Abstract
Despite the high heritability of human growth traits, until recently little was known about the underlying genes and genetic variants which explain normal variation of growth. In the past few years, genome-wide association studies have successfully identified hundreds of genetic variants that are associated with human growth traits. These variants have implicated many novel genes in the regulation of birthweight and pubertal timing through to final adult height, and are providing new insights into the biology of growth. For example, 180 genetic loci have been robustly shown to influence variation in final adult height. Despite this success, the effect sizes of these variants are small and, even in combination, have left the majority of heritable genetic variation of growth traits unexplained. In this review, I discuss the successes of the genome-wide association approach and some of the novel insights into the biology of growth that have come from these studies. I will also discuss what these studies have not told us and what the future holds for genetic studies of human growth.

Introduction
It has long been known that variation in the rate and extent that humans grow is largely due to inherited genetic variation. The proportion of variation explained by heritable factors, as opposed to environmental factors, for different aspects of growth range from around 40% for birthweight [1] to around 80% for adult height [2, 3]. Until recently, however, the specific genetic variants that explain this heritability were largely unknown.

There has been substantial success in identifying the specific genetic mutations which cause some extreme growth syndromes. For example, loss-of-function
mutations of the \textit{IGF1R} gene (insulin-like growth factor-I) have been shown to cause intrauterine and postnatal growth retardation [4]. Another example comes from Rauch et al. [5] who demonstrated that loss of function mutations of \textit{PCNT}, the gene encoding pericentrin, which is involved in centromere function in the cell cycle, are a cause of primordial dwarfism. The number of such genes implicated in extreme growth is large and growing (see OMIM, http://www.ncbi.nlm.nih.gov/omim). However, the genetic mutations that explain these extremes of growth are rare in the population and cannot explain normal variation in growth, and until recently there were no common genetic variants that were robustly shown to influence variation in any growth-related trait. This has all changed with the advent of genome-wide association (GWA) studies.

The traditional linkage and candidate gene approaches that are extremely powerful in identifying rare, highly penetrant mutations causing disease were not successful for identifying the variants that explain normal variation in growth. The linkage approach has not been a success because, compared with association methods, it is not a powerful approach to detect variants of small effect size, even when large sample sizes are used [6]. Whereas the candidate gene approach is unsuccessful because of a combination of the large number of potential candidate genes (there are \(\sim\)20,000 genes in the human genome, many potentially involved in growth and development processes), and also the small sample sizes that many studies used (typically just a few hundred samples, when many thousands are required). Instead, the development of genotyping microarray chips, which allow most of the common genetic variation (that is the \(>5\) million variants which occur in the population at \(>5\%\) frequency) to be assayed simultaneously in an individual, combined with very large sample sizes and collaborative efforts has led to the large-scale GWA study approach. These efforts have started to uncover the genetic variants that explain normal variation in growth.

In this review, I will discuss the recent successes of GWA studies for human growth traits. I will describe the hundreds of loci that have been identified for phenotypes ranging from birthweight to final adult height. I will discuss what these findings have told us about the genetic architecture of these traits as well as the novel biological insights that have been gained. I will also discuss what these studies have not told us – there is a large fraction of heritability that remains to be explained – and what the future directions are for genetic studies of growth.

\textbf{The First Successes for GWA Studies of Growth Traits – FTO and HMGA2}

In 2007, as part of the Wellcome Trust Case Control Consortium (WTCCC) [7], we reported a GWA study of type 2 diabetes, where we identified seven loci where common genetic variants influenced risk of the disease [8]. We went on to show that one of these genes, \textit{FTO}, was increasing risk of type 2 diabetes
through a primary impact on increasing weight [9]. We demonstrated that the effect was seen from the age of 7, and was specifically due to an increase in fat mass. The FTO finding provides an excellent example of the power of GWA studies to identify new biology. Before our study, FTO was a poorly described gene in an unknown pathway. A large amount of work has since been put into understanding the function of FTO and how it is affecting weight. It has been shown, for example, that it is a nucleic acid demethylase [10]. However, the underlying biological mechanism is still not entirely clear. Inactivation of FTO in a mouse model produced mice that were lean, primarily due to an effect on basal metabolic rate [11], while a study which overexpressed FTO in mice suggested a primary effect on appetite [12]. The appetite effect is supported by human studies [13–15]. Recently, in a longitudinal follow-up study of children from early infancy until age 13 years, the FTO variant that increases BMI in adults was shown to have the opposite effect in infants below the age of 2.5 years [16]. Further elucidation of the role of FTO during development is likely to provide important new insights into the biology of obesity and growth.

The successful identification of genes for type 2 diabetes and BMI spurred on efforts to identify genes for other traits. Adult height is a classic genetic trait which was used as a model trait by Galton [17] and Fisher [18] in the early days of human genetics. It is a model trait because of its high heritability and because it is easily, accurately and widely measured. We therefore performed an initial GWA study of adult height in 2,000 individuals (from the same WTCCC type 2 diabetes study) and identified no significant associations. It was only when we meta-analyzed our GWA association data with 3,000 individuals from the Diabetes Genetic Initiative study, that a significant association was observed [19]. The variants that associated with height occurred in the HMGA2 gene, which is a high mobility group protein involved in chromatin remodeling. HMGA2 was an excellent candidate to influence height because the knockout of this gene causes the pygmy mouse phenotype [20], and an inversion of this gene was shown to cause an 8-year-old boy to be more than 5 standard deviations taller than expected for his age [21]. The effect size of the HMGA2 variant was small – homozygotes for the variant allele, which has a frequency of ~50% in Europeans, were on average only a 1 cm taller than the reference homozygotes, and the variant explained only 0.3% of the variation in height. There were clearly many more variants contributing to normal variation in height.

**Increasing Sample Size by Collaboration and Meta-Analysis Identifies Hundreds of Loci for Adult Height and Other Anthropometric Traits**

The only way to find these additional loci was to increase the sample size of GWA studies. We therefore collaborated with other groups from the WTCCC and increased our discovery sample size to 13,665 individuals and used 16,482
for replication of the most associated variants. This led to the identification of 19 further loci convincingly associated with adult height. Other groups were performing similar studies [22, 23], and in 2008 a total of 52 loci were reported to be associated with final adult height. In combination, however, these loci explained only ~5% of the normal variation in European populations. It was clear that even further expansion and combination of data would yield even more associated loci, and so the GIANT (Genetic Investigation of Anthropometric Traits) consortium was set up. The GIANT consortium analysis for height included 183,727 participants from 45 individual GW A studies (from a range of different disease areas) [24]. Analysis of this huge number of individuals led to the identification of 180 height-associated loci. In combination, these variants explained ~10% of the population variation in final adult height (and so ~12% of the heritable variation).

The biological insights that can be gained from GW A studies are limited by an inability to pinpoint the exact causal mutation due to correlation with nearby variants that are co-inherited. This correlation is useful in that it allows us to identify the associations in the first place (by reducing the number of variants that need to be genotyped); but, once we have identified the association it prevents us from, at least genetically, narrowing down the causal variant from the many other, often tightly, linked, variants. So much further work is needed to fine-map the associations identified from GW A studies to provide the new insights into the biology of growth that these GW A studies have promised.

Nevertheless, we demonstrated that there is an overrepresentation of strong candidate genes at the 180 height-associated loci (i.e. those where rare, severe, mutations cause extreme growth), and that the loci we have identified tend to contain genes which are biologically connected and which cluster in biological pathways. These pathways are both known and novel and include the hedgehog signaling, growth hormone and histone modification pathways [24].

While GW A studies for height have been particularly successful, a similar pattern has been observed for other growth traits. In addition to height, the GIANT consortium also analyzed other anthropometric phenotypes, including BMI and waist circumference. For BMI, 32 associated loci were recently reported [25]. These variants, however, account for only ~1% of the population variation in BMI – with FTO accounting for the majority of this. Variants associated with waist circumference (independently of BMI) have also been identified – and interestingly strong sex-specific differences were observed [26], whereas for height and BMI no such gender interactions were found [24, 25].

**GWA Studies of Early Growth**

In addition to adult phenotypes, substantial effort has been put into trying to map genes for earlier stages of growth. For example, there has been recent success in identifying genes for birthweight. Freathy et al. [27] performed a GW
study in 10,623 children, and identified two loci strongly associated with birthweight. Interestingly, one of these, *ADCY5*, is also a variant associated with type 2 diabetes. This fits with earlier work showing that genetic variants that associate with reduced birthweight also increase risk of type 2 diabetes [27, 28], and is consistent with the fetal insulin hypothesis proposed by Hattersley and Tooke [29] that states that the strong epidemiological correlation between lower birthweight and increased risk of adult-onset diabetes could, at least partially, be explained by a shared genetic cause rather than the fetal programming proposed by Barker [30].

So far, none of the variants associated with adult height have been robustly associated with birthweight or length. Although, a recent study suggests that it is likely that a subset of the 180 height-associated variants will affect birth length and that, in combination, they explain more of the variation in growth as age increases (for example, explaining 5% of population variation of height at age 10, compared to 10% of adult height) [31]. It will be interesting to observe in future studies what the overlap between variants associated with height and fetal and postnatal stages of growth is – and it is likely that at least some of the adult height variants will be primarily fetal growth genes.

**GWA Studies of Pubertal Growth**

GWA studies of later stages of growth, in particular aspects of puberty, have also yielded associations. One particularly interesting association from these studies was at the *LIN28B* locus [32, 33]. *LIN28B* is a regulator of microRNA processing, and the same variant of this gene was also shown to be associated with adult height – with the allele that associates with earlier age at menarche associated with shorter height. Ong et al. [33] went on to show that this is a pubertal timing variant in both sexes. The allele that reduces menarche age in girls was also associated with earlier breast development, and in boys was associated with earlier voice breaking. Widen et al. [34] performed a GWA study of growth trajectories in 5,038 Finnish children and came to similar conclusions, and also found a second, independent signal at the same locus which appears to be a sex-specific effect. In this case, the causality of the *LIN28B* gene at this locus is supported by the mouse knockout of *LIN28A* (a homolog of *LIN28B*) which demonstrates consistent effects on reduced growth and early puberty phenotypes [35].

As with other traits, a larger scale meta-analysis (including 87,802 women) yielded further associations, with a total of 32 loci now being robustly associated with age at menarche [36]. These loci contain genes that fall into pathways related to energy homeostasis (*BSX, CRTCI* and *MCHR2*), hormonal regulation (*INHBA, PCSK2* and *RXRG*) pathways and coenzyme A and fatty acid biosynthesis. The strongest overlap, however, was for genes which had previously been reported to be associated with BMI – for example, the allele that is strongly
associated with increased BMI at the FTO locus is associated with reduced age at menarche, and presumably earlier puberty. This pattern was seen across all the known BMI loci, except for the MC4R locus. MC4R primarily influences weight through increasing muscle mass, and this suggests that the effect of these variants on reducing age at menarche is primarily through an effect on fat mass. This is consistent with earlier studies demonstrating that increased adiposity reduces time to puberty for girls [37]. The association of the height loci with age at menarche was more complex, with some associations (such as the LIN28B locus) having a direction consistent with epidemiological studies (earlier menarche is associated with reduced height), but other loci had the opposite direction of effect [36].

The Genetics of Growth Is More Complex Than Previously Appreciated

The large number of loci identified from these GWA studies of growth allows us to make some general conclusions about the genetic architecture of normal growth. The most obvious conclusion is that a very large number of genes and variants are responsible for normal variation in development. That multiple variants influence height is not surprising – in 1918 Ronald Fisher in his classic paper showed that this was likely the case [18]. What is surprising is the sheer number of variants. We have estimated that 697 variants would be identified with a sample size of 500,000 but that this would only explain ~20% of heritability [24] – this indicates that many thousands of variants in many thousands of genes may ultimately be responsible for normal variation in height – and millions of individuals would be needed to identify them all. In hindsight, this is perhaps not surprising given that adult height is a combination of a huge number of developmental and other processes.

What GWA Studies Have Not Told Us – The Missing Heritability

GWA studies have been a success – in many cases they have identified the first genetic variants to influence variation of a common disease or trait (a full list of GWA study associations across all traits is available at http://www.genome.gov/26525384), which have provided new insights into biology and pathophysiology (e.g. FTO and LIN28B). The follow-up to these studies is likely to provide many new insights as functional variants and genes involved in these traits are fine mapped. Although, given the difficulty of these efforts and the large number of loci, it may be that the focus will be on identifying novel pathways in which genes at these loci cluster.

One aspect that has been disappointing is the small effect size of the variants identified and the limited amount of heritability that they explain. This means
that the clinical utility of these markers for prediction is limited for most traits. For example, for height it has been shown that knowing parental height is by far a better way of predicting a child’s height than using variants identified from GWA studies – and it is likely to be the case for some time to come [38]. The relatively small amount of heritability explained also means that there is a lot of biology left to be understood. So, one of the major questions in complex traits genetics currently is: where is this missing heritability? There are several possible sources [39]. It could be that common variants will explain a substantial amount of the missing heritability – it is just that exceptionally large (i.e. millions of individuals) samples will be needed to detect these variants. The fall-off in effect sizes of newly discovered variants as larger sample sizes are analyzed suggests that there is a limit to the amount that common genetic variants can explain with reasonable sample sizes (see fig. 1). Although, work by Yang et al. [40] suggests that at least 45% of the heritability of height can be explained by genetic variants captured on current versions of genotyping chips. However, even if this proves true, this is likely to leave a substantial amount of heritability unexplained.

Another possibility is that lower frequency (0.5–5% population frequency) and rare (<0.5%) variants may explain much of the remaining heritability. These lower frequency variants are not well captured on current versions of genotyping chips (which have focused on variants with frequencies >5%). Studies in type 1 diabetes, where low frequency, but reasonably penetrant (odds ratios ~3) variants were found in the IFIH1 gene using a next generation sequencing approach, demonstrate this type of variant exists and demonstrates that it can be used to identify conclusively the gene at an associated loci [41]. We can now systematically test for association for this type of variant because of the

![Fig. 1. Cumulative proportion of variation explained for the 180-height associated SNPs. The data are based on stage 2 of the GIANT height study [24].](image-url)
huge leaps in sequencing technologies that have occurred in the past few years. These technologies now allow a high-quality whole human genome sequence to be completed within a week and for less than USD 5,000 [42]. This is compared to the 10 years and USD >1 billion required to complete the reference sequence just 10 years ago [43]. It is also possible to sequence just the 1% of the genome which codes for proteins using sequence capture methods [44]. This whole exome sequencing approach allows for a more economic analysis of the most readily interpretable regions of the genome – with prices USD <1,000 per exome. With more and more groups performing these types of study, the focus will move away from GWA chips and towards a next generation of GWA studies using whole genome and whole exome sequencing data. It is likely that cohorts of the size currently used for GWA will be needed to fully exploit these data, but there is hope that this will lead many new exciting discoveries and fill in much of the gap in unexplained heritability.

While rare variants may be of major importance, other explanations of the missing heritability include gene-gene and gene-environment interaction, structural variants and epigenetic effects. No evidence has so far been found for gene-gene interaction for any growth-related trait. For example, we assessed all 180 height-associated loci for interaction, and found no evidence of any deviation from an additive effects model [24]. Although, even though we analyzed 100,000 individuals, this may be due to lack of power due to subtle interaction effects, or that interacting variants do not present with strong main effects. Structural variants will almost certainly be a source of the missing heritability – and the detection of these should be facilitated by next-generation sequencing technologies. While common copy number variants are unlikely to have a disproportionate role (as demonstrated by the recent WTCCC study [45]), rarer structural variants will almost certainly play a role. One recent example comes from the demonstration that a low-frequency 600-kb deletion of 16p11.2 causes obesity, whereas duplication of this region causes underweight [25, 46]. Epigenetic effects will also play a role. There is already an example of this from a parent of origin analysis reported by Kong et al. [47] that demonstrated that the effect of three type 2 diabetes variants was dependent on the parent from which the allele was inherited. The importance of each of these potential sources of the missing heritability will become clearer over the next few years.

**Conclusions**

GWA studies have been successful. They have identified the first common genetic variants associated with normal variation in growth. Based on these findings, new and important insights into the biology of growth have already been made – and more will come with ongoing and future follow-up studies. The associated variants have, however, left a substantial amount of heritable
variation unexplained. With developments in sequencing technologies and large-scale collaborations, genetic studies of growth should fill in this gap over the coming years.

References

7 The Wellcome Trust Case Control Consortium: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661.


