

Identification of Metabolic Modifiers That Underlie Phenotypic Variations in Energy-Balance Regulation

Chia Lin Chang,¹ James J. Cai,^{2,3} Po Jen Cheng,¹ Ho Yen Chueh,¹ and Sheau Yu Teddy Hsu⁴

OBJECTIVE—Although recent studies have shown that human genomes contain hundreds of loci that exhibit signatures of positive selection, variants that are associated with adaptation in energy-balance regulation remain elusive. We reasoned that the difficulty in identifying such variants could be due to heterogeneity in selection pressure and that an integrative approach that incorporated experiment-based evidence and population genetics-based statistical judgments would be needed to reveal important metabolic modifiers in humans.

RESEARCH DESIGN AND METHODS—To identify common metabolic modifiers that underlie phenotypic variation in diabetes-associated or obesity-associated traits in humans, or both, we screened 207 candidate loci for regulatory single nucleotide polymorphisms (SNPs) that exhibited evidence of gene–environmental interactions.

RESULTS—Three SNPs (rs3895874, rs3848460, and rs937301) at the 5' gene region of human *GIP* were identified as prime metabolic-modifier candidates at the enteroinsular axis. Functional studies have shown that *GIP* promoter reporters carrying derived alleles of these three SNPs (haplotype *GIP*^{-1920A}) have significantly lower transcriptional activities than those with ancestral alleles at corresponding positions (haplotype *GIP*^{-1920G}). Consistently, studies of pregnant women who have undergone a screening test for gestational diabetes have shown that patients with a homozygous *GIP*^{-1920A/A} genotype have significantly lower serum concentrations of glucose-dependent insulinotropic polypeptide (GIP) than those carrying an ancestral *GIP*^{-1920G} haplotype. After controlling for a *GIPR* variation, we showed that serum glucose concentrations of patients carrying *GIP*^{-1920A/A} homozygotes are significantly higher than that of those carrying an ancestral *GIP*^{-1920G} haplotype (odds ratio 3.53).

CONCLUSIONS—Our proof-of-concept study indicates that common regulatory *GIP* variants impart a difference in GIP and glucose metabolism. The study also provides a rare example that identified the common variant-common phenotypic variation pattern based on evidence of moderate gene–environmental interactions. *Diabetes* 60:726–734, 2011

Taking advantage of the availability of genome information from diverse human populations, recent studies have revealed that human genomes contain hundreds of loci that exhibit signatures of positive selection (1,2). A number of these variants have been shown to describe phenotypic variations in appearance, physiologic adaptations, and pathologic responses to diseases, thereby opening doors to human history, unexpected involvement of risk alleles, and novel prognosis power for select diseases (3–10). These recent findings on gene selection have partially affirmed the “common variants underlie common phenotypic variation” paradigm.

Complex multifactorial diseases such as diabetes and obesity have been hypothesized to be a reflection of maladaptation of previously advantageous alleles to current environments, and the divergent manifestation of different metabolic syndromes could be associated with adaptation in response to heterogeneous changes in human culture (11,12). Therefore, the revelation of adaptive variants that are associated with energy-balance regulation could reveal novel regulatory mechanisms underlying diverse metabolic syndromes and how such variants contribute to phenotypic variation. However, such variants could be subject to selection pressures that fluctuate over time or reverse in a short time in response to shifting human culture; in this way, these variants are more likely to be associated with incomplete signatures of selection (13,14). Thus, approaches for detecting signatures of complete selection may lack the power to detect metabolic modifiers that confer recent adaptations in energy-balance regulation or that are associated with phenotypic variation in diabetes- and obesity-related traits. We consequently reasoned that an integrative approach that embraces both wet-laboratory experiment-based evidence and statistical judgments of newly available population genetics data are needed to reveal important metabolic modifiers.

With this understanding, we applied two intersecting criteria to explore metabolic modifiers that have been subject to selection since Eurasians split from Africans. First, we generated a set of a priori diabetes-related and/or obesity-related candidate loci based on previous clinical investigations, genome-wide association (GWA) studies, and quantitative trait loci analyses (6,15–20).

Second, we conducted a variety of genomic scans of the candidate genic regions for evidence of integrative signals of positive selection. Genetic variants in genic regions are major focuses of the scans because most contributions of common variation to complex traits are regulatory or non-synonymous in nature (21). In addition, we focused the analysis on variants that have a derived allele frequency >30% in the overall HapMap population because these variants presumably would affect most of the population and have a greater power for the detection of phenotypic association.

From the ¹Department of Obstetrics and Gynecology, Chang Gung Memorial Hospital Linkou Medical Center, Chang Gung University, Kweishan, Taoyuan, Taiwan; the ²Department of Biology, Stanford University, Stanford, California; the ³Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, Texas; and the ⁴Reproductive Biology and Stem Cell Research Program, Department of Obstetrics and Gynecology, Stanford University School of Medicine, Stanford, California.

Corresponding author: Sheau Yu Teddy Hsu, teddyhsu@stanford.edu.

Received 22 September 2010 and accepted 12 December 2010.

DOI: 10.2337/db10-1331

This article contains Supplementary Data online at <http://diabetes.diabetesjournals.org/lookup/suppl/doi:10.2337/db10-1331/-/DC1>.

© 2011 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. See <http://creativecommons.org/licenses/by-nc-nd/3.0/> for details.

Using this two-step approach, we identified common variants in the 5' gene region of *CDKAL1*, *CYB5R4*, *GAD2*, *PPARG*, and *GIP* as candidates for metabolic modifiers. Among these variants, those in *GIP* are of particular interest because they are highly linked to each other as well as to a functional nonsynonymous mutation (rs2291725) in exon 4 of *GIP*; that is, in a high linkage-disequilibrium (LD) block (22).

Glucose-dependent insulinotropic polypeptide (GIP) secreted from duodenal and jejunal K-cells is one of the two incretin hormones (i.e., glucagon-like peptide-1 [GLP-1] and GIP) that stimulate insulin release after food intake in humans (16). In addition, we know that patients with type 2 diabetes or in late pregnancy have a depressed β -cell response to GIP compared with healthy individuals, and GIP antagonism has been proposed as a strategy for the treatment of obesity (23). Thus, the characterization of potential modifiers resulting from prior *GIP*-environment interactions is relevant to a better understanding of molecular mechanisms that underlie diabetes and obesity as well as phenotypic variations that are associated with these diseases. Here, based on an integrative screening approach and proof-of-concept study of *GIP* variants, we show that it is possible to identify metabolic modifiers by studying genes that exhibit evidence of moderate gene-environment interactions, and that regulatory *GIP* variants impart phenotypic variation in GIP response and glucose metabolism.

RESEARCH DESIGN AND METHODS

Patients. This study was approved by the institutional ethics committee review boards of Chang Gung Memorial Hospital Linkou Medical Center and Chang Gung University. We recruited 123 patients, and all patients gave written informed consent to participate in the study. The screening glucose challenge test for gestational diabetes was performed as previously described (24).

Genotyping. Genomic DNA samples of participants were extracted and purified from anticoagulated blood with the DNeasy Blood & Tissue Kit (Qiagen, Venlo, Netherlands). Genotyping of SNPs was performed using the TaqMan Validated SNP Genotyping Assays (Applied Biosystems, Foster City, CA). The genotyping analysis had a >97% success rate and >99% reproducibility.

GIP, insulin, and C-peptide enzyme-linked immunosorbent assay. Blood samples were collected from patients for hormone measurement 1 h after administration of the oral glucose tolerance test. Total GIP, insulin, and C-peptide levels in human serum were measured using sandwich ELISA kits (Millipore, Billerica, MA; Mercodia, Uppsala, Sweden; and Calbiotech, Spring Valley, CA).

In vitro GIP promoter reporter analysis. A 2.15-kb fragment of human *GIP* promoter with the A allele at rs3895874, rs3809770, rs3848460, and rs937301 (−2073 bp to +77 bp) was chemically synthesized (Genescript Inc., Piscataway, NJ) and subcloned into the pGL4.2 luciferase reporter vector (Promega

Corp., Madison, WI). Promoter fragments with ancestral haplotypes were obtained using the site-directed mutagenesis. In the promoter reporter study, each experiment was conducted at least three times with three or four replicates for each treatment.

F_{ST} estimation, LD plot, and hitchhiking mapping for GIP locus and neighboring regions. The population genetic differentiation statistic F_{ST} was performed using the PGEToolbox (25). LD and haplotype plots were generated using HaploView 4.1 (26).

Statistical analysis. Patients' glucose challenge responses and serum hormone profiles were compared with the χ^2 test or the Student *t* test with Welch correction. Ratios of patients with glucose levels exceeding the threshold were analyzed with the χ^2 test. All *P* values were two-sided. All data were presented as mean \pm SEM, and the statistical significance cutoff value was 0.05.

RESULTS

SNPs at the GIP locus are targets of environmental selection in humans. To systematically identify putative metabolic modifiers, we curated and screened SNPs in 207 gene loci that have previously been implicated in the regulation of diabetes-related and/or obesity-related traits using F_{ST} (Supplementary Table 1). The empiric distribution of the F_{ST} statistic has been used to detect genomic regions that have rapidly increased in frequency as a result of local selective pressures (27). Of these 207 genes, 59 carried genic variants with F_{ST} values in the top 5% bracket in comparisons between the corresponding HapMap II populations YRI (Yoruba from Ibadan), CEU (U.S. residents with northern and western European ancestry), and ASN (pooled samples of Chinese from Beijing [CHB] and Japanese from Tokyo [JPT]; Supplementary Table 1). The genic region in 29 of these 59 genes also contained another indication of local selection: long haplotypes in one of the HapMap populations.

On the basis of the presence of highly divergent allele frequencies between Eurasians and Africans, the presence of LD and extended haplotypes in Eurasians, and a >30% minor allele frequency in the overall Eurasian population, we identified seven SNPs in the 5' gene region of *CDKAL1*, *CYB5R4*, *GAD2*, *GIP*, and *PPARG* as potential common metabolic modifiers (Table 1). In earlier GWA studies, select variants in *CDKAL1*, *CYB5R4*, *GAD2*, and *PPARG* were associated with type 2 diabetes-related or obesity-related traits (28–31). By contrast, there has been no report of linkage of *GIP* variants in GWA studies. Because the three *GIP* variants (rs3895874, rs3848460, and rs937301) are highly linked compared with those that appeared alone in other candidate genes (32), and because these *GIP* variants are partially linked with a nonsynonymous *GIP* SNP

TABLE 1
SNPs from the 5' gene region that exhibited signals of partial selection

Gene (candidate SNPs)	Association with disease risk	Position	Frequencies of derived alleles*			Associated with LD & long haplotypes
			YRI	CEU	ASN	
<i>CDKAL1</i> (rs9368197)	T2D, ulcerative colitis, GD, Crohn, OB	Chr. 6, 20641362	0.051	0.233	0.511	ASN
<i>CYB5R4</i> (rs1325471)	Near a T2D susceptibility locus on Chr. 6	Chr. 6, 84624394	0.143	0.958	0.715	CEU & ASN
<i>GAD2</i> (rs2839670)	OB, alcohol dependence, schizophrenia	Chr. 10, 26544346	0.217	0.817	0.578	CEU
<i>GIP</i> (rs3895874)	None	Chr. 17, 44402867	0.193	0.517	0.753	CEU & ASN
<i>GIP</i> (rs3848460)	None	Chr. 17, 44402113	0.193	0.517	0.753	CEU & ASN
<i>GIP</i> (rs937301)	None	Chr. 17, 44401275	0.193	0.517	0.753	CEU & ASN
<i>PPARG</i> (rs2920502)	T2D, obesity, coronary artery diseases	Chr. 3, 12304195	0.026	0.357	0.753	CEU & ASN

ASN, pooled samples of Chinese from Beijing and Japanese from Tokyo; CEU, Utah residents with northern and western European ancestry; GD, gestational diabetes; LD, linkage-disequilibrium; OB, obesity; SNP, single nucleotide polymorphism; T2D, type 2 diabetes; YRI, Yoruba in Ibadan, Nigeria. *The allele frequency is significantly different among HapMap II populations (*P* < 0.01).

(rs2291725) (22), these *GIP* variants have a low likelihood of being false positive (Table 1). Hence, we focused subsequent functional analysis on *GIP* variants.

Fine mapping of the *GIP* locus showed that these variants are linked with more than three dozen neighboring SNPs (from rs9904761 to rs3895874 on chr17: 44,311–44,402 kb), and these linked SNPs exhibited F_{ST} values in the top 2–10% bracket in comparisons between YRI and ASN (Supplementary Fig. 1) (33). A plot of a 250-kb region of genotypes around *GIP* in the three HapMap II populations showed that whereas genotypes surrounding *GIP* in YRI exhibited a high degree of homozygosity for ancestral alleles (Fig. 1A, upper panel), in the same region, genotypes of ASN and CEU exhibited a high degree of homozygosity for the derived alleles (Fig. 1A, middle and lower panels). Consistently, analysis of genic SNPs at the *GIP* locus between 11 populations from the HapMap III project (34) showed that F_{ST} values were the highest for comparisons between East Asian and African populations (Fig. 1B). This result is consistent with our recent finding that a nonsynonymous variant in the exon IV of *GIP* (rs2291725) and a 71-kb haplotype block surrounding this variant were positively selected in Eurasians in the last 2,000 to 11,800 years (22).

Interestingly, analyses of LD and haplotype block diversity of the genomic region that encompassed the three *GIP* variants at the 5' gene region showed that a 91-kb LD block was represented by five inferred haplotypes in CEU and four in and ASN chromosomes, respectively (Fig. 2;

Supplementary Fig. 2, left panel; Supplementary Fig. 3, SNP No. 36–79). By contrast, the same region was represented by 58 haplotypes in YRI. In addition, these analyses showed that the high F_{ST} values and the major shift in allele frequencies of *GIP* variants at the 5' gene region between Eurasian and African populations could be attributed to the increase of a derived haplotype (haplotype 50 in Fig. 2) from less than 5% in Africans to more than 50% in Eurasians. Because the three 5' gene region variants are completely linked with the positively selected rs2291725 in ASN (22), we inferred that rs3895874, rs3848460, and rs937301 at the human *GIP* promoter region were positively selected in East Asian populations as well.

Importantly, we found that the 5' gene region variants (i.e., rs3895874, rs3809770, rs3848460, and rs937301 at positions –1920, –1650, –1158, and –320 of *GIP*, respectively) are located in a haplotype block that is separated from the one containing the nonsynonymous variant rs2291725 by a hotspot for recombination in CEU and YRI (Supplementary Fig. 2, right panel, and Supplementary Fig. 3). In this haplotype block, a tetranucleotide polymorphism was represented by three inferred haplotypes in all three populations (referred to as derived $GIP^{-1920AAA}$, ancestral $GIP^{-1920GAGG}$, and ancestral $GIP^{-1920GGGG}$ haplotypes in the following text; Table 2). The derived $GIP^{-1920AAA}$ haplotype, which was found in 18.3% of YRI chromosomes, has become the dominant haplotype and has a frequency exceeding 50 and 75% in CEU and ASN, respectively (Table 2). By contrast, the dominant ancestral

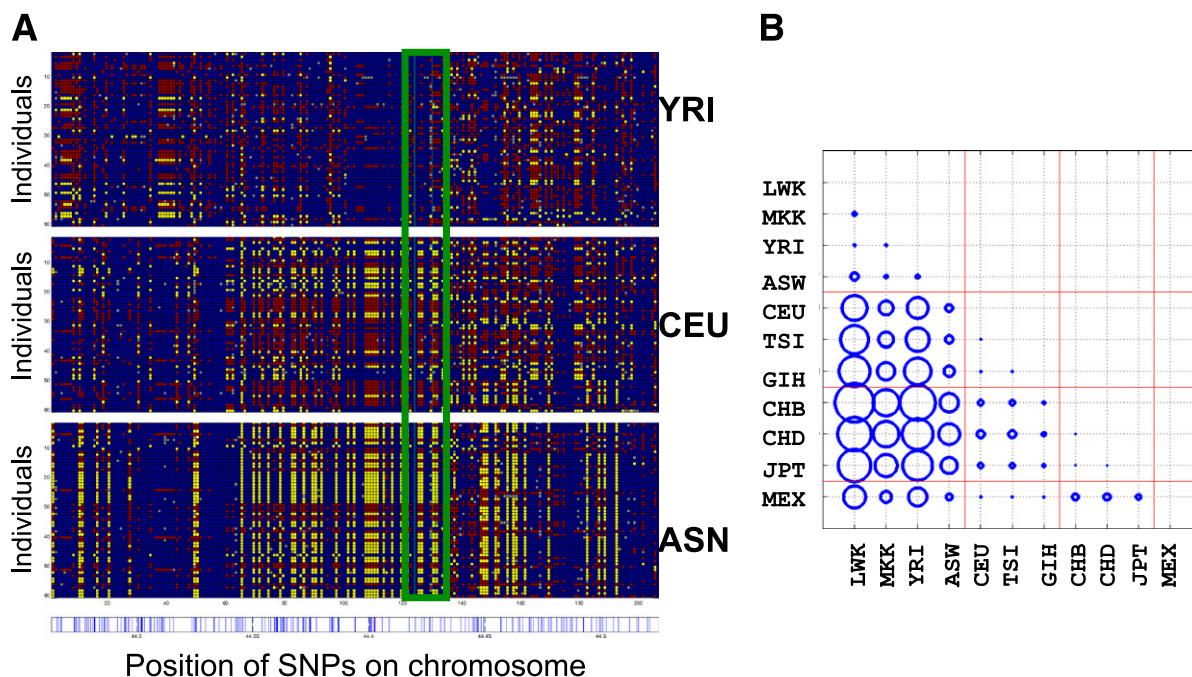


FIG. 1. Differential distribution of polymorphic alleles at the *GIP* locus in human populations. **A:** Visual depiction of genotypes within a 250-kb region around *GIP* in YRI, CEU, and ASN populations. Polymorphic sites are color coded according to their allelic state. Individuals with a homozygous genotype with ancestral allele (AA) and a homozygous genotype with derived allele (aa) are represented by blue and yellow dots, respectively. The red dots indicate individuals with a heterozygous genotype (Aa). The location of SNPs in the *GIP* gene region is indicated by a green rectangle. The position of SNPs within the genomic region is indicated by blue vertical lines in the bottom panel. The position of rs3895874 at GIP^{-1920} is indicated by a red vertical line. **B:** Average F_{ST} for genic SNPs at the *GIP* locus between 11 populations of the HapMap phase III dataset. The diameter of circles is proportional to the average F_{ST} . Population descriptors: ASW, African ancestry in southwest U.S.; CEU, Utah residents with northern and western European ancestry from the Centre d'Etude du Polymorphisme Humain (CEPH) collection; CHB, Han-Chinese in Beijing, China; CHD, Chinese in metropolitan Denver, CO; GIH, Gujarati Indians in Houston, TX; JPT, Japanese in Tokyo, Japan; LWK, Luhya in Webuye, Kenya; MEX, Mexican ancestry in Los Angeles, CA; MKK, Maasai in Kinyawa, Kenya; TSI, Tuscans in Italy; YRI, Yoruba in Ibadan, Nigeria. The highest population differentiation was found between Asian (Han-Chinese and Japanese) and African (Luhya, Maasai, or Yoruba) populations. (A high-quality color representation of this figure is available in the online issue.)

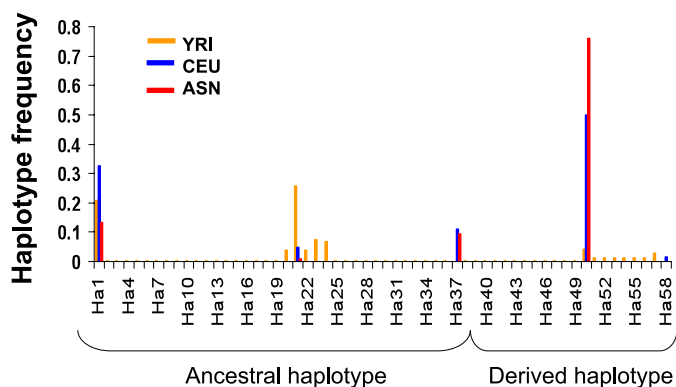


FIG. 2. Haplotype diversity in the *GIP* locus is population-dependent. Frequencies of the 58 inferred haplotypes found in the 91-kb LD region (from rs9904761 to rs3895874) surrounding *GIP* in YRI, CEU, and ASN. Most of the CEU and ASN chromosomes contained derived haplotypes (with AAAA residues at positions 76–79), whereas YRI chromosomes exhibited significant divergence in the haplotype composition. The 38 ancestral haplotypes all contained GGGG or GGAG residues at positions 76–79. Population descriptors: ASN, pooled samples of Chinese from Beijing and Japanese from Tokyo; CEU, Utah residents with northern and western European ancestry from the Centre de'Etude du Polymorphisme Humain (CEPH) collection; YRI, Yoruba in Ibadan, Nigeria. (A high-quality color representation of this figure is available in the online issue.)

haplotype in YRI chromosomes ($GIP^{-1920GAGG}$, 50.8%) was only found in 1.1% of ASN chromosomes. Therefore, these 5' gene region variants and the nonsynonymous variant rs2291725 could be selected differentially in the overall HapMap populations and represent causal mutations independent of the nonsynonymous variant rs2291725.

The *GIP* promoter reporter activity is haplotype-dependent. Because only functional investigations can convincingly demonstrate causal mutations, we sought to obtain direct evidence that these 5' gene region SNPs represent causal mutations for the population genetics observation. We constructed and tested *GIP* promoter reporters with each of the three major haplotypes in transfected human embryonic kidney (HEK) 293T cells (Fig. 3A). Measurement of promoter reporter activities showed that constructs with an ancestral haplotype ($GIP^{-1920GAGG}$ or $GIP^{-1920GGGG}$) exhibited luciferase reporter activity 25–45% higher than that of a derived haplotype ($GIP^{-1920AAAA}$),

TABLE 2
Frequency of inferred *GIP* haplotypes and rs3895874 genotypes at the 5' gene region in HapMap II populations

Inferred haplotype*	ASN (n = 174)	CEU (n = 120)	YRI (n = 114)
$GIP^{-1920GAGG}$	0.011	0.050	0.508
$GIP^{-1920GGGG}$	0.228	0.434	0.283
$GIP^{-1920AAAA}$	0.761	0.500	0.183
Genotype at rs3895874			
$GIP^{-1920G/G}$	0.081	0.233	0.386
$GIP^{-1920G/A}$	0.333	0.500	0.614
$GIP^{-1920A/A}$	0.586	0.267	0.000

ASN, pooled samples of Chinese from Beijing and Japanese from Tokyo; CEU, Utah residents with northern and western European ancestry from the Centre de'Etude du Polymorphisme Humain (CEPH) collection; YRI, Yoruba in Ibadan, Nigeria. *Frequencies of *GIP* haplotypes were significantly different among HapMap II populations ($P < 0.0001$).

$P < 0.01$; Fig. 3B). Because the *GIP* promoter region contains elements that are important for regulation by transcriptional factors, including PAX6 and GATA4 (23), we also determined whether the *GIP* promoter reporter activity was haplotype-dependent in the presence of these transcription factors. As expected, coexpression of PAX6 or GATA4 increased the basal reporter activities by 1.5–2.5-fold and 0.7–0.8-fold, respectively (Fig. 3B). Importantly, we found that reporters with an ancestral haplotype consistently exhibited significantly higher activities than those containing the derived haplotype in the presence of PAX6 or GATA4 ($P < 0.01$). Together, these results suggested that derived alleles at rs3895874, rs3848460, and rs937301—but not rs3809770—represent functional mutations.

The *GIP* haplotype is associated with serum *GIP* and glucose levels after an oral glucose challenge test in pregnant women. Given that earlier genome studies have not reported an association between *GIP* variants and any trait, we speculated that the partial selection of *GIP* variants in Eurasian populations could be associated with adaptation at a life stage that is vulnerable to environmental changes and that has not been specifically studied. Because pregnancy represents a critical life stage that subjects individuals to excessive metabolic load, and because its success has a major impact on reproductive fitness, we hypothesized that studies of phenotypic variation during pregnancy may provide a sensitive model to investigate the role of *GIP* haplotypes.

To test for an allelic effect of *GIP* variants, we studied East Asian patients for proof-of-concept testing because selection pressures are most likely ongoing in populations that exhibit the most significant evidence. We genotyped a panel of 123 unrelated Han-Chinese women who underwent a screening glucose challenge test for gestational diabetes during the 23rd to the 29th weeks of pregnancy, and these patients were assigned to three genotype groups ($GIP^{-1920G/G}$, $GIP^{-1920G/A}$, or $GIP^{-1920A/A}$) based on alleles at rs3895874, rs3848460, and rs937301. The frequency of rs3895874 in these patients was similar to that of the ASN population and was in the Hardy–Weinberg equilibrium; frequencies of $GIP^{-1920G/G}$, $GIP^{-1920G/A}$, and $GIP^{-1920A/A}$ genotypes were 0.089, 0.480, and 0.431, respectively. In addition, alleles at rs3895874 in these patients were in absolute LD with those at rs3848460 and rs937301.

Measurements of serum *GIP* and glucose levels showed that circulating *GIP* and glucose at 1 h after the challenge test were 20.6–219.9 pg/mL, and 72–230 mg/dL, respectively. Consistent with in vitro promoter reporter assays, *GIP* levels in patients carrying the ancestral GIP^{-1920G} haplotype ($GIP^{-1920G/A}$ heterozygote and $GIP^{-1920G/G}$ homozygote) were significantly higher than those with a homozygous $GIP^{-1920A/A}$ genotype (Fig. 4A, Table 3). By contrast, serum levels of glucose, insulin, and C-peptide, as well as age and BMI, were not significantly different among patients (Table 3).

Because two linked *GIP* receptor (*GIPR*) variants rs10423928 and rs1800437 (referred to as the $GIPR^{1159C/G}$ mutation in Table 3) have recently been shown to be associated with glucose and insulin levels after oral glucose challenge tests as well as the incretin effect in nondiabetic individuals in GWA studies (35,36), we also genotyped these variants and sought to isolate the potential confounding effect of *GIPR* variants. We found no association between *GIPR* SNPs and serum glucose or hormone

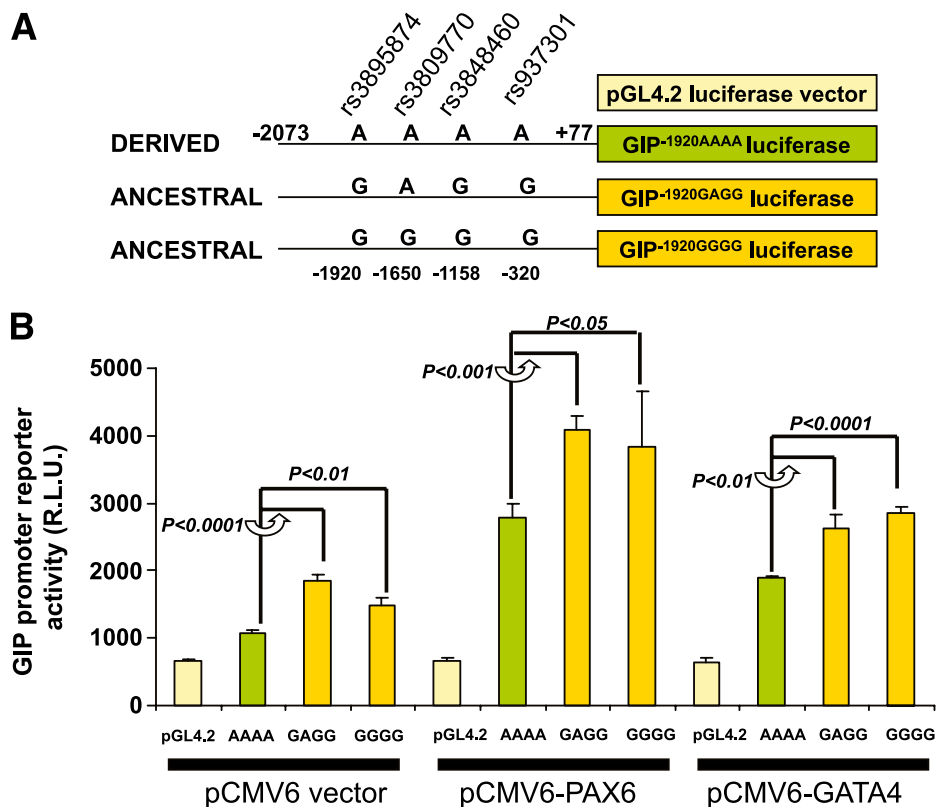


FIG. 3. The *GIP* promoter reporter activity is haplotype-dependent. **A:** Schematic representation of luciferase reporters containing one of the three inferred haplotypes (derived: *GIP*^{-1920AAAA}; ancestral: *GIP*^{-1920GAGG} and *GIP*^{-1920GGGG}) found in a 2.15-kb fragment of the *GIP* promoter. The genomic fragment from position -2073 to +77 bp of *GIP* contains four SNPs (from rs3895874 to rs937301), and three of them are linked. Genomic fragments that contained each of the three haplotypes were subcloned in the pGL4.2 luciferase reporter. **B:** The luciferase reporter activity is shown in HEK293T cells transfected with combinations of reporter and expression vectors for PAX6 and GATA4. The reporter activity is haplotype-dependent in the absence or presence of select transcription factors. To compare *GIP* promoter reporter activities, HEK293T cells in 6-well culture dishes were transfected with different combinations of transcription factor expression vector and a select *GIP* promoter reporter using Lipofectamine 2000 (Life Technology Inc., Carlsbad, CA). In each well, equal amounts of pCMV and pGL4.2 expression vectors and a one-tenth aliquot of a β -galactosidase expression vector were transfected. At 48 h after transfection, cells were free-thawed once in lysis buffer, and luciferase activity in supernatant was assayed using the Steady-Glo Luciferase Assay System (Promega Corp.) and a Lumimark microplate reader (Bio-Rad Corp., Hercules, CA). The reporter activity has been expressed as the ratio of relative luciferase unit/ β -galactosidase activity in transfected cells. Similar results were obtained in four separate experiments. (A high-quality color representation of this figure is available in the online issue.)

levels, but levels of serum glucose, in addition to GIP, were significantly different between *GIP*^{-1920A/A} homozygotes, and heterozygotes and the *GIP*^{-1920G/G} homozygotes combined within the pool of patients with the dominant *GIPR*^{1159G/G} genotype (Fig. 4B). Moreover, we noticed that after controlling for the variation at *GIPR*¹¹⁵⁹, the derived *GIP*^{-1920A/A} genotype was associated with increased risk of having a glucose level that exceeded the 140 mg/dL threshold (odds ratio 3.53 [95% CI 1.25–9.92]; $P = 0.015$; Table 3). Among patients with a *GIPR*^{1159G/G} genotype, 48.3% of patients with *GIP*^{-1920A/A} homozygotes exhibited glucose levels that exceeded the threshold, whereas only 20.9% of the remaining patients did. Therefore, the homozygous *GIP*^{-1920A/A} genotype could be associated with a reduced GIP response and reduced capability of maintaining glucose homeostasis.

DISCUSSION

On the basis of studies of signatures of selection, in vitro promoter assays, and glucose challenge tests in humans, we show that it is possible to identify causal variants related to energy-balance regulation by focusing on genic SNPs that were subject to environmental selection in a subset of candidate genes. Specifically, we demonstrated

that *GIP* variants at the 5' gene region represent metabolic modifiers that contribute to phenotypic variation in GIP response and glucose metabolism. Further characterization of these causal variants would open a new venue for understanding the molecular mechanisms that underlie phenotypic variations in energy-balance regulation and improve our ability to stratify and interpret clinical outcomes associated with the GIP signaling pathway.

For decades, adaptive selection was assumed to be rare; however, recent studies have demonstrated that adaptive substitution is pervasive in human genomes (1,7,37). Despite this progress, it is obvious that population differentiation characteristics of human genomes have yet to be fully explored because physiologic consequences of almost all of these past gene–environmental interactions remain to be verified experimentally (1,2). On the other hand, because many selection pressures could be heterogeneous or reversible in a short time, the signature of selection may have eroded in genes that were responsive to cultural selection pressures (12,38) compared with those shielded from heterogeneous selection (e.g., the adaptation to environmental oxygen levels and altitude) (5,39). We therefore reasoned that important metabolic modifiers could be hidden in the trove of SNPs that showed limited evidence of positive selection and that this

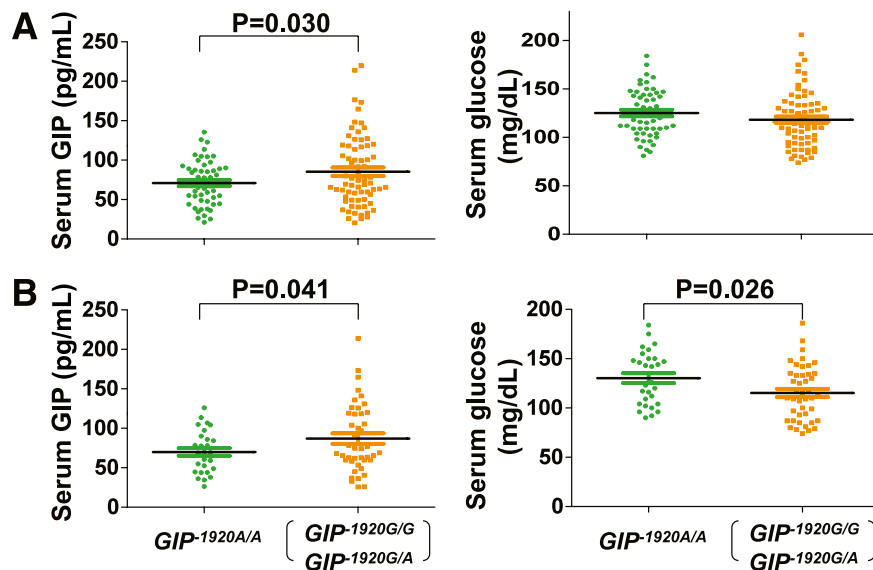


FIG. 4. *GIP* haplotypes impart a difference in *GIP* response and glucose metabolism after glucose challenge tests. **A:** Measurements of serum *GIP* and glucose levels in 123 patients during the 23rd to the 29th weeks of pregnancy at 1 h after the 50-g glucose challenge test. The *GIP* level was significantly higher in patients carrying a *GIP*^{-1920G} haplotype compared with those homozygous for the *GIP*^{-1920A} haplotype. **B:** In 72 patients with the dominant *GIP*^{-1920G} genotype, both serum *GIP* and glucose levels were both significantly different between *GIP*^{-1920A/A} homozygotes and those carrying a *GIP*^{-1920G} haplotype.

limitation could be particularly pertinent to modifiers associated with adaptations in response to shifts of subsistence cultures.

Consistent with this hypothesis, a survey of earlier studies of genome-wide or chromosome-wide positive selection using the so-called outlier approaches—in which candidate loci are identified in the extreme tails of empiric distributions (40)—showed that *GIP* variants have not been reported as positively selected (1,7). The selection of *GIP* variants was inferred after we focused the analysis on local genomic regions and assessed the significance of integrated haplotype score using coalescent simulations (22). Therefore, our investigation provided a proof-of-concept study for identifying causal mutations that underlie phenotypic variation of complex disease-related traits. This approach could open new venues for improving the translation of common variant association signals into biologic mechanisms that underlie physiologic variability or disease risk.

Neel hypothesized that mismatches between prior adaptations and new environments, or a “conflict of adaptations,” could lead to changes in fitness or health risks (11). Because ancient variants could have been selected for the organism’s reproductive success but not for its health or longevity, the ancient alleles could confer disease risks as selection pressures change. Therefore, studies of positively selected variants that are associated with adaptations in energy-balance regulation could point not only to novel genotype–phenotype relationships but also to novel molecular mechanisms that mediate the potential phenotypic variation, thereby providing much-needed insight into how and which phenotypic variations in energy-balance regulation can be attributed to the selected variants. In support of the thrifty genotype hypothesis, human *CAPN10* and house-mouse insulin genes have been shown to exhibit characteristics of adaptive evolution after the emergence of agricultural societies (41,42). Conceptually, the high *GIP* response associated with the ancestral *GIP*^{-1920G} haplotype could have been a beneficial energy-conserving mechanism

when the food supply was irregular. The ancestral haplotype could become deleterious in the last 10 millenniums as agricultural practice became widespread. One possible deleterious effect of the ancestral *GIP*^{-1920G} haplotype in an environment that supplies abundant high-starch food resources is the hypersecretion of insulin and insulin resistance (43,44).

On the other hand, a reduced *GIP* response associated with the derived *GIP*^{-1920A} haplotype could be protective by decreasing the extent of insulin secretion in the face of oversupply of energy inputs (45,46). In support of this speculation, it has been well documented that in the absence of modern medicine, diabetes-associated complications and, possibly, obesity posed detrimental effects on survival when human culture shifted (47), even though type 2 diabetes is generally considered a chronic disease in modern society. Alternatively, an elevated glucose level associated with the derived haplotype may have improved the survival of fetuses if the population faced serious famine—an event frequently experienced by agricultural societies (48)—despite the reality that an impaired glucose-tolerance response represents a risk to both the mother and fetus under normal circumstances. Moreover, the derived haplotype could have been beneficial by reducing the obesity-promoting effect of *GIP* (16). In vitro and in vivo studies have shown that *GIP* promotes fatty liver and other obesity-associated metabolic disorders, whereas *GIP* antagonists suppress lipid accumulation induced by a high-fat diet (23). Therefore, the derived *GIP*^{-1920A} haplotype could be selected for its effects on the enteroadipocyte or the enteroinsular axis, or both.

Although we speculated that the derived *GIP*^{-1920A} haplotype may have provided protective effects in famine-plagued agricultural societies, the observation that the derived haplotype has not been fixed in any population suggests that the selection of the derived *GIP* haplotype(s) (e.g., cycles of famine) could be opposed (balance selection) as populations experienced temporal changes in selection pressure (e.g., resumption of population growth

with stable food supply). Alternatively, the derived *GIP* haplotype could simply be too young to become fixed, or the spread could be limited by the transgenerational effects associated with abnormal gestational glucose metabolism, which raise the risk of macrosomia and diabetes in the offspring (11).

Although an association between *GIP* variants and glucose-metabolism regulation has not been reported, *GIPR* variants were associated with glucose and insulin levels after challenge tests as well as with BMI in GWA studies that evaluated >29,000 individuals (35,36). The finding that patients with a homozygous *GIP*^{-1920A/A} genotype have significantly higher glucose levels compared with those carrying an ancestral *GIP*^{-1920G} haplotype within the pool of *GIPR*^{I159G/G} homozygotes suggested that there is a confounding effect stemming from interactions of *GIP* and *GIPR* variants, and that *GIP* and *GIPR* variants represent novel markers for the stratification of the capability to maintain glucose homeostasis during pregnancy.

We also speculate that the significant results observed in pregnant women could be related to the fact that the success of pregnancy has a significant impact on reproductive fitness and that a major fraction of gene-environmental selections probably occurred before birth (49). Recent studies have corroborated this idea by showing that associations between many risk alleles and type 2 diabetes can be replicated with smaller sample sizes in patients with gestational diabetes mellitus (50,51).

Furthermore, given the evolutionary signatures at the *GIP* locus, the plausible molecular mechanism, and the significant results in East Asian women, we speculate that the *GIP* variant-mediated phenotypic divergence could also exist in most human populations. It is also important to note that the selection of *GIP* variants represents a unique example in which the selection process involves regulatory variants that alter the glucose-induced *GIP* response as well as a nonsynonymous variant that affects peptide bioactivity (22). Thus, the *GIP* signaling pathway could represent a hotspot for selection in recent human history and play an important role in the manifestation of phenotypic variation in energy-balance regulation among individuals.

In addition to *GIP* variants, our study identified several *CDKAL1*, *CYB5R4*, *GAD2*, and *PPARG* variants as potential metabolic modifiers. Recent studies have shown that variants in *CDKAL1*, *PPARG*, and more than two dozen genes are associated with glycemic traits in diabetic patients (10). Surprisingly, none of the *CDKAL1*, *CYB5R4*, *GAD2*, and *PPARG* variants identified here have been implicated in earlier GWA studies, which suggests that these variants could be related to novel energy-balance regulatory mechanisms that operate at certain life stages or under specific physiologic conditions that have not been specifically investigated. Future investigations of these variants could reveal additional metabolic modifiers that have arisen recently and their contributions to phenotypic variation in normal human physiology and metabolic syndrome-related traits.

In conclusion, our data demonstrated a strong association between regulatory *GIP* variants, and *GIP* response and glucose metabolism, reinforcing the indication of an important role of *GIP* signaling in diabetes-related traits from earlier GWA studies of *GIPR*. Importantly, our study also provided a novel approach to reveal metabolic modifiers by studying consequences of previous mismatches of physiologic capabilities and environments.

TABLE 3
Association between *GIP* haplotypes and *GIP* response after glucose challenge tests in 123 pregnant Han-Chinese women

Genotypes	Patients (N [%])	<i>GIP</i> (pg/mL)	Glucose (mg/dL)	>140* (%)	OR (95% CI)	<i>P</i>	Insulin (mU/L)	C-peptide (ng/mL)	Age (years)	BMI (kg/m ²)
<i>GIP</i> ^{-1920G/G} + <i>GIP</i> ^{-1920G/A}	70 (0.569)	85.39 ± 5.41	118.0 ± 3.3	20.0	—	—	35.57 ± 3.18	5.13 ± 0.28	31.6 ± 0.5	25.4 ± 0.4
<i>GIP</i> ^{-1920A/A}	53 (0.431)	70.87 ± 3.83†	125.1 ± 3.4	34.0	2.06 (0.98–4.65)	0.081	44.36 ± 5.16	5.35 ± 0.38	30.1 ± 0.7	26.3 ± 0.7
Total	123 (1.000)	79.13 ± 3.54	121.2 ± 2.5	26.0	—	—	39.38 ± 2.88	5.22 ± 0.23	31.0 ± 0.4	25.7 ± 0.4
<i>GIPR</i> ^{I159C/C}	8 (0.065)	65.92 ± 9.13	126.0 ± 8.2	37.5	—	—	30.84 ± 8.05	5.60 ± 0.86	33.0 ± 2.0	25.9 ± 1.2
<i>GIPR</i> ^{I159C/G}	43 (0.350)	79.93 ± 6.64	120.0 ± 4.0	14.0	—	—	35.77 ± 4.19	5.36 ± 0.35	32.3 ± 0.7	26.1 ± 0.6
<i>GIPR</i> ^{I159G/G}	72 (0.585)	80.12 ± 4.47	121.3 ± 3.3	31.9	—	—	42.49 ± 4.14	5.10 ± 0.32	30.2 ± 0.5	25.5 ± 0.5
Total	123 (1.000)	79.13 ± 3.54	121.2 ± 2.5	26.0	—	—	39.38 ± 2.88	5.22 ± 0.23	31.0 ± 0.4	25.7 ± 0.4
<i>GIP</i> ^{-1920G/G} + <i>IP</i> ^{-1920G/A}	43 (0.597)	86.98 ± 6.57	115.2 ± 4.2	20.9	—	—	35.91 ± 3.91	4.74 ± 0.36	31.4 ± 0.6	25.1 ± 0.6
<i>GIP</i> ^{-1920A/A}	29 (0.403)	69.95 ± 4.86†	130.2 ± 5.0†	48.3†	3.53 (1.25–9.92)	0.015	52.70 ± 8.41	5.66 ± 0.59	28.4 ± 0.9†	25.7 ± 0.8
Total	72 (1.000)	80.12 ± 4.47	121.3 ± 3.2	31.9	—	—	42.49 ± 4.14	5.11 ± 0.32	30.2 ± 0.5	25.3 ± 0.5

GIP, glucose-dependent insulinotropic polypeptide; OR, odds ratio. *Measured as mg/dL serum glucose level. †Significantly different from patients with a *GIP*^{-1920G} haplotype (χ^2 test, $P < 0.05$).

ACKNOWLEDGMENTS

S.Y.T.H. received support from a National Institutes of Health award (DK-70652) and the Avon Foundation (02-2009-054). C.L.C. received support from the Chang Gung Memorial Hospital (CMRPG34002).

No potential conflicts of interest relevant to this article were reported.

C.L.C. was involved in DNA and functional testing, was responsible for the sample collection, and primarily wrote the manuscript. J.J.C. processed the raw genome data and performed haplotype analyses, and primarily wrote the manuscript. P.J.C. and H.Y.C. were responsible for the sample collection. S.Y.T.H. was involved in DNA and functional testing, primarily wrote the manuscript, and conceived and supervised the study.

The authors thank Shripa Patel (Stanford University Pan Facility), Wei Yi (Department of Obstetrics/Gynecology, Stanford University), and An Shine Chao and Yao Lung Chang (Department of Obstetrics/Gynecology, Chang Gung Memorial Hospital Linkou Medical Center) for technical assistance. The authors thank Drs. Aaron J.W. Hsueh and Renee A. Reijo Pera (Department of Obstetrics/Gynecology, Stanford University) for helpful comments on the manuscript. C.L.C. further expresses gratitude to Dr. Yung-Kuei Soong (Department of Obstetrics/Gynecology, Chang Gung Memorial Hospital, Taiwan). J.J.C. thanks Dr. Dmitri Petrov (Department of Biology, Stanford University) for his invaluable advice and long-lasting support.

REFERENCES

- Akey JM. Constructing genomic maps of positive selection in humans: where do we go from here? *Genome Res* 2009;19:711–722
- Nei M, Suzuki Y, Nozawa M. The neutral theory of molecular evolution in the genomic era. *Annu Rev Genomics Hum Genet* 2010;11:265–289
- Ng SB, Bigham AW, Buckingham KJ, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* 2010;42:790–793
- Leslie M. Genetics. Kidney disease is parasite-slaying protein's downside. *Science* 2010;329:263
- Simonson TS, Yang Y, Huff CD, et al. Genetic evidence for high-altitude adaptation in Tibet. *Science* 2010;329:72–75
- Donnelly P. Progress and challenges in genome-wide association studies in humans. *Nature* 2008;456:728–731
- Pickrell JK, Coop G, Novembre J, et al. Signals of recent positive selection in a worldwide sample of human populations. *Genome Res* 2009;19:826–837
- Sabeti PC, Varilly P, Fry B, et al.; International HapMap Consortium. Genome-wide detection and characterization of positive selection in human populations. *Nature* 2007;449:913–918
- Tishkoff SA, Reed FA, Ranciaro A, et al. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet* 2007;39:31–40
- Voight BF, Scott LJ, Steinthorsdottir V, et al.; MAGIC investigators; GIANT Consortium. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* 2010;42:579–589
- Neel JV. Diabetes mellitus: a “thrifty” genotype rendered detrimental by “progress”? *Am J Hum Genet* 1962;14:353–362
- Laland KN, Odling-Smee J, Myles S. How culture shaped the human genome: bringing genetics and the human sciences together. *Nat Rev Genet* 2010;11:137–148
- Pritchard JK, Pickrell JK, Coop G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol* 2010;20:R208–R215
- Charlesworth D. Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet* 2006;2:e64
- Zhong H, Beaulaurier J, Lum PY, et al. Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes. *PLoS Genet* 2010;6:e1000932
- Kim W, Egan JM. The role of incretins in glucose homeostasis and diabetes treatment. *Pharmacol Rev* 2008;60:470–512
- Schmidt C, Gonzaludo NP, Strunk S, et al. A meta-analysis of QTL for diabetes-related traits in rodents. *Physiol Genomics* 2008;34:42–53
- Walley AJ, Asher JE, Froguel P. The genetic contribution to non-syndromic human obesity. *Nat Rev Genet* 2009;10:431–442
- Loos RJ. Recent progress in the genetics of common obesity. *Br J Clin Pharmacol* 2009;68:811–829
- Prokopenko I, McCarthy MI, Lindgren CM. Type 2 diabetes: new genes, new understanding. *Trends Genet* 2008;24:613–621
- Durbin RM, Abecasis GR, Altshuler DL, et al.; 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061–1073
- Chang CL, Cai JJ, Lo C, Amigo J, Park JI, Hsu SYT. Adaptive selection of an incretin gene in Eurasian populations. *Genome Res* 2011;21:21–32
- McIntosh CH, Widenmaier S, Kim SJ. Glucose-dependent insulinotropic polypeptide (Gastric Inhibitory Polypeptide; GIP). *Vitam Horm* 2009;80:409–471
- Kjos SL, Buchanan TA. Gestational diabetes mellitus. *N Engl J Med* 1999;341:1749–1756
- Cai JJ. PGEToolbox: A Matlab toolbox for population genetics and evolution. *J Hered* 2008;99:438–440
- Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263–265
- Weir BS. *Genetic Data Analysis II: Methods for Discrete Population Genetic Data*. Sunderland, MA, Sinauer Associates, 1996
- Boutin P, Dina C, Vasseur F, et al. GAD2 on chromosome 10p12 is a candidate gene for human obesity. *PLoS Biol* 2003;1:E68
- Hansen L, Ekstrøm CT, Tabanera Y, Palacios R, Anant M, Wassermann K, Reinhardt RR. The Pro12Ala variant of the PPARG gene is a risk factor for peroxisome proliferator-activated receptor-gamma/alpha agonist-induced edema in type 2 diabetic patients. *J Clin Endocrinol Metab* 2006;91:3446–3450
- Steinthorsdottir V, Thorleifsson G, Reynisdottir I, et al. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet* 2007;39:770–775
- Silander K, Scott LJ, Valle TT, et al. A large set of Finnish affected sibling pair families with type 2 diabetes suggests susceptibility loci on chromosomes 6, 11, and 14. *Diabetes* 2004;53:821–829
- Smith JM, Haigh J. The hitch-hiking effect of a favourable gene. *Genet Res* 1974;23:23–35
- The International HapMap Project, Frazer KA, Ballinger DG, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007;449:851–861
- Altshuler DM, Gibbs RA, Peltonen L, et al.; International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* 2010;467:52–58
- Ingelsson E, Langenberg C, Hivert MF, et al.; MAGIC investigators. Detailed physiologic characterization reveals diverse mechanisms for novel genetic loci regulating glucose and insulin metabolism in humans. *Diabetes* 2010;59:1266–1275
- Saxena R, Hivert MF, Langenberg C, et al.; GIANT consortium; MAGIC investigators. Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat Genet* 2010;42:142–148
- Cai JJ, Macpherson JM, Sella G, Petrov DA. Pervasive hitchhiking at coding and regulatory sites in humans. *PLoS Genet* 2009;5:e1000336
- Coop G, Pickrell JK, Novembre J, et al. The role of geography in human adaptation. *PLoS Genet* 2009;5:e1000500
- Yi X, Liang Y, Huerta-Sanchez E, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 2010;329:75–78
- Kelley JL, Madeoy J, Calhoun JC, Swanson W, Akey JM. Genomic signatures of positive selection in humans and the limits of outlier approaches. *Genome Res* 2006;16:980–989
- Vander Molen J, Frisse LM, Fullerton SM, et al. Population genetics of CAPN10 and GPR35: implications for the evolution of type 2 diabetes variants. *Am J Hum Genet* 2005;76:548–560
- Shiao MS, Liao BY, Long M, Yu HT. Adaptive evolution of the insulin two-gene system in mouse. *Genetics* 2008;178:1683–1691
- Lamont BJ, Drucker DJ. Differential antidiabetic efficacy of incretin agonists versus DPP-4 inhibition in high fat fed mice. *Diabetes* 2008;57:190–198
- Gniuli D, Calcagno A, Dalla Libera L, et al. High-fat feeding stimulates endocrine, glucose-dependent insulinotropic polypeptide (GIP)-expressing cell hyperplasia in the duodenum of Wistar rats. *Diabetologia* 2010;53:2233–2240
- Isken F, Pfeiffer AF, Nogueiras R, et al. Deficiency of glucose-dependent insulinotropic polypeptide receptor prevents variety-induced obesity in mice. *Am J Physiol Endocrinol Metab* 2008;295:E350–E355
- McClellan PL, Irwin N, Cassidy RS, Holst JJ, Gault VA, Flatt PR. GIP receptor antagonism reverses obesity, insulin resistance, and associated

- metabolic disturbances induced in mice by prolonged consumption of high-fat diet. *Am J Physiol Endocrinol Metab* 2007;293:E1746–E1755
47. LANCET T. Thrifty genotype rendered detrimental by progress? *Lancet* 1989;2:839–840
48. Prentice AM, Rayco-Solon P, Moore SE. Insights from the developing world: thrifty genotypes and thrifty phenotypes. *Proc Nutr Soc* 2005;64:153–161
49. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk of complex disease. *Curr Opin Genet Dev* 2008;18:257–263
50. Cho YM, Kim TH, Lim S, et al. Type 2 diabetes-associated genetic variants discovered in the recent genome-wide association studies are related to gestational diabetes mellitus in the Korean population. *Diabetologia* 2009;52:253–261
51. Lauenborg J, Grarup N, Damm P, et al. Common type 2 diabetes risk gene variants associate with gestational diabetes. *J Clin Endocrinol Metab* 2009;94:145–150
52. Voight BF, Kudravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol* 2006;4:e72