

Is the thrifty genotype hypothesis supported by evidence based on confirmed type 2 diabetes- and obesity-susceptibility variants?

L. Southam · N. Soranzo · S. B. Montgomery ·
T. M. Frayling · M. I. McCarthy · I. Barroso ·
E. Zeggini

Received: 10 February 2009 / Accepted: 20 May 2009 / Published online: 13 June 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract

Aims/hypothesis According to the thrifty genotype hypothesis, the high prevalence of type 2 diabetes and obesity is a consequence of genetic variants that have undergone positive selection during historical periods of erratic food supply. The recent expansion in the number of validated type 2 diabetes- and obesity-susceptibility loci, coupled with access to empirical data, enables us to look for evidence in support (or otherwise) of the thrifty genotype hypothesis using proven loci.

Methods We employed a range of tests to obtain complementary views of the evidence for selection: we determined whether the risk allele at associated ‘index’ single-nucleotide polymorphisms is derived or ancestral, calculated the

integrated haplotype score (iHS) and assessed the population differentiation statistic fixation index (F_{ST}) for 17 type 2 diabetes and 13 obesity loci.

Results We found no evidence for significant differences for the derived/ancestral allele test. None of the studied loci showed strong evidence for selection based on the iHS score. We find a high F_{ST} for rs7901695 at *TCF7L2*, the largest type 2 diabetes effect size found to date.

Conclusions/interpretation Our results provide some evidence for selection at specific loci, but there are no consistent patterns of selection that provide conclusive confirmation of the thrifty genotype hypothesis. Discovery of more signals and more causal variants for type 2 diabetes and obesity is likely to allow more detailed examination of these issues.

L. Southam and N. Soranzo contributed equally to this study.

L. Southam · M. I. McCarthy · E. Zeggini
Wellcome Trust Centre for Human Genetics,
University of Oxford,
Oxford, UK

L. Southam
Institute of Musculoskeletal Sciences, Botnar Research Centre,
Nuffield Orthopaedic Centre, University of Oxford,
Oxford, UK

N. Soranzo · S. B. Montgomery · I. Barroso · E. Zeggini (✉)
Wellcome Trust Sanger Institute,
Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1HH, UK
e-mail: eleftheria@sanger.ac.uk

N. Soranzo
Twin Research & Genetic Epidemiology Department,
King's College London,
St Thomas' Hospital Campus,
London, UK

T. M. Frayling
Genetics of Complex Traits,
Institute of Biomedical and Clinical Science,
Peninsula Medical School,
Exeter, UK

M. I. McCarthy
Oxford Centre for Diabetes,
Endocrinology and Metabolism,
Churchill Hospital,
Oxford, UK

M. I. McCarthy
Oxford National Institute for Health Research
Biomedical Research Centre,
Churchill Hospital,
Oxford, UK

Keywords Genetic association · Haplotype · Obesity · Positive selection · Thrifty genotype hypothesis · Type 2 diabetes

Abbreviations

CEU	Centre d'Etude du Polymorphisme Humain (CEPH) (Utah residents with northern and western European ancestry)
CHB	Han Chinese in Beijing, China
F_{ST}	Population differentiation statistics (fixation index)
iHS	Integrated haplotype score
JPT	Japanese in Tokyo
SNP	Single nucleotide polymorphism
YRI	Yoruba in Ibadan, Nigeria

Introduction

Type 2 diabetes and obesity are complex traits, caused by multiple environmental and genetic factors. In recent decades, there has been a dramatic rise in the prevalence of type 2 diabetes and obesity in the Western and developing world. Adaptation to powerful selective forces for genotypes that provide survival advantage has been proposed as an explanation for this observed capacity of a genetic disease to become so prevalent when unmasked by changes in environment. In 1962, James Neel suggested that exposure to periods of famine during human evolutionary history resulted in selection pressures in favour of a thrifty genotype that led to highly efficient fat storage during periods of abundance [1]. In the current climate of food overabundance and sedentary lifestyle, this thrifty genotype is suggested to lead to metabolically disadvantageous phenotypes.

Signals of positive selection resulting in reduced haplotype diversity can be identified by investigating haplotype structure and allelic architecture. For example, if the thrifty genotype hypothesis were true, we would expect to observe some of the following characteristics at disease loci: risk alleles would be derived alleles; there would be substantial differences in allele frequency across different populations; and there would be evidence that relatively recently emerging alleles have been swept to high frequency. These tests offer the possibility of detecting selection signals, operating over different time scales (ranging from recent positive selection identified through extreme integrated haplotype scores [iHSs] to the much older time frame of derived/ancestral allele status), and we would therefore not expect to obtain consistent evidence across the different tests.

The fields of type 2 diabetes and obesity genetics had until recent years met with limited success in identifying

replicating loci. The advent of large-scale, well-designed association studies, coupled with large-scale follow-up and stringent criteria for declaring reproducible association, has led to the identification of well-established type 2 diabetes and obesity loci. This enables us for the first time to carry out a systematic examination of these genomic loci for evidence of signatures of selection, and thereby seek to corroborate or refute the thrifty genotype hypothesis.

Methods

For the purposes of this study, we define a confirmed type 2 diabetes or obesity locus as one that has been robustly replicated, reaching a genome-wide significance threshold of $p < 5 \times 10^{-8}$. This criterion yields 17 loci for type 2 diabetes (in or near the *TCF7L2*, *PPARG*, *KCNJ11*, *CDKAL1*, *SLC30A8*, *IGF2BP2*, *NOTCH2*, *THADA*, *JAZF1*, *CDC123/CAMK1D*, *TSPAN8/LGR5*, *HHEX/IDE*, *CDKN2A/B*, *ADAMTS9*, *TCF2*, *WFS1* and *KCNQ1* genes) [2] and 13 for obesity (associations with BMI) (in or near the *FTO*, *TMEM18*, *MC4R*, *GNPDA2*, *SH2B1*, *KCTD15*, *MTCH2*, *NEGR1*, *PCSK1*, *LGR4/LIN7C/BDNF* [two independent single nucleotide polymorphisms {SNPs}], *ETV5/SFRS10/DGKG* and *MAF* genes) [3–8] (Tables 1 and 2). We have selected a representative (index) SNP for each of these 30 independently associated loci and have examined several characteristics of the genomic sequence that might indicate evidence for selection.

First, we determined whether the risk allele at the index SNPs is the ancestral or derived allele, using information available through dbSNP build 128 (www.ncbi.nlm.nih.gov/SNP/, accessed February 2009), based on chimpanzee/human sequence alignment.

We also calculated population differentiation statistics (fixation index F_{ST}) for the 30 loci in the three HapMap phase II populations: Centre d'Etude du Polymorphisme Humain (CEPH) (Utah residents with northern and western European ancestry) (CEU); Yoruba in Ibadan, Nigeria (YRI); and Japanese in Tokyo (JPT)+Han Chinese in Beijing, China (CHB) [9]. F_{ST} measures the proportion of total genetic variance that is caused by differences between two or more population samples. Local selection acting on a given locus can result in elevated F_{ST} values between two populations. We can identify loci that have unusually high F_{ST} values by comparing against the rest of the genome, which provides an empirical null distribution. The use of an empirical F_{ST} distribution in this case is advantageous, because it does not require assumptions about the structure of human populations, SNP ascertainment bias (which differs among the three HapMap population samples) and differences in local linkage disequilibrium patterns among different populations. We constructed an empirical F_{ST}

Table 1 Type 2 diabetes-associated risk allele characteristics

SNP	Chr	Position NCBI 36.1 (bp)	No-risk allele	Risk allele	Risk allele frequency ^b	Nearest gene(s)	iHS score ^c	F_{ST}^e global	F_{ST}^f CEU-YRI	F_{ST}^g CEU-JPT+CHB	F_{ST}^h JPT+CHB-YRI
rs864745	7	28,147,081	C	T ^a	0.518	<i>JAZF1</i>	-1.562 (11.7)	0.098 (47.3)	0.119 (35.7)	0.160 (19.7)	0 (93.3)
rs12779790	10	12,368,016	A ^a	G	0.229	<i>CDC123/CAMK1D</i>	NA	0.051 (67.4)	0.113 (37.1)	0.028 (58.7)	0.026 (71.7)
rs7961581	12	69,949,369	T ^a	C	0.233	<i>TSPAN8/LGR5</i>	-0.518 (61.1)	0 (98.3)	0 (85.1)	0 (88.9)	0 (96.4)
rs7578597	2	43,586,327	C	T ^a	0.917	<i>THADA</i>	-0.999 (32.2)	0.214 (18.8)	0.126 (33.9)	0.096 (32.7)	0.336 (11.7)
rs4607103	3	64,686,944	T	C ^a	0.808	<i>ADAMTS9</i>	0.541 (59.5)	0.060 (62.8)	0.006 (80.1)	0.103 (31.2)	0.044 (64.2)
rs10923931	1	120,319,482	G ^a	T	0.117	<i>NOTCH2</i>	2.249 (2.3)	0.258 (13.1)	0.182 (23.4)	0.069 (40.7)	0.391 (8.2)
rs10946398	6	20,769,013	A	C ^a	0.308	<i>CDKALI</i>	-0.161 (87.5)	0.122 (39.3)	0.234 (16.6)	0.009 (72.1)	0.142 (36.2)
rs5015480	10	94,455,539	T	C ^a	0.552	<i>HHEX/IDE</i>	0.479 (63.8)	0.181 (24.7)	0 (98.4)	0.236 (10.7)	0.246 (20.1)
rs10811661	9	22,124,094	C ^a	T	0.792	<i>CDKN2A/B</i>	0.328 (74.7)	0.229 (16.7)	0.199 (20.1)	0.088 (34.9)	0.373 (9.3)
rs4402960	3	186,994,381	G ^a	T	0.292	<i>IGF2BP2</i>	1.641 (9.9)	0.098 (47.3)	0.129 (33.4)	0 (94.3)	0.160 (32.8)
rs13266634	8	118,253,964	T	C ^a	0.75	<i>SLC30A8</i>	-1.869 (5.9)	0.190 (22.9)	0.123 (34.8)	0.084 (36.2)	0.314 (13.3)
rs7901695	10	114,744,078	T	C ^a	0.28	<i>TCF7L2</i>	-0.208 (83.8)	0.361 (5.2)	0.111 (37.5)	0.323 (5.2)	0.579 (2.1)
rs5215	11	17,365,206	T ^a	C	0.408	<i>KCNJ11</i>	-0.435 (66.9)	0.191 (22.7)	0.384 (5.9)	0.004 (76.4)	0.278 (16.6)
rs1801282	3	12,368,125	G	C ^a	0.925	<i>PPARG</i>	-0.571 (57.4)	0.025 (80.9)	0.065 (51.3)	0.005 (75.9)	0.026 (71.3)
rs4430796	17	33,172,153	A	G ^a	0.533	<i>TCF2</i>	0.849 (40.2)	0.098 (47.2)	0.003 (82.7)	0.096 (32.9)	0.160 (32.7)
rs10010131	4	6,343,816	A	G ^a	0.733	<i>WFS1</i>	1.461 (14.3)	0.151 (31.2)	0 (97.5)	0.241 (10.3)	0.246 (20.1)
rs2237892 ^d	11	2,796,327	T	C ^a	0.611	<i>KCNQ1</i>	-0.618 (54.3)	0.172 (26.5)	0 (89.8)	0.209 (13.4)	0.171 (30.7)

iHS scores and F_{ST} values are reported with their percentile rank in parentheses

^a Ancestral allele

^b Allele frequencies taken from HapMap data release 23a/phase II Mar08, on NCBI B36 assembly, dbSNPb126, CEU population

^c Haplotter—HapMap phase II data

^d For *KCNQ1* the JPT+CHB population iHS score is displayed and the risk allele frequency is from JPT HapMap

^e 95% quantile over 2,911,292 markers is 0.365

^f 95% quantile over 2,859,309 markers is 0.406

^g 95% quantile over 2,454,054 markers is 0.327

^h 95% quantile over 2,817,341 markers is 0.465

NA, iHS score unavailable through Haplotter

Table 2 Obesity-associated risk allele characteristics

SNP	Chr	Position NCBI 36.1 (bp)	No-risk allele	Risk allele	Risk allele frequency ^b	Nearest gene(s)	iHS score ^c	F_{ST}^d global	F_{ST}^e CEU-YRI	F_{ST}^f CEU-JPT+CHB	F_{ST}^g JPT+CHB-YRI
rs9939609	16	52,378,028	T	A ^a	0.45	<i>FTO</i>	1.991 (4.4)	0.184 (24.1)	0.005 (81.7)	0.208 (13.5)	0.290 (15.4)
rs6548238	2	624,905	T	C ^a	0.861	<i>TMEM18</i>	0.162 (87.3)	0 (96.9)	0.001 (84.3)	0.003 (79.6)	0 (97.2)
rs17782313	18	56,002,077	T ^e	C	0.283	<i>MC4R</i>	-1.166 (24.6)	0.029 (79.3)	0 (87.7)	0.022 (62.6)	0.057 (59.2)
rs10938397	4	44,877,284	A ^a	G	0.446	<i>GNPDA2</i>	-0.077 (94.0)	0.048 (69.0)	0.111 (37.6)	0.032 (56.6)	0.019 (75.2)
rs7498665	16	28,790,742	A	G ^a	0.358	<i>SH2B1</i>	0.908 (36.9)	0.073 (57.4)	0.081 (46.0)	0.120 (27.1)	0 (92.8)
rs11084753	19	39,013,977	A	G ^a	0.625	<i>KCTD15</i>	0.431 (67.2)	0.163 (28.6)	0.021 (70.7)	0.138 (23.4)	0.259 (18.6)
rs10838738	11	47,619,625	A ^a	G	0.408	<i>MTCH2</i>	-1.814 (6.8)	0.166 (27.9)	0.315 (9.6)	0 (91.4)	0.256 (18.9)
rs2815752	1	72,585,028	G ^a	A	0.65	<i>NEGR1</i>	-0.638 (53.0)	0.185 (23.9)	0.024 (69.5)	0.179 (17.0)	0.317 (13.1)
rs6235	5	95,754,654	G ^a	C	0.267	<i>PCSK1</i>	-0.294 (77.3)	0.046 (70.2)	0.089 (43.5)	0 (98.5)	0.081 (51.2)
rs7647305	3	187,316,984	T ^e	C	0.817	<i>ETV5/SFRS10/DGKG</i>	-0.554 (58.6)	0.183 (24.2)	0.072 (48.9)	0.116 (27.9)	0.324 (12.6)
rs4923461	11	27,613,486	G	A ^a	0.8	<i>LGR4/LIN7C/BDNF</i>	-0.965 (33.9)	0.123 (39.0)	0 (90.4)	0.126 (25.9)	0.169 (31.2)
rs925946	11	27,623,778	G ^a	T	0.358	<i>LGR4/LIN7C/BDNF</i>	0.542 (59.5)	0.153 (30.8)	0.006 (80.9)	0.266 (8.4)	0.179 (29.5)
rs1424233	16	78,240,252	G	A ^a	0.508	<i>MAF</i>	-0.476 (64.2)	0.052 (66.6)	0.028 (66.8)	0.102 (31.2)	0.014 (78.6)

Risk allele is the BMI-increasing allele, no-risk allele is the BMI-decreasing allele. iHS scores and F_{ST} values are reported with their percentile rank in parentheses

^a Ancestral allele

^b Allele frequencies taken from HapMap data release 23a/phase II Mar08, on NCBI B36 assembly, dbSNP126, CEU population.

^c Haplotter—HapMap phase II data

^d 95% quantile over 2,911,292 markers is 0.365

^e 95% quantile over 2,859,309 markers is 0.406

^f 95% quantile over 2,454,054 markers is 0.327

^g 95% quantile over 2,817,341 markers is 0.465

distribution using over 2.9 million SNPs, or the subset of all HapMap Phase II SNPs with genotype data available in all the three reference samples (HapMap Release 22, April 2007). We compared the observed F_{ST} values for the obesity and type 2 diabetes loci with the upper 95% tail of the distribution to obtain a one-tailed test for diversifying selection.

We additionally investigated evidence for natural selection by examining the iHS, a measure of recent positive selection for variants that have not yet reached fixation [10, 11]. This statistic identifies SNPs for which alleles have rapidly changed in frequency by comparing the haplotype background of the ancestral and derived alleles. Negative iHS values indicate that the derived allele resides on a longer haplotype, whereas positive iHS values suggest that the ancestral allele resides on a longer haplotype. For the purposes of this study, we define iHS <-1.5 and iHS >1.5 as suggestive evidence for natural selection, and iHS scores <-2 or >2 as evidence for a powerful selection signal [10]. We determined the iHS score for each locus in HapMap phase II data using Haplotter (<http://hg-wen.uchicago.edu/selection/haplotter.htm>, accessed February 2009) [10, 11].

Results

Evidence that type 2 diabetes- or obesity-associated risk alleles were more often derived than ancestral would be consistent with positive selection. In type 2 diabetes, we found the risk allele to be the derived allele at six of the 17 loci (*CDC123/CAMK1D*, *TSPAN8/LGR5*, *NOTCH2*, *CDKN2A/B*, *IGF2BP2* and *KCNJ11*) (binomial test one-sided $p=0.93$) (Table 1). Similarly, we did not observe a significant overrepresentation of derived status for the obesity-risk alleles (seven [*MC4R*, *GNPDA2*, *MTCH2*, *NEGR1*, *PCSK1*, *LGR4/LIN7C/BDNF* and *ETV5/SFRS10/DGKG*], $p=0.50$) (Table 2). Among the type 2 diabetes loci, ten risk alleles are major and seven minor (binomial test two-sided $p=0.63$) (Table 1). Among the obesity-risk alleles, six are major and seven are minor ($p=1.00$) (Table 2).

Only one locus (rs7901695 at *TCF7L2*) showed an elevated F_{ST} value of 0.579 (2.1 percentile), between the JPT + CHB and YRI sample (previously also noted [12]), and in the comparison between CEU and JPT+CHB ($F_{ST}=0.323$, 5.2 percentile) (Table 1). SNP rs5215 at *KCNJ11* demonstrated an elevated F_{ST} value of 0.384 between CEU and YRI (5.9 percentile) (Table 1).

Among the type 2 diabetes-associated loci, the *NOTCH2* rs10923931 index SNP demonstrated an elevated iHS value (2.249, 2.3 percentile) for the protective, ancestral allele (Table 1). Among the BMI-associated SNPs, the strongest signal of positive selection was obtained for the *FTO* locus, with an iHS value of 1.991 (4.4 percentile) (Table 2). No

general enrichment for high F_{ST} or long haplotypes was observed for the set of diabetes- or obesity-associated SNPs (using Mann–Whitney significance testing).

Discussion

We have not observed significant evidence for overrepresentation of ancestral/derived status or for minor/major frequency at type 2 diabetes- or obesity-risk alleles. Only one locus (at the type 2 diabetes *TCF7L2* locus) demonstrates large allele frequency differences across populations. Although this is consistent with chance, we note that *TCF7L2* represents the strongest effect size to be identified in type 2 diabetes to date and, as such, may have been more susceptible to selection forces. Notably, we did not find strong evidence for high differentiation of rs2237892 at *KCNQ1* between the European and East Asian sample ($F_{ST}=0.209$, 13.3 percentile of the empirical distribution). The risk allele C at this locus has frequencies close to 90% in the CEU and YRI HapMap samples and close to 60% in the two East Asian samples.

Our analyses indicate the presence of extended haplotypes at the *FTO* locus, the largest effect size for obesity found to date. However, we have not identified any consistent footprint of selection across the loci that would support the notion of a universal mechanism to explain the high prevalence of type 2 diabetes and obesity. The number of robustly replicating type 2 diabetes and obesity loci identified is poised to grow, offering the promise of an extended established disease locus list. In addition, expansion of association studies to populations of non-European descent is likely to broaden the spectrum of robustly associated allelic variation and may help identify loci with prominent evidence for population differentiation, for example where risk alleles at a SNP have rapidly changed in frequency since population separation. Importantly, the truly causal, functional variants for the majority, if not all, of established type 2 diabetes- and obesity-susceptibility loci have not been determined yet. We have therefore been restricted to studying index SNPs, representative of the replicating associations, which could have an effect on the variant-specific analyses we have carried out, as these may provide only indirect glimpses of the history of the causal mutations.

This study has been exhaustive in terms of comprehensively considering all known, well-established type 2 diabetes- and BMI-susceptibility variants. Some loci appear to have more ‘thrifty gene’ characteristics than others, but there is no clear globally consistent transpiring picture. Further emerging insights into the genetic aetiology of these complex traits are likely to help us distinguish between apparent and real signals for positive selection.

Acknowledgements This work was funded by the Wellcome Trust (WT088885/Z/09/Z and WT077016/Z/05/Z), MRC (G0601261), EU FP6 grant LSHM-CT-2006-037197 (INTERACT) and the Oxford NIHR Biomedical Research Centre. E. Zeggini is a Wellcome Trust Research Career Development Fellow. L. Southam is supported by EC Framework 7 Programme Grant 200800 (TREAT-OA).

Duality of interest The authors declare that there is no duality of interest associated with this manuscript.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Neel JV (1962) Diabetes mellitus: a ‘thrifty’ genotype rendered detrimental by ‘progress’? *Am J Hum Genet* 14:353–362
2. McCarthy MI, Zeggini E (2009) Genome-wide association studies in type 2 diabetes. *Curr Diab Rep* 9:164–171
3. Frayling TM, Timpson NJ, Weedon MN et al (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 316:889–894
4. Loos RJ, Lindgren CM, Li S, Wheeler E et al (2008) Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet* 40:768–775
5. Willer CJ, Speliotes EK, Loos RJ et al (2009) Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet* 41:25–34
6. Thorleifsson G, Walters GB, Gudbjartsson DF et al (2009) Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet* 41:18–24
7. Benzinou M, Creemers JW, Choquet H et al (2008) Common nonsynonymous variants in PCSK1 confer risk of obesity. *Nat Genet* 40:943–945
8. Meyre D, Delplanque J, Chèvre JC et al (2009) Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nat Genet* 41:157–159
9. International HapMap Consortium (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449:851–861
10. Voight BF, Kudravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome. *PLoS Biol* 4:e72
11. Kudravalli S, Veyrieras JB, Stranger BE, Dermitzakis ET, Pritchard JK (2009) Gene expression levels are a target of recent natural selection in the human genome. *Mol Biol Evol* 26:649–658
12. Myles S, Davison D, Barrett J, Stoneking M, Timpson N (2008) Worldwide population differentiation at disease-associated SNPs. *BMC Med Genomics* 1:22