Human Molecular Genetics, 2012, Vol. 21, No. 23 **5193–5201**
doi:10.1093/hmg/dds347
Advance Access published on August 21, 2012

Synthesizing genome-wide association studies and expression microarray reveals novel genes that act in the human growth plate to modulate height

Julian C. Lui[1,*], Ola Nilsson[2], Yingleong Chan[3,4,5], Cameron D. Palmer[3,4,5], Anenisia C. Andrade[2], Joel N. Hirschhorn[3,4,5] and Jeffrey Baron[1]

[1]Program in Developmental Endocrinology and Genetics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA, [2]Center for Molecular Medicine and Pediatric Endocrinology Unit, Department of Women's and Children's Health, Karolinska Institutet and Karolinska University Hospital, SE-171 76 Stockholm, Sweden, [3]Divisions of Genetics and Endocrinology and Program in Genomics, Children's Hospital, Boston, MA 02115, USA, [4]Metabolism Initiative and Program in Medical and Population Genetics, Broad Institute, Cambridge, MA 02142, USA and [5]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

Received July 20, 2012; Revised August 13, 2012; Accepted August 15, 2012

Previous meta-analysis of genome-wide association (GWA) studies has identified 180 loci that influence adult height. However, each GWA locus typically comprises a set of contiguous genes, only one of which presumably modulates height. We reasoned that many of the causative genes within these loci influence height because they are expressed in and function in the growth plate, a cartilaginous structure that causes bone elongation and thus determines stature. Therefore, we used expression microarray studies of mouse and rat growth plate, human disease databases and a mouse knockout phenotype database to identify genes within the GWAS loci that are likely required for normal growth plate function. Each of these approaches identified significantly more genes within the GWA height loci than at random genomic locations ($P < 0.0001$ each), supporting the validity of the approach. The combined analysis strongly implicates 78 genes in growth plate function, including multiple genes that participate in PTHrP-IHH, BMP and CNP signaling, and many genes that have not previously been implicated in the growth plate. Thus, this analysis reveals a large number of novel genes that regulate human growth plate chondrogenesis and thereby contribute to the normal variations in human adult height. The analytic approach developed for this study may be applied to GWA studies for other common polygenic traits and diseases, thus providing a new general strategy to identify causative genes within GWA loci and to translate genetic associations into mechanistic biological insights.

INTRODUCTION

To identify loci associated with adult height, a meta-analysis of genome-wide association (GWA) data was previously performed, which comprises 46 studies and 133 653 individuals (1). This meta-analysis detected 207 loci. Of these, 180 were further verified in a subsequent analysis (1). Each locus presumably implies the presence of a nearby gene (or genes) that plays a biological role in human linear growth. However, at almost all loci, there are multiple genes that are sufficiently close to account for the linkage to height, and it is unclear which of these genes actually modulates height and which are innocent neighbors. For a few loci, there is an obvious candidate gene based on previous biological or clinical data, but for most of the loci, there is not.

Height is primarily determined by the elongation of the long bones and vertebrae, a process that occurs at the cartilaginous growth plate. Bone elongation at the growth plate requires orderly spatial differentiation of chondrocytes from the progenitor cells of the resting zone to the transit amplifying

*To whom correspondence should be addressed at: CRC, Room 1-3330, 10 Center Drive, MSC 1103, Bethesda, MD 20892-1103, USA. Tel: +1 3014968049; Fax: +1 3014020574; Email: luichunk@mail.nih.gov

Published by Oxford University Press 2012.

Downloaded from https://academic.oup.com/hmg/article/21/23/5193/593394 by guest on 17 April 2024

cells of the proliferative zone and to the terminally differentiated cells of the hypertrophic zone (2). In addition, the growth plate undergoes important temporal development—embryonic formation, followed by a gradual, postnatal, programmed involution and loss of function, which causes the rate of bone growth to decrease and approach zero as the organism approaches its adult size, a process known as growth plate senescence (3). Some genes influence these growth plate processes indirectly, through circulating factors such as growth hormone (GH) and insulin-like growth factor-1 (IGF1), which regulate the growth plate (4). However, prior studies in cell culture, animal models and human diseases indicate that a large number of genes are expressed in the growth plate and act locally to influence bone growth and therefore overall body length (5).

Because growth plate function is the primary determinant of height, we reasoned that many of the causative genes within the height-associated GWA loci are expressed and function in the growth plate. Therefore, to help distinguish causative genes from the many non-causative neighbors in the GWA loci, we searched the loci for genes that are expressed in the growth plate and act there to influence chondrogenesis.

RESULTS

First, we generated a list of 833 genes (GWA list, Supplementary Material, Table S1) in close linkage proximity (see Materials and Methods) to the 207 loci defined by the GWA meta-analysis (180 published loci and 27 unpublished loci, Table 1) (1). This list presumably comprises genes that are involved locally in growth plate chondrogenesis, genes that indirectly affect growth plate function through systemic mechanisms, e.g. growth-regulating hormones, and many genes that do not influence growth but are simply located in close proximity to causative genes. To help identify those genes that act locally in the growth plate, we first used microarray analysis to detect (i) genes that have higher expression in growth plate cartilage than in other tissues; (ii) genes that are spatially regulated across different zones in the growth plate, suggesting a role in chondrocyte differentiation and/or proliferation; or (iii) genes that are temporally regulated in the growth plate, suggesting a role in growth plate senescence.

Affymetrix Mouse Genome Array 430 2.0 was used to compare expression between the whole growth plate and three different soft tissues (lung, kidney, heart) in 1-week-old male mice ($n = 5$) (6,7). This analysis identified 1754 genes that are more highly expressed ($\geq$2.0-fold) in the growth plate than in all three soft tissues [false discovery rate (FDR) $<$1%]. Similarly, spatially or temporally regulated genes in the growth plate were identified by analyzing rat microarray data (Affymetrix Rat Genome Array 230 2.0) previously collected from microdissected growth plate (8). We found 1564 genes that are spatially regulated across different zones (resting versus proliferative or proliferative versus hypertrophic, $\geq$2.0-fold, FDR $<$ 5%, 1-week-old male rats, $n = 5$) and 309 genes that are temporally regulated as growth slows with age (3-week-old versus 12-week-old proliferative zone from male rats, $\geq$2.0-fold, FDR $<$ 5%, $n = 5$).

A comparison among these three microarray analyses showed evidence of internal consistency (Fig. 1). The overlap among the three gene sets was greater than that expected by chance (Pearson's χ^2 test, $P < 0.0001$ for all three pairs of gene sets), indicating that these analyses tend to identify similar genes. We therefore focused our attention on genes that met at least two of these three criteria, in order to decrease the likelihood of spurious associations. Of the 13 254 genes in the genome that are annotated in both the rat and mouse microarray, 427 genes (GP expression set, Supplementary Material, Table S2) met at least two of the three expression criteria.

If the expression analysis correctly identified genes that function in the growth plate and if human height variation is caused by polymorphisms in genes that also affect growth plate function, then the GWA gene list should be enriched in GP expression genes. Furthermore, the genes that lie closest to the single-nucleotide polymorphism (SNP) identified by GWA should show the greatest enrichment because these genes are the most likely to account for the association with height. Consistent with these predictions, of the 427 GP expression genes, 38 were found on the GWA list, and 22 of these represented the gene closest to the SNP identified by GWA (Table 2, top panel; Supplementary Material, Table S3). These overlap frequencies indicate strong enrichment of the GP expression genes in the GWA list compared with the whole genome (genome, 3.2%; GWA list, 8.5%; GWA-closest subset, 15.7%), much greater than would be expected by chance ($P < 0.0001$, Pearson's χ^2 test, Table 2, top panel).

To supplement the expression analysis and identify additional genes from the GWA list that are likely required for growth plate function, we used bioinformatic searches to find genes for which mutations cause a growth plate phenotype in mice. To this end, we first utilized the Mouse Genome Informatics (MGI) database (9) to generate a list of 224 genes that have a skeletal phenotype with underlying growth plate dysfunction in knockout animals (mouse GP phenotype genes, see Materials and Methods and Supplementary Material, Table S4). We then asked whether these genes are enriched in the GWA list. We found that of the 224 mouse GP phenotype genes, 38 genes were found in the GWA list, 24 of which were either closest to or actually contained the sequence variation identified by GWA. Thus, the GWA list showed striking enrichment for these mouse GP phenotype genes compared with the whole genome (Table 2, middle panel), and the subset of GWA genes that lie closest to the height-associated SNPs showed even greater enrichment (genome, 1.0%; GWA list, 4.6%; GWA closest subset, 11.6%, $P < 0.0001$).

Similarly, we searched for genes in which mutations cause a growth plate phenotype in humans. Combining data from the Online Mendelian Inheritance in Man (OMIM) database and a list of genes that cause human skeletal dysplasias, recently published by the International Skeletal Dysplasia Society (10), we generated a list of 173 genes (human GP phenotype genes) in which mutations cause abnormal skeletal growth in humans and which are likely due to underlying local growth plate dysfunction (see Materials and Methods and Supplementary Material, Table S5). Genes that impair growth through a

Downloaded from https://academic.oup.com/hmg/article/21/23/5193/593394 by guest on 17 April 2024

Table 1. Association results for the 27 additional independent signals that were included both in our analyses and in the previously published pathway analysis (1,30)

SNP	Chr.	Position[a]	Nearest gene (distance from SNP)	A1/A2	A1 frequency[b]	Stage 1 *P*-value	Number of samples
rs12047268	1	103 246 082	*COL11A1* (intragenic)	C/G	0.24	3.53E − 07	130 616
rs2120003	1	145 157 259	*FMO5* (intragenic)	T/C	0.18	1.42E − 06	133 531
rs12086448	1	158 660 529	*VANGL2* (intragenic)	A/G	0.47	1.87E − 06	132 037
rs9425569	1	181 208 825	*SHCBP1L* (20 kb)	G/A	0.45	2.65E − 06	130 500
rs7601531	2	19 831 425	*FLJ12334* (101 kb)	T/C	0.65	4.35E − 07	131 995
rs2166898	2	121 329 129	*GLI2* (intragenic)	G/A	0.83	3.99E − 06	133 362
rs6772112	3	11 616 535	*VGLL4* (intragenic)	T/C	0.94	1.58E − 06	133 633
rs2718423	3	115 691 287	*ZBTB20* (intragenic)	C/T	0.82	4.49E − 06	133 834
rs9818941	3	159 169 151	*SHOX2* (127 kb)	G/A	0.22	2.13E − 07	133 848
rs2634464	4	56 225 553	*NMU* (28 kb)	G/A	0.25	3.50E − 06	130 488
rs6824258	4	122 989 417	*BBS7* (intragenic)	C/T	0.27	9.84E − 07	133 758
rs17017854	4	144 681 963	*SMARCA5* (intragenic)	C/G	0.12	3.99E − 06	132 053
rs34651	5	72 179 761	*TNPO1* (intragenic)	C/T	0.10	2.50E − 07	133 760
rs12374649	6	27 451 705	*ZNF204P* (0.6 kb)	C/T	0.93	1.72E − 06	128 313
rs16876369	6	47 790 908	*GPR115* (intragenic)	C/T	0.97	3.81E − 06	128 275
rs9459531	6	166 253 789	*C6orf176* (4 kb)	A/G	0.47	1.38E − 06	133 555
rs17172694	7	46 403 679	*IGFBP3* (476 kb)	G/T	0.96	1.88E − 07	133 795
rs4738736	8	59 998 555	*TOX* (intragenic)	T/C	0.26	1.04E − 06	133 742
rs16892729	8	120 627 259	*ENPP2* (11 kb)	C/T	0.86	1.29E − 06	133 794
rs17122670	12	58 263 968	*SLC16A7* (105 kb)	A/G	0.11	4.40E − 06	133 342
rs2408058	12	88 776 631	*LOC338758* (147 kb)	G/A	0.78	2.20E − 08	132 285
rs11835818	12	120 979 192	*BCL7A* (intragenic)	C/T	0.44	3.59E − 06	133 803
rs316618	15	39 583 790	*LTK* (intragenic)	T/A	0.84	1.00E − 06	132 039
rs11855027	15	57 313 211	*MYO1E* (intragenic)	G/T	0.54	3.79E − 06	133 831
rs528045	17	35 840 398	*TOP2A* (13 kb)	G/A	0.79	2.98E − 06	133 807
rs2158917	17	67 437 704	*SOX9* (191 kb)	T/C	0.22	2.37E − 06	133 797
rs913000	20	54 269 761	*MC3R* (11 kb)	T/C	0.34	1.54E − 06	133 779

These SNPs had strongly suggestive evidence of association to height [$P < 5 \times 10^{-8}$ before a conservative double-genomic control correction (31), $P < 5 \times 10^{-6}$ after double correction] in the initial stage 1 meta-analysis of Lango *et al.* (1).
Chr., chromosome; A1, allele 1; A2, allele 2.
[a]SNP basepair position on Build 36.
[b]Allele frequency in the total sample.

systemic mechanism (such as GH deficiency) were excluded. Comparing the lists of human GP phenotype genes and GWA genes, we found that 27 of 173 human GP phenotype genes were present in the GWA list (Table 2, bottom panel). This represents a strong enrichment of human GP phenotype genes in loci associated with adult height (27 out of 833 GWA genes, 3.2%) compared with the whole genome (173 out of approximately 23 000 genes in genome, 0.75%). The enrichment is again even stronger in the subset of GWA genes that lie closest to the GWA sequence variant (14 out of 207 GWA closest subset, 6.8%, $P < 0.0001$). The current algorithm, which focused on the growth plate, identified more genes in these loci with a human growth phenotype than prior OMIM-based analysis (1).

We further tested whether the GWA list was enriched with growth plate-related genes by permutation. We generated 1000 gene sets, each comprising 207 random loci designed to match the GWA list and then counted the number of GP expression genes, mouse GP phenotype genes and human GP phenotype genes in each set. These 1000 matched gene sets contained a median of 17 GP expression genes (range 4–30), far fewer than the observed number of 38 GP expression genes present in the GWA list ($P < 0.001$: 0 of 1000 matching sets contained 38 or more GP expression genes; Supplementary Material, Table S6 and Fig. S1). Similarly, the 1000 permutations contained significantly fewer mouse GP phenotype genes (median = 11) and human GP phenotype genes (median = 7) than did the GWA list (both $P < 0.001$, Supplementary Material, Table S6 and Fig. S1). A similar comparison using percentage overlap, rather than absolute number of overlapping genes, also demonstrated enrichment of GP expression genes, mouse GP phenotype genes and human GP phenotype genes in the GWA list (all $P < 0.001$).

Together, these three different approaches implicated 78 genes from the GWA meta-analysis as having a growth plate function based on expression patterns, mouse knockout phenotypes or human skeletal diseases (Supplementary Material, Table S7). Remarkably, 46 of the 78 genes are located closest to the GWA sequence variant, and in 24 of these cases, the height-associated SNP is located within the growth plate gene itself. This distribution differs significantly from a random allocation (Fig. 2, $P < 0.0001$, χ^2 test).

A further confirmation of the validity of this approach is that these 78 genes were distributed among the 207 loci non-randomly. Each locus is likely to contain only one gene that affects linear growth and contributes to height variation, whereas other genes in that locus are likely innocent neighbors. Ideally, therefore each of the 78 genes implicated in growth plate function would reside in a different locus. Of the 78 growth plate genes, 64 of them occupy separate loci,

Downloaded from https://academic.oup.com/hmg/article/21/23/5193/593394 by guest on 17 April 2024

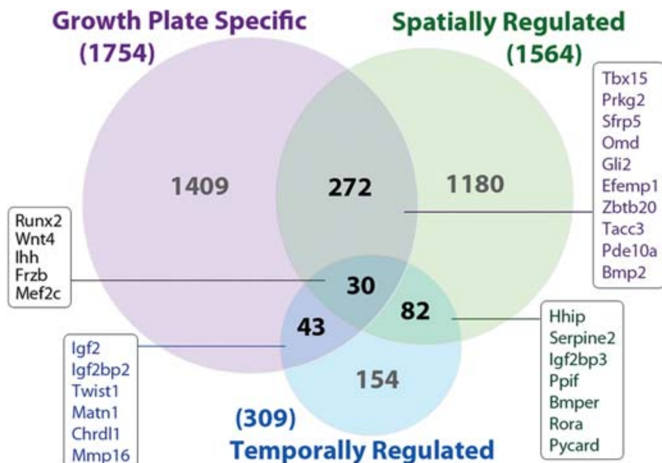

Figure 1. Venn diagram showing the number of genes with expression that is growth plate-specific, spatially regulated in growth plate or temporally regulated in growth plate by microarray. 'Growth plate specific' indicates expression in growth plate ≥2.0-fold greater than kidney, lung and heart in 1-week-old mouse ($n = 5$, FDR < 1%, Genome Array 430 2.0, Affymetrix). 'Spatially regulated' indicates up- or downregulated from resting to proliferative zone (≥2.0-fold, FDR < 5%) or from proliferative to hypertrophic zone (≥2.0-fold, FDR < 5%) in 1-week-old rat ($n = 6$, Genome Array 230 2.0, Affymetrix). 'Temporally regulated' indicates expression that is up- or downregulated from 3 weeks to 12 weeks of age in rat proliferative zone (≥2.0-fold, FDR < 5%, $n = 6$, Rat Genome Array 230 2.0, Affymetrix,). Examples of genes in each overlapping subset are shown in boxes.

whereas 7 other loci each contain 2 genes. Although not absolute, this tendency toward one growth plate gene per locus differs significantly from a random distribution ($P = 0.029$, based on simulation with 10 000 sets of 78 genes randomly assigned to the 207 loci, Supplementary Material, Fig. S2). The identification of two genes in each of seven loci may indicate false-positive designations. It is also possible that two neighboring genes, if functionally related or co-regulated in the growth plate, might both be identified by our analysis, irrespective of whether they both contribute to height determination. Conversely, there may well be other genes in the GWA list that affect growth plate function and thus height but were not detected by either expression or bioinformatic functional analysis. For example, *RUNX3* and *STC2* are both present in the GWA list and, when ablated *in vivo*, produce a bone growth phenotype (11,12). However, they were not identified by our structured search in the MGI database (see Supplementary Material, Table S7).

DISCUSSION

Meta-analysis of GWA studies has identified 180 loci that influence adult height. However, in most of these loci, it is not clear which genes are causative. Because height is determined by growth plate chondrogenesis, we used expression microarray studies of mouse and rat growth plate, human disease databases and a mouse knockout phenotype database to identify genes within the GWAS loci that regulate growth plate chondrogenesis and thereby control human height. Previous analysis of the height-associated loci, employing GRAIL text-mining algorithm and gene ontology/pathway analyses (1), identifies subsets of genes with known functional interrelationships. The current approach complements this prior method by focusing the search on growth plate chondrogenesis, the biological process responsible for height determination, and by capitalizing on large growth plate expression and functional data sets. As a result, this analysis can identify even solitary genes that have no known relationship to other genes in the loci and do not participate in well-studied molecular pathways. We also increased the scope of the analysis by including 27 height-associated loci that were not included in the prior study.

The validity of this approach was supported by multiple statistical analyses. First, each of these approaches identified significantly more genes within the GWA height loci than within random matched gene sets ($P < 0.0001$ each). Second, of the 78 genes identified by the combined analysis, 46 are located closest to the GWA sequence variant, significantly more than would be expected from a random allocation ($P < 0.0001$). For 24 of the implicated genes, the height-associated SNP is located within the growth plate gene itself. Third, the identified genes are distributed non-randomly within the loci, with a statistically significant, though not absolute, tendency toward having one growth plate-related gene per locus.

Our approach was only able to implicate causative genes in approximately one-third of the 207 loci associated with human height by GWAS. Some of the remaining causative genes are likely to act in endocrine organs and thereby indirectly affect growth plate function, for example, *GH1*, which encodes GH, and *GHSR*, which encodes the GH secretagogue receptor, both of which are present in the GWA loci. In addition, it is likely that many of the remaining causative genes still act in the growth plate but were not picked up by our analysis. Genes may have escaped our phenotype analysis either because inactivating mutations have not yet been characterized in the mouse or human or because such mutations are lethal in the embryo. Causative genes that act in the growth plate may also have escaped our expression analysis. For example, genes that serve to promote growth in multiple tissues may not show higher expression levels in growth plate than in non-skeletal tissues. The use of rodent growth plate expression data might also have limited the diagnostic approach because of differences in growth plate regulation among different mammals. If human growth plate expression data of similar quality become available in the future, it might improve the analysis of GWA height data. It will also be valuable to reapply our analytic approach as more height-associated genes emerge from larger GWA meta-analyses. The current GWA data explain only ~12% of the heritable variation in height (1). Thus, larger studies may reveal additional loci, although many growth plate-regulating genes may not be detected by GWA because functional variants are not sufficiently common in the human population.

Many of the 78 implicated genes participate in molecular pathways that are known to be important for growth plate chondrogenesis in the mouse. A large number participate in signaling pathways that regulate growth plate chondrocyte proliferation and differentiation, including Indian hedgehog, parathyroid hormone-related peptide, bone morphogenetic protein/transforming growth factor-beta superfamily, C-type

Downloaded from https://academic.oup.com/hmg/article/21/23/5193/593394 by guest on 17 April 2024

Table 2. Number of genes in the Genome, GWA list or GWA closest list (genes closest to GWA SNPs) that are implicated in growth plate function by different analytic approaches

	Observed[a]	Total	% Observed[b]	Expected[c]	χ^2 *P*-value
GP expression genes list[d]					
Whole genome	427	13 254[e]	3.22		
GWA list	38	446	8.52	14	<0.0001
GWA closest list	22	140	15.71	5	<0.0001
Mouse GP phenotype gene list[f]					
Whole genome	224	~23 000[g]	0.97		
GWA list	38	833	4.56	8	<0.0001
GWA closest list	24	207	11.59	2	<0.0001
Human GP phenotype gene list[h]					
Whole genome	173	~23 000[i]	0.75		
GWA list	27	833	3.24	6	<0.0001
GWA closest list	14	207	6.76	2	<0.0001

[a]Number of genes implicated in growth plate function by the indicated analysis (expression pattern, mouse phenotype or human phenotype).
[b](Observed/total) × 100.
[c]Number of genes expected by chance in a random gene list, based on the observed frequency in the whole genome = (% observed/100) × total.
[d]Genes with expression patterns by microarray fulfilling at least two of the following three criteria: (i) growth plate specificity, (ii) spatial regulation, (iii) temporal regulation; see text for details.
[e]Number of genes annotated in both mouse and rat array.
[f]Genes with a reported growth plate phenotype in mice.
[g]Estimated number of genes in the mouse genome.
[h]Genes with a reported growth plate phenotype in humans.
[i]Estimated number of genes in the human genome.

Downloaded from https://academic.oup.com/hmg/article/21/23/5193/593394 by guest on 17 April 2024

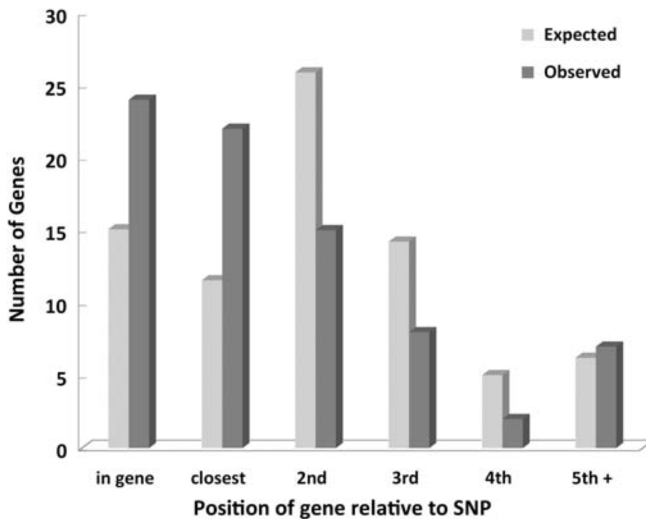

Figure 2. Observed and expected positional distribution of 78 growth plate-related genes relative to SNP. Seventy-eight genes in the GWA loci were implicated in growth plate function by expression analysis, mouse phenotype or human phenotype. Of these 78, 46 are located closest to the height SNP. In 24 of these cases, the strongly associated height SNP is located within the growth plate gene itself. This distribution differs significantly ($P < 0.0001$, χ^2 test) from a random allocation calculated based on the overall distribution of 833 genes in 207 loci. Thus, the growth plate-related genes tend to lie closer to the SNP identified by GWA than would be expected by chance.

natriuretic peptide, fibroblast growth factor and IGF signaling (Table 3, Fig. 3). Other implicated genes are involved in the formation of growth plate cartilage extracellular matrix, including structural proteins in the cartilage matrix (*ACAN*, *COL11A1*), as well as genes that encode enzymes involved in matrix biosynthesis [*GFPT2*, *SERPINH1* (13,14)] and degradation [*GALNS*, *ADAMTS10* (15,16)]. The analysis also identified multiple genes involved in cell-cycle processes, including DNA synthesis (*TOP2A*), mitotic spindle formation [*CENPW*, *TACC3* (17,18)], cell-cycle progression (*E2F1*) and translation of replication-dependent histones [*SLBP* (19)]. Altering the functions of these genes might affect cell proliferation not just in growth plate but also in other tissues. As might be expected, these genes were typically found to be expressed at high levels in the proliferative zone and to be subsequently downregulated in the hypertrophic zone, in which chondrocyte replication ceases.

One of the most exciting outcomes of this analysis is that it implicates many genes not previously known to regulate either the mouse or human growth plate. Of the 78 implicated genes, 27 were already known to function in the human growth plate because human mutations affect the growth plate. An additional 23 genes were previously known to be important in the mouse growth plate based on a knockout phenotype and thus the current study suggests that these genes are also important in the human growth plate. However, the most novel insights come from the remaining 28 genes, for which a growth plate-related phenotype has not been described in mice or humans. For example, *IGF2BP2* was implicated because it shows higher expression in growth plate cartilage than in other tissues and because its expression is downregulated during growth plate senescence. Similarly, *IGF2BP3* was implicated because its expression is upregulated during hypertrophic differentiation and downregulated during growth plate senescence. Neither gene has a known human or mouse phenotype, but targeted ablation of the third member of the gene family, *IGF2BP1*, causes shortened bones with advanced mineralization (20). Thus, taken together, the data suggest that this family of proteins plays an important regulatory role in both

Table 3. Signaling pathways and cellular functions implicated by the combined GWA expression–phenotype analysis

Functional category[a]	Gene[b]	SNP[c]	Proximity[d]
PTHrP-IHH feedback loop	*GLI2*	rs2166898	Intragenic
	IHH	rs12470505	Second (11 kb)
	HHIP	rs7689420	Intragenic
	PTCH1	rs473902	Intragenic
	PTHLH	rs2638953	Second (409 kb)
BMP/TGF superfamily signaling	*TGFB2*	rs6684205	Intragenic
	BMP6	rs3812163	Closest (1 kb)
	LTBP3	rs3782089	Second (11 kb)
	NOG	rs227724	Second (106 kb)
	BMP2	rs2145272	Closest (123 kb)
	GDF5	rs143384	Intragenic
C-type natriuretic peptide signaling	*NPPC*	rs2580816	Closest (7 kb)
	PRKG2	rs788867	Closest (24 kb)
	NPR3	rs1173727	Intragenic
GH-IGF1 signaling	*IGF2BP2*	rs720390	Closest (6 kb)
	IGF2BP3	rs12534093	Intragenic
	IGF1R	rs2871865	Intragenic
Paracrine signaling from perichondrium	*FGF18*	rs12153391	Second (319 kb)
	RUNX2	rs9472414	Second (443 kb)
	TWIST1	rs4470914	Fourth (459 kb)
Cell-cycle-related genes	*SLBP*	rs2247341	Intragenic
	TACC3	rs2247341	Third (21 kb)
	CENPW	rs1490384	Closest (181 kb)
	TOP2A	rs528045	Closest (13 kb)
	E2F1	rs7274811	Third (59 kb)
Cartilage matrix-related genes	*COL11A1*	rs12047268	Intragenic
	GFPT2	rs6879260	Intragenic
	SERPINH1	rs634552	Intragenic
	ACAN	rs16942341	Intragenic
	GALNS	rs8052560	Fifth (103 kb)
	ADAMTS10	rs4072910	Closest (1 kb)

[a]Molecular pathways or other functional role in the growth plate.
[b]Genes identified by the combined GWA expression–phenotype analysis that participate in known signaling pathways or cellular functions in the growth plate.
[c]SNPs associated with height in GWA studies.
[d]'Closest' indicates that the candidate gene is the gene that lies closest to the GWA SNP in physical distance. 'Second' indicates that one other gene in the locus lies closer to the SNP, and so on.

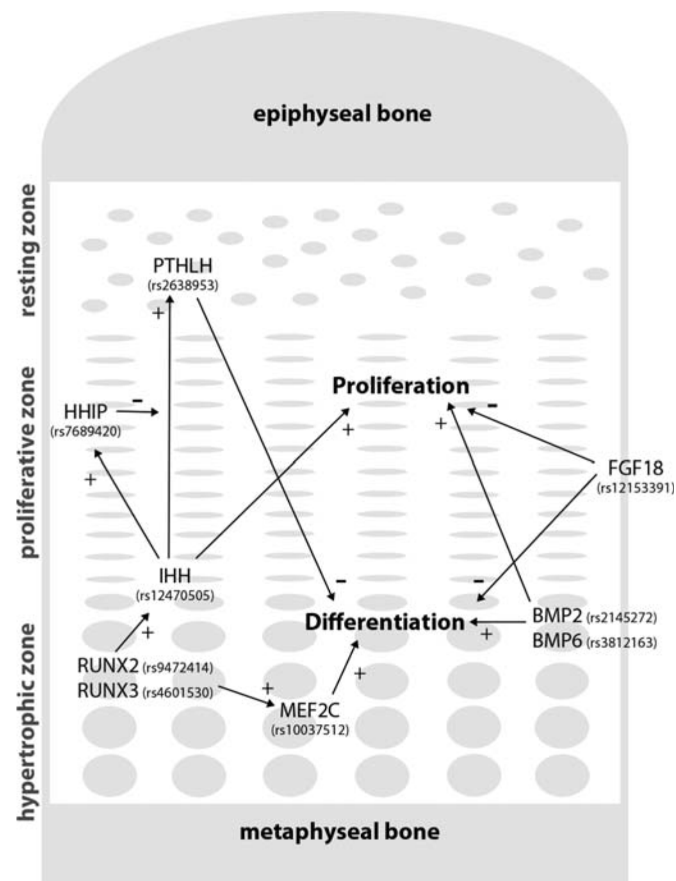

Figure 3. Examples of genes from GWA analysis that have a well-established role in chondrogenesis. PTHrP (encoded by *PTHLH*) produced in the resting zone inhibits differentiation of proliferative chondrocytes. The production of PTHrP is in turn stimulated by IHH from the hypertrophic zone. IHH stimulates chondrocyte proliferation in the proliferative zone and the expression of HHIP, which inhibits and modulates the action of IHH in the resting zone. Together, IHH and PTHrP form a negative feedback loop that determines the height of the proliferative zone. BMP2 and BMP6, produced by hypertrophic chondrocytes, stimulate proliferation and differentiation of chondrocytes, which is partially antagonized by an inhibitory effect of FGFs, such as FGF18. RUNX2 and RUNX3 are expressed in the hypertrophic zone and stimulate hypertrophic differentiation through one of their downstream targets MEF2C, which is also highly expressed in the hypertrophic zone. RUNX2 also interacts with the IHH-PTHrP feedback loop by inducing IHH expression. +, stimulatory effect; −, inhibitory effect.

the rodent and human growth plates, perhaps through their ability to bind specific mRNAs, including *IGF2* (21). Another example of genes not previously known to regulate the growth plate is *RSPO3*. A recent study indicates that *RSPO3* activates *WNT5A* signaling (22), which is known to regulate growth plate chondrocyte proliferation and differentiation (23).

Another important implication of the current analysis is that it identifies strong candidates for the etiology of human growth disorders, including isolated short stature, skeletal dysplasias and short stature that is part of a larger dysmorphic syndrome. Our analysis suggests that, for many of the height-associated genes, common polymorphisms subtly affect gene expression or function in the growth plate and thus influence adult height. Uncommon mutations in these genes that have a greater effect on expression or function would therefore be expected to cause greater disturbances in growth plate chondrogenesis, resulting in short stature or a skeletal dysplasia. Indeed, the height-associated loci were highly enriched with genes that cause skeletal dysplasias when mutated in humans. Hence, we think it is likely that other growth plate-related genes identified by this analysis are also excellent candidates for causing short stature and skeletal dysplasias of unknown etiology.

Because identification of a phenotype-associated SNP by GWA rarely isolates the causative gene, several methods have been developed to provide biological insights and to prioritize genes for further studies (24), including the use of gene ontology analysis (25), gene set enrichment analyses (26) and protein–protein interaction mapping (27). In the current study, we combined the human height GWA data with gene expression profiling of the growth plate to identify genes that play a significant role in longitudinal bone growth and consequently height determination. This idea of combining genome-wide expression with GWA data, to our knowledge, has only been conceptually explored for prostate cancer risk (28), but not

Downloaded from https://academic.oup.com/hmg/article/21/23/5193/593394 by guest on 17 April 2024

for any common polygenic traits. Therefore, our current findings suggest that this powerful analytic approach can be applied more widely to GWA studies for common polygenic traits and diseases.

In summary, GWA meta-analysis provides a powerful tool that has identified many loci that are associated with the normal variation in human height. However, this GWA analysis does not pinpoint a single gene at each locus but instead identifies a set of contiguous genes—only one of which is presumably causative. Reasoning that many of the causative genes would have a local action on growth plate chondrogenesis, we used growth plate expression and functional data to identify 78 genes from these loci that likely play a role in human growth plate biology and thereby affect longitudinal bone growth and adult height. This analysis identified many genes involved in signaling pathways and cellular functions known to be important in the rodent growth plate but also many novel genes that have not previously been implicated in growth plate chondrogenesis. These novel genes suggest new avenues for mechanistic studies to explore the underlying cellular pathways. Thus, this analytic approach provides a new strategy to translate genetic associations into mechanistic biological insights.

MATERIALS AND METHODS

Animals

All animal procedures were approved by the National Institute of Child Health and Human Development Animal Care and Use Committee. C57BL/6 mice were obtained from Charles River (Wilmington, MA, USA) and provided with food and water *ad libitum*. Animals were sacrificed at 1 week of age and growth plates were dissected from the proximal tibia.

Microarray analysis

Total RNA was extracted from 1-week-old mouse whole growth plate following microdissection, processed and hybridized to Affymetrix microarray as previously described (8). Each microarray chip (mouse genome 430 2.0 array) was hybridized to labeled RNA derived from one animal (total of five animals used). To identify growth plate-specific genes, we combined this growth plate array data with our previously published array data (6,7) of 1-week-old C57BL/6 mouse (Charles River) kidney, lung and heart. Microarray signals were analyzed using the Affymetrix RMA algorithm. Analysis of variance was performed and FDR reports were generated using the Partek Pro software (Partek, St Charles, MO, USA). Spatially and temporally regulated genes in the growth plate were identified based on data from our previous microarray analyses (GEO number: GSE16981) (8). The comparison of growth plate expression with that of other tissues was performed in mice to take advantage of the well-annotated microarray databases for the mouse, whereas the spatial and temporal expression analyses were performed in rats because the larger growth plate allows more accurate microdissection of individual zones and the thin growth plates of older animals. The analysis of temporal expression compared rats at 3 weeks of age, during active longitudinal bone growth, with rats at 12 weeks of age, when longitudinal bone growth has slowed markedly (29).

GWA genes

We generated a list of 833 genes from the 207 height-associated loci [180 previously published (1)]. Annotation was based on the NCBI Build 36.1 (hg18). Because the causal sequence variant in each locus is not necessarily the lead SNP itself but is likely in strong linkage disequilibrium with the lead SNP, we used a genomic boundary with linkage disequilibrium cut-off $r^2 > 0.5$ (HapMap CEU) to generate the list of possible causal genes. For SNPs with two or more genes within the genomic boundary, only these genes were included. For SNPs with less than two genes within this region, the genomic boundaries were extended on both sides by a quarter of the original region.

Human GP phenotype genes

We generated a list of 173 genes implicated in abnormal skeletal growth in human that are likely due to underlying local growth plate dysfunction. The list was initially obtained by searching the OMIM database using the keywords ‘short stature’, ‘overgrowth’ and ‘skeletal dysplasia’. The result was then combined with a list of genes causing skeletal dysplasias published by the International Skeletal Dysplasia Society (10). The combined list was then manually curated without reference to the GWA results, including only genes that cause overgrowth or short stature without a known underlying systemic disorder, such as GH deficiency, or failure of a major organ system. Enrichment of human GP phenotype genes in the GWA list was assessed by Pearson’s χ^2 test, comparing the number of human GP phenotype genes in the GWA list with the number that would be expected by chance in a random gene list, based on the observed frequency in the whole genome.

Mouse GP phenotype genes

We searched the MGI database (9) and identified 224 genes that have a skeletal phenotype with underlying growth plate dysfunction in knockout mice (mouse GP phenotype genes). The gene list was initially obtained searching for the keywords ‘chondrocytes’, ‘cartilage’ or ‘growth plate’ in the MGI web site ‘Genes and Markers Query Form’ and was manually curated without reference to GWA results to select genes that alter growth plate chondrogenesis in the long bones and/or alter the length of long bones. Genes that only have an effect on chondrocytes from craniofacial, ribs, sternum and articular cartilage were discarded. Enrichment of mouse GP phenotype genes in the GWA list was assessed by Pearson’s χ^2 test, comparing the number of mouse GP phenotype genes in the GWA list with the number that would expected by chance in a random gene list, based on the observed frequency in the whole genome.

Downloaded from https://academic.oup.com/hmg/article/21/23/5193/593394 by guest on 17 April 2024

Permutation analysis

To test whether the GWA list was enriched with growth plate-related genes, we generated 1000 gene sets, each comprising 207 random loci, selected to match the actual height-associated SNPs based on gene proximity and presence in HapMap. The same linkage disequilibrium cutoffs in generating the GWA list (see above) was used to determine genes to be included in each set. The number and percentage of GP expression genes, mouse GP phenotype genes and human GP phenotype genes in each set were then counted and compared with the actual overlap in the GWA list to generate an empirical *P*-value.

Simulation modeling

Seventy-eight genes implicated in growth plate function were distributed among 71 of the 207 height-associated loci, with 7 loci bearing 2 genes. Simulation was used to assess whether this distribution differed significantly from a random allocation. Using a computer algorithm, 78 genes were randomly selected from 833 genes arranged in 207 loci. Each locus contained 1–56 genes according to the actual distribution in the GWA list (see Supplementary Material, Table S1). The number of loci containing more than one selected gene was then counted. This random selection was repeated 10 000 times.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank the height working group and the GIANT consortium for providing data from their GWA studies. We thank Matthew Baron for help with the simulation modeling.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), US National Institutes of Health (NIH); grant 1R01DK075787 from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), NIH; and grant 6-FY09-507 from March of Dimes Foundation.

REFERENCES

1. Lango, A.H., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., Willer, C.J., Jackson, A.U., Vedantam, S., Raychaudhuri, S. *et al.* (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, **467**, 832–838.
2. Kronenberg, H.M. (2003) Developmental regulation of the growth plate. *Nature*, **423**, 332–336.
3. Nilsson, O. and Baron, J. (2004) Fundamental limits on longitudinal bone growth: growth plate senescence and epiphyseal fusion. *Trends Endocrinol. Metab.*, **15**, 370–374.
4. van der Eerden, B.C., Karperien, M. and Wit, J.M. (2003) Systemic and local regulation of the growth plate. *Endocr. Rev.*, **24**, 782–801.
5. Wuelling, M. and Vortkamp, A. (2010) Transcriptional networks controlling chondrocyte proliferation and differentiation during endochondral ossification. *Pediatr. Nephrol.*, **25**, 625–631.
6. Finkielstain, G.P., Forcinito, P., Lui, J.C., Barnes, K.M., Marino, R., Makaroun, S., Nguyen, V., Lazarus, J.E., Nilsson, O. and Baron, J. (2009) An extensive genetic program occurring during postnatal growth in multiple tissues. *Endocrinology*, **150**, 1791–1800.
7. Lui, J.C., Forcinito, P., Chang, M., Chen, W., Barnes, K.M. and Baron, J. (2010) Coordinated postnatal down-regulation of multiple growth-promoting genes: evidence for a genetic program limiting organ growth. *FASEB J.*, **24**, 3083–3092.
8. Lui, J.C., Andrade, A.C., Forcinito, P., Hegde, A., Chen, W., Baron, J. and Nilsson, O. (2010) Spatial and temporal regulation of gene expression in the mammalian growth plate. *Bone*, **46**, 1380–1390.
9. Blake, J.A., Bult, C.J., Kadin, J.A., Richardson, J.E. and Eppig, J.T. (2011) The Mouse Genome Database (MGD): premier model organism resource for mammalian genomics and genetics. *Nucleic Acids Res.*, **39**, D842–D848.
10. Warman, M.L., Cormier-Daire, V., Hall, C., Krakow, D., Lachman, R., LeMerrer, M., Mortier, G., Mundlos, S., Nishimura, G., Rimoin, D.L. *et al.* (2011) Nosology and classification of genetic skeletal disorders: 2010 revision. *Am. J. Med. Genet. A*, **155A**, 943–968.
11. Yoshida, C.A., Yamamoto, H., Fujita, T., Furuichi, T., Ito, K., Inoue, K., Yamana, K., Zanma, A., Takada, K., Ito, Y. and Komori, T. (2004) Runx2 and Runx3 are essential for chondrocyte maturation, and Runx2 regulates limb growth through induction of Indian hedgehog. *Genes Dev.*, **18**, 952–963.
12. Johnston, J., Ramos-Valdes, Y., Stanton, L.A., Ladhani, S., Beier, F. and Dimattia, G.E. (2010) Human stanniocalcin-1 or -2 expressed in mice reduces bone size and severely inhibits cranial intramembranous bone growth. *Transgenic Res.*, **19**, 1017–1039.
13. Hugenberg, S.T., Brandt, K.D. and Cole, C.A. (1993) Effect of sodium salicylate, aspirin, and ibuprofen on enzymes required by the chondrocyte for synthesis of chondroitin sulfate. *J. Rheumatol.*, **20**, 2128–2133.
14. Dafforn, T.R., Della, M. and Miller, A.D. (2001) The molecular interactions of heat shock protein 47 (Hsp47) and their implications for collagen biosynthesis. *J. Biol. Chem.*, **276**, 49310–49319.
15. Tomatsu, S., Montano, A.M., Gutierrez, M., Grubb, J.H., Oikawa, H., Dung, V.C., Ohashi, A., Nishioka, T., Yamada, M., Yamada, M. *et al.* (2007) Characterization and pharmacokinetic study of recombinant human N-acetylgalactosamine-6-sulfate sulfatase. *Mol. Genet. Metab.*, **91**, 69–78.
16. Gruber, H.E., Hoelscher, G.L., Ingram, J.A., Bethea, S., Zinchenko, N. and Hanley, E.N. Jr. (2011) Variations in aggrecan localization and gene expression patterns characterize increasing stages of human intervertebral disk degeneration. *Exp. Mol. Pathol.*, **91**, 534–539.
17. Kinoshita, K., Noetzel, T.L., Pelletier, L., Mechtler, K., Drechsel, D.N., Schwager, A., Lee, M., Raff, J.W. and Hyman, A.A. (2005) Aurora A phosphorylation of TACC3/maskin is required for centrosome-dependent microtubule assembly in mitosis. *J. Cell Biol.*, **170**, 1047–1055.
18. Kim, H., Lee, M., Lee, S., Park, B., Koh, W., Lee, D.J., Lim, D.S. and Lee, S. (2009) Cancer-upregulated gene 2 (CUG2), a new component of centromere complex, is required for kinetochore function. *Mol. Cell*, **27**, 697–701.
19. Sullivan, K.D., Mullen, T.E., Marzluff, W.F. and Wagner, E.J. (2009) Knockdown of SLBP results in nuclear retention of histone mRNA. *RNA*, **15**, 459–472.
20. Hansen, T.V., Hammer, N.A., Nielsen, J., Madsen, M., Dalbaeck, C., Wewer, U.M., Christiansen, J. and Nielsen, F.C. (2004) Dwarfism and impaired gut development in insulin-like growth factor II mRNA-binding protein 1-deficient mice. *Mol. Cell. Biol.*, **24**, 4448–4464.
21. Christiansen, J., Kolte, A.M., Hansen, T.O. and Nielsen, F.C. (2009) IGF2 mRNA-binding protein 2: biological function and putative role in type 2 diabetes. *J. Mol. Endocrinol.*, **43**, 187–195.
22. Ohkawara, B., Glinka, A. and Niehrs, C. (2011) Rspo3 binds syndecan 4 and induces Wnt/PCP signaling via clathrin-mediated endocytosis to promote morphogenesis. *Dev. Cell*, **20**, 303–314.
23. Yang, Y., Topol, L., Lee, H. and Wu, J. (2003) Wnt5a and Wnt5b exhibit distinct activities in coordinating chondrocyte proliferation and differentiation. *Development*, **130**, 1003–1015.
24. Raychaudhuri, S. (2011) Mapping rare and common causal alleles for complex human diseases. *Cell*, **147**, 57–69.

Downloaded from https://academic.oup.com/hmg/article/21/23/5193/593394 by guest on 17 April 2024

25. Holmans, P., Green, E.K., Pahwa, J.S., Ferreira, M.A., Purcell, S.M., Sklar, P., Owen, M.J., O'Donovan, M.C. and Craddock, N. (2009) Gene ontology analysis of GWA study data sets provides insights into the biology of bipolar disorder. *Am. J. Hum. Genet.*, **85**, 13–24.
26. Segre, A.V., Groop, L., Mootha, V.K., Daly, M.J. and Altshuler, D. (2010) Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.*, **6**, e1001058.
27. Rossin, E.J., Lage, K., Raychaudhuri, S., Xavier, R.J., Tatar, D., Benita, Y., Cotsapas, C. and Daly, M.J. (2011) Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet.*, **7**, e1001273.
28. Gorlov, I.P., Gallick, G.E., Gorlova, O.Y., Amos, C. and Logothetis, C.J. (2009) GWAS meets microarray: are the results of genome-wide association studies and gene-expression profiling consistent? Prostate cancer as an example. *PLoS One*, **4**, e6511.
29. Marino, R., Hegde, A., Barnes, K.M., Schrier, L., Emons, J.A., Nilsson, O. and Baron, J. (2008) Catch-up growth after hypothyroidism is caused by delayed growth plate senescence. *Endocrinology*, **149**, 1820–1828.
30. Raychaudhuri, S., Plenge, R.M., Rossin, E.J., Ng, A.C., International Schizophrenia Consortium, Purcell, S.M., Sklar, P., Scolnick, E.M., Xavier, R.J., Altshuler, D. and Daly, M.J. (2009) Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.*, **5**, e1000534.
31. Yang, J., Weedon, M.N., Purcell, S., Lettre, G., Estrada, K., Willer, C.J., Smith, A.V., Ingelsson, E., O'Connell, J.R., Mangino, M. *et al.* (2011) Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.*, **19**, 807–812.

Downloaded from https://academic.oup.com/hmg/article/21/23/5193/593394 by guest on 17 April 2024