Population Genetics and Pharmacogenetics

Ken Ong

Medical Research Council Epidemiology Unit and Department of Paediatrics, University of Cambridge, Cambridge, UK

With 10 million single nucleotide polymorphisms (SNPs) to choose from it is little wonder that the vast majority of candidate-SNP- or candidate-gene-driven genetic association studies turned up negative or misleading false-positive results. The leading investigators and funding bodies in this field have therefore now committed their resources to very large studies in the range of 5,000–10,000 individuals and also employing Genomewide association SNP chips that genotype hundreds of thousands of SNPs, which in turn ‘tag’, or are markers for, the majority of all the common variants in the HapMap [1]. These ‘high-stakes’ approaches are now paying dividends. In the last year alone:

- a common SNP in a T-cell transcription factor gene TCF7L2 has been repeatedly confirmed in numerous large studies as a consistent genetic marker of type 2 diabetes (T2DM) risk (see Helgason et al. below)
- the first Genomewide association study for T2DM claims to have explained 70% of the disease risk (see Sladek et al. below)
- the first Genomewide association study for type 1 diabetes has led to convincing statistical evidence for a sixth susceptibility locus at the interferon inducer IFIH1 (see Smyth et al. in Chapter 7 [2])
- most recently a Genomewide association study for T2DM unearthed a major common variant associated with obesity at the TFO gene (see Frayling et al. below).

It may be debatable whether such findings will truly generate new drug targets. However, as the number of such validated genetic markers increases, we will hopefully be able to understand much more about the genetic epidemiology of various diseases, including much better refinement of individual disease risk prediction, how genetic susceptibility interacts with environmental factors, and eventually individual-level selection of optimal drug therapies or lifestyle interventions.

New hope – genomewide association studies

A genome-wide association study identifies novel risk loci for type 2 diabetes

Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P

Department of Human Genetics, McGill University and Genome Quebec Innovation Centre, Montreal, Que., Canada
Polychronakos@McGill.ca

Nature 2007;44:881–885

Background: Type 2 diabetes (T2DM) results from the interaction between various environmental factors with a number of genetic variants, most of which are yet unknown. The recent development of high-density arrays allows the genotyping of hundreds of thousands of polymorphisms, and thereby allows a systematic search for T2DM susceptibility variants.

Methods: 392,935 single-nucleotide polymorphisms were tested in a French case-control cohort of 1,363 individuals. 59 markers with the most significant difference in genotype frequencies between cases of T2DM and controls were fast-tracked for testing in a second case-control study of 5,511 individuals.

Results: In addition to confirming the known association with the TCF7L2 gene, the second stage study identified four further loci containing variants that confer T2DM risk. These loci include a non-synonymous polymorphism in the zinc transporter SLC30A8, which is expressed exclusively in insulin-producing β -cells, and two linkage disequilibrium blocks that contain genes potentially involved in β -cell development or function (IDE-KIF11-HHEX and EXT2-ALX4).

Conclusions: These four novel genetic loci together with TCF7L2 explained a substantial 70% of the risk of T2DM. These findings constitute proof of principle for the genome-wide approach to the elucidation of complex genetic traits.

This is the first reported Genomewide association study in T2DM and it demonstrates the advantages of a hypothesis-free approach to SNP selection. The authors initially tested 392,935 SNPs in the smaller case-control study, where cases were enriched for genetic susceptibility by having a first-degree family history of diabetes, early age at onset and non-obese range BMI. They then took forward to the second stage larger study in more representative cases only those SNPs that reached Genomewide association after stringent correction for multiple testing. In addition to extremely convincing evidence for the T-cell transcription factor gene TCF7L2 (p value = 1.5×10^{-34}), the study found four new genes or loci related to T2DM risk. These include SLC30A8 on chromosome 8, which encodes a zinc transporter expressed only in the secretory granules of β -cells, and HHEX, a homeodomain protein which is essential for hepatic and pancreatic development. Further studies need to confirm these novel findings and results from other Genomewide association studies for T2DM are expected shortly. Interestingly, zinc deficiency that often accompanies fetal growth retardation, which in turn has been linked to later development of T2DM in adulthood, was associated with hypomethylation and hyperacetylation of genomic DNA in brain and liver of IUGR fetal and juvenile rats [3].

The authors speculate that the results may lead to the development of new dietary and pharmacological approaches to treating and preventing diabetes, for example with zinc supplementation. Other notable findings of this groundbreaking study were that these genetic markers explained 70% of the population-attributable risk of diabetes, which far exceeds previous estimates of the heritability of T2DM. Lastly, in 7 of the 8 SNPs associated with diabetes, the susceptibility allele was the ancestral, or older, allele. This supports the hypothesis that ancestral alleles were adapted to promote survival in the environments of ancient human history, but predispose to disease risk in today's affluent and obesogenic environments.

A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity

Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW, Shields B, Harries LW, Barrett JC, Ellard S, Groves CJ, Knight B, Patch AM, Ness AR, Ebrahim S, Lawlor DA, Ring SM, Ben-Shlomo Y, Jarvelin MR, Sovio U, Bennett AJ, Melzer D, Ferrucci L, Loos RJ, Barroso I, Wareham NJ, Karpe F, Owen KR, Cardon LR, Walker M, Hitman GA, Palmer CN, Doney AS, Morris AD, Davey-Smith G; The Wellcome Trust Case Control Consortium; Hattersley AT, McCarthy MI
Institute of Biomedical and Clinical Science, Peninsula Medical School, Exeter, UK

Andrew.Hattersley@pms.ac.uk

Science 2007 Apr 12 [Epub ahead of print]

Background: Obesity is a serious international health problem that increases the risk of several common diseases. The genetic factors predisposing to obesity are poorly understood.

Methods and Results: A genome-wide association study for type 2 diabetes-susceptibility genes identified a common variant SNP rs9939609 in the FTO gene that predisposes to diabetes through an effect on increasing body mass index (BMI). An additive association of the variant with BMI was replicated in 13 cohorts with 38,759 participants ($p = 3 \times 10^{-35}$). The 16% of adults who are homozygous for the risk allele weighed ~3 kg more and had a 1.67-fold higher risk of obesity compared to those not inheriting a risk allele. This association was observed from age 7 years upwards and reflects a specific increase in fat mass.

Conclusions: Understanding how variation in the FTO gene region is associated with BMI and adiposity may provide insights into novel pathways involved in the control of adiposity.

This study surely represents the first true genetic signal for common obesity in adults and children. Family-based studies have long indicated a large genetic contribution to obesity, yet attempts to confirm initial promising reports of association with common variants (e.g. in GAD, ENPP1 and INSIG2 genes) have been disappointing. In this study, a Genomewide association study for T2DM firstly found that SNPs in the FTO gene region on chromosome 16 were strongly associated with T2DM. The diabetes-risk alleles were also strongly associated with increased BMI, and this completely explained the susceptibility to T2DM. They then studied the FTO gene variant in an additional 19,424 white

European adults and 10,172 white European children and confirmed the association with BMI in all 13 cohorts to an unprecedented level of significance ($p = 3 \times 10^{-35}$).

FTO is a gene of yet unknown function in an unknown pathway. It was originally cloned in a fused toe (Ft) mutant mouse arising from a large chromosomal deletion. Heterozygous animals are characterized by fused toes on the forelimbs and thymic hyperplasia, but no alteration in body weight or adiposity. However, the Ft mutant is a poor model for altered Fto activity as multiple genes are deleted. In humans, FTO is widely expressed in fetal and adult tissues with expression highest in the brain and pancreatic islet. Future studies will hopefully be quick to confirm its association with obesity in other populations and to explore this potentially novel causal pathway to obesity.

New paradigms – TCF7L2 in T2DM

Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution

Helgason A, Palsson S, Thorleifsson G, Grant SF, Emilsson V, Gunnarsdottir S, Adeyemo A, Chen Y, Chen G, Reynisdottir I, Benediktsson R, Hinney A, Hansen T, Andersen G, Borch-Johnsen K, Jorgensen T, Schafer H, Faruque M, Doumatey A, Zhou J, Wilensky RL, Reilly MP, Rader DJ, Bagger Y, Christiansen C, Sigurdsson G, Hebebrand J, Pedersen O, Thorsteinsdottir U, Gulcher JR, Kong A, Rotimi C, Stefansson K
deCODE genetics, Reykjavik, Iceland
agnar@decode.is
Nat Genet 2007;39:218–225

Background: This group recently reported association between risk of type 2 diabetes (T2DM) and common single nucleotide polymorphisms (SNPs) in the transcription factor 7-like 2 gene (TCF7L2), which explained a population-attributable risk (PAR) of 17%–28% in three populations of European ancestry.

Methods: In order to refine the definition of the TCF7L2 diabetes susceptibility haplotype ‘HapB(T2D)’, replication studies were performed in West African and Danish T2DM case-control studies and in an expanded Icelandic study.

Results: The T2DM susceptibility haplotype was narrowed to a single SNP, the ancestral T allele of rs7903146. These new data also identified another TCF7L2 variant ‘HapA’ that shows evidence of positive selection in East Asian, European and West African populations. Notably, HapA shows a suggestive association with body mass index and altered concentrations of the hunger-satiety hormones ghrelin and leptin in males.

Conclusions: Selective advantage of HapA may have been mediated through effects on energy metabolism.

Since the original publication in January 2006 [4], TCF7L2, encoding the transcription factor 7-like 2, has proved to be the first consistent genetic marker of T2DM susceptibility. This association has been confirmed but only in very large studies, but also in various populations from around the world including Japanese, Indians and Mexican-Americans. The original study by this group identified strong association with an intron 3 microsatellite ‘DG105478’. While this has proved to be a good marker of disease risk, this was unlikely to be the actual causal or functional variation. The authors therefore went on in this study to try to refine the association with T2DM and identify the causal variation by genotyping a more genetically diverse West African population. Overall, these results rule out DG105478 X and rs12255372 T as causal variants but identify rs7903146 T as either the risk variant itself or its closest known correlate.

The authors then went on to look at evolutionary selection of the various TCF7L2 haplotypes, however these results raise more questions than answers. Using HapMap data, they found two major lineages, HapA and HapB (which predisposes to diabetes). The frequency of HapA increased dramatically from African to European through to the East Asian populations, suggesting a strong selection pressure. There was weaker evidence that HapA predisposes to higher BMI, yet the strongest relationship was observed for HapA with ghrelin and the ghrelin-to-leptin ratio. These findings do not explain why East Asians should have lower BMI, and develop diabetes at lower levels of obesity compared to other populations, nor indeed what the survival advantage is of having HapA.

TCF7L2-gene polymorphisms confer an increased risk for early impairment of glucose metabolism and increased height in obese children

Korner A, Berndt J, Stumvoll M, Kiess W, Kovacs P

University Hospital for Children and Adolescents, Department of Internal Medicine III, University of Leipzig, Leipzig, Germany

J Clin Endocrinol Metab 2007 Feb 20 [Epub ahead of print]

Background: Variants in the transcription factor 7-like 2 (TCF7L2) gene have been associated with increased risk for type 2 diabetes (T2DM) in adults. This study evaluated whether TCF7L2 variants confer a higher risk for obesity and early impairment of glucose metabolism in children.

Methods: The five reported TCF7L2 risk variants were genotyped in a representative cohort of 1,029 Caucasian children and an independent cohort of 283 obese children.

Results: There was a significantly lower prevalence of the rs11196205 and rs7895340 T2DM risk alleles in the obese ($n = 283$) compared to lean ($n = 672$) children (0.40 vs. 0.45, $p = 0.02$). However, there were no significant associations between these genotypes and quantitative traits of obesity in either the normal or the obesity cohort. Obese children were significantly taller than lean children. This increase in height was independently associated with T2DM risk variants of the TCF7L2-gene, while in the normal representative cohort height appeared to be decreased in carriers of the minor alleles. In the obese cohort, three risk alleles (rs7901695, rs7903146, rs1225572) were significantly associated with higher fasting and 120-minute blood glucose levels independent of sex, age, pubertal stage, and BMI. Fasting and peak insulin levels and HOMA-IR appeared with a similar tendency but were not statistically significant.

Conclusions: These data indicate for the first time that TCF7L2 variants are associated with increased risk for early impairment of glucose metabolism in obese children. These findings are consistent with adult studies identifying TCF7L2 as a major diabetes susceptibility gene.

These authors found that the main genetic markers at TCF7L2 that confer T2DM susceptibility in adults are also associated with higher glucose levels in obese children. Such findings may help to discriminate which obese children are at most risk for diabetes and therefore need oral glucose tolerance testing.

Furthermore, such studies in young non-diabetic subjects may also help to understand the functional consequences of variants in this gene that lead to T2DM risk. Unfortunately, despite looking at this in a large and representative childhood cohort no clear answers to this question were apparent. In adults, TCF7L2 variants show increasing associations with reduced insulin secretion and defective insulin processing rather than increased insulin resistance or greater obesity risk [5]. In contrast, in the current study, non-significant trends towards both higher fasting and peak insulin levels were seen. Two SNPs were weakly associated with reduced obesity risk, but these variants showed least association with glucose levels and the findings were not confirmed in the more powerful quantitative trait analyses. Finally, some associations were seen with childhood heights, but intriguingly these findings were in the opposite directions in their obese and representative cohorts. Further studies in other childhood cohorts with more detailed birth weight and growth data might help the identification of early pathways to T2DM.

New paradigms – childhood models of adult T2DM

Common variants in maturity-onset diabetes of the young genes contribute to risk of type 2 diabetes in Finns

Bonnycastle LL, Willer CJ, Conneely KN, Jackson AU, Burrill CP, Watanabe RM, Chines PS, Narisu N, Scott LJ, Enloe ST, Swift AJ, Duren WL, Stringham HM, Erdos MR, Riebow NL, Buchanan TA, Valle TT, Tuomilehto J, Bergman RN, Mohlke KL, Boehnke M, Collins FS

Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Md., USA

francisc@mail.nih.gov

Diabetes 2006;55:2534–2540

Background: Previous studies have suggested that variants in the genes for maturity-onset diabetes of the young (MODY) may confer susceptibility to type 2 diabetes (T2DM). However, the results have been

conflicting and coverage of the MODY genes has been incomplete. To complement previous studies of HNF4A, the other five known MODY genes were examined for association with T2DM in Finnish individuals.

Methods: The following single nucleotide polymorphism (SNP) selection criteria were used for each of the five genes: (1) tagging in linkage ($r^2 < 0.8$) with other SNPs from the HapMap database or another linkage disequilibrium map, (2) SNPs previously associated with T2DM, and (3) non-synonymous coding SNPs. In total, 128 SNPs were tested for association with T2DM in 786 index cases from T2DM families and 619 normal glucose-tolerant controls. 35 of the most significant SNPs were genotyped in a second study of 384 cases and 366 controls from Finland. Previously reported HNF4A results were extended by genotyping 12 SNPs on additional Finnish samples.

Results: After correcting for multiple testing, there was evidence of T2DM association with SNPs in five of the six known MODY genes: GCK, HNF1A, HNF1B, NEUROD1, and HNF4A.

Conclusions: These studies suggest that common variants in five of the six known MODY genes play a modest role in T2DM susceptibility.

Assessment of the role of common genetic variation in the transient neonatal diabetes mellitus region in type 2 diabetes: a comparative genomic and tagging single nucleotide polymorphism approach

Gloyn AL, Mackay DJ, Weedon MN, McCarthy MI, Walker M, Hitman G, Knight BA, Owen KR, Hattersley AT, Frayling TM

Diabetes Research Laboratories, Oxford Centre for Diabetes, Endocrinology and Metabolism, Oxford, UK
frayling@pms.ac.uk

Diabetes 2006;55:2272–2276

Background: Recent evidence supports the strong overlap between genes implicated in monogenic diabetes and susceptibility to type 2 diabetes (T2DM). Several T2DM linkage studies have reported linkage to chromosome 6q22–25. Transient neonatal diabetes mellitus (TNDM) is a rare disorder associated with overexpression of genes at a paternally expressed imprinted locus on chromosome 6q24. There are two overlapping genes in this region: the transcription factor zinc finger protein associated with cell cycle control and apoptosis (ZAC also known as PLAGL1) and HYMA1, which encodes an untranslated mRNA. This study tested the hypothesis that common genetic variants at this TNDM region may be associated with T2DM susceptibility.

Methods: In 47 individuals, new single nucleotide polymorphisms (SNP) in the TNDM gene region were identified by resequencing the ZAC coding regions and also regions of high conservation between human and mouse ZAC and 20-kb flanking sequences. The resulting tag SNPs were genotyped in a large UK Caucasian case-control ($n = 3,594$) and family-based ($n = 1,654$) study.

Results: 26 SNPs were identified. 15 tag SNPs were successfully genotyped, which captured 92% of the 26 SNPs with $r^2 > 0.5$. There was no evidence of association or overtransmission of any SNP to affected offspring or of a parent-of-origin effect.

Conclusions: In a study sufficiently powered to detect odds ratios of <1.2 , common variation in the TNDM region does not appear to play an important role in the genetic susceptibility to T2DM in UK Caucasians.

These two papers explore the principle that the detailed study of rare childhood disorders may help the understanding of the etiology of common adult disease. The candidate gene approach to genetic association studies of common disease seems to have had greatest success by choosing genes in which rare, usually deleterious, mutations have already been shown to cause monogenic forms of those conditions. The above two studies have studied common variations in genes involved in MODY and TNDM for association with T2DM in adults. Interestingly, five of the six known MODY genes showed evidence of association with T2DM. One further variant in HNF4A associated with earlier age at diagnosis. However, the effect sizes were modest, and the authors cited personal communications in the discussion that the findings with the MODY genes were not confirmed in a large cohort of Swedish, Finnish, Canadian and Polish individuals. Further large studies and meta-analyses may also conclude that the current findings are false positives. However it is not long ago that these different forms of MODY were clinically classified together with T2DM. Similarly, a subtype of early onset type 1 diabetes has recently been recognized to have a distinct monogenic cause and dramatic response to oral sulfonyleurea

therapy [6]. In addition to the large-study and large-scale genotyping approach, careful recognition of further clinical subtypes of disease are necessary to unravel the genetics of common diseases.

New genetic associations: with childhood obesity

A common genetic variant is associated with adult and childhood obesity

Herbert A, Gerry NP, McQueen MB, Heid IM, Pfeufer A, Illig T, Wichmann HE, Meitinger T, Hunter D, Hu FB, Colditz G, Hinney A, Hebebrand J, Koberwitz K, Zhu X, Cooper R, Ardlie K, Lyon H, Hirschhorn JN, Laird NM, Lenburg ME, Lange C, Christman MF

Department of Genetics and Genomics, Boston University Medical School, Boston, Mass., USA

aherbert@bu.edu

Science 2006;312:279–283

Background: Obesity is a heritable trait and is an important risk factor for several common diseases such as type 2 diabetes, heart disease, and hypertension.

Methods: In a dense whole-genome scan, 116,204 SNPs were genotyped in 694 participants from the Framingham Heart Study offspring cohort. After exclusions, 86,604 SNPs were tested for association with BMI.

Results: One SNP (rs7566605) near the INSIG2 gene reached overall significance for association with obesity at a genome-wide level. This association was subsequently replicated in four separate populations composed of individuals of Western European ancestry, African-Americans, and children. The obesity-predisposing genotype is present in 10% of individuals.

Conclusions: This study suggests that common genetic polymorphisms are important determinants of obesity.

This Genomewide association study using the Affymetrix 100K SNP chip found strong evidence for a new obesity-associated common genetic variant. A meta-analysis of all their case-control samples showed that the rs7566605 CC genotype conferred a mild increase in obesity risk in a recessive model with odds ratio of 1.22 (95% CI 1.05–1.42, $p = 0.008$). This variant lies 10 kb upstream of the insulin induced gene 2 (INSIG2), which includes a protein that inhibits synthesis of fatty acid and cholesterol, and is therefore an attractive biological candidate for further functional studies.

However, while the authors went to great lengths to show the consistency of their finding a wide range of other populations, the subsequent correspondence relating to this paper only a few months later in *Science* reveals that other investigators have failed to support the main finding of the original study. Loos et al. [7] if anything found the opposite tendency with BMI in two large UK Caucasian cohorts totalling 6,599 individuals; Dina et al. [8] also found no evidence for this association in 10,265 Caucasian individuals, and Rosskopf et al. [9] found no association with BMI in 4,089 German individuals. A major contribution of INSIG2 rs7566605 to the genetic risk of obesity in the West European population is therefore unlikely, however it remains possible that INSIG2 contributes to BMI variation in children and other ethnic groups.

The enormous analytical and computational capacity of modern technology requires special attention to old rules of statistics. Tossing a cube 100 thousand times, one would expect just by chance one gene with a p value of 10^{-5} .

ENPP1 variants and haplotypes predispose to early onset obesity and impaired glucose and insulin metabolism in German obese children

Bottcher Y, Korner A, Reinehr T, Enigk B, Kiess W, Stumvoll M, Kovacs P

Medical Department III, University of Leipzig, Leipzig, Germany

peter.kovacs@medizin.uni-leipzig.de

J Clin Endocrinol Metab 2006;91:4948–4952

Background: ENPP1 (nucleotide pyrophosphatase/phosphodiesterase-1) encodes a membrane-bound glycoprotein that inhibits the insulin-receptor tyrosine kinase activity, resulting in reduced insulin sensitivity. Common variants in this gene have previously been associated with obesity and insulin resistance. This

study explored the role of ENPP1 genetic variants in obesity and related traits in a representative population of Caucasian children and in cohorts of obese children with detailed metabolic assessments.

Methods: The K121Q, IVS20delT-11, and A/G+1044TGA variants in ENPP1 were genotyped in 712 schoolchildren (346 boys and 366 girls; mean age 12 ± 3 years; BMI SD score 0.09 ± 0.04) and in independent cohorts of 205 obese children from Leipzig and 195 obese children from Datteln, Germany.

Results: Compared to a lean control group, there was an increased risk of obesity in Leipzig children carrying the 121Q variant (adjusted odds ratio 1.82; 95% CI 1.30–2.56, $p = 0.0005$) or the [Q-delT-G] haplotype [1.75 (1.17–2.62), $p = 0.006$]. These findings were replicated in another independent obesity/overweight cohort from Leipzig as well as obese children from Datteln. In addition, among obese children from Leipzig, the [Q-delT-G] haplotype was associated with higher 2-hour glucose levels, whereas the [K-delT-G] and [K-insT-A] haplotypes were associated with improved insulin sensitivity and glucose metabolism (all $p < 0.05$ after adjusting for age, gender, and BMI).

Conclusions: This study suggests a potential role of the K121Q polymorphism and ENPP1 haplotypes in conferring increased susceptibility to obesity and early impairment of glucose tolerance in children.

Recent studies summarized in *Yearbook 2006* [1] had raised the hope that ENPP1 might represent the first common genetic determinant of both obesity and T2DM, and might help to identify which obese individuals are at high risk of metabolic complications. However, as is often the case, subsequent large studies in adult populations have since found no evidence of the original associations [10, 11], and even findings in the opposite direction [12], indicating that the results may be false positive or are not easily generalizable.

In contrast to adults, the above study in German children does provide some support for a consistent association in children. The ENPP1 Q121 variant, proposed to impair insulin signaling (or a haplotype of Q121 plus two other variants), was associated with increased risk of obesity and glucose intolerance, while haplotypes with the K121 variant conferred improved insulin sensitivity and glucose homeostasis. These findings strongly resemble those published in French children [13] and obviously need to be confirmed in further large and well-characterized childhood cohorts. Early-onset obesity and childhood or adolescent T2DM may be determined by different or more specific risk factors compared to their adult counterparts, and the confirmation of their genetic risk factors should not be reliant on large-scale studies performed only in adults.

New genetic associations: with birth weight

Polymorphism in maternal LRP8 gene is associated with fetal growth

Wang L, Wang X, Laird N, Zuckerman B, Stubblefield P, Xu X

Program for Population Genetics, Harvard School of Public Health, Boston, Mass., USA

xin_xu@harvard.edu

Am J Hum Genet 2006;78:770–777

Background: Fetal growth restriction (FGR) affects >200,000 pregnancies in the USA annually and is associated with increased perinatal mortality and morbidity, as well as poorer long-term health for the infants. FGR appears to be a complex trait, and the role of genetic factors is largely unknown.

Methods: A candidate-gene association study of birth weight and FGR was performed in two independent study samples obtained at the Boston Medical Center. Firstly, association between maternal genotypes of 68 SNPs from 41 candidate genes and fetal growth was investigated in a sample of 204 black women selected for a previous study of preeclampsia, 92 of whom had preeclampsia (characterized by high blood pressure and the presence of protein in the urine).

Results: SNP rs2297660 in the LRP8 gene was significantly associated with birth weight. In a subsequent replication study, there was a similar association between LRP8 and FGR in a larger independent sample of 1,094 black women: the 'A' allele at rs2297660 was associated with a higher standardized birth weight and a lower risk of FGR. In an additive model, each additional copy of the 'A' allele reduced the risk of FGR by 33% ($p < 0.05$).

Conclusions: Results from two independent samples of black women provide consistent evidence that SNP rs2297660 in LRP8 is associated with fetal growth.

Many childhood cohorts are being studied in relation to the etiology of long-term health outcomes associated with low birth weight and related postnatal growth patterns. This study reminds us that the predisposition to producing low birth weight offspring is a heritable maternal trait and that candidate underlying factors may therefore reside in mother's genotype.

In this study all the candidate genes were selected based on biological plausibility, as they were potentially important in preeclampsia and chronic hypertension, which is a major pathogenic pathway of FGR. The author's definition of FGR was synonymous with small gestational age (SGA), being defined as birth weight below the 10th centile for sex and gestation. In their pooled analysis, each A allele at rs2297660 was associated with a 0.18 SD increase in birth weight and 0.67 relative risk of FGR.

LRP8 encodes the apolipoprotein E (ApoE) receptor 2, which mediates the cellular recognition and internalization of ApoE-containing lipoproteins and is highly expressed in the placenta. The authors did not directly report whether this variant was associated with maternal hypertension or preeclampsia, however further putative effects of the transmitted variant in the fetus could also contribute to links between low birth weight and long-term risk for adult disease.

A common haplotype of the glucokinase gene alters fasting glucose and birth weight: association in six studies and population-genetics analyses

Weedon MN, Clark VJ, Qian Y, Ben-Shlomo Y, Timpson N, Ebrahim S, Lawlor DA, Pembrey ME, Ring S, Wilkin TJ, Voss LD, Jeffery AN, Metcalf B, Ferrucci L, Corsi AM, Murray A, Melzer D, Knight B, Shields B, Smith GD, Hattersley AT, Di Rienzo A, Frayling TM

Institute of Biomedical and Clinical Science, Peninsula Medical School, Exeter, UK

Am J Hum Genet 2006;79:991–1001

Background: Fasting glucose is associated with future type 2 diabetes and ischemic heart disease risks and is tightly regulated despite considerable variations in food intake. During pregnancy, maternal fasting glucose concentration is an important determinant of offspring birth weight. The key determinant of fasting glucose is the enzyme glucokinase (GCK). Rare mutations of GCK cause fasting hyperglycemia and alter birth weight. The extent to which common variation of GCK explains normal variation of fasting glucose and birth weight is not known.

Methods: In order to comprehensively define the role of common genetic variation in GCK in the regulation of fasting glucose and birth weight, a tagSNP approach was studied in a total of 19,806 individuals from six population-based studies.

Results: Out of 22 tagSNPs, only the promoter variant rs1799884 was independently associated with fasting glucose levels. This association was seen at all ages in the normal population and exceeded genomewide levels of significance ($p = 10^{-9}$). rs3757840 was also highly significantly associated with fasting glucose ($p = 8 \times 10^{-7}$), but this was explained by linkage disequilibrium ($r^2 = 0.2$) with rs1799884. A maternal A allele at rs1799884 was associated with a 32-g (95% CI 11–53 g) increase in offspring birth weight ($p = 0.002$). Genetic variation influencing birth weight may have conferred a selective advantage in human populations. However, extensive population-genetics analyses revealed no evidence of recent positive natural selection on patterns of GCK variation.

Conclusions: This comprehensive analysis of common variation of the glucokinase gene shows that this is the first gene to be reproducibly associated with fasting glucose and fetal growth.

During pregnancy, maternal glucose concentration is an important determinant of offspring birth weight. Rare mutations in GCK cause the mild hyperglycemia phenotype of maturity-onset diabetes of the young type 2 (MODY2). The original study of the effects of these rare mutations on birth weight provided the original model for the fetal insulin hypothesis [14]; maternal mutations leading to gestational hypoglycemia increased birth weight, while fetal mutations leading to fetal insulin undersecretion lowered birth weight.

The authors have now comprehensively tested the hypothesis that common variants in GCK might contribute to the normal variation in glucose levels and birth weight. They performed new sequencing of the exons and surrounding gene region to identify the common tagSNPs, which were then genotyped

in a huge combined population of 19,806 individuals. Overall minor allele carriers of rs1799884 had an average 0.06 mmol/l increase in fasting glucose ($p = 10^{-9}$), and mothers with this variant had babies with 32 g higher birth weights. However, no independent effect of fetal genotype was seen. The frequency of the variant was lower in a Sub-Saharan African population than in non-African populations cut, however further analyses did not show a significant deficit in haplotype variation. The relevance of these findings remains to be fully established, however the study adds support the fruitfulness of studying common variants in genes where rare mutations are known to be deleterious.

A quantitative trait locus on chromosome 6q influences birth weight in two independent family studies

Arya R, Demerit E, Jenkinson CP, Goring HH, Puppala S, Farook V, Fowler S, Schneider J, Granato R, Resendez RG, Dyer TD, Cole SA, Almasy L, Comuzzie AG, Siervogel RM, Bradshaw B, DeFronzo RA, MacCluer J, Stern MP, Towne B, Blangero J, Duggirala R

Division of Clinical Epidemiology, University of Texas Health Science Center, San Antonio, Tex., USA

arya@uthscsa.edu

Hum Mol Genet 2006;15:1569–1579

Background: Low birth weight is an important cause of infant mortality and morbidity. Birth weight has also been shown to correlate inversely with adult obesity, type 2 diabetes and cardiovascular disease. However, little is known about the genetic factors influencing variation in birth weight and its association with diseases in later life.

Methods: A genome-wide linkage study was performed to identify genes that influence birth weight in Mexican-Americans using the data from the San Antonio Family Birth Weight Study participants ($n = 840$).

Results: Heritability of birth weight was estimated as $72.0 \pm 8.4\%$ ($p < 0.0001$) after adjusting for sex and gestation. Multipoint linkage analysis yielded the strongest evidence for linkage of birth weight ($LOD = 3.7$) between the markers D6S1053 and D6S1031 on chromosome 6q. This finding was replicated ($LOD = 2.3$) in an independent European-American population. Together, these findings provide substantial evidence ($LOD_{adj} = 4.3$) for a major locus influencing variation in birth weight. Positional candidate genes in this region include: chorionic gonadotropin, α -chain, collagen, type XIX, α_1 , and protein-tyrosine phosphatase, type 4A-1. In addition, potential evidence for linkage ($LOD > \text{or} = 1.2$) was found on chromosomes 1q, 2q, 3q, 4q, 9p, 19p and 19q with LODs ranging from 1.3 to 2.7.

Conclusions: This study found strong evidence for a major gene on chromosome 6q that influences variation in birth weight in both Mexican- and European-Americans.

This study complements the above two studies by searching for 'positional candidate genes' for birth weight in the mother's genotype. Not to be confused with the more recent Genomewide Association studies, this is an example of a Genomewide Linkage study where highly polymorphic markers are genotyped at a much greater interval of 10 cM density. These studies also differ by being based in family pedigrees, rather than unrelated individuals. The study found strong evidence for a common genetic variant on chromosome 6q and suggests that further studies of the candidate genes in this region may be fruitful. Interestingly, the candidate region on chromosome 6q contains the *hCG* gene, which stimulates progesterone secretion and acts as a growth factor that facilitates endometrial receptivity.

With the advent of Genomewide association studies, the further utility of Genomewide Linkage analyses is debated. The T2DM susceptibility gene *TCF7L2* was discovered as a positional candidate gene through the latter approach, however the current Genomewide association studies also quickly detected its very strong signal ($p \text{ value} = 1.5 \times 10^{-34}$ in Sladek et al. above). One clear further benefit of the family-based approach is the ability to estimate heritability. Indeed in the above study the heritability of birth weight as a continuous trait was very high at 72%, suggesting that further studies to identify the genetics of birth weight using either approach will be worthwhile.

Global variation in copy number in the human genome

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzalez JR, Gratacos M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodward C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME

The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK

steve@genet.sickkids.on.ca

Nature 2006;444:444–454

Background: Copy number variation (CNV) of DNA sequences is functionally significant but has yet to be fully ascertained. This report describes the construction of a first-generation CNV map of the human genome based on the HapMap population of 270 individuals from four populations with ancestry in Europe, Africa or Asia.

Methods: DNA from these individuals was screened for CNV using two complementary technologies: SNP genotyping arrays and clone-based comparative genomic hybridization.

Results: A total of 1,447 copy number variable regions (CNVRs) covering 360 megabases (12% of the genome) were identified in these populations. These CNVRs encompassed overlapping or adjacent gains or losses, and contained hundreds of genes, disease loci, functional elements and segmental duplications. Notably, the CNVRs encompassed more nucleotide content per genome than SNPs, underscoring the importance of CNV in genetic diversity and evolution.

Conclusions: These findings delineate linkage disequilibrium patterns for many CNVs, and reveal marked variation in copy number among populations. They also demonstrate the utility of this resource for genetic disease studies.

A comprehensive analysis of common copy-number variations in the human genome

Wong KK, deLeeuw RJ, Dosanjh NS, Kimm LR, Cheng Z, Horsman DE, MacAulay C, Ng RT, Brown CJ,

Eichler EE, Lam WL

Department of Cancer Genetics and Developmental Biology, University of British Columbia, Vancouver, B.C., Canada

kwong@bccrc.ca

Am J Hum Genet 2007;80:91–104

Background: Segmental copy-number variations (CNVs) in the human genome are associated with developmental disorders and susceptibility to diseases. More importantly, CNVs may represent a major genetic component of human phenotypic diversity.

Methods: A whole-genome array comparative genomic hybridization assay was performed in 95 individuals.

Results: 3,654 autosomal segmental CNVs were identified, 800 of which appeared at a frequency of at least 3%. Of these frequent CNVs, 77% are novel. In the 95 individuals analyzed, the two most diverse genomes differed by at least 9 Mb in size or varied by at least 266 loci in content. Approximately 68% of the 800 polymorphic regions overlap with genes, which may reflect human diversity in senses (smell, hearing, taste, and sight), rhesus phenotype, metabolism, and disease susceptibility. Intriguingly, 14 polymorphic regions harbor 21 of the known human microRNAs, raising the possibility of the contribution of microRNAs to phenotypic diversity in humans.

Conclusions: This in-depth survey of CNVs across the human genome provides a valuable baseline for studies involving human genetics.

The first map of CNV in the human genome has been published. Copy CNVs are defined as DNA segments >1 kb that are present in variable copy number. Such variation can arise through various ways including: deletions, insertions, duplications and complex multisite variants such as gains or losses of homologous sequences at multiple sites in the genome. Several of these CNV-associated genes are related to the senses, including a group of olfactory receptor genes, multiple taste-receptor genes, and several genes related to sight or hearing.

It is known that rare genomic rearrangements are a cause of human disease, for example MODY5 [15], while other large structural variations have been described to occur commonly and may even confer some survival advantage or positive selection [16]. The current studies now help describe the full extent of CNVs in the human genome. The estimated average total length of CNV regions per genome was more than 20 million base pairs, which is far greater than the total variation due to SNPs alone. Furthermore, 285 out of the 1,961 (14.5%) genes in the OMIM database of genetic diseases overlapped with CNVs, indicating numerous possible disease associations.

Concepts revised: the insulin gene VNTR

Insulin gene variable number of tandem repeats (INS VNTR) genotype and metabolic syndrome in childhood obesity

Santoro N, Cirillo G, Amato A, Luongo C, Raimondo P, D'Aniello A, Perrone L, Miraglia del Giudice E

Department of Pediatrics, F. Federe Seconda Università di Napoli, Napoli, Italy

emanuele.miraglia@unina2.it

J Clin Endocrinol Metab 2006;91:4641–4644

Background: The insulin gene variable number of tandem repeats (INS VNTR) polymorphism has been associated with insulin levels in obese children. Hyperinsulinemia is a pivotal factor in the development of metabolic syndrome, an emerging complication in childhood obesity. This study tested the associations between INS VNTR genotype and the metabolic syndrome in juvenile-onset obesity.

Methods: A marker for INS VNTR allele class was genotyped in 320 obese children (152 girls, mean age 11.2 ± 2.3 years, mean z-score BMI 3.6 ± 1.1). All of them underwent a standard oral glucose tolerance test (OGTT); baseline measurements included blood pressure, plasma lipid profile and fasting insulin levels. Whole-body insulin sensitivity and the insulinogenic index were calculated from the OGTT data.

Results: The prevalence of metabolic syndrome was 39%. There was no difference in INS VNTR genotype distribution between obese and 200 lean, age- and sex-matched children ($p = 0.7$). However, among obese children, those who were homozygous for the class I allele showed higher insulin levels and insulinogenic index but lower whole-body insulin sensitivity than other obese children. Accordingly, among obese children the odds ratio for developing the metabolic syndrome was 2.5-fold higher (95% CI 1.5–3.9, $p = 0.006$) in obese children carrying the I/I genotype, compared to other obese children.

Conclusions: Class I/I homozygotes for the INS VNTR insulin promoter predisposes obese children to develop the metabolic syndrome.

The INS VNTR locus does not associate with smallness for gestational age (SGA) but interacts with SGA to increase insulin resistance in young adults

Vu-Hong TA, Durand E, Deghmoon S, Boutin P, Meyre D, Chevenne D, Czernichow P, Froguel P, Levy-Marchal C

Institut National de la Santé et de la Recherche Médicale, Unit 690, Robert Debré Hôpital, Paris, France

J Clin Endocrinol Metab 2006;91:2437–2440

Background: Both adverse intrauterine events and genetic background have been suggested to confer insulin resistance in individuals born small for gestational age (SGA). Insulin is a candidate gene that could potentially influence both fetal growth and glucose metabolism. However, the potential effect of the insulin gene VNTR (INS) on birth weight is yet controversial.

Methods: This study tested the contribution of the INS VNTR locus on birth weight and on the metabolic profile of young adults born SGA (mean age 22 years). Two groups were selected based their birth weights: SGA (birth weight <10th percentile, $n = 735$), and appropriate for gestational age (AGA; birth weight between 25th and 75th percentiles, $n = 886$). All individuals were genotyped for the rs689 A/T SNP, which is in complete linkage disequilibrium with the INS VNTR classes I and III, respectively.

Results: Class I allele frequencies were similar in the two groups (70% in AGA; 72% in SGA, $p = 0.42$), and the INS VNTR was not associated with mean birth weight in either SGA ($p = 0.99$) or AGA groups ($p = 0.18$). Although the INS VNTR genotype did not associate with indices of insulin resistance

in the AGA group, in the SGA group, INS VNTR class III allele was associated with higher insulin resistance (quantitative insulin sensitivity check index = 0.38 vs. 0.39, $p = 0.05$). Furthermore, there was a significant interaction between the SGA/AGA status and INS VNTR genotype on insulin resistance indices ($p = 0.01$) in a multivariate analysis.

Conclusions: The INS VNTR locus was not directly associated with SGA or birth weight in this French population. However, these data support an interaction between fetal growth restriction and INS VNTR genotype, which influences insulin resistance in young adults born SGA.

The INS VNTR polymorphism has occupied the literature for many years. It is a functional polymorphism that regulates insulin and insulin-like growth factor-2 transcription and class I alleles have been consistently shown to predispose to type 1 diabetes risk. In contrast, studies for T2DM and obesity risks in large adult populations have been inconsistent and are overall unconvincing. However, in addition to the above two recent studies, several childhood cohorts have also repeatedly shown that the genotype effects on insulin levels and adiposity in children are highly modified by birth weight, postnatal weight gain and childhood obesity [17–19].

These findings in children either represent multiple false-positive reports in childhood populations that are often much smaller than their adult population counterparts, or there are important interactive effects of early growth on the expression of certain genotypes, which could have important relevance not only to childhood growth and development, but also for adult disease risks. Very large childhood cohort studies, or multicenter collaborations such as those that have been set up between adult studies [20], are urgently needed in pediatric genetic epidemiology.

Food for thought: the human epigenome project

DNA methylation profiling of human chromosomes 6, 20 and 22

Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, Burton J, Cox TV, Davies R, Down TA, Haefliger C, Horton R, Howe K, Jackson DK, Kunde J, Koenig C, Liddle J, Niblett D, Otto T, Pettett R, Seemann S, Thompson C, West T, Rogers J, Olek A, Berlin K, Beck S
Epigenomics AG, Berlin, Germany
florian.eckhardt@epigenomics.com
Nat Genet 2006;38:1378–1385

Background: DNA methylation is the most stable type of epigenetic modification modulating the transcriptional plasticity of mammalian genomes.

Methods: This study reports the high-resolution methylation profiles of human chromosomes 6, 20 and 22, using bisulfite DNA sequencing. The data provide a resource of around 1.9 million CpG methylation values derived from 43 samples from 12 different tissues.

Results: Analysis of six annotation categories showed that evolutionarily conserved regions are the predominant sites for differential DNA methylation and that a core region surrounding the transcriptional start site is an informative surrogate for promoter methylation. 17% of the 873 analyzed genes were differentially methylated in their 5' UTRs, and around one-third of the differentially methylated 5' UTRs are inversely correlated with levels of transcription. Despite the fact that this study included factors reported to affect DNA methylation such as sex and age, no significant attributable effects were found.

Conclusions: These data suggest DNA methylation to be ontogenetically more stable than previously thought.

Completion of the Human Genome Project gave us the blueprint of the genetic sequence. However, many processes that regulate gene expression are governed by additional layers of epigenetic information that are not directly encoded in the DNA sequence itself but take the form of chemical modifications such as DNA methylation. The Human Epigenome Project aims to identify, catalogue and interpret genome-wide DNA methylation profiles of all human genes in all major tissues.

The current study reports the results of that project for chromosomes 6, 20 and 22 based on 43 samples derived from 12 different healthy tissues. Almost 25% of all amplicons showed tissue-specific methylation. Most genetic epidemiology studies are based on DNA samples derived from peripheral lymphocytes or buccal cells. Future studies should therefore consider the possibility of collecting

other tissue samples. Recent developments in epigenetics are also described in other chapters this year [21].

References

1. Heude B, Ong KK: Population genetics and pharmacogenetics; in Carel J-C, Hochberg Z (eds): Yearbook of Pediatric Endocrinology 2006. Basel, Karger, 2006, pp 117–130.
2. Chiarelli F: Type 1 diabetes: clinical and experimental; in Carel J-C, Hochberg Z (eds): Yearbook of Pediatric Endocrinology 2007. Basel, Karger, 2007.
3. Simmons RA: Developmental origins of β -Cell failure in type 2 diabetes: the role of epigenetic mechanisms. *Pediatr Res* 2007 Mar 15 [Epub ahead of print].
4. Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, et al: Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* 2006;38:320–323.
5. Loos RJ, Franks PW, Francis RW, Barroso I, Gribble FM, Savage DB, et al: TCF7L2 polymorphisms modulate proinsulin levels and β -cell function in a British Europid population. *Diabetes* 2007 Apr 6 [Epub ahead of print].
6. Pearson ER, Flechtner I, Njolstad PR, Malecki MT, Flanagan SE, Larkin B, et al: Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations. *N Engl J Med* 2006;355:467–477.
7. Loos RJ, Barroso I, O'Rahilly S, Wareham NJ: Comment on 'A common genetic variant is associated with adult and childhood obesity'. *Science* 2007;315:187.
8. Dina C, Meyre D, Samson C, Tichet J, Marre M, Jouret B, et al: Comment on 'A common genetic variant is associated with adult and childhood obesity'. *Science* 2007;12;315:187.
9. Rosskopf D, Bornhorst A, Rimmbach C, Schwahn C, Kayser A, Kruger A, et al: Comment on 'A common genetic variant is associated with adult and childhood obesity'. *Science* 2007;315:187.
10. Weedon MN, Shields B, Hitman G, Walker M, McCarthy MI, Hattersley AT, et al: No evidence of association of ENPP1 variants with type 2 diabetes or obesity in a study of 8,089 UK Caucasians. *Diabetes* 2006;55:3175–3179.
11. Lyon HN, Florez JC, Bersaglieri T, Saxena R, Winckler W, Almgren P, et al: Common variants in the ENPP1 gene are not reproducibly associated with diabetes or obesity. *Diabetes* 2006;55:3180–3184.
12. Prudente S, Trischitta V: Editorial: The pleiotropic effect of the ENPP1 (PC-1) gene on insulin resistance, obesity, and type 2 diabetes. *J Clin Endocrinol Metab* 2006;91:4767–4768.
13. Meyre D, Bouatia-Naji N, Tounian A, Samson C, Lecoœur C, Vatin V, et al: Variants of ENPP1 are associated with childhood and adult obesity and increase the risk of glucose intolerance and type 2 diabetes. *Nat Genet* 2005;37:863–867.
14. Hattersley AT, Tooke JE: The fetal insulin hypothesis: an alternative explanation of the association of low birthweight with diabetes and vascular disease. *Lancet* 1999;353:1789–1792.
15. Bellanne-Chantelot C, Clauiin S, Chauveau D, Collin P, Daumont M, Douillard C, et al: Large genomic rearrangements in the hepatocyte nuclear factor-1 β (TCF2) gene are the most frequent cause of maturity-onset diabetes of the young type 5. *Diabetes* 2005;54:3126–3132.
16. Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J, et al: A common inversion under selection in Europeans. *Nat Genet* 2005;37:129–137.
17. Heude B, Petry CJ, Pembrey M, Dunger DB, Ong KK: The insulin gene variable number of tandem repeat: associations and interactions with childhood body fat mass and insulin secretion in normal children. *J Clin Endocrinol Metab* 2006;91:2770–2775.
18. Le Stunff C, Fallin D, Schork NJ, Bougneres P: The insulin gene VNTR is associated with fasting insulin levels and development of juvenile obesity. *Nat Genet* 2000;26:444–446.
19. Dos Santos C, Fallin D, Le Stunff C, LeFur S, Bougneres P: INS VNTR is a QTL for the insulin response to oral glucose in obese children. *Physiol Genomics* 2004;16:309–313.
20. Ioannidis JP, Gwinn M, Little J, Higgins JP, Bernstein JL, Boffetta P, et al: A road map for efficient and reliable human genome epidemiology. *Nat Genet* 2006;38:3–5.
21. Hochberg Z, Carel J-C: The year in science and medicine; in Carel J-C, Hochberg Z (eds): Yearbook of Pediatric Endocrinology 2007. Basel, Karger, 2007.