ARTICLE

Associations of *IGF2* *ApaI* RFLP and *INS* VNTR class I allele size with obesity

Sandra D O'Dell¹, Sarah R Bujac², George J Miller² and Ian NM Day¹

¹Human Genetics Research Division, Southampton University School of Medicine

²MRC Epidemiology and Medical Care Unit, Wolfson Institute of Preventive Medicine, St Bartholomew's and The Royal London School of Medicine and Dentistry, Queen Mary and Westfield College, University of London, UK

Body mass index (BMI) is an established epidemiological predictor of coronary disease, diabetes and hypertension. In a previous study of 2560 healthy British Caucasoid males aged 50–61 years (Northwick Park Heart Study II; NPHSII), we showed that *IGF2* *ApaI* AA homozygotes display a mean body weight 3.3 kg lower than GG homozygotes ($P = 0.0002$) independent of height. Two RFLPs in the insulin (*INS*) gene, +1127/*PstI* shown previously and –23/*HphI* in this study, both of which are in strong linkage disequilibrium with class I/III alleles of the *INS* 5' variable number tandem repeat (VNTR), are not associated with weight or BMI. The *IGF2* *ApaI* polymorphism therefore appears to mark an effect independent of *INS* VNTR class I vs class III. We now show by regression that there is a positive correlation of BMI with *INS* VNTR class I allele size, with an average 0.33% (95% CI = 0.13%, 0.50%) increase in BMI per extra tandem repeat ($P < 0.0001$) representing variation of 4.8% over the allele size range. However, an alternative interpretation is of 'step' rather than 'slope', the small class I subclass allele group (mode 669 bp) being lighter than the large subclass group (mode 814 bp). This small effect would not be evident as an association between *INS* VNTR class I/III genotype and BMI. The *IGF2* *ApaI* association and *INS* VNTR class I subclass regression association account for at least 1.1% of population BMI variance. Neither, both, or a third site may be aetiological.

Keywords: association; BMI; *IGF2*; *INS* VNTR; population obesity

Introduction

Adult obesity is an important determinant of coronary disease risk, type II diabetes, hyperlipidaemia, hypertension and other disease susceptibility in later life.¹ In epidemiological studies, the Quetelet body mass index (BMI), defined as weight (kg)/height (m²) has been widely used as a standard economically obtained

surrogate for obesity. Used carefully, (not between genders nor across ethnic groups) it permits broad comparisons of fat mass. 40% excess weight confers a 5–7 fold increased risk, but there is a continuous and steepening gradient of increased risk for BMI > 30 kg/m². Central (abdominal) obesity is thought to be the most distinctive mark of a 'metabolic syndrome' which is characterised by a clustering of many insulin-resistant cardiovascular risk factors such as hypertension, hypertriglyceridaemia, low HDL-cholesterol, abnormal glucose metabolism and hyperinsulinaemia.²

Evidence for genetic control of human body weight and composition is well established from adoption, twin and family studies (eg Stunkard *et al*³). Human obesity

Correspondence: Sandra D O'Dell, Human Genetics Research Division, Southampton University School of Medicine, Duthie Building (MP808), Southampton SO17 6YD, UK. Tel: +44 (0)1703 794141; Fax: +44 (0)1703 794264; E-mail: S.D.O'Dell@soton.ac.uk
Received 16 March 1999; revised 27 May 1999; accepted 18 June 1999

is a polygenic disorder, with only a small variance in phenotype accounted for by individual genes. Despite this, the proportion of people affected due to such genes may be large because they are quite frequent in the population, which makes them of public health significance. These genes seem to operate through modifying susceptibility; the occurrence of the gene increases the risk of developing a characteristic but is not essential for its expression, nor is it itself sufficient to explain the development of the disease. Differences in genetic susceptibility within a population may therefore determine which individuals are most likely to become obese in any given set of environmental circumstances. No human obesity gene has yet been characterised, although a number of candidate genes have been identified which are associated with the obese phenotype.^{4,5}

The polygenic basis of obesity traits may best be determined by association, but this demands candidate selection based on linkage, known function or animal studies. The insulin-like growth factors (IGF-I and IGF-II) play a key role in mammalian growth, influencing foetal cell division and differentiation, and possibly metabolic regulation.⁶ In the human, unlike in the mouse, *IGF2* gene expression continues post-natally, but the significance of this is unknown. However, based on the premise that factors essential for foetal development may also have a role in later life, together with an established association between IGF-I levels and obesity,⁷ we previously investigated the relationship between an *ApaI* polymorphic marker in the *IGF2* gene⁸ and BMI in over 2500 middle-aged males in the Northwick Park Heart Study II (NPHSII).⁹ *IGF2 ApaI* rare AA homozygotes showed a mean body weight 3.3 kg lower than *ApaI* common GG homozygotes with heterozygotes intermediate ($P = 0.0002$). We believe this to be the first large cohort study of adults to establish a positive association between a common variation of a genomic region and obesity in the general population. Continuing the theme of population-scale association study, we have extended our initial observations and establish here novel relationships between markers in the *tyrosine hydroxylase-insulin-IGF2* gene region of chromosome 11p15 and population obesity.

The *INS* promoter variable number tandem repeat (VNTR) class I/III genotype has been shown by others to associate with type I diabetes,¹⁰ obesity,¹¹ polycystic ovary syndrome obesity¹² and growth in early life,¹³ and at least for type I diabetes is aetiological.¹⁰ The *INS* VNTR locus is 596 bp 5' to the *INS* start site, and arises

from tandem repetition of a 14–15 bp oligonucleotide consensus sequence.¹⁴ The VNTR has been divided into three distinct classes I, II and III according to the number of repeats of the oligonucleotide, measured by Southern blotting.¹⁵ Extreme rarity of class II alleles in Caucasoid populations effectively reduces the VNTR to a biallelic site. Since there is 99.6% concordance between the $-23/HphI$ alleles and the *INS* VNTR alleles in people of European descent,¹⁶ $-23/HphI$ genotype implies *INS* VNTR class I/III genotype.

Here we have examined *INS* + 1227/*PstI* RFLP, *INS* $-23/HphI$ RFLP (class I/III VNTR) and *INS* VNTR class I size subclass by cohort study of adult weight and BMI. Polymorphisms are designated by their position with respect to the first nucleotide of the initiating codon of the *INS* gene + 1, following Julier *et al.*¹⁷ We aimed to find out first whether an association between *INS* VNTR class I/III alleles and BMI existed in the NPHSII cohort and secondly to examine the possibility that *INS* VNTR class I allele size associates independently with weight or BMI.

Materials and Methods

Study Samples

The Northwick Park Heart Study II (NPHSII) sample comprises healthy adult males aged 50–61 years,¹⁸ with the following exclusions: a history of unstable angina or MI, regular anticoagulant medication, cerebrovascular disease, malignancy, or other conditions precluding informed consent. An unselected group of 2331 from nine widespread UK general practices was studied here.

$-23/HphI/INS$ ARMS PCR and Genotyping

The $-23/HphI$ polymorphic site is situated at nucleotide 4462 on GenBank sequence L15440. The rare allele (non-cutting) sequence TCTCC is given in GenBank; the transition T > A creates the common allele (cutting) sequence TCACC. Specific A alleles were PCR-amplified using forward primer HPHA1: 5'-CTTCAGCCTGCCTCAGCCCTGCCTGTGA-3', and specific T alleles were PCR-amplified using forward primer HPHA2: 5'-CTTCAGCCTGCCTCAGCCCTGCC-TGTTT-3', where italicised bases are destabilising mismatches with target, and bases in bold type match the specific allele. The common reverse primer was HPHC: 5'-GGCGGGTCTTGGGTGTGTAGAAGAAGCC-3', giving an ARMS product of size 216 bp. Primers for the control PCR reaction amplified a 360 bp region in the human α_1 -antitrypsin (*AAT*) gene as follows:¹⁹ forward primer AATF: 5'-CCCCACCTTCCCCTCTCTCCAGGCAAATGGG-3'; reverse primer AATR: 5'-GGCCTCAGTCCCAACATGGCTAAGAGGGTG-3'. Primers were obtained from Genosys (Cambridge, UK). PCR amplification (1 cycle at 94°C for 6 min, 35 cycles at 94°C for 1 min, 60°C for 1 min, 72°C for 1 min, 1 cycle at 72°C for 10 min) was in 96-well Omniplates (Hybaid, Teddington, UK), each 10 μ l reaction containing approximately 40 ng of DNA, 1.2 mM MgCl₂,

50 mM KCl, 10 mM Tris pH 8.3, 0.01% gelatin, 200 μ M each dNTP, 8 pmol each primer and 0.2 U of *Taq* polymerase (Gibco, Paisley, UK).

PCR products were electrophoresed on a 5% polyacrylamide microplate array diagonal gel electrophoresis (MADGE) gel,²⁰ in TBE buffer at 150V for 20 min. Gels were post-stained in ethidium bromide and examined using a FluorImager® 595 fluorescent scanning system (Molecular Dynamics, Sunnyvale, CA, USA). The images were viewed using Image Quant™ fragment analysis software (Molecular Dynamics) and printed on a Model HP-4M Laserprinter (Hewlett Packard, Palo Alto, CA, USA).

INS 5' VNTR Class I PCR and Allele Size Determination

Class I/III status was confirmed with reference to the -23/*HphI* genotype. Specific *INS* class I VNTR alleles differing by a repeat unit (14–15 bp) were PCR-amplified according to the method of Bennett *et al*,¹⁰ using forward primer 5'-TCAGGCTGGACCTCCAGGTGCCTGTTCTGT-3', reverse primer 5'-TCGTCAGCACCTCTTCCTCAGGACCA-GC-3'. Primers were obtained from Genosys (Cambridge, UK). PCR amplification (1 cycle at 94°C for 10 min, and then 94°C for 1 min and 72°C for 5 min for 30 cycles) was in 96-well Omniplates (Hybaid, Teddington, UK), each 20 μ l reaction containing approx 80 ng of DNA, 6.7 mM MgCl₂, 2 μ l 10 $\times$ ammonium based reaction buffer (Bioline, London, UK), 10% glycerol, 1.5 mM each dNTP, 10 pmol each primer and 2.5 U of *Taq* polymerase (Gibco, UK). Each PCR product was reduced in length by 450 bp by digesting 8 μ l of each reaction with 2 U of *NcoI* (New England Biolabs, Beverly, Md, USA) in a total volume of 10 μ l for 2 h at 37°C according to the supplier's instructions.

NcoI digest samples were diluted 1:9 in water. The gel loading buffer contained formamide dye mix (98% v/v formamide, 50 mM EDTA, and xylene cyanol) and ϕ X 174 DNA-Hae III digest ladder (New England Biolabs) diluted 1:499 in water and mixed in the ratio of 2:1 v/v dye:ladder. A volume of 5 μ l of 1/10 dilution of *NcoI* digest was mixed with 3 μ l loading buffer, and 5 μ l of this was loaded on to the gel. We used the 96-well MADGE polyacrylamide gel system²⁰ with electrophoretic conditions which enable resolution of the 15 common *INS* VNTR alleles over a short (less than 24 mm) track length. Using FluorImager 595® technology, the internal size markers, and sensitive fluorescent post-staining, we have produced resolution and accurate sizing of bands representing differences of one tandem repeat.

Statistical Analysis

Hardy-Weinberg equilibrium and significance of difference between *INS* 5' VNTR class I allele size frequency distribution in different samples were examined by χ^2 test. Comparison of continuous variables in the NPHSII sample with respect to *IGF2* and *INS* RFLP genotypes was by one-way ANOVA. Association between mean ($\log_e$ BMI) and *INS* VNTR class I allele size was examined by backward stepwise regression. Table 2 displays how variables were coded for each *INS* VNTR genotype and describes the regression modelling used. The contribution of VNTR class I alleles to the variation

in the BMI was estimated using the value of R^2 from the final regression model.

Results

Tests of association between the *INS* -23/*HphI* genotype weight and body mass index (BMI) in the NPHSII sample are shown in Table 1, together with results for *INS* +1227/*PstI* and *IGF2* *ApaI* obtained previously.⁹ There was no significant difference in these parameters

Table 1 Northwick Park Heart Study II (NPHSII): Mean weight and body mass index (BMI) with respect to genotype *IGF2* *ApaI* genotype ($n=2560$). *INS*-23/*HphI* genotype ($n=2331$) and *INS*+1227/*PstI* genotype ($n=563$) are shown. Results are presented as geometric means $\pm$ sd in parentheses

<i>IGF2</i> <i>ApaI</i> genotype	mean weight (kg)	mean BMI (kg/m ²)
1/1 (<i>GG</i>) ($n=1362$)	80.6 (11.1)	26.4 (3.3)
1/2 (<i>GA</i>) ($n=1019$)	79.5 (11.5)	26.2 (3.4)
2/2 (<i>AA</i>) ($n=179$)	77.1 (11.0)	25.3 (3.2)
ANOVA	$P=0.0002$	$P=0.0004$
<i>INS</i> -23/ <i>HphI</i> genotype	mean weight (kg)	mean BMI (kg/m ²)
1/1 ($n=1244$)	79.8 (11.3)	26.1 (3.4)
1/2 ($n=888$)	79.7 (11.1)	26.2 (3.3)
2/2 ($n=199$)	80.2 (11.8)	26.6 (3.5)
ANOVA	$P=0.84$	$P=0.14$
<i>INS</i> +1227/ <i>PstI</i> genotype	mean weight (kg)	mean BMI (kg/m ²)
1/1 ($n=366$)	80.1 (10.9)	26.1 (3.2)
1/2 ($n=178$)	79.6 (11.2)	26.3 (3.2)
2/2 ($n=19$)	81.8 (10.3)	26.3 (3.5)
ANOVA	$P=0.86$	$P=0.64$

1=common allele, 2=rare allele.

with respect to $-23/HphI$ genotype, and by inference with *INS* VNTR class I/III genotype, consistent with the absence of linkage disequilibrium found between *IGF2* *ApaI* and *INS* $-23/HphI$ sites.

Lengths of *INS* VNTR class I alleles form a quasi-continuous distribution with three distinct modes,²¹ a finding corroborated in other studies.^{10,22} In the NPHSII cohort, the *INS* 5' VNTR class I allele size frequency distribution conformed to a pattern characteristically similar to that observed by Bennett *et al*¹⁰ in parents of diabetics, and therefore the same nomenclature of subclasses is used here.

The association between *INS* VNTR class I allele size and BMI was investigated using backward stepwise regression of $Y = (\log_e \text{BMI})$ on X (Table 2). The mean ($\log_e \text{BMI}$) in the non-carriers of the class I allele, ie class III/III homozygotes is 3.282 units or a mean BMI of 26.64 kg/m². For class I/III heterozygotes, the single class I allele lowers mean BMI compared with the non-carriers, to an extent determined by the size of the class I allele ($P < 0.0001$). The smaller the allele size, the greater the difference in mean BMI between heterozygotes and non-carriers. For homozygotes, the effect of possessing two class I alleles is also to lower mean BMI, dependent on allele size, but possession of one class I allele is almost identical to possession of two class I alleles of that same size or averaging the same size. The model provides evidence that the class I effect is dominant rather than co-dominant ($P < 0.0001$) and shows that this effect on BMI depends on the size of the class I alleles involved ($P = 0.0002$).

Figure 1 shows the relationship between mean ($\log_e \text{BMI}$) and class I allele size. All class I alleles are considered, such that same size class I homozygotes contribute twice to the mean ($\log_e \text{BMI}$) determination. The gradient of the line of best fit is 0.000234 $\log_e \text{BMI}$ units per bp. Hence a 14 bp tandem repeat increase in allele size corresponds to a 0.00327 unit increase in $\log_e \text{BMI}$ or a 0.33% (95% confidence interval CI = 0.13%, 0.50%) increase in BMI. Over the whole range of class I allele sizes, (641 to 843 bp), this corresponds to a mean increase of 0.0472 units in $\log_e \text{BMI}$. With mean estimated BMI of 25.32 kg/m² at 641 bp and 26.55 kg/m² at 843 bp this corresponds to an increase in absolute BMI of 4.83% (95% CI = 2.29%, 7.44%) over the range of class I allele sizes. The sizes of standard errors on the mean reflect the relative frequencies of the alleles. The distribution of points may be alternatively construed as bimodal, and the implications of this are considered below.

Table 2 Regression modelling

Y	mean ($\log_e \text{BMI}$)
X ₁	additive class effect, code 2 for class I/I, 1 for I/III, 0 for III/III
X ₂	dominance of class I, code 1 for class I/I, 0 for others
X ₃	allele size effect, code 0 for III/III, using single allele size for I/III and average of a pair of class I alleles for I/I
X ₄	X ₂ × X ₃ , dominance for class I allele size
X ₅	dominant class I allele, code 1 for I/I and I/III, 0 for III/III, replacing X ₁ and X ₂
<i>Tests:</i>	
Regress	Y or X stepwise
Hypothesis ₀	no effect
Hypothesis ₁	X ₁ : additive class effect: positive signal implies difference between class I and class III alleles
Hypothesis ₂	X ₁ and X ₃ effects: positive signal implies class size additivity effect
Hypothesis ₃	X ₁ , X ₃ and X ₂ or X ₄ : positive signal implies dominance of allele size effect

Treating X₁ as continuous assumes codominance of class I and III alleles; X₂ tests class I dominance; X₃ tests linear effect of class I allele size and X₄ tests dominance of class I allele size. Starting with all four explanatory variables in the model, backward stepwise regression selects the model with just X₁, X₂ and X₃, eliminating X₄ at either $P=0.1$ or 0.05 for rejection. For class I/III heterozygotes, possession of one class I allele lowers mean BMI compared with the non-carriers. The smaller the allele size, the greater the difference in mean BMI between heterozygotes and non-carriers ($P < 0.0001$ for X₂). From the regression model, the estimated mean ($\log_e \text{BMI}$) for carriers of one class I allele is: $3.282539 + (-0.201162) + 0.000234 \times \text{allele size} = 3.081377 + 0.000234 \times \text{allele size}$. For homozygotes, two class I alleles also lowers mean BMI dependent on allele size. Estimated mean ($\log_e \text{BMI}$) for carriers of two class I alleles is: $3.282539 + 2(-0.201162) + 0.201606 + 0.000234 \times \text{average allele size} = 3.081821 + 0.000234 \times \text{average allele size}$. One class I allele has almost identical effect to two class I alleles of that same size or averaging the same size. Comparing like alleles, X₂ effectively cancels out the additive allele effect constrained by X₁, ie evidence that the class I effect is dominant ($P < 0.0001$). This allows the replacement of X₁ and X₂ with just one variable, X₅. Replacing X₁ and X₂ with X₅ in the model provides evidence for a dominant effect of class I allele size on BMI ($P < 0.0001$). The significance of X₃ in the model ($P=0.0002$) shows that this effect depends on the size of the class I alleles involved.

Discussion

We have shown that there is no association between *INS* VNTR class I/III alleles and BMI in a large sample of middle-aged men. The *IGF2* *ApaI* polymorphism which we investigated previously⁹ is not in linkage disequilibrium with *INS* VNTR class I/class III alleles, and therefore must represent an independent association with BMI. However, in this study we have shown

Figure 1 Association between INS 5' VNTR Class I allele size and mean ($\log_e$ BMI). Bars represent standard errors of mean ($\log_e$ BMI). Equivalent BMI values are indicated on a separate axis. The number of alleles of each size is indicated in parenthesis. In contrast to the regression analysis in which each subject was included once (see Table 2), all VNTR class I alleles were used to estimate the class-specific mean ($\log_e$ BMI) in this figure. Thus subjects with two class I alleles contributed once to two classes when class I heterozygote (different size alleles), and twice to one class when class I homozygote (same size alleles). Therefore the mean values presented here are not independent.

an association between INS VNTR class I allele size and BMI which parallels that which we previously found with *IGF2 ApaI*. The absence of association between INS VNTR class I/III genotype and BMI in the presence of association between INS VNTR class I allele size and BMI is compatible with the expectation that VNTR class I/III genotype would not reveal an association with BMI which depends on class I allele size associated with 5% variation. A positive signal could be evident in a much larger study, but our investigation, based on over 2500 individuals demonstrates a within-class I effect more striking than one between class I and III.

The association between INS VNTR class I allele size and BMI is evident in each of the three *IGF2 ApaI* genotype groups. Post *hoc* examination of the combinatoric subgroups does not identify any single haplotype with distinctive weight effect. However, it is apparent that a haplotype containing tyrosine hydroxylase (*TH*) *HUMTH01* microsatellite allele 9, which is in strong linkage disequilibrium with the INS VNTR class I small subclass alleles²¹ and with which *IGF2 ApaI* A alleles are preferentially associated²² may be important. This raises an alternative explanation of the data, which is that the 'noisy' regression line observed is instead a 'noisy' step (ie bimodality), and that this haplotype contains an aetiological site for low weight. The simplest hypothesis would then be an as yet

unidentified single nucleotide polymorphism accounting for the weight variance observed. Many have considered the VNTR aetiological of various phenotypes,^{10–13} but recently others, including the original groups, have come to question the evidence that the VNTR is an aetiological site.^{23,24} The uncertainty of a conclusion of linear regression vs bimodal step obliges examination of a wider genomic region. We have analysed the distribution of *IGF2 ApaI* alleles with respect to INS VNTR class I alleles, and found an apparent clustering of *IGF2 ApaI* AA genotypes with the VNTR class I small size subclass (not shown). Taken as a whole our results suggest that there is no partitioning of *IGF2 ApaI* alleles with respect to INS VNTR class I/III genotype, only with respect to class I allele size. Recently Sharma *et al*²⁵ have reported a positive association between *HUMTH01* allele 9 and low blood pressure in a case-control association study of hypertension. Although they did not study obesity phenotypes this would be consistent with our finding of *IGF2 ApaI* A allele and INS VNTR class I small subclass association with low BMI, since central obesity is distinctive of a 'metabolic syndrome' characterised by several cardiovascular risk factors which include hypertension.²

Several examples exist wherein the number of repeat units within a gene influence function. Triplet repeat length correlates with disease severity and explains anticipation in several loci causing severe neurological disorders²⁶ and length of a triplet repeat in the androgen receptor gene also influences prostatic cancer risk and age of onset.²⁷ Gavish *et al*²⁸ showed that the variable number of kringle IV-like domains encoded by the gene encoding apo (a) is the main factor determining plasma concentration of lipoprotein (a), and with it susceptibility to cardiovascular disease. The correlation of INS VNTR class I repeat number with BMI represents the first apparent quantitative identification of a correlation of a human minisatellite length with a quantitative phenotype. Previous studies have mainly been dichotomous classifications, disease vs control and class I vs class III. *Pur-1* transcription factor is known to bind the repeat,²⁹ but the possible mechanism of the length correlation remains obscure. The simplest hypothesis based on the known biological effects is that lengthening VNTR increases *Pur-1* driven INS transcription, increasing insulin levels and therefore promoting lipogenesis. A comparable hypothesis has been invoked for the differential effects of class I vs class III seen in the qualitative state of type 1 diabetes, although

no consistent quantitative effects dependent on exact class I size were observed in IDDM.

Paquette *et al*³⁰ found higher steady state *IGF2* mRNA levels were associated with *INS* VNTR class I than class III alleles in human placenta and human liver HepG2 cell line, and proposed that the VNTR may act as a long range control element affecting the expression of both *INS* and *IGF2*. However, Vafiadis *et al*³¹ found the *IGF2* *Apal* G allele (associated with increased BMI in our studies) was associated with significantly higher *IGF2* mRNA levels than the A allele, but unlike Paquette *et al* in placenta and liver cells, found no difference in expression of leukocyte *IGF2* with respect to *INS* VNTR class I/III genotype, as they had previously established in thymus and pancreas.³²

Since our results show no association between *INS* VNTR I vs III classification and BMI, but substantial association with class I small subclass alleles (in coupling with *IGF2* *Apal* A alleles) versus large subclass alleles (in coupling with *IGF2* *Apal* G alleles), a precise aetiological chain between *IGF2* and *INS* mRNA and hormone expression and protection from obesity (the *IGF2* *Apal* A/*INS* VNTR class I small subclass haplotype) cannot yet be inferred, although its genetic susceptibility is already measurable. The most plausible hypothesis may be that there is increased average transcription and IGF-II secretion from liver. At least 1.1% of total population BMI variance is accounted for by the present study. As high throughput genome analysis programs gather pace, in conjunction with the genetic study of cohorts, other comparable genomic regions will eventually be identified. Fifty such genetic markers acting in conjunction could account for much of individual and hence population weight variance. This would be a genotype analogue of the use of cholesterol assays in cardiovascular multiple risk assessment and the selection of subjects for risk reduction.³³ A molecular classification of obesity susceptibility pathways (metabolism, regulation, central nervous system etc) may define new targets for therapy and subgroups of the population most likely to be responsive to tailored approaches in preventive medicine.

Acknowledgements

Professor Newton E Morton is thanked for helpful discussions on the regression model. This work was funded by the British Heart Foundation (Grant No PG/98158), the UK Medical Research Council, the US National Institutes of Health (Grant No NHLBI33014) and DuPont Pharma,

Wilmington, USA. The Wessex Medical Trust is thanked for support. INMD is a Lister Institute professor.

References

- 1 Lamon-Fava S, Wilson PWF, Schaefer EJ: Impact of body mass index on coronary heart disease risk factors in men and women: the Framingham offspring study. *Arterioscler Thromb Vasc Biol* 1996; **16**: 1509–1515.
- 2 Vanhala MJ, Pitkajarvi TK, Kumpusalo EA, Takal JK: Obesity type and clustering of insulin-resistance associated cardiovascular risk factors in middle-aged men and women. *Int J Obesity* 1998; **22**: 369–374.
- 3 Stunkard AJ, Harris JR, Pedersen NL, McClearn GE: The body mass index of twins who have been reared apart. *New England J Med* 1990; **322**: 1483–1487.
- 4 Beales P, Kopelman P: Obesity genes. *Clin Endocrinol* 1996; **45**: 373–378.
- 5 Chagnon Y, Perusse L, Bouchard C: Familial aggregation of obesity, candidate genes and quantitative trait loci. *Curr Opin Lipidol* 1997; **8**: 205–211.
- 6 O'Dell SD, Day INM: Molecules in focus: Insulin-like growth factor II (IGF-II). *Int J Biochem Cell Biol* 1998; **30**: 767–771.
- 7 Marin P, Kvist H, Lindstedt G, Sjostrom L, Bjorntorp P: Low concentrations of insulin-like growth factor-I in abdominal obesity. *Int J Obesity* 1993; **17**: 83–89.
- 8 Tadokoro K, Fujii H, Inoue T, Yamada M: Polymerase chain reaction (PCR) for detection of *Apal* polymorphism at the insulin-like growth factor II gene (*IGF2*). *Nucleic Acids Res* 1991; **19**: 6967.
- 9 O'Dell SD, Miller GJ, Cooper JA *et al*: *Apal* polymorphism in insulin-like growth factor II (*IGF2*) gene and weight in middle-aged males. *Int J Obesity* 1997; **21**: 822–825.
- 10 Bennett ST, Lucassen AM, Gough SCL *et al*: Susceptibility to human type 1 diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus. *Nat Genet* 1995; **9**: 284–292.
- 11 Weaver JU, Kopelman P, Hitman GA: Central obesity and hyperinsulinaemia in women are associated with polymorphism in the 5' flanking region of the human insulin gene. *Eur J Clin Invest* 1992; **22**: 265–270.
- 12 Waterworth DM, Bennett ST, Gharani N *et al*: Linkage and association of insulin gene VNTR regulatory polymorphism with polycystic ovary syndrome. *Lancet* 1996; **349**: 986–990.
- 13 Dunger DB, Ong KKL, Huxtable SJ *et al*: Association of the *INS* VNTR with size at birth. *Nat Genet* 1998; **19**: 98–100.
- 14 Bell GI, Selby MJ, Rutter WJ: The highly polymorphic region near the insulin gene is composed of simple tandemly repeating sequences. *Nature* 1982; **295**: 31–35.
- 15 Bell GI, Horita S, Karam JH: A polymorphic locus near the human insulin gene is associated with human diabetes mellitus. *Diabetes* 1984; **33**: 176–183.
- 16 Bennett ST, Todd JA: Human type 1 diabetes and the insulin gene: principles of mapping polygenes. *Ann Rev Genet* 1996; **30**: 343–370.

- 17 Julier C, Hyer RN, Davies J *et al*: Insulin-IGF2 region on chromosome 11p encodes a gene implicated in HLA-DR4-dependent diabetes susceptibility. *Nature* 1991; **354**: 155–159.
- 18 Miller GJ, Bauer KA, Barzegar S, Cooper JA, Rosenberg RD: Increased activation of the haemostatic system in men at high risk of fatal coronary heart disease. *Thromb Haemost* 1996; **75**: 767–771.
- 19 Little S: ARMS analysis of point mutations. In: Taylor GR (ed). *Laboratory Methods for the Detection of Mutations and Polymorphisms in DNA*, CRC Press, Boca Raton, 1997, pp 45–51.
- 20 Day INM, Humphries SE: Electrophoresis for genotyping: microplate array diagonal gel electrophoresis on horizontal polyacrylamide gels, Hydrolink or agarose. *Anal Biochem* 1994; **222**: 391–394.
- 21 McGinnis RE, Spielman RS: Insulin gene 5' flanking polymorphism. Length of class 1 alleles in number of repeat units. *Diabetes* 1995; **44**: 1296–1302.
- 22 Awata T, Kurihara S, Kikuchi C *et al*: Evidence for association between the class I subset of the insulin gene minisatellite (IDDM2 locus) and IDDM in the Japanese population. *Diabetes* 1997; **46**: 1637–1642.
- 23 Merriman TR, Eaves IA, Twells RCJ *et al*: Transmission of haplotypes of microsatellite markers rather than single marker alleles in the mapping of a putative type 1 diabetes susceptibility gene (IDDM6). *Hum Mol Genet* 1998; **7**: 517–524.
- 24 Doria A, Lee J, Warram JH, Krolewski AS: Diabetes susceptibility at IDDM2 cannot be positively mapped to the VNTR locus of the insulin gene. *Diabetologia* 1996; **39**: 549–549.
- 25 Sharma P, Hingorani A, Jia H *et al*: Positive association of tyrosine hydroxylase microsatellite marker to essential hypertension. *Hypertension* 1998; **32**: 676–682.
- 26 La Spada AR: Trinucleotide repeat instability: genetic features and molecular mechanisms. *Brain Pathol* 1997; **7**: 943–963.
- 27 Ingles SA, Ross RK, Yu MC *et al*: Association of prostate cancer risk with genetic polymorphisms in vitamin D receptor and androgen receptor. *J Natl Cancer Inst* 1997; **89**: 166–170.
- 28 Gavish D, Zrolan N, Breslow JL: Plasma Lp(a) concentration is inversely correlated with the ratio of kringle IV/kringle V encoding domains in the Apo(a) gene. *J Clin Invest* 1989; **84**: 2021–2027.
- 29 Kennedy GC, German MS, Rutter WJ: The minisatellite in the diabetes susceptibility locus IDDM2 regulates insulin transcription. *Nat Genet* 1995; **9**: 293–298.
- 30 Paquette J, Giannoukakis N, Polychronakos C, Vafiadis P, Deal C: The INS 5' variable number of tandem repeats is associated with IGF2 expression in humans. *J Biol Chem* 1998; **273**: 14158–14164.
- 31 Vafiadis P, Bennett ST, Todd JA, Grabs R, Polychronakos C: Divergence between genetic determinants of IGF2 transcription levels in leucocytes and of IDDM2-encoded susceptibility to type 1 diabetes. *J Endocrinol Metab* 1998; **83**: 2933–2939.
- 32 Vafiadis P, Grabs R, Goodyer CG, Colle E, Polychronakos C: A functional analysis of the role of IGF2 in IDDM2-encoded susceptibility to type 1 diabetes. *Diabetes* 1998; **47**: 831–836.
- 33 Shepherd J: Preventing coronary artery disease in the West of Scotland: implications for primary prevention. *Am J Cardiol* 1998; **82**: 57T–59T.