics reported 27 loci associated with height; most of the GIANT loci had strong independent support from the deCODE data with p values below 10^{-4} [54], and many of the deCODE loci were also identified in one of the two GIANT publications. Thus, these loci have strong evidence for an association with height.

Many of the newly identified loci highlight pathways (groups of functionally related genes), including some pathways not usually thought of as central to human growth. Although it is sometimes difficult to know which nearby gene(s) is affected by a particular associated variant, the association studies identified several apparently related groups of genes. These include chromatin factors *(HMGA1*, *HMGA2*, *DOT1L*, the two histone clusters, *SCMH1)*, hedgehog signalling pathway genes *(HHIP, IHH, PTCH1)*, targets of the *let-7* microRNA *(HMGA2, DOT1L, CDK6, LIN28B)*, noggin/BMP signalling *(NOG, BMP2, BMP6)*, extracellular matrix proteins and proteases possibly related to fibulin and/or aggrecan biology *(ACAN, ADAMTSL3, ADAMTS17, EFEMP1, FBLN5)* and

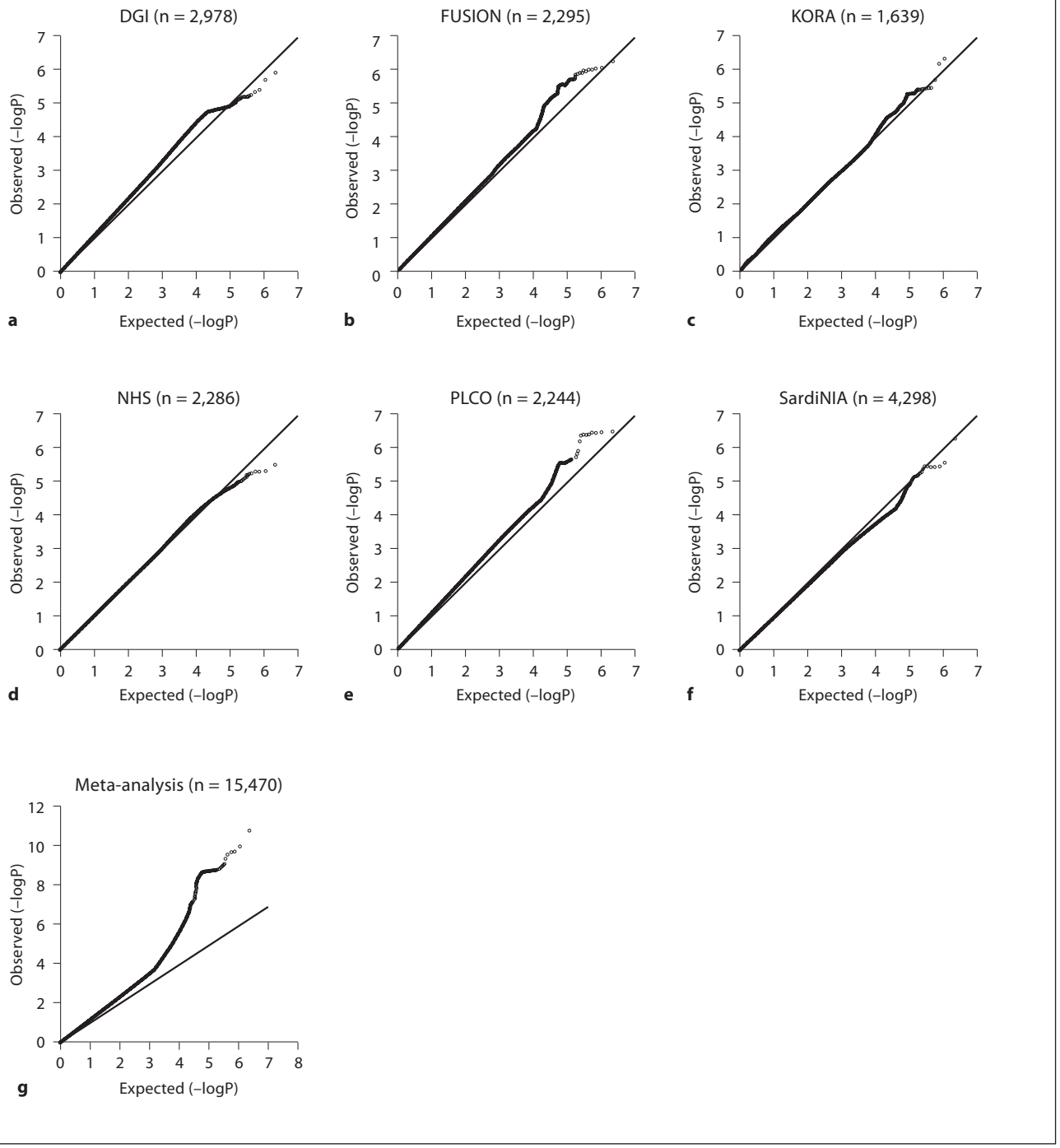

Fig. 1. Quantile–quantile (Q–Q) plots illustrate the power of collaboration and increased sample size through meta-analysis in uncovering novel loci associated with height. For six studies participating in the GIANT consortium, Q–Q plots are shown for each individual study (**a–f**) and then for a meta-analysis of all six studies (**g**). For each SNP in each Q–Q plot, the p values are ranked in descending order (increasing significance), and the negative $\log_{10}$ of the p value is plotted against the expected p value under the null model of no association (under the null model, the k-th lowest p value of a total of n is expected to be k/n). Points at the right end of a Q–Q plot that fall above the diagonal line have lower p values than expected by chance and hence are more likely to represent valid associations. In general, p values below the range of 10^{-7}–10^{-8} are also more likely to be validated. Reproduced, with permission, from Lettre et al. [52].

Table 1. Genetic loci for which common variants have been associated with height, their effect on adult height and their association with syndromes of abnormal skeletal growth

Chromosomal location	Nearby gene(s)	SNP(s)	Effect per height-increasing allele, cm	Human skeletal growth phenotype?	References
1p34	*SCMHI*	rs6686842	0.3		53
1p12	*SPAG17*	rs12735613	0.4		53
1q21	*SV2A, SF3B4, MTMR11,* Histone class 2A	rs11205277	0.3		54
1q24	*DNM3*	rs678962	0.3		54
1q25	*C1orf19, GLT25D2*	rs2274432	0.3		54
1q42	*ZNF678*	rs1390401	0.4		53
2p16	*EFEMP1*	rs3791679, rs3791675	0.4		53, 54
2q35	*IHH, NHEJ1*	rs6724465	0.4	Yes *(IHH)*	53
3q22	*ANAPC13, CEP63*	rs10935120	0.4		53
3q23	*ZBTB38*	rs724016, rs6440003, rs6763931	0.4		52–54
4p15	*LCORL*	rs6830062, rs16896068	0.4		53, 54
4q31	*HHIP*	rs1492820, rs6854783, rs1812175	0.4		52–54
6p24	*BMP6*	rs12198986	0.4		54
6p21	Histone class 1	rs10946808	0.4		52, 54
6p21	*HLA* class III	rs2844479, rs3130050, rs185819	0.4		54
6p21	*HMGA1*	rs1776897	0.6		54
6p21	*PPARD, FANCE, RPL10A, TEAD3*	rs4713858	0.4		54
6p21	*C6orf106*	rs2814993	0.6		53
6q21	*LIN28B*	rs314277	0.4		52
6q22	*LOC387103*	rs4549631	0.4		53
6q24	*GPR126*	rs4896582, rs3748069	0.4		52, 54
7p22	*GNA12*	rs798544	0.4		54
7q21	*CDK6*	rs2040494, rs2282978, rs2282978	0.3		52–54
8q12	*CHCHD7, RDHE2*	rs9650315, rs10958476	0.4		52, 54
8q21	*PXMP3, ZFHX4*	rs7846385	0.3		54
9q22	*PTCH1*	rs10512248	0.4		53
9q31	*ZNF462*	rs4743034	0.3		54
9q34	*FUBP3*	rs7466269	0.3		52, 53
12q14	*HMGA2*	rs1042725	0.4	Yes	42
12q22	*SOCS2*	rs11107116	0.3		53
13q14	*DLEU7*	rs3116602	0.4		53
14q32	*FBLN5, TRIP11, ATXN3*	rs8007661, rs7153027	0.4		52, 54
15q24	*SH3GL3, ADAMTSL3*	rs2562784, rs10906982	0.3		52, 53
15q26	*ADAMTS17*	rs4533267	0.4		54
15q26	*ACAN*	rs8041863	0.3	Yes	53
17q11	*ATAD5, C17orf42, CENTA2, RNF135*	rs3760318	0.4	Yes *(RNF135)*	54
17q23	*NOG, C17orf67, DGKE, TRIM25, COIL*	rs4794665	0.2	Yes *(NOG)*	54
17q23	*BCAS3, TBX2, C17orf82, TBX4*	rs757608	0.3	Yes *(TBX4)*	54
18q11	*RBBP8, CABLES1, C18orf45*	rs4800148	0.4		54
18q21	*DYM*	rs8099594	0.3	Yes	53
19p13	*DOT1L*	rs12986413	0.3		52
20p12	*BMP2*	rs967417	0.3	Yes	54
20q11	*GDF5*	rs6060369	0.4	Yes	47

cancer/cell-cycle-related genes *(CDK6, HMGA1, HMGA2, DLEU7, FUBP3, ANAPC13)*. Several of the genes found through these unbiased searches of the genome are also known to be mutated in syndromes of abnormal skeletal growth (table 1).

Together, these studies provided support for several ideas. First, height is clearly a polygenic trait, and there are common variants in the genome that can be shown to influence height reproducibly if large enough samples are studied. Second, several of these variants are in or near

Table 2. Potential basic, clinical and therapeutic implications

Many genetic variants are involved in regulating adult height, and genome-wide association studies have identified the first group of >40 such variants
These variants as yet provide limited predictive power but highlight biological pathways involved in human skeletal growth
Some pathways were previously unsuspected as playing a role in growth, and therefore have revealed novel biology
Many of the genes highlighted by genome-wide association studies are also mutated in syndromes of abnormal skeletal growth, suggesting that the genes from genome-wide association studies are also candidates for skeletal dysplasias or other syndromes of abnormal skeletal growth for which there is not yet any known genetic aetiology

genes that also underlie syndromes with short or tall stature or skeletal dysplasia as main features (table 1), indicating that the same loci can harbour both common variants with mild effects and rare variants with stronger phenotypic consequences. Third, these new genes outline several pathways, including some not previously strongly implicated in human growth, demonstrating that unbiased genetic approaches can generate new biological insights and hypotheses. Finally, some of the height variants had additional phenotypic consequences, suggesting that discovering functional variants through their effect on height could help studies of other traits and diseases.

Importantly, these initial studies are just the beginning of this phase of discovery. The variants discovered explain perhaps 5% of the total variation in adult height. Because heritability, the fraction of variation explained by inherited factors, is estimated at 80–90%, genetic factors explaining up to 75–85% of population variation remain to be identified. One consequence is that these SNPs do not accurately predict height [52, 53]. As much larger datasets become available (efforts to perform a meta-analysis of over 100,000 genome-wide association studies is underway), more loci will emerge. Furthermore, the genes identified so far likely harbour additional variants (common or rare) that influence height. Finally, no-one has yet investigated the role of copy number variation in height. These efforts will no doubt raise the fraction of height variation that can be explained by genetic factors, but it remains to be seen whether genetic variants discovered in the near future will together provide substantial predictive information.

Next Steps

As described above, it is very possible that some of the newly identified genes also underlie syndromes of abnormal skeletal growth where the responsible gene has not yet been identified. A logical next step would be to resequence these genes (or at a minimum the exons of these genes) in patients with such syndromes and appropriate controls to see if severe mutations in these genes are responsible for a particular clinical picture. In addition, resequencing of these genes in healthy individuals with different heights would help establish whether additional variation in these genes plays a role in height in the general population. Finally, more work is needed to discover new height-related genes, including larger consortia, additional analyses of new and existing data sets (e.g. tests of copy number variants, sex-specific, recessive or population-specific analyses, analyses of effects on childhood growth, birth length and other clinical phenotypes, and searches for gene–gene interactions).

The past year has seen remarkable progress in understanding the genetics of human growth. As the number of new height loci continues to increase, so too will our understanding of the inherited basis of stature and the biology of human growth. It remains to be seen how much of human height can be accounted for by genetic variants discovered using these approaches, or whether these variants have other predictive value, such as predicting response to GH treatment. However, prediction, in our opinion, is not the primary goal of these studies. Rather, the main impact of these genetic discoveries is that they uncover previously unsuspected genes as playing a role in human growth, expanding our understanding of this basic phenomenon (table 2). In addition, the lessons learned in studies of height, a model polygenic trait, may be valuable in informing studies of other quantitative traits and common diseases.

Disclosure Statement

Joel Hirschhorn is on the Scientific Advisory Board of Correlagen, Inc.

References

1 Galton F: Regression towards mediocrity in hereditary stature. J Anthropol Inst 1886;15:246–263.
2 Pearson K, Lee A: On the laws of inheritance in man. I. Inheritance of physical characteristics. Biometrika 1903;2:357–462.
3 Fisher RA: The correlation between relatives on the supposition of Mendelian inheritance. Trans R Soc Edinb 1918;52:399–433.
4 Perola M, Sammalisto S, Hiekkalinna T, Martin NG, Visscher PM, Montgomery GW, Benyamin B, Harris JR, Boomsma DI, Willemsen G, Hottenga J-J, Christensen K, Kyvik KO, Sørensen TI, Pedersen NL, Magnusson PK, Spector TD, Widen E, Silventoinen K, Kaprio J, Palotie A, Peltonen L: Combined genome scans for body stature in 6,602 European twins: evidence for common Caucasian loci. PLoS Genet 2007;3:e97.
5 Silventoinen K, Sammalisto S, Perola M, Boomsma DI, Cornes BK, Davis C, Dunkel L, de Lange M, Harris JR, Hjelmborg JV, Luciano M, Martin NG, Mortensen J, Nisticò L, Pedersen NL, Skytthe A, Spector TD, Stazi MA, Willemsen G, Kaprio J: Heritability of adult body height: a comparative study of twin cohorts in eight countries. Twin Res 2003;6:399–408.
6 Palmert MR, Hirschhorn JN: Genetic approaches to stature, pubertal timing, and other complex traits. Mol Genet Metab 2003;80:1–10.
7 Phillips JA, III, Hjelle BL, Seeburg PH, Zachmann M: Molecular basis for familial isolated growth hormone deficiency. Proc Natl Acad Sci USA 1981;78:6372–6375.
8 Amselem S, Duquesnoy P, Attree O, Novelli G, Bousnina S, Postel-Vinay MC, Goossens M: Laron dwarfism and mutations of the growth hormone-receptor gene. N Engl J Med 1989;321:989–995.
9 Godowski PJ, Leung DW, Meacham LR, Galgani JP, Hellmiss R, Keret R, Rotwein PS, Parks JS, Laron Z, Wood WI: Characterization of the human growth hormone receptor gene and demonstration of a partial gene deletion in two patients with Laron-type dwarfism. Proc Natl Acad Sci USA 1989;86:8083–8087.
10 Wajnrajch MP: Genetic disorders of human growth. J Pediatr Endocrinol Metab 2002;15(suppl 2):701–714.
11 Kofoed EM, Hwa V, Little B, Woods KA, Buckway CK, Tsubaki J, Pratt KL, Bezrodnik L, Jasper H, Tepper A, Heinrich JJ, Rosenfeld RG: Growth hormone insensitivity associated with a *STAT5B* mutation. N Engl J Med 2003;349:1139–1147.
12 Domené HM, Bengolea SV, Martínez AS, Ropelato MG, Pennisi P, Scaglia P, Heinrich JJ, Jasper HG: Deficiency of the circulating insulin-like growth factor system associated with inactivation of the acid-labile subunit gene. N Engl J Med 2004;350:570–577.
13 Rimoin DL, Cohn D, Krakow D, Wilcox W, Lachman RS, Alanay Y: The skeletal dysplasias: clinical–molecular correlations. Ann NY Acad Sci 2007;1117:302–309.
14 Superti-Furga A, Unger S: Nosology and classification of genetic skeletal disorders: 2006 revision. Am J Med Genet A 2007;143:1–18.
15 Hirschhorn JN, Lindgren CM, Daly MJ, Kirby A, Schaffner SF, Burtt NP, Altshuler D, Parker A, Rioux JD, Platko J, Gaudet D, Hudson TJ, Groop LC, Lander ES: Genomewide linkage analysis of stature in multiple populations reveals several regions with evidence of linkage to adult height. Am J Hum Genet 2001;69:106–116.
16 Perola M, Öhman M, Hiekkalinna T, Leppavuori J, Pajukanta P, Wessman M, Koskenvuo M, Palotie A, Lange K, Kaprio J, Peltonen L: Quantitative-trait-locus analysis of body-mass index and of stature, by combined analysis of genome scans of five Finnish study groups. Am J Hum Genet 2001;69:117–123.
17 Deng H-W, Xu F-H, Liu Y-Z, Shen H, Deng H, Huang Q-Y, Liu Y-J, Conway T, Li J-L, Davies KM, Recker RR: A whole-genome linkage scan suggests several genomic regions potentially containing QTLS underlying the variation of stature. Am J Med Genet 2002;113:29–39.
18 Wiltshire S, Frayling TM, Hattersley AT, Hitman GA, Walker M, Levy JC, O'Rahilly S, Groves CJ, Menzel S, Cardon LR, McCarthy MI: Evidence for linkage of stature to chromosome 3p26 in a large U.K. family data set ascertained for type 2 diabetes. Am J Hum Genet 2002;70:543–546.
19 Liu Y-Z, Xu F-H, Shen H, Deng H, Liu Y-J, Zhao L-J, Dvornyk V, Conway T, Li J-L, Huang Q-Y, Davies KM, Recker RR, Deng H-W: Confirmation linkage study in support of the X chromosome harbouring a QTL underlying human height variation. J Med Genet 2003;40:825–831.
20 Hirschhorn JN, Daly MJ: Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 2005;6:95–108.
21 McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN: Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. Nat Rev Genet 2008;9:356–369.
22 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.
23 Newton-Cheh C, Hirschhorn JN: Genetic association studies of complex traits: design and analysis issues. Mutat Res 2005;573:54–69.
24 Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N: Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. J Natl Cancer Inst 2004;96:434–442.
25 Campbell CD, Ogburn EL, Lunetta KL, Lyon HN, Freedman ML, Groop LC, Altshuler D, Ardlie KG, Hirschhorn JN: Demonstrating stratification in a European American population. Nat Genet 2005;37:868–872.
26 Clayton DG, Walker NM, Smyth DJ, Pask R, Cooper JD, Maier LM, Smink LJ, Lam AC, Ovington NR, Stevens HE, Nutland S, Howson JM, Faham M, Moorhead M, Jones HB, Falkowski M, Hardenbol P, Willis TD, Todd JA: Population structure, differential bias and genomic control in a large-scale, case-control association study. Nat Genet 2005;37:1243–1246.
27 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904–909.
28 Price AL, Butler J, Patterson N, Capelli C, Pascali VL, Scarnicci F, Ruiz-Linares A, Groop L, Saetta AA, Korkolopoulou P, Seligsohn U, Waliszewska A, Schirmer C, Ardlie K, Ramos A, Nemesh J, Arbeitman L, Goldstein DB, Reich D, Hirschhorn JN: Discerning the ancestry of European Americans in genetic association studies. PLoS Genet 2008;4:e236.
29 Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P: A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 2007;445:881–885.
30 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 M-R, 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, Smith GD, Hattersley AT, McCarthy MI: A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. Science 2007;316:889–894.

31 Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson Boström K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Råstam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjögren M, Sterner M, Surti A, Svensson M, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn G-W, Ma Q, Parikh H, Richardson D, Ricke D, Purcell S: Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 2007;316:1331–1336.

32 Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding C-J, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li X-Y, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M: A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 2007; 316:1341–1345.

33 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, Hattersley AT: Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 2007;316:1336–1341.

34 Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, Healey CS, Bowman R, Meyer KB, Haiman CA, Kolonel LK, Henderson BE, Le Marchand L, Brennan P, Sangrajrang S, Gaborieau V, Odefrey F, Shen C-Y, Wu P-E, Wang H-C, Eccles D, Evans DG, Peto J, Fletcher O, Johnson N, Seal S, Stratton MR, Rahman N, Chenevix-Trench G, Bojesen SE, Nordestgaard BG, Axelsson CK, Garcia-Closas M, Brinton L, Chanock S, Lissowska J, Peplonska B, Nevanlinna H, Fagerholm R, Eerola H, Kang D, Yoo K-Y, Noh D-Y, Ahn S-H, Hunter DJ, Hankinson SE, Cox DG, Hall P, Wedren S, Liu J, Low Y-L, Bogdanova N, Schürmann P, Dörk T, Tollenaar RA, Jacobi CE, Devilee P, Klijn JG, Sigurdson AJ, Doody MM, Alexander BH, Zhang J, Cox A, Brock IW, MacPherson G, Reed MW, Couch FJ, Goode EL, Olson JE, Meijers-Heijboer H, van den Ouweland A, Uitterlinden A, Rivadeneira F, Milne RL, Ribas G, Gonzalez-Neira A, Benitez J, Hopper JL, McCredie M, Southey M, Giles GG, Schroen C, Justenhoven C, Brauch H, Hamann U, Ko Y-D, Spurdle AB, Beesley J, Chen X, Mannermaa A, Kosma V-M, Kataja V, Hartikainen J, Day NE, Cox DR, Ponder BA: Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007;447:1087–1093.

35 Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, Bailey R, Nejentsev S, Field SF, Payne F, Lowe CE, Szeszko JS, Hafler JP, Zeitels L, Yang JH, Vella A, Nutland S, Stevens HE, Schuilenburg H, Coleman G, Maisuria M, Meadows W, Smink LJ, Healy B, Burren OS, Lam AA, Ovington NR, Allen J, Adlem E, Leung H-T, Wallace C, Howson JM, Guja C, Ionescu-Tîrgovişte C, Simmonds MJ, Heward JM, Gough SC, Dunger DB, Wicker LS, Clayton DG: Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat Genet 2007;39:857–864.

36 Menzel S, Garner C, Gut I, Matsuda F, Yamaguchi M, Heath S, Foglio M, Zelenika D, Boland A, Rooks H, Best S, Spector TD, Farrall M, Lathrop M, Thein SL: A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. Nat Genet 2007;39:1197–1199.

37 Scuteri A, Sanna S, Chen W-M, Uda M, Albai G, Strait J, Najjar S, Nagaraja R, Orrú M, Usala G, Dei M, Lai S, Maschio A, Busonero F, Mulas A, Ehret GB, Fink AA, Weder AB, Cooper RS, Galan P, Chakravarti A, Schlessinger D, Cao A, Lakatta E, Abecasis GR: Genome-wide association scan shows genetic variants in the *FTO* gene are associated with obesity-related traits. PLoS Genet 2007;3: e115.

38 Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S, Depner M, von Berg A, Bufe A, Rietschel E, Heinzmann A, Simma B, Frischer T, Willis-Owen SA, Wong KC, Illig T, Vogelberg C, Weiland SK, von Mutius E, Abecasis GR, Farrall M, Gut IG, Lathrop GM, Cookson WO: Genetic variants regulating *ORMDL3* expression contribute to the risk of childhood asthma. Nature 2007;448: 470–473.

39 Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, Wacholder S, Wang Z, Welch R, Hutchinson A, Wang J, Yu K, Chatterjee N, Orr N, Willett WC, Colditz GA, Ziegler RG, Berg CD, Buys SS, McCarty CA, Feigelson HS, Calle EE, Thun MJ, Hayes RB, Tucker M, Gerhard DS, Fraumeni JF Jr, Hoover RN, Thomas G, Chanock SJ: A genome-wide association study identifies alleles in *FGFR2* associated with risk of sporadic postmenopausal breast cancer. Nat Genet 2007;39:870–874.

40 Hakonarson H, Grant SF, Bradfield JP, Marchand L, Kim CE, Glessner JT, Grabs R, Casalunovo T, Taback SP, Frackelton EC, Lawson ML, Robinson LJ, Skraban R, Lu Y, Chiavacci RM, Stanley CA, Kirsch SE, Rappaport EF, Orange JS, Monos DS, Devoto M, Qu H-Q, Polychronakos C: A genome-wide association study identifies *KIAA0350* as a type 1 diabetes gene. Nature 2007;448:591–594.

41 Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, Helgason A, Rafnar T, Bergthorsson JT, Agnarsson BA, Baker A, Sigurdsson A, Benediktsdottir KR, Jakobsdottir M, Xu J, Blondal T, Kostic J, Sun J, Ghosh S, Stacey SN, Mouy M, Saemundsdottir J, Backman VM, Kristjansson K, Tres A, Partin AW, Albers-Akkers MT, Godino-Ivan Marcos J, Walsh PC, Swinkels DW, Navarrete S, Isaacs SD, Aben KK, Graif T, Cashy J, Ruiz-Echarri M, Wiley KE, Suarez BK, Witjes JA, Frigge M, Ober C, Jonsson E, Einarsson GV, Mayordomo JI, Kiemeney LA, Isaacs WB, Catalona WJ, Barkardottir RB, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K: Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat Genet 2007;39:631–637.

42 Weedon MN, Lettre G, Freathy RM, Lindgren CM, Voight BF, Perry JR, Elliott KS, Hackett R, Guiducci C, Shields B, Zeggini E, Lango H, Lyssenko V, Timpson NJ, Burtt NP, Rayner NW, Saxena R, Ardlie K, Tobias JH, Ness AR, Ring SM, Palmer CN, Morris AD, Peltonen L, Salomaa V, Davey Smith G, Groop LC, Hattersley AT, McCarthy MI, Hirschhorn JN, Frayling TM: A common variant of *HMGA2* is associated with adult and childhood height in the general population. Nat Genet 2007;39:1245–1250.

43 Zhou X, Benson KF, Ashar HR, Chada K: Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C. Nature 1995;376:771–774.

44 Ligon AH, Moore SD, Parisi MA, Mealiffe ME, Harris DJ, Ferguson HL, Quade BJ, Morton CC: Constitutional rearrangement of the architectural factor *HMGA2*: a novel human phenotype including overgrowth and lipomas. Am J Hum Genet 2005;76:340–348.

45 Menten B, Buysse K, Zahir F, Hellemans J, Hamilton SJ, Costa T, Fagerstrom C, Anadiotis G, Kingsbury D, McGillivray BC, Marra MA, Friedman JM, Speleman F, Mortier G: Osteopoikilosis, short stature and mental retardation as key features of a new microdeletion syndrome on 12q14. J Med Genet 2007;44:264–268.

46 Fusco A, Fedele M: Roles of HMGA proteins in cancer. Nat Rev Cancer 2007;7:899–910.

47 Sanna S, Jackson AU, Nagaraja R, Willer CJ, Chen W-M, Bonnycastle LL, Shen H, Timpson N, Lettre G, Usala G, Chines PS, Stringham HM, Scott LJ, Dei M, Lai S, Albai G, Crisponi L, Naitza S, Doheny KF, Pugh EW, Ben-Shlomo Y, Ebrahim S, Lawlor DA, Bergman RN, Watanabe RM, Uda M, Tuomilehto

J, Coresh J, Hirschhorn JN, Shuldiner AR, Schlessinger D, Collins FS, Smith GD, Boerwinkle E, Cao A, Boehnke M, Abecasis GR, Mohlke KL: Common variants in the *GDF5-UQCC* region are associated with variation in human height. Nat Genet 2008;40:198–203.

48 Polinkovsky A, Robin NH, Thomas JT, Irons M, Lynn A, Goodman FR, Reardon W, Kant SG, Brunner HG, van der Burgt I, Chitayat D, McGaughran J, Donnai D, Luyten FP, Warman ML: Mutations in *CDMP1* cause autosomal dominant brachydactyly type C. Nat Genet 1997;17:18–19.

49 Thomas JT, Kilpatrick MW, Lin K, Erlacher L, Lembessis P, Costa T, Tsipouras P, Luyten FP: Disruption of human limb morphogenesis by a dominant negative mutation in *CDMP1*. Nat Genet 1997;17:58–64.

50 Thomas JT, Lin K, Nandedkar M, Camargo M, Cervenka J, Luyten FP: A human chondrodysplasia due to a mutation in a TGF-β superfamily member. Nat Genet 1996;12:315–317.

51 Miyamoto Y, Mabuchi A, Shi D, Kubo T, Takatori Y, Saito S, Fujioka M, Sudo A, Uchida A, Yamamoto S, Ozaki K, Takigawa M, Tanaka T, Nakamura Y, Jiang Q, Ikegawa S: A functional polymorphism in the 5′ UTR of *GDF5* is associated with susceptibility to osteoarthritis. Nat Genet 2007;39:529–533.

52 Lettre G, Jackson AU, Gieger C, Schumacher FR, Berndt SI, Sanna S, Eyheramendy S, Voight BF, Butler JL, Guiducci C, Illig T, Hackett R, Heid IM, Jacobs KB, Lyssenko V, Uda M, Boehnke M, Chanock SJ, Groop LC, Hu FB, Isomaa B, Kraft P, Peltonen L, Salomaa V, Schlessinger D, Hunter DJ, Hayes RB, Abecasis GR, Wichmann HE, Mohlke KL, Hirschhorn JN: Identification of ten loci associated with height highlights new biological pathways in human growth. Nat Genet 2008;40:584–591.

53 Weedon MN, Lango H, Lindgren CM, Wallace C, Evans DM, Mangino M, Freathy RM, Perry JR, Stevens S, Hall AS, Samani NJ, Shields B, Prokopenko I, Farrall M, Dominiczak A, Johnson T, Bergmann S, Beckmann JS, Vollenweider P, Waterworth DM, Mooser V, Palmer CN, Morris AD, Ouwehand WH, Zhao JH, Li S, Loos RJ, Barroso I, Deloukas P, Sandhu MS, Wheeler E, Soranzo N, Inouye M, Wareham NJ, Caulfield M, Munroe PB, Hattersley AT, McCarthy MI, Frayling TM: Genome-wide association analysis identifies 20 loci that influence adult height. Nat Genet 2008;40:575–583.

54 Gudbjartsson DF, Walters GB, Thorleifsson G, Stefansson H, Halldorsson BV, Zusmanovich P, Sulem P, Thorlacius S, Gylfason A, Steinberg S, Helgadottir A, Ingason A, Steinthorsdottir V, Olafsdottir EJ, Olafsdottir GH, Jonsson T, Borch-Johnsen K, Hansen T, Andersen G, Jorgensen T, Pedersen O, Aben KK, Witjes JA, Swinkels DW, Heijer MD, Franke B, Verbeek AL, Becker DM, Yanek LR, Becker LC, Tryggvadottir L, Rafnar T, Gulcher J, Kiemeney LA, Kong A, Thorsteinsdottir U, Stefansson K: Many sequence variants affecting diversity of adult human height. Nat Genet 2008;40:609–615.