Review

Cell PRESS

Reaching new heights: insights into the genetics of human stature

Michael N. Weedon and Timothy M. Frayling

Genetics of Complex Traits, Institute of Biomedical and Clinical Science, Peninsula College of Medicine and Dentistry, Magdalen Road, Exeter, EX1 2LU, UK

Human height is a highly heritable, classic polygenic trait. Until recently, there had been limited success in identifying the specific genetic variants that explain normal variation of human height. The advent of large-scale genome-wide association studies, however, has led to dramatic progress. In the past 18 months, the first robust common variant associations were identified and there are now 44 loci known to influence normal variation of height. In this review, we summarize this exciting recent progress, discuss implicated biological pathways, the overlap with monogenic growth and skeletal dysplasia syndromes, links to disease and insights into the genetic architecture of this model polygenic trait. We also discuss the strong probability of finding several hundred more such loci in the near future.

Height as a model genetic trait

Height is a trait that has long fascinated scientists. It is among the most visible of human characteristics, is easily and accurately measured and is highly heritable. In many countries, the average height of the population has increased substantially over the past few generations. For example, Dutch males are now, on average, almost 20 cm taller than 150 years ago [1]. This clearly demonstrates that non-genetic factors influence height; however, within a given population at a given time ~80% of height variation can be explained by genetic differences [2–6]. These features make height a model genetic trait from which general conclusions about quantitative traits can be drawn. This was first appreciated by Galton >120 years ago, when he investigated the genetic contribution to stature variation, famously demonstrating that mid-parental height predicts offspring height [7]. Stature was also used by Fisher in his classic 1918 paper, in which he showed that the inheritance pattern of quantitative traits could be explained by the combination of many 'Mendelian factors', each explaining only a modest proportion of the overall heritability (see Glossary) of the trait [8].

The height of an individual is the result of many growth and development processes and greater stature is not just a result of increasing bone length; tissue and organ sizes are usually also proportionately increased. In addition to being a model trait, a major driver for the search for height genes is to provide novel insights into human growth and development. Normal variation of height is associated with a range of diseases, including various cancers [9,10] (whereby taller people tend to be at increased risk) and type 2 diabetes [11,12] (whereby shorter people tend to be at increased risk); therefore, a better understanding of the gene variants underlying height differences might also provide novel insights into these disorders.

In this review, we summarize some of the exciting recent progress made in identifying the particular 'Mendelian factors' that underlie normal height variation and discuss some of the methodological and biological insights that have come from these studies.

Hunting height 'genes'

Until recently, there had been limited success in identifying the genetic variants influencing normal variation of human height. Before 2007, the hunt for genes had followed a familiar path of genome-wide linkage and candidate-gene association studies. As was the case for most common traits, these approaches had limited success. Perola *et al.* [6] recently summarized the results from all published height linkage studies and demonstrated that there was little overlap in the regions of the genome nominally linked to stature. (Figure 1). The most likely reason for the lack of success of the linkage approach is that height variation is explained by many variants, each of which explains only a small proportion of the heritability of the trait. In this scenario, association has been shown to be a much more powerful approach to gene identification than linkage [13]. Until recently, it was only feasible to genotype a small subset (tens or hundreds) of the >10 million common variants that occur across the human genome in association studies, so only a small number of candidate genes (implicated through prior biological knowledge) could be assessed. Candidate-gene association studies, however, have been similarly unsuccessful at finding robustly associated variants, even when the candidacy of the genes was excellent and the study comprehensive and large (e.g. see Ref. [14]). Now, however, with technological advances giving researchers the ability to efficiently and cost-effectively assay a large proportion (>2/3) of the common variation across the human genome [15], progress has been rapid.

Glossary

Heritability: this refers to the proportion of variation of a phenotype (in a given population at a given time) that is a result of genetic variation.
Linkage disequilibrium: the association between alleles at two or more loci in a population.

Corresponding author: Weedon, M.N. (Michael.Weedon@pms.ac.uk).

0168-9525/$ – see front matter doi:10.1016/j.tig.2008.09.006 Available online xxxxxx

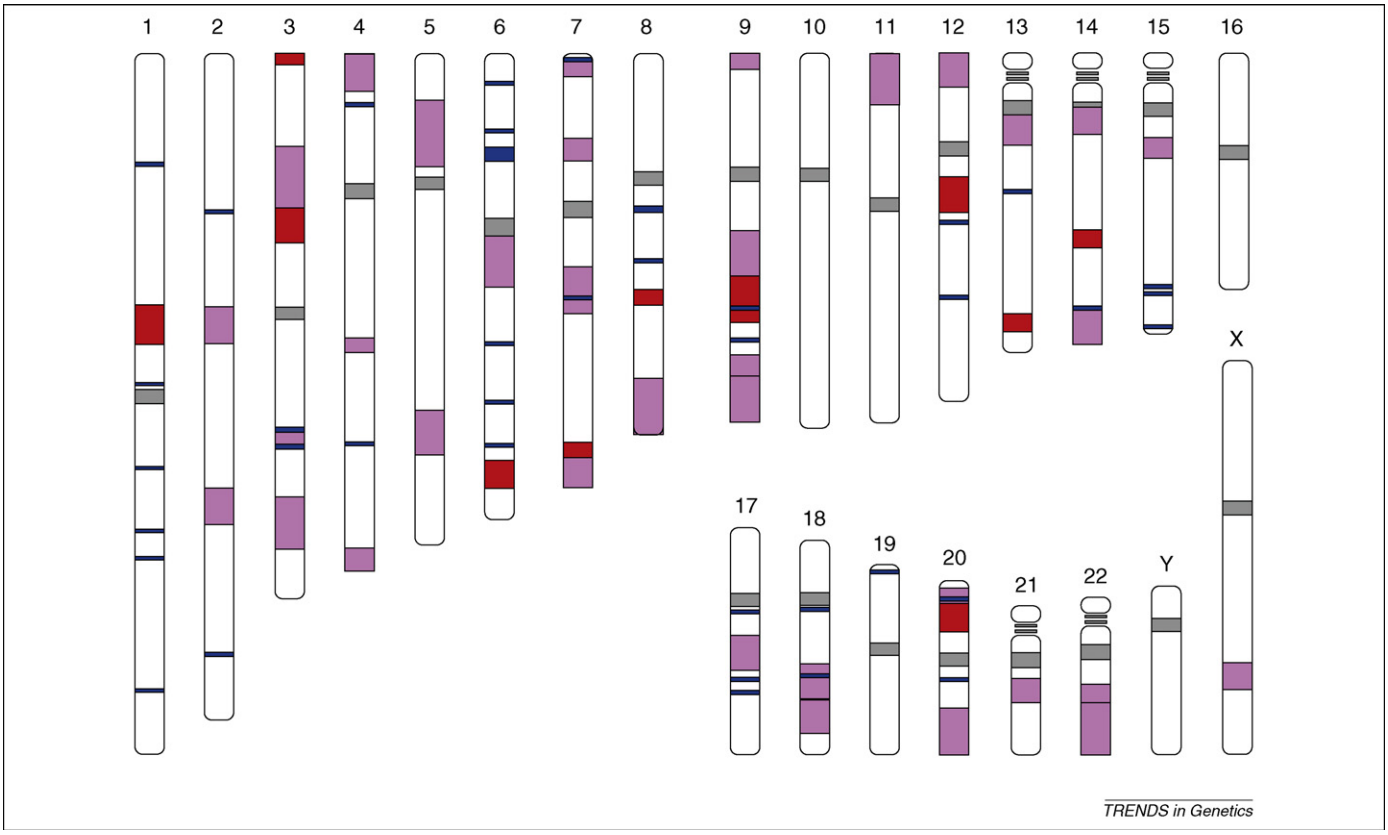

Figure 1. Regions of the genome associated and/or linked with height. The red and pink regions are chromosome bands that have been linked (at logarithms of the odds [LOD] > 3 and >2, respectively) with height from genome-wide linkage studies. The blue lines represent loci that have been convincingly associated with height from genome-wide association studies [17–21]. The associated regions are spread across the chromosomes and do not occur in linked regions more than would be expected by chance. For example, only one of the loci overlaps the LOD > 3 linked regions and, given the proportion of the genome covered by these regions, we would have expected one by chance. Data used in figure adapted from Ref. [6].

The genome-wide association study (GWAS) approach and its success at finding loci for a range of complex traits has been described in detail recently [16]. Five recent studies using genome-wide association data have identified a total of 44 independent (HapMap [http://www.hapmap.org/] $r^2 < 0.1$) variants that are robustly associated with adult height in the general population. An initial GWAS of 4921 individuals, together with replication in 29 098 individuals, identified a variant in the high mobility group AT hook 2 (*HMGA2*) gene associated with height at the robust levels of statistical significance required for GWAS (to take into account the multiple hypotheses being tested) [17]. This was followed by a similar study that used 6669 GWAS individuals and 28 801 replication samples, which identified a height locus containing the growth/differentiation factor 5 (*GDF5*) gene [18]. Following on from these initial successes, three studies, each with GWAS data from >13 000 individuals and up to 16 500 replication samples, identified a further 42 loci [19–21].

Implicated genes and biological processes

The causal gene and variants at each of these loci has not yet been proven – hence, the genes are implicated. For half the loci, there is a strong, although far from definitive, case for a particular gene being causal (Table 1). In the remaining cases, the likely gene or genes involved are less obvious (Table 2). Despite this, the implicated genes can give us some insights into genes and molecular processes that are important in normal human growth. An overview of the biological pathways and processes in which the implicated genes occur is given in Figure 2. There is a clear over-representation of genes involved in processes known to influence bone and cartilage development and there are several specific pathways worth highlighting.

Skeletal development signaling pathways

It is encouraging that many of the implicated genes are components of signaling pathways that are known to be important in skeletal growth and development (whereby they function as something of a 'positive control'). Most of these genes seem to act at the growth plate (the cartilage at the end of growing bones; see Figure 3 for an overview). In particular, three of these genes (Indian hedgehog [*IHH*], patched 1 [*PTCH1*] and hedgehog-interacting protein [*HHIP*]) occur in the Hedgehog signaling pathway, an intensively studied signaling cascade that is crucial for vertebrate patterning and development [22]. Others are bone morphogenic proteins (BMPs; implicated genes: *BMP2*, *BMP6* and *GDF5*) and their regulators (*Noggin*) [23]. BMPs are growth factors and cytokines that are able to induce formation of bone and cartilage [23].

Extracellular matrix proteins

Bone and cartilage mostly consist of an extra-cellular matrix (ECM). Genes encoding ECM components, therefore, represent another obvious set of candidates for influ-

Table 1. Height loci in instances in which there is a strong case for the causality of a specific gene.

Index SNP[a]	Chr	Position (bp)[b]	Recombination region (kb)[c]	N genes[d]	Implicated gene[e]	Human growth or skeletal monogenic syndrome [f]	Mouse knockout causes skeletal or growth defects[g]	Expression evidence
rs6686842	1	41303458	287	3	*SCMH1*		Skeletal defects	
rs6724465	2	219652090	194	6	*IHH*	Brachydactyly	Skeletal defects	
rs10935120	3	135715782	186	3	*ANAPC13*			Yes [29]
rs6440003	3	142576899	299	2	*ZBTB38*			Yes [21]
rs1812175	4	145794294	134	1	*HHIP*		Skeletal defects	
rs12198986	6	7665058	54	1	*BMP6*		Skeletal and growth	
rs1776897	6	34302989	72	2	*HMGA1*		Growth and size	
rs2814993	6	34726871	813	10	*PPARD*		Growth and size	
rs4713858	6	35510763	85	4	*PPARD*		Growth and size	
rs798544	7	2729628	189	2	*GNA12*			Yes [29]
rs2282978	7	92102346	300	5	*CDK6*		Growth and size	Yes [29]
rs10958476	8	57258362	137	2	*PLAG1*		Growth and size	
rs9650315	8	57318152	137	2	*PLAG1*		Growth and size	
rs7846385	8	78322734	162	0	*PXMP3*	Zellweger syndrome	Skeletal defects	
rs10512248	9	97299524	176	1	*PTCH1*	Gorlin syndrome, holoprosencephaly	Skeletal defects	
rs1042725	12	64644614	84	1	*HMGA2*	Tall stature	Growth and size	
rs11107116	12	92502635	47	1	*SOCS2*		Growth and size	
rs8041863	15	87160693	25	1	*ACAN*	Spondyloepiphyseal dysplasia type Kimberley	Skeletal and growth	
rs3760318	17	26271841	226	4	*RNF135*	Overgrowth syndrome		
rs4794665	17	52205328	108	2	*NOG*	Various skeletal defects	Skeletal and growth	
rs757608	17	56852059	65	3	*TBX2*		Skeletal defects	
rs8099594	18	45245158	498	3	*DYM*	Dyggve-Melchior-Clausen		
rs967417	20	6568893	49	0	*BMP2*		Skeletal and growth	
rs6060369	20	33370575	673	13	*GDF5*	Various skeletal defects	Skeletal defects	

[a]In cases in which the locus was identified by more than one of the three studies, the index SNP is the one that had the largest effect size across the studies.
[b]Genome co-ordinates are based on NCBI build 36.
[c]Recombination hotspots are based on HapMap Phase II data. The data and methods used to derive them are available from the HapMap website (http://www.hapmap.org/) [15].
[d]N genes is the number of RefSeq genes between flanking recombination hotspots.
[e]A gene is considered to have a strong case for being the causal gene if the variant affects the mRNA expression of the gene, or if it occurs between flanking recombination hotspots or in a 500-kb window of the index SNP and has a monogenic growth or skeletal defect phenotype in humans or knockout mice.
[f]Human monogenic phenotype is based on OMIM data.
[g]Mouse knockout data are from the Jackson laboratory (http://www.jax.org/).

Table 2. Loci in instances in which there is limited evidence for a particular gene being causal.

SNP[a]	Chromosome	Position (bp)[b]	Recomb region (kb)[c]	N genes[d]	Implicated gene[e]	Other genes in recombination region
rs12735613	1	118685496	141	0	*SPAG17*	
rs11205277	1	148159496	177	14	*Histone cluster 2*	
rs678962	1	170456512	109	1	*DNM3*	
rs2274432	1	182287568	149	2	*TSEN15*	*GLT25D2*
rs1390401	1	225864573	301	4	*ZNF678*	*JMJD4; MPN2; c1orf142*
rs3791679	2	55950396	100	1	*EFEMP1*	
rs16896068	4	17553938	252	3	*NCAPG*	*LCORL; c4orf30*
rs10946808	6	26341366	110	16	*Histone cluster 2*	
rs2844479	6	31680935	105	11	*HLA Class III*	
rs185819	6	32158045	508	42	*HLA Class III*	
rs314277	6	105514355	114	1	*LIN28B*	
rs4549631	6	127008001	455	1	*C6orf173*	
rs4896582	6	142745570	226	1	*GPR126*	
rs4743034	9	108672174	120	1	*ZNF462*	
rs3116602	13	50009356	104	0	*DLEU7*	
rs8007661	14	91529711	175	3	*FBLN5*	*TRIP11; ATXN3*
rs2562784	15	82077496	220	2	*ADAMSTSL3*	*SH3GL3*
rs4533267	15	98603794	56	1	*ADAMSTS17*	
rs4800148	18	18978326	97	1	*CABLES1*	
rs12986413	19	2121954	149	4	*DOT1L*	*AP3D1; PLEKHJ1; SF3A2*

[a]In cases in which the locus was identified by more than one of the three studies, the index SNP is the one that had the largest effect size across the studies.
[b]Genome co-ordinates are based on NCBI build 36.
[c]Recombination hotspots are based on HapMap Phase II data. The data and methods used to derive them are available from the HapMap website (http://www.hapmap.org/) [15].
[d]N genes is the number of RefSeq genes between flanking recombination hotspots.
[e]Implicated genes are either the nearest gene or a gene (within the recombination region or a 500-kb window) that occurs in one of the pathways that cluster with growth and development.

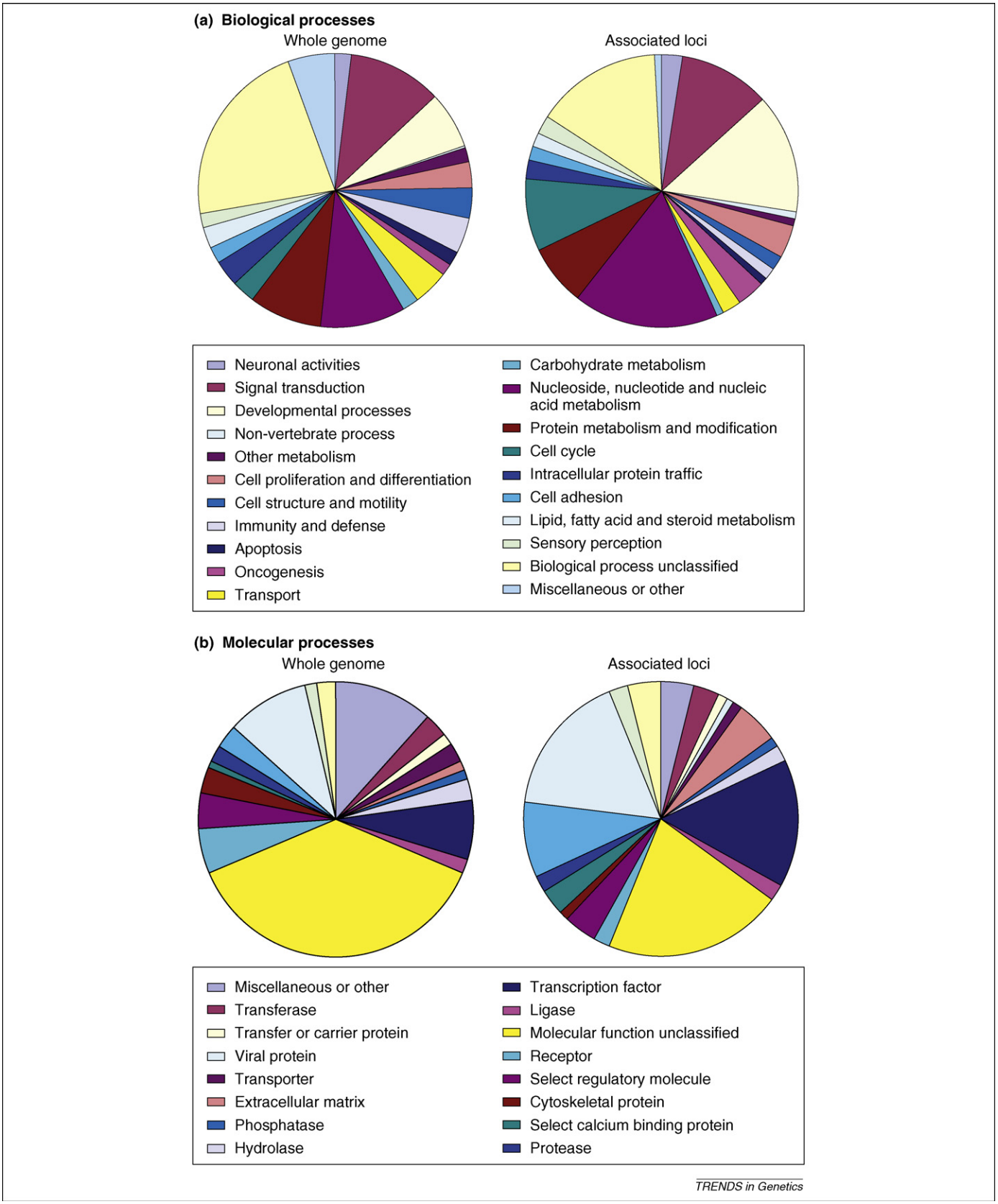

Figure 2. The biology of height. **(a)** Implicated biological and **(b)** molecular processes for genes at the associated loci based on Panther classification [47] compared with all genes in the genome. Genes at the associated loci are defined as all RefSeq genes between flanking recombination hotspots at the 44 associated loci (including a total of 77 genes), excluding the HLA region loci and including only a single gene from each of the histone cluster regions. The whole genome data are based on all 18 308 RefSeq genes. There is a clear over-representation of genes in the cell cycle, developmental, nucleic acid (including the chromatin structure and re-modeling sub-pathway) and extracellular matrix pathways at the height-associated loci compared to the whole genome.

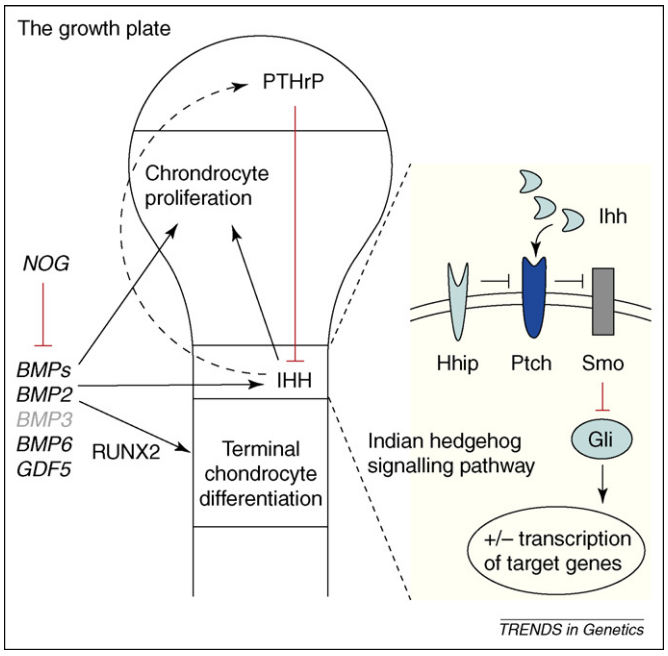

Figure 3. Many of the implicated genes function at the growth plate. The figure represents the growth plate, which is the cartilage at the end of growing bones where bone growth occurs and demonstrates that many of the implicated genes function here. The inset represents the Indian Hedgehog (Ihh) signaling pathway, which is a key pathway for growth plate development. The genes in black are those that have been robustly associated with height. The genes in grey are 'greyzone' genes, these are genes that did not reach robust levels of statistical significance in any of the individual studies, but were associated at $P < 1 \times 10^{-5}$ in two of the three studies. The red lines represent inhibition and the black arrows represent activation effects. Abbreviations: PTHrP, parathyroid hormone-like hormone; RUNX2, runt-related transcription factor 2; Smo, smoothened. Figure based on Refs [22,48].

encing height. Implicated genes from the GWAS studies include those encoding structural components (e.g. ACAN, which encodes aggrecan, a proteoglycan and one of the most important structural components of cartilage); ECM-stabilizing proteins (fibulin 5 [*FBLN5*] and *EFEMP1* [also known as fibulin-3]) binding to, amongst other components, elastic fibers and basement membranes [24]; and proteins involved in the turnover of ECM components (the ADAMTS [a disintegrin and metalloproteinase with thrombospondin motif] and ADAMTSL proteins, a class of matrix mellatonporteinases) [25,26].

Chromatin structure and regulation

Six of the implicated genes are involved in chromatin structure and its regulation. Unlike the ECM and skeletal development pathways, which were more obvious candidates, these genes highlight a more general and fundamental biological process as being important in normal human growth. The identification of common variants near genes important in chromatin structure implicates the altered expression of many downstream genes. The importance of histones, the main protein component of chromatin, is demonstrated by the association at histone clusters 1 and 2. Furthermore, the importance of the regulation of the histones is suggested by DOT1-like (*DOT1L*), a histone methyl-transferase [27] and sex comb on midleg 1 (*SCMH1*), a gene encoding a polycomb group protein [28]. Non-histone chromosomal proteins involved in chromatin rearrangement are also implicated by the association of variants in two high mobility group A (HMGA) genes, *HMGA1* and *HMGA2*.

Cell-cycle regulation and mitosis

Four of the implicated genes are involved in the cell cycle and its regulation. All three of the larger GWAS found association of variants of *CDK6*, and mRNA expression data indicate that it is the likeliest causal gene at the locus [29]. CDK6 is a cyclin-dependent kinase (CDK). CDKs regulate progression of mammalian cells through the cell cycle and CDK6 is thought, in combination with other CDKs, to regulate progression through the G1 phase of mitosis [30]. Regulators of the CDKs are also represented by CDK5 and Abl enzyme substrate (*CABLES1*), a cyclin-dependant kinase-binding protein. Other implicated genes include anaphase promoting complex 1 (*ANAPC13*), which has strong claims for causality based on mRNA expression data [29] and *NCAPG* (a regulatory subunit of the condensin complex, required for the condensing of chromosomes from interphase chromatin), which are subunits of complexes that are important for mitosis [31,32].

Novel pathways and mechanisms

For many of the implicated genes, it is far from clear how they influence height. This is either because the function of the gene has not been classified or the gene has not previously been implicated in developmental pathways and processes. For example, a variant of Zinc finger and BTB domain containing (*ZBTB38*) is the most strongly associated across all the GWAS performed to date and there is evidence from mRNA expression data to indicate that it is the causal gene at the locus [21]. The fact that it is a methyl-DNA-binding transcriptional repressor gene indicates that it has a role in regulation of expression through an effect on chromatin structure [33]. In rats, the homolog of *ZBTB38* has been shown to regulate the transcription of the tyrosine hydroxylase gene, which is the rate-limiting enzyme in catecholamine biosynthesis [34]. Elucidating the mechanism by which *ZBTB38* variants (as well as other genes such as *C6orf173*, a hypothetical gene, but the only gene at the locus) impact on height should provide novel insights into human growth.

Many of the loci contain genes mutated in monogenic growth and skeletal development syndromes

Although the GWAS have identified the first common variants to associate with normal variation of height in the general population, many individual genes have been identified that, when mutated, severely affect growth and/or skeletal development in humans. Online Mendelian Inheritance in Man® (OMIM®) (http://www.ncbi.nlm.nih.gov/omim/) provides an up-to-date list of the hundreds of these rare single-gene syndromes. There is much overlap between the genes, pathways and biological processes implicated by the GWAS studies and those identified through these monogenic human and animal model studies (Table 1). For example, mutations in genes involved in skeletal development signaling pathways [35] (e.g. *IHH* mutations and brachydactyly type A1 [36]), extra-cellular matrix (e.g. *ACAN* and spondyloepiphyseal dysplasia, a syndrome presenting with short stature and severe early

onset osteoarthritis [37]) and genes involved in chromatin organization (e.g. mutations in the histone methyltransferase *NSD1* cause Sotos syndrome, which presents with extreme tall stature [38]), have all been shown to severely affect growth and/or skeletal development in humans. The role of cell-cycle and mitosis-related genes has recently been highlighted by the identification of a gene for primordial dwarfism, a severe stature disorder in which patients present with an average adult height of only 100 cm, various bone abnormalities and small brain size (although near-normal intellect). Mutations of the pericentrin (*PCNT*) gene were shown to be the cause of the disease [39]. *PCNT* encodes a protein that localizes specifically to the centromere throughout the cell cycle and its absence results in disorganized mitotic spindles and mis-segregation of chromosomes. The overlap between the common variants explaining normal variation of human height and rare monogenic growth and development syndrome genes indicates a complementary approach to further gene identification. Genes identified by monogenic studies will be excellent candidates for harboring the variants that explain normal variation in height (the GWAS results indicate that the major reason for the failure of previous candidate gene studies was insufficient sample size and inadequate gene coverage), whereas polygenic genes will be excellent candidates for being involved in rare monogenic conditions presenting with extreme stature, but in which the gene has not yet been mapped.

Links to disease and pleiotropic effects

Many of the monogenic growth syndromes often present with pleiotropic effects, most notably with cancer. As more common variants influencing a particular trait are identified by GWAS, pleiotropic effects are being observed. For example, a variant of *TCF2* increases risk of prostate cancer but decreases risk of type 2 diabetes [40]. So far, there are three examples of common height-associated variants that have pleiotropic effects on disease; the same allele of *GDF5* that associates with greater height also associates with reduced risk of osteoarthritis [41,42]; the allele at *CDK6* that associates with greater height also associates with increased risk of rheumatoid arthritis [43]; and an allele of melanocortin 4 receptor (*MC4R*) that associates with increased risk of obesity is also strongly associated with greater height (the association with height does not reach genome-wide levels of significance and is not among the 44 loci discussed here, but the finding is consistent with that seen when the *MC4R* gene is mutated in monogenic obesity) [44]. The *MC4R* finding is particularly interesting because it is an example of a variant that influences not just linear growth but also overall body shape. It is likely that more variants with pleiotropic effects will soon be documented, which might help explain some of the associations between height and disease seen in epidemiological studies [9–12].

Polygenic human traits – insights from GWAS

The GWAS approach has been particularly successful at tracking genes for height, at least in terms of the number of loci identified. In addition to other factors, such as the accuracy with which height can be measured and its constancy over adult life, one of the major reasons for this success is that height is measured in almost all studies on which GWAS are performed. Investigators can, therefore, combine studies from across a variety of cohorts (both disease- and population-based) through meta-analysis (as was done for the five published height GWAS [17–21]) so that extremely large sample sizes can be analyzed. Such sample sizes are not so easily achievable for many other traits (e.g. ranging from blood pressure to IQ to specific biochemical measurements, such as detailed glucose tolerance tests) that require more detailed measurement and even then are likely to be less accurately measured and less stable over adult life compared with height. What do the GWAS results tell us about the genetic architecture of this model complex trait? And what has it taught us about the identification of genes for polygenic traits in general? Extrapolation of these findings to other traits needs to be done with caution – for example, there might be less scope for gene–environment interaction in a trait in which such a large proportion of the variation is due to genetic variation.

Many common genetic variants of small effect contribute to height variation

Height is ~80% heritable but, in total, the common variants identified so far explain only ~5% of this, with the most strongly associated variant across the GWAS studies explaining only ~0.3% [17–21]. There has been no robust evidence of linkage to any particular chromosomal region [6,45] (and the GWAS-associated loci do not cluster in linked regions; Figure 1). As predicted by Fisher [8], this strongly indicates that the heritability of height is explained by many variants of individually small effect. Some rarer variants of large effect might be identified by future studies, but it is unlikely that these variants (given the absence of linkage evidence) will individually explain a substantial proportion of the population variation of height.

None of the variants identified to date show any evidence of deviation from an additive mode of inheritance either within or between loci, even for genes occurring in the same biological pathway (e.g. Hedgehog signaling genes) [17–21]. Similarly, there is currently no evidence that any of the associated variants have sex-specific effects. Such effects might have been expected given the sexual dimorphism of height, the clear gender difference in growth trajectories and the influence of sex hormones. So far, however, all the GWAS studies have been based on a single-point additive model, not stratified by gender and will, therefore, have been biased away from identifying such effects. A comprehensive genome-wide association view of the importance of these phenomena is therefore needed.

More loci will be identified by larger studies, but how many?

Most of the heritability of height remains to be explained. The results from the three large GWA studies of height indicate that the common variant-based approach will identify many more loci as sample size increases. One indication of this comes from the lack of overlap between

the results of the GWA studies [46]. Only 11 of the 44 loci were identified in at least two out of the three studies and only four were identified in all three studies. The most likely reason for this is that studies of 15 000–30 000 individuals were underpowered to detect many of the loci. This meant that chance had a large role in determining whether or not a variant was identified (Figure 4). Further evidence comes from the large excess of independent loci in each GWAS, which reached strong but not individually convincing levels of statistical significance [19–21].

It is very likely that the next year will see GWAS meta-analyses expand to sample sizes of >100 000. To estimate the number of loci that it might be possible to detect with the current single-nucleotide polymorphism (SNP)-based GWAS approach using realistic GWAS sample sizes, we can perform some simple extrapolations from the existing studies. We meta-analyzed data from six different GWAS studies and, based on these data, each doubling of the sample size led to an approximate doubling of the number of loci with $P < 1 \times 10^{-5}$ (for example, four such loci when N = 1914 to 27 loci when N = 13665, although we would only expect four by chance under the null distribution, thereby indicating that most of these are real associations) [19]. Consistent with this study, Gudbjartsson *et al.* [21] used an effective sample size of ~27 000 individuals with GWAS data (~2 × the total N of Weedon *et al.* [19]) and observed 53 independent signals at $P < 1 \times 10^{-5}$ (~2 × the total N of independent signals at the same statistical cut-off). If such a trend continues, these extrapolations indicate that, using 100 000 individuals with GWAS data, we could identify ~200 associated loci. Such simple linear predictions are of course highly speculative and do not take into account how much of the variation in height is explained by common variation, the distribution of effect sizes and where the inevitable 'plateau' will occur. Despite this, such extrapolations indicate that even with these very large (but feasible) sample sizes, common variation captured on existing chips is unlikely to explain much more than 15–20% of population height variation.

Height variants occur in genes, outside of genes and a long way from genes

The identification of a large number of variants associated with height means we can start to assess the positions in which quantitative trait variants lie in relation to genes and other features of the genome. This might give us some insight into the feasibility of narrowing down causal variants and genes for polygenic traits. This is important so that the discovery of variants can lead to novel biological insights. Figure 5 summarizes the relative locations of index SNPs at the 44 confirmed height loci in relation to sequence and genomic features. 59% (26/44) of the index SNPs lie within a gene, two are non-synonymous variants, one occurs in the 3′ untranslated region and the remaining 23 in introns. The remaining 41% (18) lie between 2 and 297 kb from the nearest gene. However, these figures are not overly helpful because the correlation between variants owing to linkage disequilibrium means that the causal variant could be many kilobases away from the index SNP. It is more useful to define the region that is most likely to contain the causal variant. Here, we can be more precise – it is highly likely (although not certain) that causal variants will lie within regions defined by recombination hotspots. These regions can be generated from HapMap data. The recombination regions (defined by flanking recombination hotspots) containing the 44 height variants range in size from 25 to 813 kb (median 145 kb). The number of genes that occur in these regions ranges from 0 to 42 (median two genes; Figure 5). In the four situations in which there are no genes in a recombination region (the nearest genes are between 105 and 118 kb from a flanking hotspot), there is a strong indication that the variant is a regulatory variant.

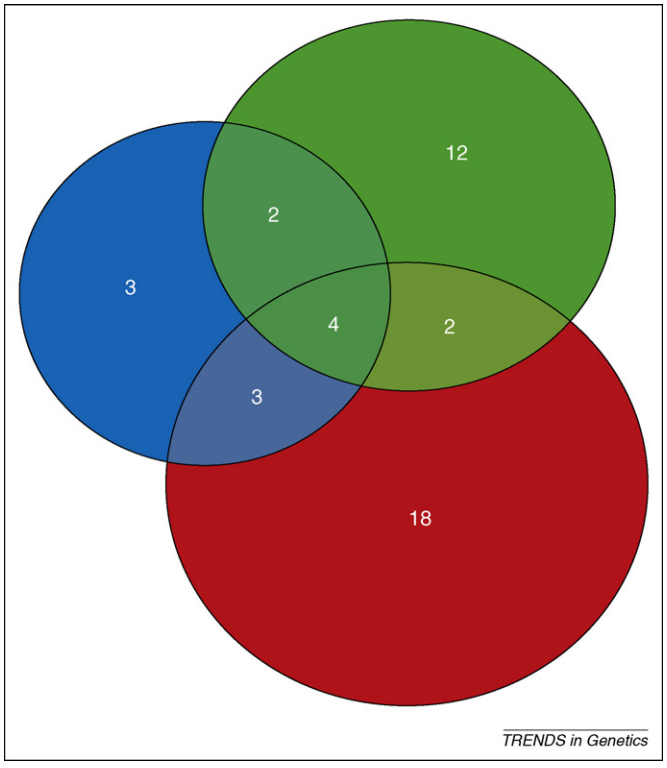

Figure 4. Comparing results across studies. The figure shows a Venn diagram for the 44 independent loci identified across the three large genome-wide association studies. Numbers represent robustly identified loci in Weedon *et al.* [19] (N = 13 665; green circle); Lettre *et al.* [20] (N = 15 821; blue circle) and Gudbjartsson *et al.* [21] (N ~27,000, red circle) and the overlap of the loci. The lack of overlap is most likely caused by the small effects sizes that these variants confer, together with the large number of such loci. This means that for many variants a study size of 14 000–27 000 individuals is not well powered to detect specific 'real' variants at the levels of significance needed to justify follow-up of GWAS results (typically $P < 1 \times 10^{-5}$). For example, of the 20 variants identified by Weedon *et al.* [19], eight had effect sizes in replication samples that would only be detected at $P < 1 \times 10^{-5}$ in 16 000 individuals, in two in ten studies. If there is only 20% power to detect one particular variant (and we assume each of the three GWAS had equivalent power) this means that only one in 100 times (0.2 × 0.2 × 0.2) would all three studies find the variant. The fact that individual variants with these small effect sizes are being identified from these GWAS must, therefore, mean that there is a large pool of such loci with similar effect sizes and whether or not a variant rises to the top in any one study then depends on sampling error. This is a statistical term that describes how sampling different datasets from the same population (in this case Europeans) will, purely by chance, result in different effect size estimates. For example, if the real effect size of a height allele is 0.5 cm, then simply by chance, some studies will estimate its effect as 0.6 cm, some at 0.4. This will explain most of the lack of overlap between the height GWAS results.

Concluding remarks and future perspectives

After many years of searching, the 'Mendelian factors' explaining normal variation of human height are being uncovered. The results have provided numerous novel

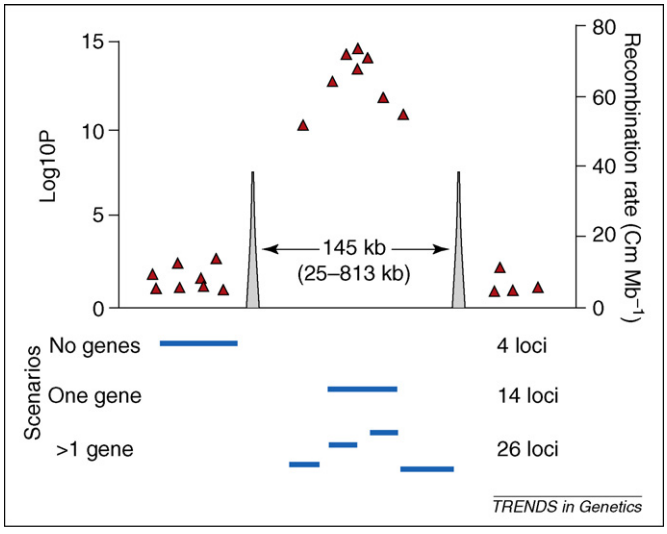

Figure 5. Size and number of genes in height-associated regions that are defined by flanking recombination hotspots. The plot shows that the majority of the associated loci contain more than one gene, such that we cannot be certain which gene is causal. Fine-mapping is required to determine the functional gene, so that novel biological insights can be gleaned. It also demonstrates the scale of re-sequencing efforts required to fine-map the causal variant at the associated loci, with a median distance between recombination hotspots of 145 kb (range: 25–813 kb). Recombination hotspots are based on HapMap Phase II data. The data and methods used to derive these hotspots are available from the HapMap website (http://www.hapmap.org/). See also Ref. [15]. The large grey triangles represent recombination hotspots. Red triangles represent single-point genome-wide association results. The blue lines represent genes.

insights into growth and development processes, disease and the genetic architecture of quantitative traits and many more such insights are likely in the near future. First, it is likely that large meta-analyses of GWAS data will result in the identification of >100 common variants robustly associated with height. These efforts are being led by large consortia such as the Genetic Investigation of ANthropometric Traits (GIANT) consortium, which currently consists of ~33 000 individuals from 13 different studies with GWAS data and which is likely to expand to include >100 000 individuals in the near future. Second, although far from a trivial task, especially for variants with such small effect sizes, attempts at fine-mapping the causal genes and variants by deep re-sequencing the associated loci might enable us to be more certain of the causal genes and variants. Such efforts are necessary, however, not only to provide novel biological insights, but they might also enable us to explain substantially more of the heritability of height. This is because the causal variant (which in most cases will not have been genotyped) can only be an equal or (perhaps substantially) better predictor of height than the marker variants we are using to identify it and also because it might enable the identification of additional independent common variants at the same loci. We already have one example of this, with two robust independent signals (defined as a lack of correlation between SNPs: HapMap $r^2 = 0.003$) occurring in the *PLAG1* gene. Third, it seems unlikely that even meta-analyses and fine-mapping efforts on this scale will result in a set of loci that explain much more than 20–25% of the genetic contribution to height. This means height might be an ideal phenotype to use to try and gauge the extent to which complex variation (e.g. copy number polymorphisms) and rare variation (the many millions of variants that occur in < 1% of individuals) contribute to polygenic traits. Such studies are now becoming feasible with the advent of next-generation sequencing and the ability to re-sequence a substantial amount of sequence in a large number of individuals. Fourth, it will be intriguing to assess the role of known height variants in relation to disease. Here, height variants might be used as important fine-mapping tools. For example, the correlation between variants that is caused by linkage disequilibrium is substantially reduced in individuals of African ancestry [15]. This means that African populations might be useful (depending on differences in allele frequencies and/or genetic architecture) for narrowing down causal variants. However, often it might be difficult to collect large numbers of disease cases and controls of African ancestry, especially for diseases that are more common in western environments. If, however, a height variant also alters disease risk, height could be used as the 'fine-mapping' phenotype. Finally, the genes implicated are likely to continue to be a mixture of those already known to have a key role in growth and development and those which offer new insights into these processes.

Acknowledgements

M.W. is a Vandervell Foundation Research Fellow. The authors would like to thank Cecilia Lindgren and Rachel Freathy for their careful reading and helpful comments on drafts of this article.

References

1 Cole, T.J. (2003) The secular trend in human physical growth: a biological view. *Econ. Hum. Biol.* 1, 161–168
2 Macgregor, S. *et al.* (2006) Bias, precision and heritability of self-reported and clinically measured height in Australian twins. *Hum. Genet.* 120, 571–580
3 Preece, M.A. (1996) The genetic contribution to stature. *Horm. Res.* 45, 56–58
4 Silventoinen, K. *et al.* (2000) Relative effect of genetic and environmental factors on body height: Differences across birth cohorts among Finnish men and women. *Am. J. Public Health* 90, 627–630
5 Silventoinen, K. *et al.* (2003) Heritability of adult body height: A comparative study of twin cohorts in eight countries. *Twin Res.* 6, 399–408
6 Perola, M. *et al.* (2007) Combined genome scans for body stature in 6,602 European twins: evidence for common Caucasian loci. *PLoS Genet.* 3, e97
7 Galton, F. (1885) Regression towards mediocrity in hereditary stature. *J. R. Anthropol. Inst.* 5, 329–348
8 Fisher, R.A. (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* 52, 399–433
9 Davey Smith, G. *et al.* (2000) Height and risk of death among men and women: aetiological implications of associations with cardiorespiratory disease and cancer mortality. *J. Epidemiol. Community Health* 54, 97–103
10 Gunnell, D. *et al.* (2001) Height, leg length, and cancer risk: A systematic review. *Epidemiol. Rev.* 23, 313–342
11 Lawlor, D.A. *et al.* (2002) The association between components of adult height and Type II diabetes and insulin resistance: British Women's Heart and Health Study. *Diabetologia* 45, 1097–1106
12 Lawlor, D.A. *et al.* (2004) Associations of components of adult height with coronary heart disease in postmenopausal women: the British Women's Heart and Health Study. *Heart* 90, 745–749
13 Risch, N. and Merikangas, K. (1996) The future of genetic studies of complex human diseases. *Science* 273, 1516–1517
14 Lettre, G. *et al.* (2007) Common genetic variation in eight genes of the GH/IGF1 axis does not contribute to adult height variation. *Hum. Genet.* 122, 129–139

15 Frazer, K.A. *et al.* (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449, 851–861
16 McCarthy, M.I. *et al.* (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* 9, 356–369
17 Weedon, M.N. *et al.* (2007) A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nat. Genet.* 39, 1245–1250
18 Sanna, S. *et al.* (2008) Common variants in the GDF5-UQCC region are associated with variation in human height. *Nat. Genet.* 40, 198–203
19 Weedon, M.N. *et al.* (2008) Genome-wide association analysis identifies 20 loci that influence adult height. *Nat. Genet.* 40, 575–583
20 Lettre, G. *et al.* (2008) Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat. Genet.* 40, 584–591
21 Gudbjartsson, D.F. *et al.* (2008) Many sequence variants affecting diversity of adult human height. *Nat. Genet.* 40, 609–615
22 Kronenberg, H.M. (2003) Developmental regulation of the growth plate. *Nature* 423, 332–336
23 Gazzerro, E. and Canalis, E. (2006) Bone morphogenetic proteins and their antagonists. *Rev. Endocr. Metab. Disord.* 7, 51–65
24 Argraves, W.S. *et al.* (2003) Fibulins: physiological and disease perspectives. *EMBO Rep.* 4, 1127–1131
25 Krane, S.M. and Inada, M. (2008) Matrix metalloproteinases and bone. *Bone* 43, 7–18
26 Woodward, J.K. *et al.* (2007) The roles of proteolytic enzymes in the development of tumour-induced bone disease in breast and prostate cancer. *Bone* 41, 912–927
27 Feng, Q. *et al.* (2002) Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr. Biol.* 12, 1052–1058
28 Berger, J. *et al.* (1999) The human homolog of Sex comb on midleg (SCMH1) maps to chromosome 1p34. *Gene* 237, 185–191
29 Dixon, A.L. *et al.* (2007) A genome-wide association study of global gene expression. *Nat. Genet.* 39, 1202–1207
30 Malumbres, M. and Barbacid, M. (2005) Mammalian cyclin-dependent kinases. *Trends Biochem. Sci.* 30, 630–641
31 Thornton, B.R. *et al.* (2006) An architectural map of the anaphase-promoting complex. *Genes Dev.* 20, 449–460
32 Legagneux, V. *et al.* (2004) Multiple roles of condensins: a complex story. *Biol. Cell* 96, 201–213
33 Filion, G.J. *et al.* (2006) A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Mol. Cell. Biol.* 26, 169–181
34 Kiefer, H. *et al.* (2005) ZENON, a novel POZ Kruppel-like DNA binding protein associated with differentiation and/or survival of late postmitotic neurons. *Mol. Cell. Biol.* 25, 1713–1729
35 Kornak, U. and Mundlos, S. (2003) Genetic disorders of the skeleton: a developmental approach. *Am. J. Hum. Genet.* 73, 447–474
36 Gao, B. *et al.* (2001) Mutations in IHH, encoding Indian hedgehog, cause brachydactyly type A-1. *Nat. Genet.* 28, 386–388
37 Gleghorn, L. *et al.* (2005) A mutation in the variable repeat region of the aggrecan gene (AGC1) causes a form of spondyloepiphyseal dysplasia associated with severe, premature osteoarthritis. *Am. J. Hum. Genet.* 77, 484–490
38 Kurotaki, N. *et al.* (2002) Haploinsufficiency of NSD1 causes Sotos syndrome. *Nat. Genet.* 30, 365–366
39 Rauch, A. *et al.* (2008) Mutations in the pericentrin (PCNT) gene cause primordial dwarfism. *Science* 319, 816–819
40 Gudmundsson, J. *et al.* (2007) Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat. Genet.* 39, 977–983
41 Miyamoto, Y. *et al.* (2007) A functional polymorphism in the 5′ UTR of GDF5 is associated with susceptibility to osteoarthritis. *Nat. Genet.* 39, 529–533
42 Southam, L. *et al.* (2007) 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, 2226–2232
43 Raychaudhuri, S. *et al.* (2008) Common variants at CD40 and other loci confer risk of rheumatoid arthritis. *Nat. Genet.* 40, 1216–1223
44 Loos, R.J. *et al.* (2008) Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat. Genet.* 40, 768–775
45 Visscher, P.M. *et al.* (2007) Genome partitioning of genetic variation for height from 11,214 sibling pairs. *Am. J. Hum. Genet.* 81, 1104–1110
46 Visscher, P.M. (2008) Sizing up human height variation. *Nat. Genet.* 40, 489–490
47 Thomas, P.D. *et al.* (2003) PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* 13, 2129–2141
48 Minina, E. *et al.* (2001) BMP and Ihh/PTHrP signaling interact to coordinate chondrocyte proliferation and differentiation. *Development* 128, 4523–4534