Population Genetics and Pharmacogenetics

Ken K. Ong

Medical Research Council Epidemiology Unit, Cambridge; Medical Research Council Unit for Lifelong Health and Ageing, London; Department of Paediatrics, University of Cambridge, Cambridge, UK

Genome-wide association studies (GWAS) have been hugely successful in identifying thousands of genetic loci robustly associated with hundreds of disease and other human traits. Since the first GWAS publication in an endocrinology-related topic, obesity in 2007, these studies have very rapidly increased in size and productivity. In 2011, this chapter is now largely dominated by a number of 'mega' or greatly-expanded GWAS papers, each involving over 100,000 participants in the combined initial discovery and replication phases. Hundreds of new genes, or genetic loci, are reported for adult height, lipids, BMI, waist-hip ratio, type 2 diabetes (T2DM), and age at menarche. New topics for GWAS are also described, including polycystic ovary syndrome (PCOS) and hypospadias.

Prior to this wave of expanded studies, some commentators had criticized the relatively minor effect sizes of common genetic variants. They argued that the previous GWAS papers, already involving tens of thousands of individuals, would have already found the loci with the strongest effects, and that further expansion would only find variants with effects that are too small to be of biological interest. In addition to simple identification of genetic loci, many of the current papers describe key applications and supporting experimental data that underline the new biological advances. The GWAS locus at HMGCR is held up as an important illustration of the potential application of these novel findings, being associated with only a very modest change in LDL cholesterol levels, yet it encodes the target of statins that currently benefit tens of millions of patients.

We should not forget that prior to the GWAS era we had to endure barren decades of linkage mapping and weakly replicable candidate genotyping studies. Eric Lander's review article published in February 2011 on the 10th anniversary of the Human Genome Project describes the remarkable extent of achievements in the last decade and the rapid pace of current research efforts towards the routine use of sequencing as 'a simple and inexpensive general-purpose tool throughout biomedicine'.

New 'expanded-GWAS' associations: adult height

Hundreds of variants clustered in genomic loci and biological pathways affect human height

Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, Willer CJ, Jackson AU, Vedantam S, Raychaudhuri S, Ferreira T, Wood AR, Weyant RJ, Segre AV, Speliotes EK, Wheeler E, Soranzo N, Park JH, Yang J, Gudbjartsson D, Heard-Costa NL, Randall JC, Qi L, Vernon Smith A, Magi R, Pastinen T, Liang L, Heid IM, Luan J, Thorleifsson G, Winkler TW, Goddard ME, Sin Lo K, Palmer C, Workalemahu T, Aulchenko YS, Johansson A, Zillikens MC, Feitosa MF, Esko T, Johnson T, Ketkar S, Kraft P, Mangino M, Prokopenko I, Absher D, Albrecht E, Ernst F, Glazer NL, Hayward C, Hottenga JJ, Jacobs KB, Knowles JW, Kutalik Z, Monda KL, Polasek O, Preuss M, Rayner NW, Robertson NR, Steinthorsdottir V, Tyrer JP, Voight BF, Wiklund F, Xu J, Zhao JH, Nyholt DR, Pellikka N, Perola M, Perry JR, Surakka I, Tammesoo ML, Altmajer EL, Amin N, Aspelund T, Bhangale T, Boucher G, Chasman DI, Chen C, Coin L, Cooper MN, Dixon AL, Gibson Q, Grundberg E, Hao K, Juhani Junttila M, Kaplan LM, Kettunen J, König IR, Kwan T, Lawrence RW, Levinson DF, Lorentzon M, McKnight B, Morris AP, Müller M, Suh Ngwa J, Purcell S, Rafelt S, Salem RM, Salvi E, Sanna S, Shi J, Sovio U, Thompson JR, Turchin MC, Vandenput L, Verlaan DJ, Vitart V, White CC, Ziegler A, Almgren P, Balmforth AJ, Campbell H, Citterio L, De Grandi A, Dominiczak A, Duan J, Elliott P, Elosua R, Eriksson JG, Freimer NB, Geus EJ, Glorioso N, Haing S, Hartikainen AL, Havulinna AS, Hicks AA, Hui J, Igl W, Illig T, Julia A, Kajantie E, Kilpeläinen TO, Koiranen M, Kolcic I, Koskinen S, Kovacs P, Laitinen J, Liu J, Lokki ML, Marusic A, Maschio A, Meitinger T, Mulas A, Pare G, Parker AN, Peden JF, Petersmann A, Pichler I, Pietiläinen KH, Pouta A, Ridderstrale M, Rotter JJ, Sambrook JG, Sanders AR, Schmidt CO, Sinisalo J, Smit JH, Stringham HM, Bragi Walters G, Widen E, Wild SH, Willemsen G, Zagato L, Zgaga L, Zitting P, Alavere H, Farrall M, McArdle WL, Nelis M, Peters MJ, Ripatti S,

van Meurs JB, Aben KK, Ardlie KG, Beckmann JS, Beilby JP, Bergman RN, Bergmann S, Collins FS, Cusi D, den Heijer M, Eiriksdottir G, Gejman PV, Hall AS, Hamsten A, Huikuri HV, Iribarren C, Kahonen M, Kaprio J, Kathiresan S, Kiemeny L, Kocher T, Launer LJ, Lehtimäki T, Melander O, Mosley TH, Jr., Musk AW, Nieminen MS, O'Donnell CJ, Ohlsson C, Oostra B, Palmer LJ, Raitakari O, Ridker PM, Rioux JD, Rissanen A, Rivolta C, Schunkert H, Shuldiner AR, Siscovick DS, Stumvoll M, Tonjes A, Tuomilehto J, van Ommen GJ, Viikari J, Heath AC, Martin NG, Montgomery GW, Province MA, Kayser M, Arnold AM, Atwood LD, Boerwinkle E, Chanoock SJ, Deloukas P, Gieger C, Gronberg H, Hall P, Hattersley AT, Hengstenberg C, Hoffman W, Lathrop GM, Salomaa V, Schreiber S, Uda M, Waterworth D, Wright AF, Assimes TL, Barroso I, Hofman A, Mohlke KL, Boomsma DI, Caulfield MJ, Cupples LA, Erdmann J, Fox CS, Gudnason V, Gyllenstein U, Harris TB, Hayes RB, Jarvelin MR, Mooser V, Munroe PB, Ouwehand WH, Penninx BW, Pramstaller PP, Quertermous T, Rudan I, Samani NJ, Spector TD, Volzke H, Watkins H, Wilson JF, Groop LC, Haritunians T, Hu FB, Kaplan RC, Metspalu A, North KE, Schlessinger D, Wareham NJ, Hunter DJ, O'Connell JR, Strachan DP, Wichmann HE, Borecki IB, van Duijn CM, Schadt EE, Thorsteinsdottir U, Peltonen L, Uitterlinden AG, Visscher PM, Chatterjee N, Loos RJ, Boehnke M, McCarthy MI, Ingelsson E, Lindgren CM, Abecasis GR, Stefansson K, Frayling TM, Hirschhorn JN
Genetics of Complex Traits, Peninsula College of Medicine and Dentistry, University of Exeter, UK.
Nature 2010;467:832–8.

Background: Most human traits and common diseases display a polygenic pattern of inheritance: DNA sequence variants at many loci contribute to each phenotype. Genome-wide association (GWA) studies have identified more than 600 variants associated with human traits, but these yet explain only little phenotypic variation, questioning the utility of further studies.

Methods: GWA and replication genotyping were performed in up to 183,727 individuals for association with adult height, a highly heritable and classic polygenic trait.

Results: Hundreds of genetic variants, in at least 180 loci, were associated with adult height. The large number of loci reveals important implications for genetic studies of common human diseases and traits. First, the 180 loci are not random, but instead are enriched for genes that are implicated in biological pathways ($p = 0.016$) and that underlie skeletal growth defects ($p < 0.001$). Second, the likely causal gene is often located near the most strongly associated variant: in 13 of 21 loci containing a known skeletal growth gene, that gene was the closest to the associated variant. Third, at least 19 loci have multiple independently associated variants, suggesting that allelic heterogeneity is common for polygenic traits, and that comprehensive explorations of already known loci should discover additional variants. Fourth, associated variants are enriched for likely functional effects, being overrepresented among variants that alter amino-acid sequence and expression levels of nearby genes.

Conclusions: These findings explain ~10% of the phenotypic variation in height, and unidentified common variants of similar effect sizes are estimated to increase this figure to ~16% of phenotypic variation (or 20% of its heritability). Although additional approaches are needed to more fully dissect the genetic architecture of polygenic human traits, these findings indicate that GWA studies identify large numbers of loci that uncover biologically relevant genes and pathways.

In a year of large GWAS papers, this study of adult height takes top billing in terms of both sample size and number of genetic loci identified (and probably also for the number of authors!). Adult height has been most amenable to GWAS for a number of reasons: it is invariably measured or at least self-reported with reasonable accuracy; it is a relatively stable trait; and it shows the highest heritability of all body measures – 80% of the population variation in adult height is estimated to be genetic in origin.

While the initial adult height GWAS papers were notable for the absence of genes implicated in hormone pathways [1, 2], it is very reassuring to see that many of the major genes involved in the GH-IGF1 axis, and other hormonal pathways, were represented among the 180 adult height loci, including PITX1, GH1, GH5R, IGF1R, INSR1, ESR1, CYP19A1 (the aromatase gene), IGF2BP2, and IGF2BP3. These findings confirm the increasing recognition that hormonal modulation of such pathways, even in healthy children, may significantly impact on adult height. In total, 21 height loci were found to be located near to or in abnormal skeletal/growth genes. As for T2DM, there is increasing overlap between the genetic loci that underlie rare disease and common variation. It is remarkable to consider how many different biological pathways may impact on adult height. Further understanding of the timing of action of these variants on childhood growth will help to inform the basis of recognized links between adult height and various disease risks.

Biological, clinical and population relevance of 95 loci for blood lipids

Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, Johansen CT, Fouchier SW, Isaacs A, Peloso GM, Barbalic M, Ricketts SL, Bis JC, Aulchenko YS, Thorleifsson G, Feitosa MF, Chambers J, Orho-Melander M, Melander O, Johnson T, Li X, Guo X, Li M, Shin Cho Y, Jin Go M, Jin Kim Y, Lee JY, Park T, Kim K, Sim X, Twee-Hee Ong R, Croteau-Chonka DC, Lange LA, Smith JD, Song K, Hua Zhao J, Yuan X, Luan J, Lamina C, Ziegler A, Zhang W, Zee RY, Wright AF, Witteman JC, Wilson JF, Willemsen G, Wichmann HE, Whitfield JB, Waterworth DM, Wareham NJ, Waeber G, Vollenweider P, Voight BF, Vitart V, Uitterlinden AG, Uda M, Tuomilehto J, Thompson JR, Tanaka T, Surakka I, Stringham HM, Spector TD, Soranzo N, Smit JH, Sinisalo J, Silander K, Sijbrands EJ, Scuteri A, Scott J, Schlessinger D, Sanna S, Salomaa V, Saharinen J, Sabatti C, Ruokonen A, Rudan I, Rose LM, Roberts R, Rieder M, Psaty BM, Pramstaller PP, Pichler I, Perola M, Penninx BW, Pedersen NL, Pattaro C, Parker AN, Pare G, Oostra BA, O'Donnell CJ, Nieminen MS, Nickerson DA, Montgomery GW, Meitinger T, McPherson R, McCarthy MI, McArdle W, Masson D, Martin NG, Marroni F, Mangino M, Magnusson PK, Lucas G, Luben R, Loos RJ, Lokki ML, Lettre G, Langenberg C, Launer LJ, Lakatta EG, Laaksonen R, Kyvik KO, Kronenberg F, König IR, Khaw KT, Kaprio J, Kaplan LM, Johansson A, Jarvelin MR, Janssens AC, Ingelsson E, Igl W, Kees Hovingh G, Hottenga JJ, Hofman A, Hicks AA, Hengstenberg C, Heid IM, Hayward C, Havulinna AS, Hastie ND, Harris TB, Haritunians T, Hall AS, Gyllenstein U, Guiducci C, Groop LC, Gonzalez E, Gieger C, Freimer NB, Ferrucci L, Erdmann J, Elliott P, Ejebe KG, Doring A, Dominiczak AF, Demissie S, Deloukas P, de Geus EJ, de Faire U, Crawford G, Collins FS, Chen YD, Caulfield MJ, Campbell H, Burt NP, Bonnycastle LL, Boomsma DI, Boekholdt SM, Bergman RN, Barroso I, Bandinelli S, Ballantyne CM, Assimes TL, Quertermous T, Altshuler D, Seielstad M, Wong TY, Tai ES, Feranil AB, Kuzawa CW, Adair LS, Taylor HA, Jr., Borecki IB, Gabriel SB, Wilson JG, Holm H, Thorsteinsdottir U, Gudnason V, Krauss RM, Mohlke KL, Ordovas JM, Munroe PB, Kooner JS, Tall AR, Hegele RA, Kastelein JJ, Schadt EE, Rotter JI, Boerwinkle E, Strachan DP, Mooser V, Stefansson K, Reilly MP, Samani NJ, Schunkert H, Cupples LA, Sandhu MS, Ridker PM, Rader DJ, van Duijn CM, Peltonen L, Abecasis GR, Boehnke M, Kathiresan S

Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, Mich., USA.

Nature 2010;466:707–13.

Background: Circulating concentrations of total cholesterol, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol and triglycerides are among the most important risk factors for coronary artery disease (CAD) and are targets for drug therapy.

Methods: More than 100,000 individuals of European ancestry had their genomes screened for common variants associated with plasma lipid concentrations.

Results: 95 genetic loci were significantly associated with lipid traits ($p < 5 \times 10^{-8}$), with 59 showing genome-wide significant association for the first time. The new loci include those near to known lipid regulators (e.g. CYP7A1, NPC1L1 and SCARB1) as well as many loci not previously implicated in lipoprotein metabolism. The 95 loci also contribute to extreme lipid phenotypes and influenced lipid traits in three non-European populations (East Asians, South Asians and African-Americans). Several novel plasma lipid loci were also associated with CAD. Finally, three of the novel genes (GALNT2, PPP1R3B and TTC39B) were experimentally validated in mouse models.

Conclusions: These findings provide the foundation for a broader biological understanding of lipoprotein metabolism and the identification of new therapeutic targets for the prevention of CAD.

The impressive results of this greatly expanded GWAS explained 10–12% of the total variation in lipid levels in adults, corresponding to around 25–30% of the genetic variance of each lipid marker; many of these loci overlapped with 18 genes already implicated in mendelian lipid disorders. But the authors didn't stop there. They also achieved their other stated aims by showing that: (i) these findings in populations of white European descent are relevant to non-Europeans relevance; (ii) these markers of normal variation in lipid levels are relevant to risks for both extreme lipid phenotypes and coronary artery disease, and (iii) the 'hypothesis-free' genome-wide approach identifies new areas of biology that can be confirmed in experimental models.

In summary, their results revealed a common genetic architecture of lipid traits across European, South Asian, East Asian and African-American populations. Secondly, they showed that additive effects across multiple common variants contribute to the risk of having an extreme phenotype, and so are very relevant to understanding disease etiology in our patients. Thirdly, mouse gene overexpression and knockdown models confirmed the importance of novel lipid genes GALNT2, PPP1R3B, and TTC39B. These represent new pathways for potential modulation of lipid levels in humans.

Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index

Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, Allen HL, Lindgren CM, Luan J, Magi R, Randall JC, Vedantam S, Winkler TW, Qi L, Workalemahu T, Heid IM, Steinthorsdottir V, Stringham HM, Weedon MN, Wheeler E, Wood AR, Ferreira T, Weyant RJ, Segre AV, Estrada K, Liang L, Nemesh J, Park JH, Gustafsson S, Kilpelainen TO, Yang J, Bouatia-Naji N, Esko T, Feitosa MF, Kutalik Z, Mangino M, Raychaudhuri S, Scherag A, Smith AV, Welch R, Zhao JH, Aben KK, Absher DM, Amin N, Dixon AL, Fisher E, Glazer NL, Goddard ME, Heard-Costa NL, Hoessler V, Hottenga JJ, Johansson A, Johnson T, Ketkar S, Lamina C, Li S, Moffatt MF, Myers RH, Narisu N, Perry JR, Peters MJ, Preuss M, Ripatti S, Rivadeneira F, Sandholt C, Scott LJ, Timpson NJ, Tyrer JP, van Wingerden S, Watanabe RM, White CC, Wiklund F, Barlassina C, Chasman DI, Cooper MN, Jansson JO, Lawrence RW, Pellicka N, Prokopenko I, Shi J, Thiering E, Alavere H, Alibrandi MT, Almgren P, Arnold AM, Aspelund T, Atwood LD, Balkau B, Balmforth AJ, Bennett AJ, Ben-Shlomo Y, Bergman RN, Bergmann S, Biebermann H, Blakemore AI, Boes T, Bonnycastle LL, Bornstein SR, Brown MJ, Buchanan TA, Busonero F, Campbell H, Cappuccio FP, Cavalcanti-Proenca C, Chen YD, Chen CM, Chines PS, Clarke R, Coin L, Connell J, Day IN, Heijer M, Duan J, Ebrahim S, Elliott P, Elosua R, Eiriksdottir G, Erdos MR, Eriksson JG, Facheris MF, Felix SB, Fischer-Posovszky P, Folsom AR, Friedrich N, Freimer NB, Fu M, Gaget S, Gejman PV, Geus EJ, Gieger C, Gjesing AP, Goel A, Goyette P, Grallert H, Grassler J, Greenawald DM, Groves CJ, Gudnason V, Guiducci C, Hartikainen AL, Hassanali N, Hall AS, Havulinna AS, Hayward C, Heath AC, Hengstenberg C, Hicks AA, Hinney A, Hofman A, Homuth G, Hui J, Igl W, Iribarren C, Isomaa B, Jacobs KB, Jarick I, Jewell E, John U, Jorgensen T, Jousilahti P, Jula A, Kaakinen M, Kajantie E, Kaplan LM, Kathiresan S, Kettunen J, Kinnunen L, Knowles JW, Kolcic I, Konig IR, Koskinen S, Kovacs P, Kuusisto J, Kraft P, Kvaloy K, Laitinen J, Lantieri O, Lanzani C, Launer LJ, Lecoeur C, Lehtimäki T, Lettre G, Liu J, Lokki ML, Lorentzon M, Luben RN, Ludwig B, Manunta P, Marek D, Marre M, Martin NG, McArdle WL, McCarthy A, McKnight B, Meitinger T, Melander O, Meyre D, Midtthjell K, Montgomery GW, Morken MA, Morris AP, Mulic R, Ngwa JS, Nelis M, Neville MJ, Nyholt DR, O'Donnell CJ, O'Rahilly S, Ong KK, Oostra B, Pare G, Parker AN, Perola M, Pichler I, Pietiläinen KH, Platou CG, Polasek O, Pouta A, Rafelt S, Raitakari O, Rayner NW, Ridderstrale M, Rief W, Ruokonen A, Robertson NR, Rzehak P, Salomaa V, Sanders AR, Sandhu MS, Sanna S, Saramies J, Savolainen MJ, Scherag S, Schipf S, Schreiber S, Schunkert H, Silander K, Sinisalo J, Siscovick DS, Smit JH, Soranzo N, Sovio U, Stephens J, Surakka I, Swift AJ, Tammesoo ML, Tardif JC, Teder-Laving M, Teslovich TM, Thompson JR, Thomson B, Tonjes A, Tuomi T, van Meurs JB, van Ommen GJ, Vatin V, Viikari J, Visvikis-Siest S, Vitart V, Vogel CI, Voight BF, Waite LL, Wallaschofski H, Walters GB, Widen E, Wiegand S, Wild SH, Willemssen G, Witte DR, Witteman JC, Xu J, Zhang Q, Zgaga L, Ziegler A, Zitting P, Beilby JP, Farooqi IS, Hebebrand J, Huikuri HV, James AL, Kahonen M, Levinson DF, Macciardi F, Nieminen MS, Ohlsson C, Palmer LJ, Ridker PM, Stumvoll M, Beckmann JS, Boeing H, Boerwinkle E, Boomsma DI, Caulfield MJ, Chanock SJ, Collins FS, Cupples LA, Smith GD, Erdmann J, Froguel P, Gronberg H, Gyllenstein U, Hall P, Hansen T, Harris TB, Hattersley AT, Hayes RB, Heinrich J, Hu FB, Hveem K, Illig T, Jarvelin MR, Kaprio J, Karpe F, Khaw KT, Kiemeny LA, Krude H, Laakso M, Lawlor DA, Metspalu A, Munroe PB, Ouwehand WH, Pedersen O, Penninx BW, Peters A, Pramstaller PP, Quertermous T, Reinehr T, Rissanen A, Rudan I, Samani NJ, Schwarz PE, Shuldiner AR, Spector TD, Tuomilehto J, Uda M, Uitterlinden A, Valle TT, Wabitsch M, Waeber G, Wareham NJ, Watkins H, Wilson JF, Wright AF, Zillikens MC, Chatterjee N, McCarroll SA, Purcell S, Schadt EE, Visscher PM, Assimes TL, Borecki IB, Deloukas P, Fox CS, Groop LC, Haritunians T, Hunter DJ, Kaplan RC, Mohlke KL, O'Connell JR, Peltonen L, Schlessinger D, Strachan DP, van Duijn CM, Wichmann HE, Frayling TM, Thorsteinsdottir U, Abecasis GR, Barroso I, Boehnke M, Stefansson K, North KE, McCarthy ML, Hirschhorn JN, Ingelsson E, Loos RJ

Metabolism Initiative and Program in Medical and Population Genetics, Broad Institute, Cambridge, Mass., USA.
Nat Genet 2010;42:937–48.

Background: Obesity is highly prevalent globally and is highly heritable. However, its underlying genetic factors remain largely elusive.

Methods: To identify genetic loci for obesity susceptibility, associations were examined between body mass index and ~2.8 million SNPs in up to 123,865 adult individuals. 42 SNPs were subsequently selected for follow-up in up to 125,931 additional individuals.

Results: The 14 already known obesity susceptibility loci were confirmed. 18 new loci were associated with body mass index ($p < 5 \times 10^{-8}$), one of which includes a copy number variant near GPRC5B. Some loci (at MC4R, POMC, SH2B1 and BDNF) localize near to key hypothalamic regulators of energy balance. Another new locus was near GIPR, an incretin receptor.

Conclusions: These findings of many newly associated loci provide new insights into the regulation of human body weight.

It seems only a short while ago since Frayling et al. described variants in FTO as the first genetic locus for common obesity [3]. Since then there have been two intermediate ‘waves’ of GWAS, each with increasing sample sizes. The current study was envisaged as the ‘mega-GWAS’ for BMI and obesity, by including as much as possible of the available GWAS data from around world. While the current ‘haul’ of 32 obesity loci is certainly impressive, they only explain 1.45% of the interindividual variation in adult BMI. Further expansion of sample sizes, fine mapping of existing loci to identify the causal variants (both underway using Cardio-Metabo BeadChip genotyping), and identification of less common variants by whole-genome or whole-exome sequencing are among the current options.

While the findings of previous GWAS had suggested an almost invariable brain-centered regulation of BMI, a notable finding of the current study was the locus GIPR, which encodes a receptor for the incretin hormone ‘gastric inhibitory peptide’. Variants at MC4R, BDNF and SH2B1 overlap with monogenic obesity loci. However, there was also surprisingly little overlap with previous GWAS findings for early-onset or adult morbid obesity (e.g. NPC1, MAF, PTER and TNKS-MSRA) which showed limited or no evidence of association in this huge study. Possibly this might provide a clue that some variants are active only in some populations or only in certain settings, however it will be challenging to demonstrate these modifications robustly and without trawling up numerous false positive interactions.

New ‘expanded-GWAS’ associations: waist-hip ratio

Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution

Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, Thorleifsson G, Zillikens MC, Speliotes EK, Magi R, Workalemahu T, White CC, Bouatia-Naji N, Harris TB, Berndt SI, Ingelsson E, Willer CJ, Weedon MN, Luan J, Vedantam S, Esko T, Kilpelainen TO, Kutalik Z, Li S, Monda KL, Dixon AL, Holmes CC, Kaplan LM, Liang L, Min JL, Moffatt MF, Molony C, Nicholson G, Schadt EE, Zondervan KT, Feitosa MF, Ferreira T, Allen HL, Weyant RJ, Wheeler E, Wood AR, Estrada K, Goddard ME, Lettre G, Mangino M, Nyholt DR, Purcell S, Smith AV, Visscher PM, Yang J, McCarroll SA, Nemesh J, Voight BF, Absher D, Amin N, Aspelund T, Coin L, Glazer NL, Hayward C, Heard-Costa NL, Hottenga JJ, Johansson A, Johnson T, Kaakinen M, Kapur K, Ketkar S, Knowles JW, Kraft P, Kraja AT, Lamina C, Leitzmann MF, McKnight B, Morris AP, Ong KK, Perry JR, Peters MJ, Polasek O, Prokopenko I, Rayner NW, Ripatti S, Rivadeneira F, Robertson NR, Sanna S, Sovio U, Surakka I, Teumer A, van Wingerden S, Vitart V, Zhao JH, Cavalcanti-Proenca C, Chines PS, Fisher E, Kulzer JR, Lecoecur C, Narisu N, Sandholt C, Scott LJ, Silander K, Stark K, Tammesoo ML, Teslovich TM, Timpson NJ, Watanabe RM, Welch R, Chasman DI, Cooper MN, Jansson JO, Kettunen J, Lawrence RW, Pellikka N, Perola M, Vandenput L, Alavere H, Almgren P, Atwood LD, Bennett AJ, Biffar R, Bonnycastle LL, Bornstein SR, Buchanan TA, Campbell H, Day IN, Dei M, Dorr M, Elliott P, Erdos MR, Eriksson JG, Freimer NB, Fu M, Gaget S, Geus EJ, Gjesing AP, Grallert H, Grassler J, Groves CJ, Guiducci C, Hartikainen AL, Hassanali N, Havulinna AS, Herzig KH, Hicks AA, Hui J, Igl W, Jousilahti P, Julia A, Kajantie E, Kinnunen L, Kolcic I, Koskinen S, Kovacs P, Kroemer HK, Krzeli V, Kuusisto J, Kvaloy K, Laitinen J, Lantieri O, Lathrop GM, Lokki ML, Luben RN, Ludwig B, McArdle WL, McCarthy A, Morken MA, Nelis M, Neville MJ, Pare G, Parker AN, Peden JF, Pichler I, Pietilainen KH, Platou CG, Pouta A, Ridderstrale M, Samani NJ, Saramies J, Sinisalo J, Smit JH, Strawbridge RJ, Stringham HM, Swift AJ, Teder-Laving M, Thomson B, Usala G, van Meurs JB, van Ommen GJ, Vatin V, Volpato CB, Wallaschowski H, Walters GB, Widen E, Wild SH, Willemsen G, Witte DR, Zgaga L, Zitting P, Beilby JP, James AL, Kahonen M, Lehtimäki T, Nieminen MS, Ohlsson C, Palmer LJ, Raitakari O, Ridker PM, Stumvoll M, Tonjes A, Viikari J, Balkau B, Ben-Shlomo Y, Bergman RN, Boeing H, Smith GD, Ebrahim S, Froguel P, Hansen T, Hengstenberg C, Hveem K, Isomaa B, Jorgensen T, Karpe F, Khaw KT, Laakso M, Lawlor DA, Marre M, Meitinger T, Metspalu A, Midtjell K, Pedersen O, Salomaa V, Schwarz PE, Tuomi T, Tuomilehto J, Valle TT, Wareham NJ, Arnold AM, Beckmann JS, Bergmann S, Boerwinkle E, Boomsma DI, Caulfield MJ, Collins FS, Eirisdottir G, Gudnason V, Gyllenstein U, Hamsten A, Hattersley AT, Hofman A, Hu FB, Illig T, Iribarren C, Jarvelin MR, Kao WH, Kaprio J, Launer LJ, Munroe PB, Oostra B, Penninx BW, Pramstaller PP, Psaty BM, Quertermous T, Rissanen A, Rudan I, Shuldiner AR, Soranzo N, Spector TD, Syvanen AC, Uda M, Uitterlinden A, Volzke H, Vollenweider P, Wilson JF, Witteman JC, Wright AF, Abecasis GR, Boehnke M, Borecki IB, Deloukas P, Frayling TM, Groop LC, Haritunians T, Hunter DJ, Kaplan RC, North KE, O’Connell JR, Peltonen L, Schlessinger D, Strachan DP, Hirschhorn JN, Assimes TL, Wichmann HE, Thorsteinsdottir U, van Duijn CM, Stefansson K, Cupples LA, Loos RJ, Barroso I, McCarthy MI, Fox CS, Mohlke KL, Lindgren CM

Background: Waist-hip ratio (WHR) is used to estimate body fat distribution and can predict metabolic disease risks independent of overall adiposity. WHR is heritable, but few genetic variants for this trait have been identified.

Methods: A meta-analysis of 32 genome-wide association studies for WHR adjusted for body mass index (comprising up to 77,167 participants) was performed. Subsequently, 16 loci were selected for follow-up in a further 29 studies (comprising up to 113,636 subjects).

Results: In addition to the known signal at LYPLAL1, 13 new loci were identified in or near RSPO3, VEGFA, TBX15-WARS2, NFE2L3, GRB14, DNM3-PIGC, ITPR2-SSPN, LY86, HOXC13, ADAMTS9, ZNRF3-KREMEN1, NISCH-STAB1 and CPEB4 (all $p < 5 \times 10^{-8}$). Seven of these loci exhibited marked sexual dimorphism, all with a stronger effect on WHR in women than men (p for sex difference = 1.9×10^3 to $p = 1.2 \times 10^{13}$).

Conclusions: These findings identify multiple loci that influence body fat distribution independent of overall adiposity and reveal strong gene-by-sex interactions.

Waist-hip ratio (WHR) and BMI sound like complementary traits. Indeed previous studies had shown some overlap between BMI and waist circumference loci. However, these results from the GIANT international consortium were deliberately distinct from the BMI results in the accompanying paper by Speliotes et al. because they analyzed 'WHR adjusted for BMI'. The meaning of this phenotype is not obviously clear, but in effect this analysis actually identified variants associated with lower hip circumference (rather than larger waist circumference).

It is therefore particularly notable that the loci identified here also showed significant associations with various metabolic syndrome traits, including lipid profiles, fasting insulin and glucose levels, glucose tolerance and T2DM. Indeed, these variants had greater relevance to metabolic disease risk than the BMI loci, which highlights the importance of body shape rather than overall body size.

Furthermore, several of these were found to show different patterns of expression between gluteal and abdominal fat tissue. This supports the hypothesis that fat depot-specific differences in gene expression patterns explain the association with WHR and are an important determinant of metabolic disease risk. As the authors conclude, efforts to tackle obesity through medication or lifestyle to change overall energy balance have been extremely challenging to implement. The manipulation of processes associated with beneficial patterns of fat distribution offers an alternative perspective for future drug discovery.

New 'expanded-GWAS' associations: T2DM

Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis

Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, Zeggini E, Huth C, Aulchenko YS, Thorleifsson G, McCulloch LJ, Ferreira T, Grallert H, Amin N, Wu G, Willer CJ, Raychaudhuri S, McCarroll SA, Langenberg C, Hofmann OM, Dupuis J, Qi L, Segre AV, van Hoek M, Navarro P, Ardlie K, Balkau B, Benediktsson R, Bennett AJ, Blagieva R, Boerwinkle E, Bonnycastle LL, Bengtsson Bostrom K, Bravenboer B, Bumpstead S, Burt NP, Charpentier G, Chines PS, Cornelis M, Couper DJ, Crawford G, Doney AS, Elliott KS, Elliott AL, Erdos MR, Fox CS, Franklin CS, Ganser M, Gieger C, Grarup N, Green T, Griffin S, Groves CJ, Guiducci C, Hadjadj S, Hassanali N, Herder C, Isomaa B, Jackson AU, Johnson PR, Jorgensen T, Kao WH, Klopp N, Kong A, Kraft P, Kuusisto J, Lauritzen T, Li M, Lieveer A, Lindgren CM, Lyssenko V, Marre M, Meitinger T, Midhjel K, Morken MA, Narisu N, Nilsson P, Owen KR, Payne F, Perry JR, Petersen AK, Platou C, Proenca C, Prokopenko I, Rathmann W, Rayner NW, Robertson NR, Rocheleau G, Roden M, Sampson MJ, Saxena R, Shields BM, Shrader P, Sigurdsson G, Sparso T, Strassburger K, Stringham HM, Sun Q, Swift AJ, Thorand B, Tichet J, Tuomi T, van Dam RM, van Haften TW, van Herpt T, van Vliet-Ostaptchouk JV, Walters GB, Weedon MN, Wijmenga C, Witteman J, Bergman RN, Cauchi S, Collins FS, Gloyn AL, Gyllenstein U, Hansen T, Hide WA, Hitman GA, Hofman A, Hunter DJ, Hveem K, Laakso M, Mohlke KL, Morris AD, Palmer CN,

Pramstaller PP, Rudan I, Sijbrands E, Stein LD, Tuomilehto J, Uitterlinden A, Walker M, Wareham NJ, Watanabe RM, Abecasis GR, Boehm BO, Campbell H, Daly MJ, Hattersley AT, Hu FB, Meigs JB, Pankow JS, Pedersen O, Wichmann HE, Barroso I, Florez JC, Frayling TM, Groop L, Sladek R, Thorsteinsdottir U, Wilson JF, Illig T, Froguel P, van Duijn CM, Stefansson K, Altshuler D, Boehnke M, McCarthy MI
Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Mass., USA.
Nat Genet 2010;42:579–89.

Background: Type 2 diabetes (T2D) is a heritable and common disease associated with defects in both β -cell function and insulin action.

Methods: Genome-wide association data from 8,130 individuals were combined and compared to 38,987 controls of European descent. Follow-up of previously unidentified meta-analysis signals was performed in a further 34,412 cases and 59,925 controls.

Results: This analysis identified 12 new T2D association loci each with a combined $p < 5 \times 10^{-8}$. These new loci include a second independent signal at KCNQ1; the first report, to our knowledge, of an X-chromosome GWAS association (near DUSP9), and a further example of overlap between loci implicated in both monogenic and multifactorial forms of diabetes (at HNF1A).

Conclusions: The identified loci affect both β -cell function and insulin action. Overall, T2D loci show evidence of enrichment for genes involved in cell cycle regulation. A high proportion of T2D susceptibility loci harbor independent association signals influencing apparently unrelated traits.

This study is by far the largest genetic study of T2DM to date. By contrast, the previous multicenter GWAS meta-analysis in 2008 comprised 10,128 individuals (including 4,549 cases) [4]. The results of the current study add a further 12 new ‘hits’ to the previous list of ~26 genome-wide significant T2DM susceptibility loci. The current extended list of 38 T2DM susceptibility loci contains some interesting insights.

Five common susceptibility loci overlap with genes where mutations cause rare monogenic forms of diabetes, including MODY, indicating the mutual benefits of studying rare and common disease. Risk alleles at 10 loci were associated with reduced β -cell function, while three were associated with reduced insulin sensitivity (PPARG, IRS1 and KLF14). 13 loci were ‘colocalized’ in the same genomic region as (but not directly overlapping with) loci for other diseases, suggesting that variants in different regulatory domains of the same region might have tissue-specific and disease-specific effects via the same genes and biological pathways. Hence, systemic treatment of one condition could affect many others.

In general, risk estimates for T2DM loci were stronger in non-obese cases, suggesting that these have a more ‘genetic’ disease – potentially their optimal treatment or prevention strategies may differ from obese cases.

Together, these discovery efforts so far explain ~10% of the familial clustering of T2DM risk.

New ‘expanded-GWAS’ associations: age at menarche

Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies

Elks CE, Perry JR, Sulem P, Chasman DI, Franceschini N, He C, Lunetta KL, Visser JA, Byrne EM, Cousminer DL, Gudbjartsson DF, Esko T, Feenstra B, Hottenga JJ, Koller DL, Kutalik Z, Lin P, Mangino M, Marongiu M, McArdle PF, Smith AV, Stolk L, van Wingerden SH, Zhao JH, Albrecht E, Corre T, Ingelsson E, Hayward C, Magnusson PK, Smith EN, Ulivi S, Warrington NM, Zgaga L, Alavere H, Amin N, Aspelund T, Bandinelli S, Barroso I, Berenson GS, Bergmann S, Blackburn H, Boerwinkle E, Buring JE, Busonero F, Campbell H, Chanock SJ, Chen W, Cornelis MC, Couper D, Coviello AD, d’Adamo P, de Faire U, de Geus EJ, Deloukas P, Doring A, Smith GD, Easton DF, Eiriksdottir G, Emilsson V, Eriksson J, Ferrucci L, Folsom AR, Foroud T, Garcia M, Gasparini P, Geller F, Gieger C, Gudnason V, Hall P, Hankinson SE, Ferrelli L, Heath AC, Hernandez DG, Hofman A, Hu FB, Illig T, Jarvelin MR, Johnson AD, Karasik D, Khaw KT, Kiel DP, Kilpelainen TO, Kolcic I, Kraft P, Launer LJ, Laven JS, Li S, Liu J, Levy D, Martin NG, McArdle WL, Melbye M, Mooser V, Murray JC, Murray SS, Nalls MA, Navarro P, Nelis M, Ness AR, Northstone K, Oostra BA, Peacock M, Palmer LJ, Palotie A, Pare G, Parker AN, Pedersen NL, Peltonen L, Pennell CE, Pharoah P, Polasek O,

Plump AS, Pouta A, Porcu E, Rafnar T, Rice JP, Ring SM, Rivadeneira F, Rudan I, Sala C, Salomaa V, Sanna S, Schlessinger D, Schork NJ, Scuteri A, Segre AV, Shuldiner AR, Soranzo N, Sovio U, Srinivasan SR, Strachan DP, Tammesoo ML, Tikkanen E, Toniolo D, Tsui K, Tryggvadottir L, Tyrer J, Uda M, van Dam RM, van Meurs JB, Vollenweider P, Waeber G, Wareham NJ, Waterworth DM, Weedon MN, Wichmann HE, Willemssen G, Wilson JF, Wright AF, Young L, Zhai G, Zhuang W, Bierut LJ, Boomsma DI, Boyd HA, Crisponi L, Demerath EW, van Duijn CM, Econs MJ, Harris TB, Hunter DJ, Loos RJ, Metspalu A, Montgomery GW, Ridker PM, Spector TD, Streeten EA, Stefansson K, Thorsteinsdottir U, Uitterlinden AG, Widen E, Murabito JM, Ong KK, Murray A Medical Research Council (MRC) Epidemiology Unit, Institute of Metabolic Science, Cambridge, UK. *Nat Genet* 2010;42:1077–85.

Background: Age at menarche shows wide variability between girls and is highly heritable trait associated with risks of obesity and many associated disease.

Methods: A meta-analysis of 32 genome-wide association studies in 87,802 women of European descent was performed. Replication of findings was performed in up to 14,731 additional women.

Results: In addition to the already known loci at LIN28B and 9q31.2, 30 new menarche loci were identified (all $p < 5 \times 10^{-8}$). There was also suggestive evidence for a further 10 loci. The new loci included four previously associated with body mass index (in or near FTO, SEC16B, TRA2B and TMEM18), three loci in or near other genes implicated in energy homeostasis (BSX, CRTC1 and MCHR2) and three loci in or near genes implicated in hormonal regulation (INHBA, PCSK2 and RXRG). Ingenuity and gene-set enrichment pathway analyses identified coenzyme A and fatty acid biosynthesis as biological processes related to menarche timing.

Conclusions: These findings greatly extend understanding of the genetic architecture of age at menarche and identify a common genetic basis for body weight regulation and the timing of puberty in girls.

The authors of this paper had previously separately reported in 2009 four GWAS in 55,000 women that identified two loci for age at menarche, at LIN28B and at 9q31.2. One year later, this study reports a remarkable 30 further menarche loci by combining those four original samples plus including new GWAS data on a further 32,000 women. Collaboration is clearly a fruitful strategy in discovery science.

In addition to the loci highlighted as overlapping with those for BMI and in genes implicated in energy regulation, some hormone-related candidate genes were identified. These include INHBA, which encodes the protein subunit Inhibin β A, which contributes to both the dimers Inhibin A and Activin A. RXRG, the retinoic X receptor γ , forms dimers with retinoic acid, thyroid hormone and vitamin D. PCSK2 encodes a proprotein convertase that cleaves proinsulin, POMC, and also pro-LHRH.

The original findings of LIN28B as a locus for menarche timing and adult height helped to open up new areas of biological understanding of the role of microRNA regulation linking puberty timing to glucose metabolism [5] and to risks of colon and ovarian cancers. Hopefully, these new findings will also stimulate a better understanding of how nutritional status and energy balance signal to the reproductive axis.

New GWAS associations: PCOS

Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3

Chen ZJ, Zhao H, He L, Shi Y, Qin Y, Shi Y, Li Z, You L, Zhao J, Liu J, Liang X, Zhao X, Zhao J, Sun Y, Zhang B, Jiang H, Zhao D, Bian Y, Gao X, Geng L, Li Y, Zhu D, Sun X, Xu JE, Hao C, Ren CE, Zhang Y, Chen S, Zhang W, Yang A, Yan J, Li Y, Ma J, Zhao Y

Center for Reproductive Medicine, Shandong Provincial Hospital, Shandong University, Jinan, China.

Nat Genet 2011;43 55–9.

Background: Polycystic ovary syndrome (PCOS) is a common and heritable metabolic disorder in women.

Methods: To identify causative genes for PCOS, a genome-wide association study (GWAS) of PCOS was performed in Han Chinese. The initial discovery set included 744 PCOS cases and 895 controls; subse-

quent replications involved two independent cohorts comprising 2,840 PCOS cases and 5,012 controls from northern Han Chinese, and 498 cases and 780 controls from southern and central Han Chinese.

Results: Strong evidence of associations was identified between PCOS and three loci: 2p16.3 (rs13405728), odds ratio (OR) 0.71; 2p21 (rs13429458, OR 0.67), and 9q33.3 (rs2479106, OR 1.34).

Conclusions: These findings provide new insight into the pathogenesis of PCOS. Follow-up studies in these regions are recommended to identify the underlying causal genes.

It is somewhat surprising that this is the first reported GWAS for PCOS, given that it is a common condition that displays familial clustering. Possibly other studies have been performed and their underpowered results have yet to be confirmed. The PCOS cases in this study came from reproductive and gynecology clinics throughout China. It is notable that mean BMI in these cases, 23–24 kg/m², was higher than in controls, 21–23 kg/m², but is far lower than is typically seen in other settings, and this may have enhanced the power of the study to identify specific genetic factors for PCOS susceptibility.

Three loci were identified. The first locus at 2p16.3 was rich in potential candidate genes, including GTF2A1L (shows germ-cell specific expression), LHCCR (encodes a G-protein-coupled receptor for LH and HCG), and the FSH receptor gene FSHR. The second locus at 2p.21 appeared to be centered on THADA (a gene initially identified as disrupted in thyroid adenomas). The third locus at 9q33.3 is within the gene DENND1A which encodes a domain that binds to ERAP1 (endoplasmic reticulum aminopeptidase 1), whose levels are reportedly elevated in PCOS women.

New GWAS associations: hypospadias

Common variants in DGKK are strongly associated with risk of hypospadias

Van der Zanden LF, van Rooij IA, Feitz WF, Knight J, Donders AR, Renkema KY, Bongers EM, Vermeulen SH, Kiemeny LA, Veltman JA, Arias-Vasquez A, Zhang X, Markljung E, Qiao L, Baskin LS, Nordenskjold A, Roeleveld N, Franke B, Knoers NV

Department of Epidemiology, Biostatistics and HTA, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.

Nat Genet 2011;43:48–50.

Background: Hypospadias is a common congenital malformation of the external genitalia in boys.

Methods: A genome-wide association study was performed using pooled DNA from 436 individual cases with hypospadias and 494 controls of European descent. Subsequently, the highest ranked SNPs were selected for individual genotyping in the discovery sample, and also in an additional Dutch sample of 133 cases and their parents, and a Swedish sample of 266 cases and 402 controls.

Results: Individual genotyping of two SNPs (rs1934179 and rs7063116) in DGKK, encoding diacylglycerol kinase κ , produced genome-wide significant evidence for association with hypospadias in the discovery sample (allele-specific odds ratio (OR) = 2.5, and OR = 2.3, respectively) and in the Dutch (OR = 3.9, and OR = 3.8,) and Swedish (OR = 2.5, and OR = 2.2) replication samples. Expression studies showed expression of DGKK in preputial tissue of cases and controls, and these levels were lower in carriers of the risk allele of rs1934179.

Conclusions: These studies identify DGKK as a major risk gene for hypospadias.

Hypotheses about the genetic etiology of hypospadias have focused on disruption of hormonal pathways, however previous associations have not been replicated. This study demonstrates the power of the genomic approach to uncover new biological understanding. GWAS is an expensive technology and typically one chip, costing hundreds of US dollars each, is run in each individual. This study used an efficient, though slightly less precise, approach of using just four chips (the assays were run in duplicate) to compare pooled DNA samples of all cases versus all controls. The specific alleles that showed the largest differences in frequency between groups were then genotyped in each individual.

Variants in DGKK were very strongly related to hypospadias, by association, parental allele transmission and expression studies. DGKK encodes an enzyme that modulates the balance between diacylg-

lycerol and phosphatidic acid, two important signaling lipids. Remarkably, the single variant rs1934179 was calculated to explain 31–32% of the risk of hypospadias in the Dutch and Swedish populations studied. It is interesting to note that previous studies have reported familial clustering only of anterior and middle, but not posterior forms of hypospadias, and indeed the genetic associations were stronger in the former groups.

New GWAS associations: obesity in Pima Indians

Evaluation of A2BP1 as an obesity gene

Ma L, Hanson RL, Taurig MT, Muller YL, Kaur BP, Perez JM, Meyre D, Fu M, Korner A, Franks PW, Kiess W, Kobes S, Knowler WC, Kovacs P, Froguel P, Shuldiner AR, Bogardus C, Baier LJ
Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Disease, National Institutes of Health, Phoenix, Ariz., USA.
Diabetes 2010;59:2837–45.

Background: Genome-wide association study (GWAS) findings in populations of European descent do not contribute to the dramatic susceptibility to obesity of Pima Indians.

Methods: GWAS was performed in Pima Indians (n = 413). A2BP1 was genetically and functionally analyzed to assess its potential role in human obesity. Variants spanning A2BP1 were genotyped in a population-based sample of 3,234 full-heritage Pima Indians, 2,843 of whom were not part of the initial GWAS. Published GWAS data across A2BP1 were additionally analyzed in French adults (n = 1,426) and children case-control subjects (n = 1,392). Selected variants were genotyped in two additional samples of Caucasians (Amish, n = 1,149, and German children case-control subjects, n = 998) and one additional Native American (n = 2,531) sample. Small interfering RNA was used to knockdown A2bp1 expression in mouse embryonic hypothalamus cells.

Results: The GWAS identified variation in the ataxin-2 binding protein 1 gene (A2BP1) to be associated with percent body fat. No single variant in A2BP1 was reproducibly associated with obesity across the different populations. However, different variants within intron 1 of A2BP1 were associated with BMI in full-heritage Pima Indians (rs10500331, $p = 1.9 \times 10^{-7}$) and obesity in French Caucasian adult (rs4786847, $p = 1.9 \times 10^{-10}$) and children (rs8054147, $p = 9.2 \times 10^{-6}$) case-control subjects. Reduction of A2bp1 in mouse embryonic hypothalamus cells decreased expression of Atxn2, Insr, and Mc4r.

Conclusions: These findings suggest that variation in A2BP1 influences obesity susceptibility in both Pima Indians and Europeans. Functional studies suggest that A2BP1 could influence adiposity via the hypothalamic MC4R pathway.

In addition to the GWAS for BMI in populations of mainly European descent, similar studies in diverse ethnic groups could identify novel obesity-susceptibility genes. A relatively small GWAS was performed in 413 Pima Indians, full members of this famous obesity- and T2DM-prone heritage. While their top hit did not quite reach genome-wide significance levels, it localized a gene previously implicated in the neurodegenerative disorder spinocerebellar ataxia type 2 and with severe obesity and hyperphagia in humans and mice. Their experimental studies also showed that a 75% reduction in A2bp1 expression in an embryonic hypothalamic cell line led to a 53, 60, and 75% decrease in gene expression levels of Atxn2, Insr, and Mc4r, respectively, but did not alter levels of Lepr or Npy1r.

Pertinently, highly significant associations were also seen with risk of severe obesity in French adult and French children studies. However, the associations were seen with different A2BP1 variants across the different studies. The lack of reproducibility with any one specific variant may be due to the existence of an ungenotyped causal variant that has differing linkage disequilibrium patterns between ethnic groups. Alternatively, multiple rare causal variants might be present that differ between and even within a single ethnic group. While exploring associations in different ethnic groups has many benefits in terms of investigating the wider application of initial findings and for fine mapping studies to identify causal genes and causal variants, the strong possibility of different causal variants

within populations or even within pedigrees represents a major challenge for the robust assimilation of the evidence.

New paradigms: HapMap3

Integrating common and rare genetic variation in diverse human populations

Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, Peltonen L, Dermitzakis E, Bonnen PE, Altshuler DM, Gibbs RA, de Bakker PI, Deloukas P, Gabriel SB, Gwilliam R, Hunt S, Inouye M, Jia X, Palotie A, Parkin M, Whittaker P, Yu F, Chang K, Hawes A, Lewis LR, Ren Y, Wheeler D, Gibbs RA, Muzny DM, Barnes C, Darvishi K, Hurler M, Korn JM, Kristiansson K, Lee C, McCarroll SA, Nemesh J, Dermitzakis E, Keinan A, Montgomery SB, Pollack S, Price AL, Soranzo N, Bonnen PE, Gibbs RA, Gonzaga-Jauregui C, Keinan A, Price AL, Yu F, Anttila V, Brodeur W, Daly MJ, Leslie S, McVean G, Moutsianas L, Nguyen H, Schaffner SF, Zhang Q, Ghorri MJ, McGinnis R, McLaren W, Pollack S, Price AL, Schaffner SF, Takeuchi F, Grossman SR, Shlyakhter I, Hostetter EB, Sabeti PC, Adebamowo CA, Foster MW, Gordon DR, Licinio J, Manca MC, Marshall PA, Matsuda I, Ngare D, Wang VO, Reddy D, Rotimi CN, Royal CD, Sharp RR, Zeng C, Brooks LD, McEwen JE

Broad Institute, 7 Cambridge Center, Cambridge, Mass., USA.

Nature 2010;467:52–8.

Background: Despite great progress in identifying many genetic variants associated with human disease, most inherited risk remains unexplained. A more complete understanding requires genome-wide studies that fully examine less common alleles in populations with a wide range of ancestry.

Methods: To inform the design and interpretation of such studies, 1.6 million common single nucleotide polymorphisms (SNPs) were genotyped in 1,184 reference individuals from 11 global populations. In addition, ten 100-kilobase regions were sequenced in 692 of these individuals.

Results: This integrated dataset of common and rare alleles, called ‘HapMap 3’, includes both SNPs and copy number polymorphisms (CNPs). Analyses were used to characterize population-specific differences among low-frequency variants, to measure the improvement in imputation accuracy afforded by the larger reference panel, especially in imputing SNPs with a minor allele frequency of $< \text{or } \approx 5\%$, and to demonstrate the feasibility of imputing newly discovered CNPs and SNPs.

Conclusions: This expanded public resource of genome variants in global populations will allow future studies to more deeply examine genomic variation and its role in human disease, and serves as a step towards a high-resolution map of the landscape of human genetic variation.

Previous versions of HapMap have provided the basis for the powerful genome-wide association studies that have led to the vast surge in discoveries of common gene variants related to several diseases and traits in recent years. But those versions of HapMap were very limited, particularly with regard their coverage of only four ethnic populations: European-American (CEU), Chinese in Beijing (CHB), Japanese in Tokyo (JPT), and Yoruba in Ibadan, Nigeria (YRI). HapMap3 increased the numbers in those populations and extended coverage to a further 7 populations: African-American, Chinese-American, Indian (Gujarati)-American, Mexican-American, Luyha and Maasai in Kenya, and Tuscans in Italy. The importance of these population-specific references was illustrated by the finding that overall variation in the Indian and Kenyan genomes differs markedly from the genomes of the other populations. So we cannot simply assume that the variant frequencies and linkage disequilibrium (the co-inheritance of adjacent variants) in one population applies to another.

In addition to genotyping using large-scale microarrays, they also performed sequencing at ten 100-kb regions (in total 0.03% of the genome). Unlike the SNPs present on those microarrays, which are intentionally biased towards higher frequency variants, 99% of the SNPs that were newly discovered by sequencing were rare or had an allele frequency $< 5\%$. Hence the next waves of gene variant discovery studies will likely identify novel associations less common variants, and it is hoped that their effect sizes will be larger than those currently seen.

A further notable finding was that genomic diversity in all the non-African populations was largely a subset of African diversity. This observation supports the concept of an 'Out of Africa' evolutionary origin of *Homo sapiens*.

New paradigms: 1000 Genomes Project

A map of human genome variation from population-scale sequencing

Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA
Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, UK.
Nature 2010;467:1061–73.

Background: The 1000 Genomes Project aims to provide a deep characterization of human genome sequence variation as a foundation for the investigation of the relationship between genotype and phenotype. This paper describes the results of the pilot phase of the project, designed to develop and compare different strategies for genome-wide sequencing with high-throughput platforms.

Methods: Three projects were performed: (1) low-coverage whole-genome sequencing of 179 individuals from four populations; (2) high-coverage sequencing of two mother-father-child trios, and (3) exon-targeted sequencing of 697 individuals from seven populations.

Results: The results describe the location, allele frequency and haplotype structure of ~15 million single nucleotide polymorphisms, 1 million short insertions and deletions, and 20,000 structural variants, most of which were previously unknown. By measuring the vast majority of common variation, this dataset contains over 95% of the currently accessible variants found in any individual. On average, each person is found to carry approximately 250–300 loss-of-function gene variants and 50–100 variants implicated in inherited disease. From the two trios, the rate of de novo germline base substitution mutations were directly estimated to be approximately 10^{-8} per base pair per generation. Further analyses identify a marked reduction of genetic variation in the neighborhood of genes, due to selection at linked sites.

Conclusions: These analyses demonstrate how whole-genome sequencing can be used to inform association and functional studies. These methods and public data will support the next phase of human genetic research.

Just when HapMap is finishing the 'highways' map of the common patterns of human DNA sequence variation, 1000 Genomes is filling in the less well-travelled 'B roads'. Its aim is to find most genetic variants with frequencies of at least 1%. The paper reports the 3 pilot studies of this large international collaboration that has now grown to actually include 2,500 individual genomes.

The paper reports that 'functional genetic variants', i.e. those that are predicted to alter gene or protein function and thereby directly affect disease risk, are likely to be less common than the variants genotyped in current studies, and it is hoped that this will account for much of the 'missing heritability' of disease. However, future disease association studies that apply this new information will have to consider how they can efficiently include an order of magnitude greater number of variants to effectively cover the human genome at this depth.

Meanwhile, one interim approach is to use this new information to extrapolate from our current more limited GWAS data, for example on 'only' 500,000 SNPs, to 'impute' the likely genotypes at up to 10 million SNPs. The authors demonstrate that such imputation methods will help to identify new disease loci, and also to localize current GWAS signal to the correct gene. For example, by such reanalysis the T1DM locus currently assigned to ORMDL3 actually appears to represent a splice variant in GSDMB, which encodes a member of the gasdermin-domain containing protein family, of which other members are implicated in the regulation of apoptosis in epithelial cells.

It is notable that two-thirds of currently identified GWAS loci are not in LD with a common coding variant, suggesting that most have genomic regulatory functions rather than acting via protein diversity. While the more efficient exome sequencing can be used to identify rare conditions, complex disease will continue to require a consideration of the whole genome.

Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants

Li Y, Vinckenbosch N, Tian G, Huerta-Sanchez E, Jiang T, Jiang H, Albrechtsen A, Andersen G, Cao H, Korneliussen T, Grarup N, Guo Y, Hellman I, Jin X, Li Q, Liu J, Liu X, Sparso T, Tang M, Wu H, Wu R, Yu C, Zheng H, Astrup A, Bolund L, Holmkvist J, Jorgensen T, Kristiansen K, Schmitz O, Schwartz TW, Zhang X, Li R, Yang H, Wang J, Hansen T, Pedersen O, Nielsen R, Wang J

BGI-Shenzhen, Shenzhen, China.

Nat Genet 2010;42:969–72.

Background: Targeted capture combined with massively parallel exome sequencing has been suggested as a promising approach to identify genetic variants implicated in human traits.

Methods: This study reports the exome sequencing of 200 individuals from Denmark with targeted capture of 18,654 coding genes. A statistical method for SNP calling was developed to estimate allelic frequencies based on these population data.

Results: Sequence coverage of each individual exome was achieved at an average depth of 12-fold. On average, about 95% of the target regions were covered by at least one read. 121,870 SNPs in the sample population were identified, including 53,081 coding SNPs (cSNPs). The allele frequency spectrum of cSNPs with a minor allele frequency >0.02 was described. There was a 1.8-fold excess of deleterious, non-synonymous cSNPs over synonymous cSNPs in the low-frequency range (minor allele frequencies between 2 and 5%). This excess was more pronounced for X-linked SNPs.

Conclusions: This study demonstrates the use of whole exome sequencing to identify low-frequency variants. The findings suggest that deleterious substitutions are primarily recessive.

This paper illustrates some of the drawbacks of next generation sequencing technologies. Many copies of a person's DNA are broken into short pieces. These random pieces are then aligned to a reference sequence. But many areas of their genome will be covered by multiple pieces of sequence, while others may not be covered at all. To ensure full coverage of a person's genome requires sequencing of that person's DNA the equivalent of 28 times ('28X'). Deeper sequencing will also provide information on both chromosomes and allows sequencing errors to be corrected. But while this technology may be becoming feasible in the search for mutations in novel genes in individual patients with extreme conditions, for large population-based studies, such inefficiencies greatly push up costs and time.

Whole exome sequencing covers 'just' the ~5% coding regions of the genome, i.e. those regions that directly encode amino acid sequences, but hopefully with higher return as these regions contain around 85% of all known disease-causing mutations.

In this study, even at a '12X' depth of sequencing coverage, the false positive error rate was unacceptably high for variants with minor allele frequency (MAF) <2%. Variants in the low-frequency range, 2–5%, were 1.8-fold more likely to be predicted as being deleterious to gene function compared to more common variants. The authors therefore conclude that humans have evolved mechanisms to select against such deleterious variants. It is hoped that many of these low-frequency variants will in combination explain a substantial proportion of the heritability of disease.

New paradigms: microbiome

Enterotypes of the human gut microbiome

Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, de Vos WM, Brunak S, Dore J, Antolin M, Artiguenave F, Blottiere HM, Almeida M, Brechot C, Cara C, Chervaux C, Cultrone A, Delorme C, Denariac G, Dervyn

R, Foerstner KU, Friss C, van de Guchte M, Guedon E, Haimet F, Huber W, van Hylckama-Vlieg J, Jamet A, Juste C, Kaci G, Knol J, Lakhdari O, Layec S, Le Roux K, Maguin E, Merieux A, Melo Minardi R, M'Rini C, Muller J, Oozeer R, Parkhill J, Renault P, Rescigno M, Sanchez N, Sunagawa S, Torrejon A, Turner K, Vandemeulebrouck G, Varela E, Winogradsky Y, Zeller G, Weissenbach J, Ehrlich SD, Bork P
European Molecular Biology Laboratory, Heidelberg, Germany.
Nature 2011;473:174–80.

Background: Knowledge of species and functional composition of the human gut microbiome is rapidly increasing, but it is still based on very few cohorts and little is known about its possible global variation.

Methods: The fecal metagenomes of 22 individuals from four countries were sequenced and then combined with previously published data.

Results: Three robust clusters (referred to as enterotypes) were identified that were not nation- or continent-specific. The enterotypes in two published, larger cohorts were also confirmed. The enterotypes are mostly driven by species composition, but abundant molecular functions are not necessarily provided by abundant species, highlighting the importance of a functional analysis to understand microbial communities. Although individual host properties such as body mass index, age, or gender cannot explain the observed enterotypes, data-driven marker genes or functional modules can be identified for each of these host properties. For example, 12 genes significantly correlate with age and three functional modules with the body mass index, hinting at a diagnostic potential of microbial markers.

Conclusions: In conclusion, intestinal microbiota variation is generally stratified, rather than continuous. This finding supports the existence of a limited number of well-balanced host-microbial symbiotic states that might respond differently to diet and drug intake.

The wide diversity in the human gut microbiome can be revealed by deep-sequencing of faecal RNA, an onerous task as the number of genes in the microbiome outnumbers human genes by 100-to-1. While infants have distinctive and unstable microbiomes [6], thereafter your composition of gut bacterial flora appears to be remarkably stable. Indeed, we have a synergistic relationship with our microbiomes, as our bacteria flourish on our diet as their source of nutrition, while they break down complex carbohydrates and produce vitamins such as biotin, riboflavin, thiamine and folate. This large study of microbiomes in 39 individuals found surprisingly little international differences between the six countries studied, and little correlation between phylogenetic enterotypes and age (excluding infancy), gender or BMI.

However, when grouped by function rather than phylogenetic signals, a number of correlations appeared with nationality, gender and age. Furthermore, three functional modules correlated strongly with BMI, two of which are ATPase complexes, supporting the previously described link between obesity and a microbiome with an increased capacity for energy harvest [7]. The microbiome has promise as a diagnostic or prognostic marker of disease, and may even represent a target for future interventions.

New mechanisms: pharmacogenetics and metformin

Common variants near ATM are associated with glycemic response to metformin in type 2 diabetes

Zhou K, Bellenguez C, Spencer CC, Bennett AJ, Coleman RL, Tavendale R, Hawley SA, Donnelly LA, Schofield C, Groves CJ, Burch L, Carr F, Strange A, Freeman C, Blackwell JM, Bramon E, Brown MA, Casas JP, Corvin A, Craddock N, Deloukas P, Dronov S, Duncanson A, Edkins S, Gray E, Hunt S, Jankowski J, Langford C, Markus HS, Mathew CG, Plomin R, Rautanen A, Sawcer SJ, Samani NJ, Trembath R, Viswanathan AC, Wood NW, Harries LW, Hattersley AT, Doney AS, Colhoun H, Morris AD, Sutherland C, Hardie DG, Peltonen L, McCarthy MI, Holman RR, Palmer CN, Donnelly P, Pearson ER
Biomedical Research Institute, University of Dundee, Dundee, UK.
Nat Genet 2011;43:117–20.

Background: Metformin is one of the most commonly used drugs in the management of type 2 diabetes.

Methods: A genome-wide association study was performed for glycemic response to metformin in 1,024 Scottish individuals with type 2 diabetes. Replication was performed in two cohorts including 1,783 Scottish individuals and 1,113 individuals from the UK Prospective Diabetes Study.

Results: In a combined meta-analysis, the SNP rs11212617 was associated with treatment success ($n = 3,920$, $p = 2.9 \times 10^{-9}$), odds ratio = 1.35, 95% CI 1.22–1.49) at a locus containing ATM, the gene mutated in ataxia telangiectasia. In a rat hepatoma cell line, inhibition of ATM with KU-55933 attenuated the phosphorylation and activation of AMP-activated protein kinase in response to metformin.

Conclusions: In conclusion, ATM, a gene known to be involved in DNA repair and cell cycle control, was shown to modify the effect of metformin upstream of AMP-activated protein kinase, and variation in this gene alters glycemic response to metformin.

Pharmacogenetics is portrayed as the basis for the future individualization of drug treatment. Knowing your patient's genome will enable you to predict their likely treatment effectiveness and risk responses. For warfarin dosing, it is suggested that polymorphisms in two genes, VKORC1 and CYP2C9, explain 40% of dose variation between patients and allow tailoring of initial doses. This paper describes an additional utility of exploring such gene-drug interactions, in that they can shed light on the detailed mechanisms of a drug's action.

Remarkably the mechanism of action of metformin is not fully clear, despite being in clinical use for over 50 years and with increasingly widespread reported benefits on glucose control, cardiovascular disease and cancer prevention in T2DM, weight loss in obesity, fertility in PCOS, growth and body composition in premature pubarche, and many more. Metformin is known to activate AMP-activated protein kinase (AMPK) by inhibition of the mitochondrial respiratory chain, causing an increase in cellular AMP, and other recent studies reported interactions between metformin and variants in the AMPK gene *STK11* [8, 9].

These authors identified that a common variant in ATM, the ataxia-telangiectasia gene, modified the HbA1c response to metformin in adults with T2DM. They were aware that individuals with ataxia-telangiectasia have marked insulin resistance and a high risk of T2DM. ATM is activated by double-stranded DNA breaks and acts upstream of AMPK to induce cell-cycle arrest and facilitate DNA repair. Their experimental studies showed that ATM is essential for metformin's actions and their findings establish a direct link between T2DM, cancer and metformin activation of AMPK.

New concepts: changing T1DM penetrance

Stepwise or linear decrease in penetrance of type 1 diabetes with lower-risk HLA genotypes over the past 40 years

Steck AK, Armstrong TK, Babu SR, Eisenbarth GS

Barbara Davis Center for Childhood Diabetes, University of Colorado Denver, Aurora, Colo., USA.

Diabetes 2011;60:1045–9.

Background: This study aimed to test whether the proportion of new-onset diabetic subjects with the HLA-DR3/4-DQB1*0302 genotype is decreasing over time.

Methods: HLA class II genotype frequencies were analyzed over time in two large populations with type 1 diabetes diagnosed at ≤ 18 years of age. Analyses included 4,075 individuals from the Type 1 Diabetes Genetics Consortium (T1DGC) and 1,675 from the Barbara Davis Center (BDC).

Results: Both T1DGC and BDC cohorts showed a decrease of the highest-risk HLA-DR3/4-DQB1*0302 genotype over time. This decrease was greatest over time in T1DGC children with age of onset ≤ 5 years ($p = 0.004$) and onset between ages 6 and 10 years ($p = 0.002$). The proportion of HLA-DR3/4-DQB1*0302 was higher in T1DGC than BDC. There was an increase over time in the proportion of non-HLA-DR3 or -DR4 genotypes in new-onset T1DGC cases ($p = 0.003$), with a similar trend among BDC cases ($p = 0.08$). There appeared to be a large stepwise decrease in percent DR3/4 in the 1980s in T1DGC children with onset age < 5 years ($p = 0.0001$).

Conclusions: The change in frequency of multiple different genotypes and a possible stepwise decrease in percent DR3/4 suggest a temporal change in the environmental determinants of type 1 diabetes. If con-

firmed in larger studies, a stepwise change may have direct bearing on defining critical environmental determinants of type 1 diabetes.

The rise in T1DM incidence is now well documented, and especially among children under 5 years of age. The environmental factors driving this change remain unknown, but they are having a peculiar influence on the changing genetic profile of T1DM.

This study describes a near stepwise decrease in the mid-1980s in the proportion of newly diagnosed T1DM young children carrying the traditionally high-risk HLA-DR3/4 genotype. The authors discuss that this suggests that there was a relatively acute environmental change, resulting in either a lower disease penetrance in children with the high-risk genotype, or rather a much greater rise in the disease penetrance in children with traditionally low-risk non-DR3/4 genotypes.

There has been great progress in the identification of other T1DM genetic loci in recent years, and it is possible that the rise in T1DM could coincide with a rise in the proportion of other susceptibility genotypes such as IFIH1, a gene which encodes a cytoplasmic protein that recognizes RNA of picornaviruses, including enteroviruses [10]. It is possible that such genetic studies and genotype-year analyses could paradoxically shed light on the environmental determinants of these worrying secular trends.

Food for thought: hidden heritability

Common SNPs explain a large proportion of the heritability for human height

Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM

Queensland Institute of Medical Research, Brisbane, Qld., Australia.

Nat Genet 2010;42:565–9.

Background: SNPs discovered by genome-wide association studies (GWAS) explain only a small proportion of the genetic variation of complex traits in humans. Where is the remaining heritability?

Methods: A linear model analysis was used to estimate the proportion of variance in adult height explained by 294,831 SNPs genotyped on 3,925 unrelated individuals. This estimation method was validated by simulations based on the observed genotype data.

Results: The analyses predicted that 45% of the variance in adult height can be explained by considering all SNPs simultaneously. Much of the remaining heritability is due to incomplete linkage disequilibrium between causal variants and genotyped SNPs, exacerbated by causal variants having lower minor allele frequency than the SNPs that have been genotyped to date.

Conclusions: Most of the heritability of adult height is not missing, but has not yet been identified because the individual SNP effects are too small to reach stringent significance tests.

One of the keys to the success of GWAS to identify robust new loci has been the extremely stringent statistical criteria that have to be fulfilled before placing confidence in its findings. The usually cited genome-wide threshold of $p < 5 \times 10^{-8}$ was designed to reduce the possibility of false positive findings that had plagued candidate gene association studies, but it is possible that studies will never have sufficient power to enable the majority of common variants to meet this bar.

These authors modeled the possibility that many of the weaker associations for adult height are indeed true findings. Their raw estimate that all SNPs will explain 45% of the total variation in adult height represents 56% of the genetic contribution to height, and this increases to 67% if they account for the imperfect correlation between the tagging and the causal variants.

In the same journal, a related paper by Park et al. [11] estimated that to identify these additional variants will require studies of hundreds of thousands of individuals. This is truly a daunting task, but is one that is becoming a realizable prospect in light of the increasing number of national and international Biobank projects [12].

Initial impact of the sequencing of the human genome

Lander ES

Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Mass., USA.

Nature 2011 470:187–97.

Knowledge of the human genome sequence has dramatically accelerated biomedical research. This review explores its impact, in the decade since its initial publication, on our understanding of the biological functions encoded in the genome, on the biological basis of inherited diseases and cancer, and on the evolution and history of the human species. The author also discusses the road ahead to fulfill the promise of genomics for medicine.

This is a very readable, personal and yet comprehensive insight into the story of our understanding of the human genome, including the hopes and expectations, the challenges along the way, and the prospects for its future application in science and medicine. Eric Lander was first author on the paper 'Initial sequencing and analysis of the human genome' published in *Nature* on February 15, 2001 [13]. Actually, following this 'rough draft', the updated sequence was published in 2004 [14], improving coverage from 90 to 99.7% and reducing the number of gaps from 250,000 to 300, mostly in highly repetitive regions and centromeres.

Lander says that its greatest impact has been as the basis to study biology in a comprehensive and unbiased manner, using genomics and its subsequent applications such as epigenomics and transcriptomics. In 2001, fewer than 100 mendelian disease genes had been identified – a decade later we have more than 2,850, and many further loci for complex diseases and traits from GWAS. As well as the benefits of analyzing germline DNA, testing for somatic mutations will one day routinely inform cancer diagnosis, prognosis and screening, and the ability to distinguish bacterial and viral from human genomes will enable the rapid detection of infectious disease.

References

1. Lettre G, Jackson AU, Gieger C, Schumacher FR, Berndt SI, Sanna S, et al: Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat Genet* 2008;40:584–591.
2. Weedon MN, Lango H, Lindgren CM, Wallace C, Evans DM, Mangino M, et al: Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet* 2008;40:575–583.
3. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al: A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 2007;316:889–894.
4. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, et al: Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 2008;40:638–645.
5. Zhu H, Shah S, Shyh-Chang N, Shinoda G, Einhorn WS, Viswanathan SR, et al: Lin28a transgenic mice manifest size and puberty phenotypes identified in human genetic association studies. *Nat Genet* 2010;42:626–630.
6. Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO: Development of the human infant intestinal microbiota. *PLoS Biol* 2007;5:e177.
7. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI: An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 2006;444:1027–1031.
8. Lopez-Bermejo A, Diaz M, Moran E, de Zegher F, Ibanez L: A single nucleotide polymorphism in STK11 influences insulin sensitivity and metformin efficacy in hyperinsulinemic girls with androgen excess. *Diabetes Care* 2010;33:1544–1548.
9. Jablonski KA, McAteer JB, de Bakker PI, Franks PW, Pollin TI, Hanson RL, et al: Common variants in 40 genes assessed for diabetes incidence and response to metformin and lifestyle intervention in the diabetes prevention program. *Diabetes* 2010;59:2672–2681.
10. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA: Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* 2009;324:387–389.
11. Park JH, Wacholder S, Gail MH, Peters U, Jacobs KB, Chanock SJ, et al: Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat Genet* 2010;42:570–575.
12. Gaskell G, Gottweis H: Biobanks need publicity. *Nature* 2011;471:159–160.
13. Lander ES, Linton LM, Birren B, Nussbaum C, Zody MC, Baldwin J, et al: Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921.
14. International Human Genome Sequencing Consortium: Finishing the euchromatic sequence of the human genome. *Nature* 2004;431:931–45.