

# Genetic and Functional Assessment of the Role of the rs13431652-A and rs573225-A Alleles in the *G6PC2* Promoter That Are Strongly Associated With Elevated Fasting Glucose Levels

Nabila Bouatia-Naji,<sup>1,2</sup> Amélie Bonnefond,<sup>1,2</sup> Devin A. Baerenwald,<sup>3</sup> Marion Marchand,<sup>1,2</sup> Marco Bugliani,<sup>4</sup> Piero Marchetti,<sup>4</sup> François Pattou,<sup>5</sup> Richard L. Printz,<sup>3</sup> Brian P. Flemming,<sup>3</sup> Obi C. Umunakwe,<sup>3</sup> Nicholas L. Conley,<sup>3</sup> Martine Vaxillaire,<sup>1,2</sup> Olivier Lantieri,<sup>6</sup> Beverley Balkau,<sup>7</sup> Michel Marre,<sup>8</sup> Claire Lévy-Marchal,<sup>9</sup> Paul Elliott,<sup>10</sup> Marjo-Riitta Jarvelin,<sup>10,11</sup> David Meyre,<sup>1,2</sup> Christian Dina,<sup>1,2</sup> James K. Oeser,<sup>3</sup> Philippe Froguel,<sup>1,2,12</sup> and Richard M. O'Brien<sup>3</sup>

**OBJECTIVE**—Genome-wide association studies have identified a single nucleotide polymorphism (SNP), rs560887, located in a *G6PC2* intron that is highly correlated with variations in fasting plasma glucose (FPG). *G6PC2* encodes an islet-specific glucose-6-phosphatase catalytic subunit. This study examines the contribution of two *G6PC2* promoter SNPs, rs13431652 and rs573225, to the association signal.

**RESEARCH DESIGN AND METHODS**—We genotyped 9,532 normal FPG participants (FPG <6.1 mmol/l) for three *G6PC2* SNPs, rs13431652 (distal promoter), rs573225 (proximal promoter), rs560887 (3rd intron). We used regression analyses adjusted for age, sex, and BMI to assess the association with FPG and haplotype analyses to assess comparative SNP contributions. Fusion gene and gel retardation analyses characterized the effect of rs13431652 and rs573225 on *G6PC2* promoter activity and transcription factor binding.

**RESULTS**—Genetic analyses provide evidence for a strong contribution of the promoter SNPs to FPG variability at the *G6PC2* locus (rs13431652:  $\beta = 0.075$ ,  $P = 3.6 \times 10^{-35}$ ; rs573225  $\beta = 0.073$   $P = 3.6 \times 10^{-34}$ ), in addition to rs560887 ( $\beta = 0.071$ ,  $P = 1.2 \times 10^{-31}$ ). The rs13431652-A and rs573225-A alleles

promote increased NF-Y and Foxa2 binding, respectively. The rs13431652-A allele is associated with increased FPG and elevated promoter activity, consistent with the function of *G6PC2* in pancreatic islets. In contrast, the rs573225-A allele is associated with elevated FPG but reduced promoter activity.

**CONCLUSIONS**—Genetic and in situ functional data support a potential role for rs13431652, but not rs573225, as a causative SNP linking *G6PC2* to variations in FPG, though a causative role for rs573225 in vivo cannot be ruled out. *Diabetes* 59:2662–2671, 2010

The glucose-6-phosphatase catalytic subunit gene family comprises three members, *G6PC*, *G6PC2*, and *G6PC3* (1). *G6PC2*, also known as the IGRP gene (2–4), is principally expressed in the  $\beta$ -cells of pancreatic islets (5). Whether *G6PC2* accounts for the low glucose-6-phosphatase enzyme activity detected in islets is unclear (2,3,6,7); however, a global knockout of *G6pc2* results in a mild metabolic phenotype characterized by a  $\sim 15\%$  decrease in fasting blood glucose (4). This observation suggests that *G6PC2* may oppose the action of glucokinase and therefore modulate  $\beta$ -cell glycolytic flux and glucose-stimulated insulin secretion (8). This hypothesis is consistent with recent genetic studies in humans. Thus, using a genome-wide association approach to study the genetic basis for variations in fasting plasma glucose levels (FPG), we recently identified strong association signals in and around *G6PC2* (9). FPG is an important metabolic trait that is correlated with cardiovascular-associated mortality (10,11). We showed that rs560887, a common variant located in the 3rd intron of *G6PC2*, may explain  $\sim 1\%$  of the total variance in FPG. The rs560887-A allele is associated not only with elevated FPG, but also with long-term glucose regulation, as estimated by elevated glycated hemoglobin A1C levels, decreased basal insulin secretion, as assessed by Homa%B, and increased risk of incidence of impaired fasting glucose over time using prospective data from the DESIR study (9). Data from an independent study reported similar findings (12), and another meta-analysis confirmed that the *G6PC2* locus harbors the strongest genetic determinant of FPG in terms of effect size and significance (13). However, a functional link between the genetic variation in the *G6PC2* locus and variations in FPG remains to be determined.

Two genetic variants located in the *G6PC2* promoter

From the <sup>1</sup>CNRS-UMR-8199, Institut Pasteur de Lille, Lille, France; the <sup>2</sup>University Lille Nord de France, Lille, France; the <sup>3</sup>Department of Molecular Physiology and Biophysics, Vanderbilt University School of Medicine, Nashville, Tennessee; the <sup>4</sup>Department of Endocrinology and Metabolism, University of Pisa, Pisa, Italy; <sup>5</sup>INSERM U859, Université de Lille-Nord de France, Centre Hospitalier Régional et Universitaire de Lille, Lille, France; the <sup>6</sup>Institut Inter-Régional pour la Santé, La Riche, France; <sup>7</sup>INSERM U780, Villejuif; University Paris-Sud, Orsay, France; the <sup>8</sup>Department of Endocrinology, Diabetology and Nutrition, Bichat-Claude Bernard University Hospital, Assistance Publique des Hôpitaux de Paris, Paris, France; INSERM U695, Université Paris 7, Paris, France; <sup>9</sup>INSERM U690, Robert Debré Hospital, Paris; Paris Diderot University, Paris, France; the <sup>10</sup>Department of Epidemiology and Public Health, Imperial College London, London, U.K.; the <sup>11</sup>Institute of Health Sciences, University of Oulu, Department of Child and Adolescent Health, National Public Health Institute, Biocenter Oulu, University of Oulu, Oulu, Finland; and the <sup>12</sup>Department of Genomics of Common Disease, School of Public Health, Imperial College London, London, U.K.

Corresponding author: Richard M. O'Brien, richard.obrien@vanderbilt.edu. Received 22 March 2010 and accepted 27 June 2010. Published ahead of print at <http://diabetes.diabetesjournals.org> on 9 July 2010. DOI: 10.2337/db10-0389.

The present address for C.D. is INSERM UMR915 and CNRS ERL3147, Institut du Thorax, Nantes, France.

N.B.-N. and A.B. contributed equally to this work.

© 2010 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.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

region, rs13431652 and rs573225, are the only common variants (MAF >0.05) that show high LD (>0.80) with rs560087, according to HapMap phase III data ( $r^2 = 0.88$  and  $r^2 = 0.96$ , respectively), but their precise correlation with rs560087 and their contribution to the association with FPG in large and independent samples have not been investigated. In this report, we provide genetic support for the contribution of rs13431652 and rs573225 to the association signal between *G6PC2* and FPG, but we do not exclude a role for rs560087. We also characterize the effect of these two *G6PC2* promoter variants on transcription factor binding and *G6PC2* fusion gene expression. We show that the rs13431652-A allele affects binding of the transcription factor NF-Y and that it is associated with both elevated FPG and elevated fusion gene expression, a correlation that is consistent with the function of *G6PC2* (4) since elevated *G6PC2* expression would oppose the action of glucokinase leading to elevated FPG. In contrast, we show that the rs573225-A allele affects binding of the transcription factor Foxa2 and that it is associated with elevated FPG but reduced fusion gene expression. These functional data support a potential role for rs13431652, but not rs573225, as a causative single nucleotide polymorphism (SNP) linking *G6PC2* to variations in FPG.

## RESEARCH DESIGN AND METHODS

**Study participants.** We analyzed 9,532 normal fasting glucose (NG) Europeans drawn from three general population studies (DESIR,  $N = 3,483$ ; NFBC 1986,  $N = 4,372$ ; and the Haguenu Study,  $N = 1,201$ ) and one childhood obesity case population ( $N = 476$ ).

**DESIR.** The data from the Epidemiological Study on the Insulin Resistance Syndrome (DESIR) cohort is a longitudinal French general population cohort that is fully described elsewhere (9,14). Because ethnicity could not be legally documented in the DESIR study, the proportion of subjects having non-European ancestry was estimated to be low (~0.30%) (15). We also excluded all individuals born outside metropolitan France before analysis. We analyzed 3,483 normoglycemic (NG) DESIR participants successfully genotyped for all SNPs with fasting plasma glucose (FPG) available (supplementary Table 1 available in an online appendix at <http://diabetes.diabetesjournals.org/cgi/content/full/db10-0389/DC1>). The study protocol was approved by the Ethics Committee of Bicêtre Hospital (Paris, France).

**NFBC 1986.** The Northern Finland 1986 Birth Cohort (NFBC 1986) is a prospective 1-year birth cohort including all white Caucasian mothers with children whose expected dates of birth fell between July 1, 1985, and June 30, 1986, in the two northernmost provinces of Finland (16). The clinical examinations were conducted between August 2001 and June 2002. All cohort members living in Finland with known addresses were invited, and 6,798 (74%) participated. Overall, we analyzed 4,372 NG individuals of the NFBC 1986 cohort successfully genotyped for all SNPs with FPG available (supplementary Table 1). The study protocol was approved by the Ethics Committee of the Faculty of Medicine of the University of Oulu, Finland and the Finnish Ministry of Social and Health Affairs.

**Obese children.** NG obese French children with European ancestry were selected from obesity nuclear families recruited by the CNRS-UMR8090 unit (Lille, France) through an ongoing national media campaign (17). We analyzed 476 NG unrelated obese children (defined as BMI >97th percentile for age and sex according to a French population-based cohort) (18) who were successfully genotyped for all SNPs with FPG available (supplementary Table 1).

**Haguenu.** The Haguenu population is a community-based cohort of young adults that investigates the long-term consequences of being born small for gestational age and was fully described elsewhere (19). Briefly, subjects born between 1971 and 1985 were identified from a population-based registry of Haguenu, France. Non-European ancestry subjects are estimated to be <0.1% of the general population (19). At a mean age of 22 years, participants underwent a medical examination to assess anthropometric and clinical parameters. We analyzed 1,201 NG subjects successfully genotyped for all SNPs with FPG available (supplementary Table 1). The study protocol was approved by the Ethics Committee of Paris St. Louis University (Paris, France).

All study participants and parents of children signed an informed consent form. For each population, glycemic status was defined according to the 1997 American Diabetes Association criteria (20): normal glucose, defined as FPG

<6.1 mmol/l without hypoglycemic treatment; impaired fasting glucose, defined as FPG between 6.1 and 6.99 mmol/l without hypoglycemic treatment; and type 2 diabetes, defined as FPG  $\geq 7.0$  mmol/l and/or treatment with antidiabetic agents.

**Genotyping.** Genotyping was performed using either TaqMan assays or SNPlex technology (Applied Biosystems) according to the manufacturer's instructions. The genotyping success rate was at least 95% and genotype concordance between the two methods is more than 99% (internal data). No significant deviation ( $P > 0.05$ ) from the Hardy-Weinberg equilibrium was observed. To allow optimal analytical conditions, all individuals in this study were successfully genotyped for the four SNPs studied here.

**Genotype/expression correlation in human pancreatic cDNAs.** Please see the online appendix (available at <http://diabetes.diabetesjournals.org/cgi/content/full/db10-0389/DC1>).

**G6PC2 expression pattern in human tissues.** We used commercial cDNAs from the Human Multiple Tissue cDNA Panel I for lung, liver, heart, skeletal muscle, kidney, placenta, and pancreas (BD Biosciences Clontech), and RNAs that were reverse-transcribed from the brain, small intestine, and adipose tissue (Human Adult Normal 5 Donor Pool, BioChain Institute). Pancreatic islets and sorted  $\beta$ -cells were obtained from human adult brain-dead donors in accordance with French regulations and the local institutional ethical committee, as previously described (21). Briefly, pancreatic islets were isolated after ductal distension of the pancreata and digestion of the tissue with Liberase (Roche Diagnostics). Human  $\beta$ -cells were sorted by fluorescence-activated cell sorter (FACS) analysis of semipurified preparations of islet cells using Newport Green, a zinc-specific fluorescent probe (21). Total RNA was extracted using a Nucleospin RNA II kit (Macherey Nagel) according to the manufacturer's instructions. Samples were treated with DNase I (Ambion) to ensure that residual genomic contamination was removed. For each sample, 1  $\mu$ g of total RNA was used to generate cDNA by random, primed, first-strand synthesis (Applied Biosystems) according to manufacturer's protocol. Resulting cDNAs were diluted 10-fold, and 4  $\mu$ l of each sample was used in a 20- $\mu$ l quantitative RT-PCR reaction including 10  $\mu$ l of TaqMan gene expression mastermix (Applied Biosystems) and 1  $\mu$ l of the appropriate TaqMan gene expression assay (Applied Biosystems). Quantitative RT-PCR analyses were performed with the ABI 7900 HT SDS 2.3 software, and each sample was run in triplicate. *G6PC2* (assay-on-demand ID: Hs01549773\_m1, sequence context: ACAGGTCCAGGAAGTCCATCTGGCC) expression levels were obtained relative to the *POLR2A* housekeeping gene (ID: Hs00172187\_m1, sequence context: CGGATGAACTGAAGCGAATGTCTGT). Content of cDNA samples was normalized by subtracting the number of copies of the mean of the housekeeping gene from the number of copies of the target gene ( $\Delta Ct = Ct$  of target gene -  $Ct$  of housekeeping gene). Expression of the specific gene was calculated using the formula  $100 \times 2^{-\Delta Ct}$ .

**Fusion gene plasmid construction.** The construction of mouse *G6pc2*-chloramphenicol acetyltransferase (CAT) fusion genes containing the wild-type promoter sequence from -306 to +3, or the same sequence with a site-directed mutation (SDM) of the Foxa2 binding site have been previously described (22,23). The construction of a human *G6PC2*-CAT fusion gene containing the wild-type promoter sequence from -324 to +3 has also been previously described (3). A three-step PCR strategy (24) was used to introduce the rs573225-G allele (Fig. 4A) or a two base pair mutation (supplementary Table 6) into the Foxa2 binding site in the human *G6PC2* promoter within the context of the -324 to +3 promoter fragment. All promoter fragments were subcloned into the pGL3 MOD and/or pGL4 MOD luciferase expression vectors (25). pGL3 MOD and pGL4 MOD were generated by replacing the polylinker in the pGL3-Basic or pGL4.16 vectors (Promega) with a polylinker containing the following restriction endonuclease recognition sites: *KpnI*, *BamHI*, *HindIII*, *XbaI*, *XhoI*, and *BglII*.

A 8,574 base pair *Mfe* I DNA restriction fragment extending from -8,563 to +11 relative to the human *G6PC2* gene transcription start site was isolated from the previously described Pac 294 clone (3). This *Mfe* I DNA fragment was ligated into the *EcoRI* site of the pGEM7 vector (Promega) and the resulting plasmid was designated *G6PC2*Pac294MfepGEM7. The alleles of rs573225 and rs13431652 present within the Pac 294 derived -8,563 to +11 *G6PC2* promoter fragment were determined through DNA sequencing. Two additional *G6PC2* subclones were then generated in pGEM7 so as to change the sequences of rs573225 and rs13431652 to their alternative alleles by site-directed mutagenesis. One of these subclones, designated *G6PC2*Pac294SacShuttlepGEM7, was generated by isolating a *SacI* fragment from *G6PC2*Pac294MfepGEM7 that contained human *G6PC2* promoter sequence from -8,563 to -4,390. The second subclone, designated *G6PC2*Pac294XmaShuttlepGEM7, was generated by isolating an *XmaI* fragment from *G6PC2*Pac294MfepGEM7 that contained human *G6PC2* promoter sequence from -2,803 to +11. The *G6PC2*Pac294SacShuttlepGEM7 plasmid was used as a template along with mutagenesis primers 5'-TGTCAGGAGGCTGTGTCAGTGGAGGGAAG-3' and 5'-CTTCCCTCCAGTGACACAGCCTGCTGACA-3' to convert rs13431652 at position -4,405 from T to C and the

G6PC2Pac294XmaShuttlepGEM7 plasmid was used as a template with mutagenesis primers 5'-GAAATGAACTGTACAAAAAATTTCAAGCAAACATGATCCAACACTGTC-3' and 5'-GAACAGTTGGATCATGTTTGCTTGAATTTTTGTACAGTTTCATTCC-3' to convert rs573225 at position -259 from A to G. Mutagenesis was performed with these templates and primers using a QuikChange II Kit (Stratagene) as described by the manufacturer.

G6PC2-luciferase reporter plasmids, containing promoter sequence from -8,563 to +11, were generated in the pGL3-Basic and pGL4.16 vectors (Promega) by a two-step process. In the initial cloning step, a *SacI* - *HinD* III fragment was isolated from the G6PC2Pac294MfepGEM7 plasmid, using a *SacI* site in G6PC2 at -4,395 and a *HinD* III site within the pGEM7 vector, and ligated into the *SacI* - *HinD* III digested pGL3-Basic and pGL4.16 vectors to generate fusion gene constructs that contain G6PC2 promoter sequence from -4,395 to +11. In the second cloning step a *SacI* fragment was isolated from the G6PC2Pac294SacShuttlepGEM7 plasmid and ligated into the *SacI* digested -4,395 pGL3-Basic and pGL4.16 fusion gene plasmids to extend the G6PC2 promoter sequences from -4,395 to -8,563.

A *SacI* fragment from the G6PC2Pac294SacShuttlepGEM7 plasmid containing the alternate rs13431652 allele and an *XmaI* fragment from the G6PC2Pac294XmaShuttlepGEM7 plasmid with the alternate rs573225 allele were then exchanged with the corresponding *SacI* and/or *XmaI* fragments that contain the original Pac294 sequences to generate pGL3-Basic and pGL4.16 plasmids with the alternate SNP alleles within the context of the G6PC2 promoter region from -8,563 to +11.

**Cell culture, transient transfection and luciferase assays.** Hamster insulinoma tumor (HIT),  $\beta$ TC-3, and Min6 cells were cultured and transfected using lipofectamine as previously described (23). Luciferase assays were performed as previously described (23).

#### Gel retardation assays

**Labeled probes.** Sense and anti-sense oligonucleotides representing wild-type or mutant Foxa binding sites (supplementary Table 6) or the rs573225 variants (Fig. 4A) were synthesized with *Bam*HI compatible ends. Oligonucleotides representing the rs13431652 variants of the NF-Y binding site (Fig. 1A) were synthesized with *Hin*DIII compatible ends. Oligonucleotides were subsequently gel purified, annealed, and labeled with [ $\alpha$ - $^{32}$ P]dATP using the Klenow fragment of *Escherichia coli* DNA Polymerase I to a specific activity of  $\sim 2.5$   $\mu$ Ci/pmol (26).

**Nuclear extract preparation.** The preparation of  $\beta$ TC-3 and H4IIE nuclear extracts through extraction of nuclei with 800 mmol/l NaCl (designated high salt) was as previously described (24). The preparation of rat liver nuclear extract was also performed as previously described (27).

**Binding assays.** Approximately 14 fmol of radiolabeled probe ( $\sim 50,000$  cpm) was incubated with the indicated nuclear extract in a final 20  $\mu$ l reaction volume. Foxa2 binding reactions contained 1.0  $\mu$ g  $\beta$ TC-3, 2.0  $\mu$ g liver, or 2.0  $\mu$ g H4IIE nuclear extract, 20 mmol/l Hepes pH 7.9, 0.1 mmol/l EDTA, 0.1 mmol/l EGTA, 10% glycerol (v/v), 1 mmol/l dithiothreitol, 1  $\mu$ g poly(dI-dC)-poly(dI-dC), and 100 mmol/l KCl. NF-Y binding reactions contained 2.0  $\mu$ g high salt  $\beta$ TC-3 nuclear extract, 20 mmol/l Hepes pH 7.9, 0.1 mmol/l EDTA, 0.1 mmol/l EGTA, 10% glycerol (v/v), 1 mmol/l dithiothreitol, 1  $\mu$ g poly(dI-dC)-poly(dI-dC), and 50 mmol/l KCl. After incubation at room temperature for 20 min, samples were loaded onto a 6% polyacrylamide gel containing 1X TGE (25 mmol/l Tris Base, 190 mmol/l glycine, 1 mmol/l EDTA) and 2.5% (v/v) glycerol. Samples were electrophoresed for 1.5 h at 150V in 1X TGE buffer before the gel was dried and then exposed to Kodak XB film with intensifying screens. Competition and supershift experiments were performed as previously described (24). For supershift experiments, antisera raised against Foxa2 (HNF-3 $\beta$ ) (sc-65540x) and NF-Y (sc-17753x) were obtained from Santa Cruz Biotechnology.

**Statistical analysis.** The effect of SNPs on FPG was analyzed using linear regression models under the additive model adjusted for age, sex, and BMI. The estimates of the effect of each SNP on FPG and their standard errors for each separate population were combined in the meta-analysis using the weighted inverse normal method. The overall effect and its CI were estimated using the inverse variance method implemented using the meta.summaries function of the R meta package. No heterogeneity in effects was observed ( $P > 0.44$ ). All statistical analyses of genetic data were performed with R (version 2.6.1), combined with the survival and rmeta packages. Please see the online supplementary appendix for haplotype analyses.

To test the effect of a specific SNP in comparison with the effects of other SNPs inside a 2  $\times$  2 haplotype, we analyzed haplotype genetic models using the THESIAS program (28) in each population separately. This method estimates the likelihood of specified models for haplotype effects under the general linear model (GLM) framework, and allows testing of constrained models. This analysis aimed at testing 2-SNP models while accounting for the high correlation between SNPs (unlike a plain 2-loci regression model). The model M1 is a 2-SNP (labeled SNP1 and SNP2) model with each haplotype effect being allowed to vary, and 1 and 2 being the at-risk (increasing) alleles

of each SNP. The model Mno1 is a model in which the effect of the first SNP is null, or, translated in haplotype model language,  $\beta_{1-1} = \beta_{2-1}$  and  $\beta_{1-2} = \beta_{2-2}$ . Conversely, Mno2 is a model in which  $\beta_{1-1} = \beta_{1-2}$  and  $\beta_{2-1} = \beta_{2-2}$ . A significant difference between M1 and Mno2 would mean that we could not ignore the effect of SNP2 (supplementary Table 3).

The transfection data were analyzed for differences from the control values, as specified in the figure legends. Statistical comparisons were calculated using an unpaired Student *t* test. The level of significance was  $P < 0.05$  (two-sided test).

## RESULTS

**G6PC2 promoter variants strongly associate with FPG.** Only two common SNPs (minor allele frequency  $> 0.05$ ), namely rs13431652 and rs573225, are reported to be in high linkage disequilibrium (LD) with the previously identified intronic G6PC2 SNP rs560887 ( $r^2 > 0.80$ , according to HapMap data [release XX]). rs13431652 is located in the distal G6PC2 promoter, at -4,405 relative to the transcription start site, whereas rs573225 is located in the proximal promoter at -259. We have assessed the effect of these two G6PC2 promoter variants on FPG, adjusted for age, sex and BMI in 9,532 NG Europeans from four independent populations. In the meta-analyses, we show strong associations of both rs13431652 ( $\beta = 0.075$ ,  $P = 3.6 \times 10^{-35}$ ) and rs573225 ( $\beta = 0.073$ ,  $P = 3.6 \times 10^{-34}$ ) with FPG, in similar magnitude to that for rs560887 ( $\beta = 0.071$ ,  $P = 1.2 \times 10^{-31}$ ) (Table 1). Conditioned regression model analyses to assess the independency of the three SNPs studied turn out to be noninformative (data not shown), as these analyses include a high variance inflation factor because of high LD that we observed between the SNPs in our populations (supplementary Table 2).

To compensate for this limited analytical situation, we analyzed the association of FPG with multiple haplotypes of the promoter variants and rs560887, specifically individual contributions of each variant and contributions when variants were analyzed in pairs and when we constrained the effect of rs560887 to be null (see supplementary METHODS). Because of the LD discrepancies between populations (supplementary Table 2), these haplotype analyses were conducted in the four populations separately (supplementary Table 3). In DESIR, we found that rs560887 is significantly contributing to the haplotype association in combination with rs13431652 ( $P = 7.58 \times 10^{-5}$ ) and rs573225 ( $P = 0.015$ ). We also observed a significant contribution of rs13431652 ( $P = 4.09 \times 10^{-7}$ ) and rs573225 ( $P = 1.21 \times 10^{-8}$ ) to the haplotype associations, supporting independent effects of both promoter variants to the global association of the haplotype in combination with rs560887. In contrast, distinct results were obtained in the remaining cohorts in which we found that both promoter and intronic variants are not independently contributing to the haplotype associations, suggesting that promoter and intronic variants are indispensable to the two by two haplotype associations, with a modestly more significant contribution of rs560887 ( $P = 0.006$  against rs13431652 and  $P = 0.07$  against rs573225) in the NFBC 1986 cohort (supplementary Table 3). We speculate that these distinct results in some way reflect the differences in LD within the different cohorts. Despite these differences, haplotype analyses support a substantial role of promoter variants to the association signal with FPG, and do not discard the contribution of rs560887.

We have also assessed the effect of the intronic variant rs853789 located in the 19th intron of *ABCB11*, one of two SNPs, along with rs560887, that were reported by Chen et al. (12) to be associated with FPG. *ABCB11* (ATP-binding

TABLE 1  
*G6PC2* genetic variation and fasting glucose in four cohorts and one obese population

	Meta-analysis ( $n = 9,532$ )		DESIR ( $n = 3,483$ )		NFBC86 ( $n = 4,372$ )		Haguenu ( $n = 1,201$ )		Obese children ( $n = 476$ )		
	EA (freq)	$\beta$ (SE)	$P$	$\beta$ (SE)	$P$	$\beta$ (SE)	$P$	$\beta$ (SE)	$P$	$P$	
rs13431652	A (0.68)	0.075 (0.006)	$3.6 \times 10^{-35}$	0.083 (0.010)	$2.0 \times 10^{-15}$	0.069 (0.009)	$1.1 \times 10^{-14}$	0.071 (0.015)	$2.0 \times 10^{-6}$	0.086 (0.03)	0.005
rs573225	A (0.66)	0.073 (0.006)	$3.6 \times 10^{-34}$	0.080 (0.010)	$1.3 \times 10^{-14}$	0.066 (0.009)	$6.8 \times 10^{-14}$	0.073 (0.015)	$9.2 \times 10^{-7}$	0.093 (0.03)	0.002
rs560887	G (0.69)	0.071 (0.006)	$1.2 \times 10^{-31}$	0.069 (0.011)	$3.0 \times 10^{-11}$	0.072 (0.009)	$8.7 \times 10^{-16}$	0.076 (0.015)	$4.5 \times 10^{-7}$	0.062 (0.03)	0.04

Individual and meta-analyses data are displayed as regression coefficient  $\beta$  (SE) adjusted for age, sex, BMI, and associated  $P$  values.

cassette family B 11) encodes the major bile salt export pump and is expressed predominantly in the liver (29) and was hypothesized to potentially drive the association with FPG observed at this locus (12). We found a significant association between rs853789 and FPG ( $\beta = -0.062$ ,  $P = 1.1 \times 10^{-25}$ , supplementary Table 4). However, when we include either rs560887, rs13431652, or rs573225 in a regression conditional model that includes rs853789, age, sex, BMI, and cohort, the effect of rs853789 is highly reduced (supplementary Table 4), whereas rs13431652 ( $P = 1.2 \times 10^{-10}$ ), rs573225 ( $P = 1.8 \times 10^{-9}$ ), and rs560887 ( $P = 1.3 \times 10^{-7}$ ) remain strongly associated with FPG (supplementary Table 5). These findings were confirmed by the haplotype analyses that did not support a frank, independent contribution of rs853789 to the association in any of the cohorts studied (supplementary Table 3). These genetic data suggest that in our populations, the *ABCB11* rs853789 effect on FPG is driven mainly by *G6PC2* variation through moderate LD, estimated between 0.42 and 0.78, depending on the population studied (supplementary Table 5).

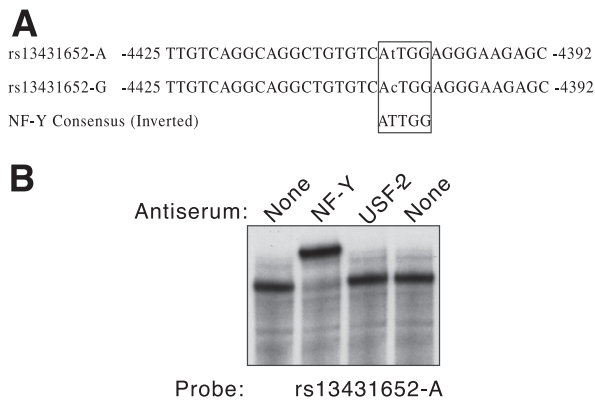
**Expression analyses.** Using quantitative real time PCR we confirmed that the expression of *G6PC2* is restricted to pancreatic tissues, including whole pancreas, islets and sorted  $\beta$ -cells, in humans (supplementary Fig. 1A). Similar results were previously reported through RNA blotting analyses of mouse (2,9) and human (3) RNA. In a limited sample ( $N = 24$ ) in which pancreatic islet cDNAs and corresponding genomic DNAs were available, we found no significant correlation between the alleles of rs13431652-A, rs573225-A, or rs560887-G, that are associated with high FPG and *G6PC2* gene expression (supplementary Fig. 1B).

Because of the limited sample sizes in the genetic and expression analyses, it was not possible to identify a causative variant among rs13431652, rs573225, and rs560887. We therefore decided to take a biochemical approach to assess the ability of the promoter variants to affect transcription factor binding and *G6PC2* promoter activity and, hence, their ability to alter FPG levels.

**NF-Y binds to the *G6PC2* promoter in vitro.** The *G6PC2* promoter region that encompasses rs13431652 was analyzed using MatInspector sequence analysis software (30) with the goal of identifying a *cis*-acting element whose binding of its cognate *trans*-acting factor was likely to be affected by the alternate rs13431652 alleles. This analysis identified a CCAAT box (Fig. 1A), an element known to bind multiple transcription factors, including NF-Y (31,32). The SNP changes the sequence CCAAT to CCAgT.

Gel retardation assays were used to investigate whether NF-Y can bind to this *G6PC2* promoter region in vitro. When a labeled double-stranded oligonucleotide, designated rs13431652-A, representing the *G6PC2* promoter sequence from  $-4,425$  to  $-4,392$  and the rs13431652-A allele (Fig. 1A), was incubated with nuclear extract prepared from  $\beta$ TC-3 cells a single protein-DNA complex was detected (Fig. 1B).

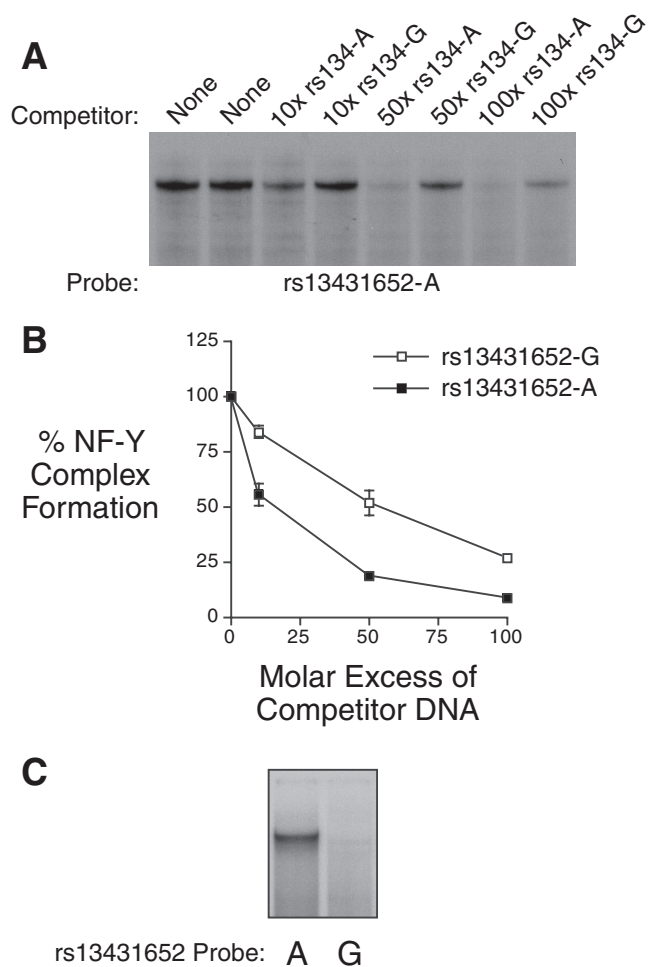
To identify the factor present in this complex, a gel retardation assay was performed in which  $\beta$ TC-3 cell nuclear extract was preincubated with antisera specific for NF-Y or, as a control, USF-2. As can be seen in Fig. 1B, addition of antibodies recognizing NF-Y resulted in a clear supershift in the migration of the complex, whereas the addition of antibodies recognizing USF-2 had no effect. This result strongly suggests that the complex represents NF-Y binding.



**FIG. 1.** NF-Y binds the  $-4,425/-4,392$  *G6PC2* promoter region in vitro. **A:** The sense strand sequences of the oligonucleotides used in gel retardation assays are shown. SNP base pairs are shown in bold lower case letters. The consensus NF-Y binding motif, CCAAT, is taken from Quandt et al. (31). **B:**  $\beta$ TC-3 nuclear extract was incubated in the absence or presence of the indicated anti-serum for 10 min on ice. A labeled oligonucleotide representing the  $-4,425/-4,392$  *G6PC2* promoter region and containing the rs13431652-A allele (rs13431652-A; panel A) was then added and incubation continued for 20 min at room temperature. Protein binding was then analyzed using the gel retardation assay as described in RESEARCH DESIGN AND METHODS. In the representative autoradiograph shown, only the retarded complexes are visible and not the free probe, which was present in excess.

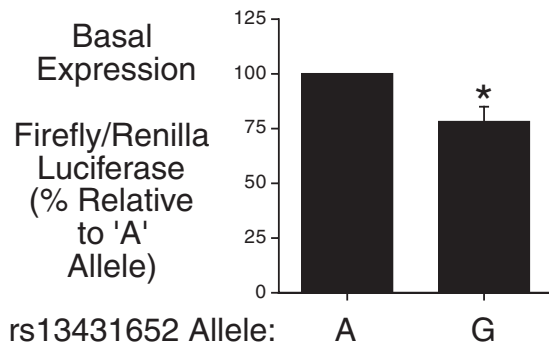
Gel retardation competition experiments, in which a varying molar excess of unlabeled DNA was included with the labeled rs13431652-A oligonucleotide, were used to compare the affinity of NF-Y binding to the rs13431652 variants of the *G6PC2* NF-Y binding site. Fig. 2A shows that the rs13431652-A oligonucleotide competed more effectively than the rs13431652-G oligonucleotide for the formation of the NF-Y-DNA complex and quantitation of the results of several experiments (Fig. 2B) showed that NF-Y binds the rs13431652-A oligonucleotide with approximately fivefold higher affinity. Interestingly, in contrast to the competition experiment data (Fig. 2B), the direct analysis of NF-Y binding to the rs13431652-A and -G oligonucleotide probes, labeled with the identical specific activity, suggested a much more dramatic difference in NF-Y binding affinity to the rs13431652-A and -G oligonucleotides (Fig. 2C). We recently reported a similar discrepancy between the results of gel retardation competition and binding experiments in studies on FOXO1 binding (33). We hypothesize that such apparent discrepancies can arise when both the association and dissociation rate of factor binding to DNA are simultaneously increased since this would result in limited change in  $K_D$  (34), but a marked difference in the dissociation of the protein-DNA complex upon separation of bound and free probe during electrophoresis.

**rs13431652 alters *G6PC2* fusion gene expression.** We next investigated the functional significance of altered NF-Y binding on human *G6PC2* promoter activity. NF-Y can function as either an activator or repressor such that the effect of NF-Y on gene transcription is context dependent (35). Fusion genes containing each of the rs13431652 alleles, generated in the context of the  $-8,563$  to  $+11$  *G6PC2* promoter region, were analyzed by transient transfection of  $\beta$ TC-3 cells. Fig. 3 shows that the rs13431652-G allele was associated with an  $\sim 25\%$  decrease in promoter activity in comparison with that observed with the rs13431652-A allele. Supplementary Fig. 2 shows that this difference was observed only when the *G6PC2* promoter



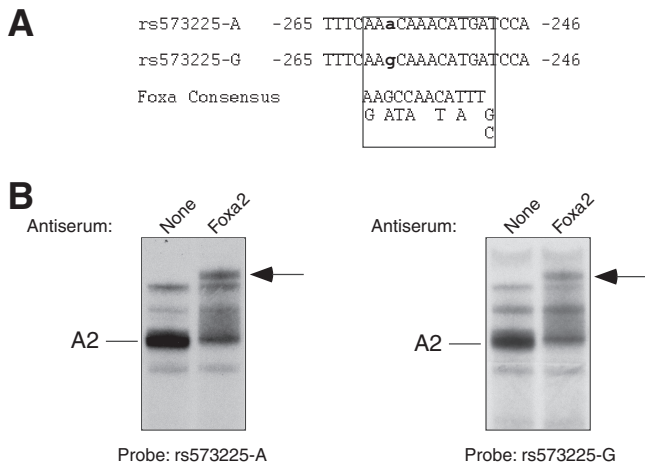
**FIG. 2.** Comparison of the affinity of NF-Y binding to the rs13431652-A and rs13431652-G variants of the *G6PC2* NF-Y binding site in vitro. **A:** The labeled rs13431652-A oligonucleotide (Fig. 1A) was incubated in the absence or presence of the indicated molar excess of the unlabeled human rs13431652-G or rs13431652-A oligonucleotide competitors (Fig. 1A) before the addition of  $\beta$ TC-3 cell nuclear extract. Protein binding was then analyzed using the gel retardation assay as described in RESEARCH DESIGN AND METHODS. In the representative autoradiograph that is shown, only the retarded complexes are visible and not the free probe, which was present in excess. **B:** Protein binding was quantified by using a Packard Instant Imager to count  $^{32}$ P associated with the retarded complex. The data represent the mean  $\pm$  SEM of three experiments. **C:** The labeled rs13431652-A and rs13431652-G oligonucleotides (Fig. 1A) were incubated with  $\beta$ TC-3 cell nuclear extract, and protein binding was analyzed using the gel retardation assay as described in RESEARCH DESIGN AND METHODS. In the representative autoradiograph shown, only the retarded complexes are visible and not the free probe, which was present in excess.

was analyzed in the context of the pGL4, but not the pGL3 vector. The former is an improved vector that lacks multiple transcription factor binding sites in the plasmid backbone and luciferase gene that are known to occasionally give rise to spurious data (23). These results indicate that NF-Y acts as an activator in the context of the *G6PC2* promoter. More importantly, these data are consistent with the reduced FPG observed in *G6pc2* knockout mice (4) in that the rs13431652-A allele is associated with both elevated FPG (Table 1) and *G6PC2* promoter activity. As such, these functional data support a potential role for rs13431652 as a SNP linking *G6PC2* to variations in FPG. **Foxa2 binds to the *G6PC2* promoter in vitro.** The *G6PC2* promoter region that encompasses rs573225 is highly conserved in the mouse *G6pc2* promoter (3). We



**FIG. 3.** The human *G6PC2* rs13431652-G allele is associated with decreased *G6PC2* promoter activity relative to the rs13431652-A allele.  $\beta$ TC-3 cells were transiently cotransfected, as described in RESEARCH DESIGN AND METHODS, using a lipofectamine solution containing various *G6PC2*-luciferase fusion genes in the pGL4 MOD vector (2  $\mu$ g) and an expression vector encoding *Renilla* luciferase (0.5  $\mu$ g). The *G6PC2*-luciferase fusion genes represented the rs13431652-A or rs13431652-G alleles present in the context of the human *G6PC2* promoter sequence located between  $-8,563$  and  $+11$ . After transfection, cells were incubated for 18–20 h in serum-containing medium. The cells were then harvested, and firefly and *Renilla* luciferase activity were assayed as described in RESEARCH DESIGN AND METHODS. Results are presented as the ratio of firefly:*Renilla* luciferase activity, expressed as a percentage relative to the value obtained with the rs13431652-A allele, and represent the mean of 12 experiments  $\pm$  SEM, each using an independent preparation of each fusion gene plasmid, assayed in triplicate. \* $P < 0.05$  vs. rs13431652-A allele.

recently characterized this conserved region in the context of the mouse *G6pc2* promoter and showed that it contains a Foxa2 binding site (23). In addition, we demonstrated, using chromatin immunoprecipitation (ChIP) assays, that Foxa2 binds the endogenous *G6pc2* promoter in mouse  $\beta$ TC-3 cells in situ (23). Fig. 4 shows that this region of the human *G6PC2* promoter also binds Foxa2 in vitro. When labeled double-stranded oligonucleotides, desig-



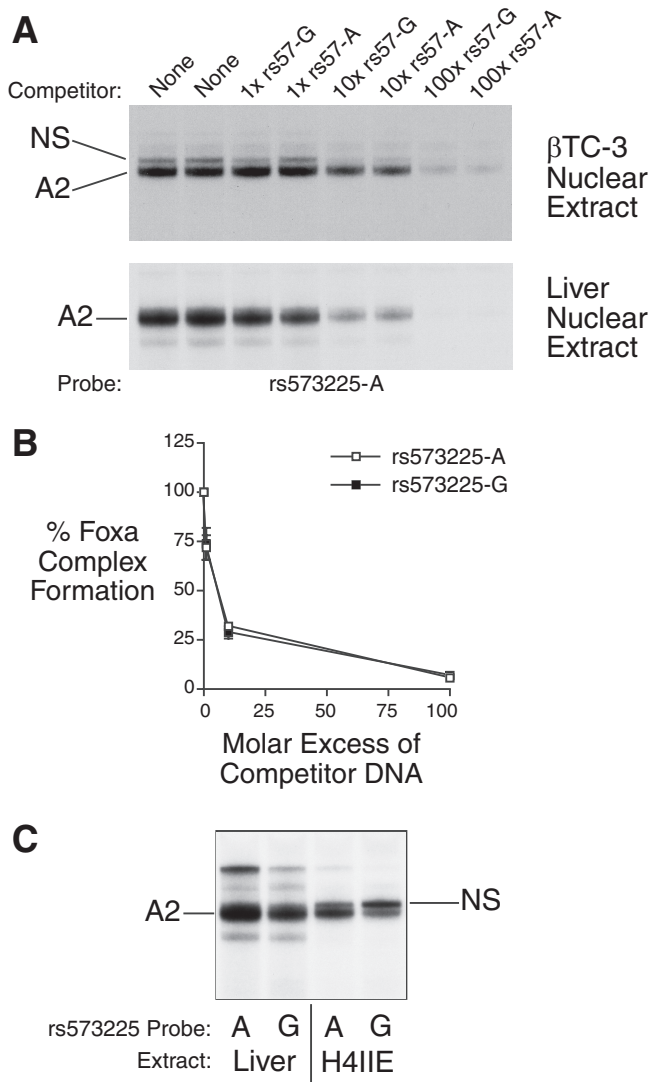
**FIG. 4.** Foxa2 binds the human *G6PC2*  $-265/-246$  promoter region in vitro. **A:** The sense strand sequences of the oligonucleotides used in gel retardation assays are shown. SNP base pairs are shown in bold lower case letters. The consensus Foxa binding motif, (A/G)A(G/A)(C/T)(C/A)AA(C/T)A(T/A)T(T/G/C), is taken from Overdier et al. (46). **B:** Liver nuclear extract was incubated in the absence (–) or presence of the indicated anti-serum for 10 min on ice. Labeled oligonucleotides representing the  $-265/-246$  *G6PC2* promoter region and containing the rs573225-A or rs573225-G alleles (rs573225-A or rs573225-G; panel A) were then added and incubation continued for 20 min at room temperature. Protein binding was then analyzed using the gel retardation assay as described in RESEARCH DESIGN AND METHODS. In the representative autoradiograph shown, only the retarded complexes are visible and not the free probe, which was present in excess. A single major complex was detected (A2) that represents Foxa2 binding. The arrow indicates a supershifted complex.

nated rs573225-A and rs573225-G (Fig. 4A), representing the *G6PC2* promoter sequence from  $-265$  and  $-246$  and the rs573225-A and -G alleles, respectively, were incubated with rat liver nuclear extract, a single major complex, designated A2, was detected with both alleles (Fig. 4B). When rat liver nuclear extract was preincubated with antisera specific for Foxa2, a clear supershift in the migration of complex A2 was observed (Fig. 4B), strongly suggesting that this complex represents Foxa2 binding. This experiment was performed using rat liver nuclear extract because, like pancreatic islets (36), it contains both Foxa1 and Foxa2 (37). In contrast,  $\beta$ TC-3 cells express only Foxa2 (23). The results indicate that rs573225 does not differentially affect binding of these related factors, instead both alleles strongly bind Foxa2.

Gel retardation competition experiments, in which a varying molar excess of unlabeled DNA was included with the labeled rs573225-A probe, were used to compare the affinity of Foxa binding to the rs573225 variants of the *G6PC2* Foxa binding site. Fig. 5A shows that when using rat liver nuclear extract, both the rs573225-A and rs573225-G oligonucleotides competed equally effectively for the formation of the Foxa2-DNA complex. Quantitation of the results of several experiments confirmed that Foxa2 binds the rs573225-A and rs573225-G oligonucleotides with equal affinity (Fig. 5B). Identical results were obtained using mouse  $\beta$ TC-3 cell nuclear extract (Fig. 5A). We have previously shown that the two complexes detected using  $\beta$ TC-3 cell nuclear extract represent the binding of Foxa2 (A2) and an unknown factor that either represents a nonspecific (NS) interaction or binding to another region of the probe (23).

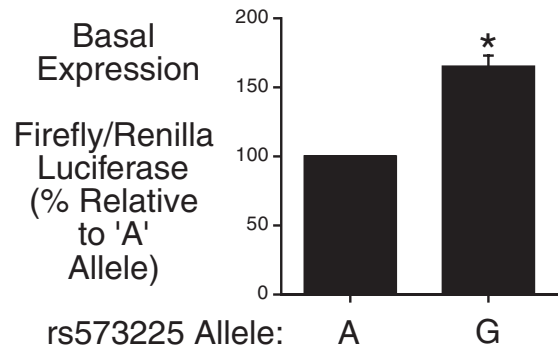
Interestingly, in contrast with the competition experiment data (Fig. 5A and B), the direct analysis of Foxa2 binding to the rs573225-A and rs573225-G oligonucleotide probes labeled with the identical specific activity showed that Foxa2 binds with slightly higher affinity to the rs573225-A allele (Figs. 4B and 5C). An identical observation was made using rat liver and rat H4IIE hepatoma cell nuclear extract (Fig. 5C). As explained above in the analysis of NF-Y binding (Fig. 2C), this suggests that rs573225 alters the association and dissociation rates of Foxa2 binding without affecting binding affinity.

**rs573225 alters *G6PC2* promoter activity in an unexpected direction.** We next investigated the functional significance of altered Foxa2 binding on human *G6PC2* promoter activity. A block mutation in the *G6PC2* Foxa binding site that abolishes Foxa2 binding (supplementary Fig. 3) also markedly reduces fusion gene expression (supplementary Fig. 4), indicating that Foxa2 is an activator in the context of the human *G6PC2* promoter, as it is in the mouse *G6pc2* promoter (23). Fusion genes containing each rs573225 variant generated in the context of the  $-324$  to  $+3$  (Fig. 6) and  $-8,563$  to  $+11$  (Fig. 7) *G6PC2* promoter regions, were analyzed by transient transfection of  $\beta$ TC-3 cells. Since the loss of Foxa2 binding reduces *G6PC2* promoter activity (supplementary Fig. 4), and because the rs573225-G allele decreases Foxa2 binding (Fig. 5C), we anticipated that the rs573225-G allele would also be associated with reduced *G6PC2* promoter activity. Surprisingly, Figs. 6 and 7 show that the rs573225-G allele was associated with an increase in promoter activity in comparison with that observed with the rs573225-A allele. This result is not explained by the de novo creation of an activator binding site by the rs573225-G allele (Fig. 4B). Supplementary Figs. 5 and 6 show that a similar effect of



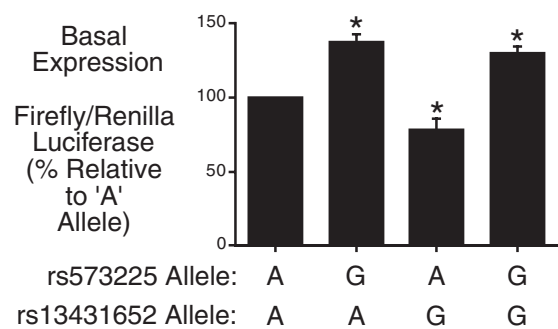
**FIG. 5.** Comparison of the affinity of Foxa binding to the rs573225-A and rs573225-G variants of the *G6PC2* Foxa binding site in vitro. **A:** The labeled rs573225-A oligonucleotide (Fig. 4A) was incubated in the absence or presence of the indicated molar excess of the unlabeled human rs573225-A or rs573225-G oligonucleotide competitors (Fig. 4A) before the addition of  $\beta$ TC-3 cell or rat liver nuclear extract. Protein binding was then analyzed using the gel retardation assay as described in RESEARCH DESIGN AND METHODS. In the representative autoradiograph shown, only the retarded complexes are visible and not the free probe, which was present in excess. With  $\beta$ TC-3 nuclear extract, complex A2 represents Foxa2 binding, and complex NS represents nonspecific binding (23). With liver nuclear extract, complex A2 represents Foxa2 binding (37). **B:** Protein binding in experiments using liver nuclear extract was quantified by using a Packard Instant Imager to count  $^{32}$ P associated with the retarded complex. The data represents the mean  $\pm$  SEM of three experiments. **C:** The labeled rs573225-A and rs573225-G oligonucleotides (Fig. 4A) were incubated with either liver or H4IIE nuclear extract, and protein binding was analyzed using the gel retardation assay as described in RESEARCH DESIGN AND METHODS. In the representative autoradiograph shown, only the retarded complexes are visible and not the free probe, which was present in excess. NS, nonspecific.

the rs573225-G allele was also seen in the HIT and Min6 cell lines and when the *G6PC2* promoter was analyzed in the context of both the pGL4 and pGL3 vectors, respectively. Since the Foxa2 binding site incorporates the target sequence for the bacterial Dam methylase (GATC), we repeated this analysis using plasmids grown in SCS110 Dam and Dcm methylase-deficient bacteria. Supplementary Fig. 7 shows that Dam methylation of the Foxa2



**FIG. 6.** The human *G6PC2* rs573225-G allele is associated with increased *G6PC2* promoter activity relative to the rs573225-A allele.  $\beta$ TC-3 cells were transiently cotransfected, as described in RESEARCH DESIGN AND METHODS, using a lipofectamine solution containing various *G6PC2*-luciferase fusion genes in the pGL3 MOD vector (2  $\mu$ g) and an expression vector encoding *Renilla* luciferase (0.5  $\mu$ g). The *G6PC2*-luciferase fusion genes represented the rs573225-A or -G alleles present in the context of the human *G6PC2* promoter sequence located between -324 and +3. After transfection, cells were incubated for 18–20 h in serum-containing medium. The cells were then harvested, and firefly and *Renilla* luciferase activity was assayed as described in RESEARCH DESIGN AND METHODS. Results are presented as the ratio of firefly:*Renilla* luciferase activity, expressed as a percentage relative to the value obtained with the rs573225-A allele, and represent the mean of three experiments  $\pm$  SEM, each using an independent preparation of each fusion gene plasmid, assayed in triplicate. \* $P$  < 0.05 vs. rs573225-A allele.

binding site did not alter the stimulatory effect of rs573225 on fusion gene expression. Although the increased promoter activity observed with the rs573225-G allele was unexpected, given the reduction in Foxa2 binding, the key observation here is that the functional data obtained with rs573225 are not consistent with the reduced FPG observed in *G6pc2* knockout mice (4) in that the rs573225-A allele is associated with elevated FPG (Table 1) but reduced *G6PC2* promoter activity. As such, in contrast with the functional data obtained with rs13431652, these functional data do not support a potential role for rs573225 as a causative SNP linking *G6PC2* to variations in FPG.



**FIG. 7.** Interaction between the *G6PC2* rs13431652-G and rs573225-G alleles on *G6PC2* promoter activity.  $\beta$ TC-3 cells were transiently cotransfected, as described in RESEARCH DESIGN AND METHODS, using a lipofectamine solution containing various *G6PC2*-luciferase fusion genes in the pGL4 MOD vector (2  $\mu$ g) and an expression vector encoding *Renilla* luciferase (0.5  $\mu$ g). The *G6PC2*-luciferase fusion genes represented the rs573225-A or rs573225-G alleles and rs13431652-A or rs13431652-G alleles present in the context of the human *G6PC2* promoter sequence located between -8,563 and +11. After transfection, cells were incubated for 18 to 20 h in serum-containing medium. The cells were then harvested, and firefly and *Renilla* luciferase activity was assayed as described in RESEARCH DESIGN AND METHODS. Results are presented as the ratio of firefly:*Renilla* luciferase activity, expressed as a percentage relative to the value obtained with the rs573225-A and rs13431652-A alleles, and represent the mean of eight experiments  $\pm$  SEM, each using an independent preparation of each fusion gene plasmid, assayed in triplicate. \* $P$  < 0.05 vs. rs573225-A and rs13431652-A alleles.

Because endogenous *G6PC2* gene expression will reflect the combined effects of multiple SNPs, we investigated the interaction between rs573225 and rs13431652 on *G6PC2* fusion gene expression (Fig. 7). The results show a trend toward the rs13431652-G allele blunting the elevated expression conferred by the rs573225-G allele, although, because of the relatively small effects involved and the inherent variability in transient transfections, the decrease did not reach statistical significance.

## DISCUSSION

Genome-wide association studies have recently provided important new insights about the genetics of common forms of type 2 diabetes and its related quantitative traits, especially for FPG (9,12,38). After this first discovery phase, attention is now beginning to focus on the study of the functional properties of the variants at the confirmed loci. In this study using a combination of genetic and functional data, we provide further genetic support for the primary contribution of *G6PC2* to variations in FPG, a conclusion supported by other genetic studies (13,39–41). Furthermore, we specifically demonstrate that 1) the 'A' allele of the common variant rs13431652 is strongly associated with elevated FPG; 2) the rs13431652-A allele enhances NF-Y binding to the *G6PC2* promoter (Figs. 1 and 2); and 3) the rs13431652-A allele increases fusion gene expression (Fig. 3). Although we failed to observe any significant correlation between endogenous *G6PC2* gene expression and rs13431652 genotypes, probably because of the limited sample analyzed ( $N = 24$ ), our in situ data are consistent with the reduced FPG observed in *G6pc2* knockout mice (4) and the hypothesis that *G6PC2* would oppose the action of glucokinase, leading to elevated FPG. These data suggest that rs13431652 is a strong candidate to be a causative SNP that contributes to the association signal between *G6PC2* and FPG, though it is important to note that our genetic data do not rule out a significant contribution of the intronic variant rs560887. Future studies will be designed to address our hypothesis that rs560887 affects *G6PC2* mRNA splicing (9).

Our study does not give support to the contribution of intronic variant rs853789 located in the 19th intron of *ABCB11* to the association signal with FPG in our European cohorts. In a recent study conducted in Asian populations, a haplotype tagging variant rs3755157 located in the 21st intron of *ABCB11* was identified as associated with FPG independently from rs560887, which may support the existence of two independent signals at this locus (42). In this study, limited attention was given to the striking differences in allele frequencies between Asians and Europeans, both at rs560887 ( $MAF_{EU} = 0.30$  vs.  $MAF_{AS} = 0.03$ ) and rs3755157 (0.12 calculated in our French population and 0.09 in HapMap CEU population vs.  $MAF_{AS} = 0.38$ ). We believe that these differences may alter the efficiency of these two SNPs to tag putative causal variant(s), as reflected by the difference in the strength of the associations of these SNPs observed in Europeans versus Asians. Nonetheless, we acknowledge that a more extensive fine mapping analysis involving a larger number of variants in this locus will be required to elucidate the contribution of *G6PC2* and *ABCB11* variants to the association signal and provide a more comprehensive genetic assessment of this important locus in the genetic determination of FPG.

Using a combination of genetic and functional data, we show that rs573225-A allele is highly associated with elevated FPG (Table 1), enhanced Foxa2 binding (Figs. 4 and 5), but lower fusion gene expression (Fig. 6), a correlation that is inconsistent with the reduced FPG observed in *G6pc2* knockout mice (4). As such, unlike the functional data obtained with rs13431652, these functional data do not support a potential role for rs573225 as a causative SNP linking *G6PC2* to variations in FPG. In contrast, Dos Santos et al. (43) recently concluded that rs573225 is the causative SNP that explains the association signal between FPG and *G6PC2*. As in our study, they observed that the rs573225-A allele is associated with elevated FPG but lower fusion gene expression, but they did not comment on the fact that these data are inconsistent with the function of *G6PC2*. Dos Santos et al. (43) also suggested that rs573225 is an epiSNP, because the rs573225-G allele is located at a GpC dinucleotide within the Foxa2 binding site and methylation of the 'C' nucleotide affected Foxa2 binding. However, in contrast with the well-studied methylation of CpG dinucleotides (44), we can find no reports of the existence of methylated GpC dinucleotides in mammals, although this modification does exist in fish (45). Finally, although our fusion gene data match that of Dos Santos et al. (43), our gel retardation results are the complete opposite. Dos Santos et al. (43) report that the rs573225-G allele binds Foxa2, whereas the rs573225-A allele completely abolishes Foxa2 binding. Their binding data therefore appear to correlate with their fusion gene data since the rs573225-G allele confers higher promoter activity. However, a detailed previous study by Overdier et al. (46) showed that both 'G' and 'A' nucleotides at this location within the Foxa2 binding site support Foxa2 binding. Our demonstration that Foxa2 binds with the same affinity to both alleles (Fig. 5A and B), though with different kinetics (Fig. 5C), is therefore consistent with the observations of Overdier et al. (46), but not with those of Dos Santos et al. (43). Moreover, given the importance of Foxa2 for *G6PC2* fusion gene expression (supplementary Fig. 4), if the rs573225-A allele were to abolish Foxa2 binding, Dos Santos et al. (43) should have detected a major difference in the level of reporter gene expression conferred by the promoters containing the rs573225-A and rs573225-G alleles instead of the ~15% difference reported.

Although the available functional data does not support a role for rs573225 as a causative SNP, a caveat with the experiments reported here is that the functional effect of rs573225 on endogenous *G6PC2* gene transcription in vivo may not be replicated by analyses involving the transient transfection of fusion genes in islet-derived cell lines. Thus, fusion gene experiments have several limitations, including: 1) the absence of chromatin structure in transient transfection assays (47); 2) the use of a truncated fusion gene that lacks the five known *G6PC2* transcriptional enhancers (48); and 3) the use of tissue culture cell lines in which the gene expression profiles do not match that of islets in vivo. All of these factors have the potential to alter the effect of rs573225 on *G6PC2* gene transcription. Finally, Foxa2 is known to act as a pioneer factor, opening chromatin to allow access of other transcription factors (49). As such, rs573225 may have different effects on *G6PC2* gene expression early in development than in adults.

In summary, our study provides genetic and functional evidence supporting an important role for the promoter



variant rs13431652 as a potentially causative SNP that contributes to the association signal between *G6PC2* and FPG, but the data do not preclude a significant contribution of the intronic variant rs560887 or other additional unidentified variants.

#### ACKNOWLEDGMENTS

Research in the laboratory of R.M.O'B. was supported by National Institutes of Health grants DA-027002 and P60 DK-20593, which funds the Vanderbilt Diabetes Center Core Laboratory. Research in the laboratory of P.F. was supported by the "Conseil Régional Nord-Pas-de-Calais Fonds Européen de Développement économique et Régional," the French Agence Nationale de la Recherche (ANR-08 GENOPAT), and the European Union (Integrated Project EuroDia LSHM-CT-2006-518153 in the Framework Programme 6 of the European-Community, to P.F.). The NFBC 1986 study is supported by the European Commission, contract number QLGI-CT-2000-01643, Biocenter, University of Oulu, Finland, and the Academy of Finland.

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

N.B.-N. supervised the design of the genetic and expression studies and wrote parts of the manuscript. A.B. contributed to the experiments and draft of the manuscript relating to the genetic and expression studies. D.A.B. performed gel retardation and fusion gene analyses. M. Marchand performed genotyping and expression experiments. M.B. prepared pancreatic islet cDNA and DNA. P.M. prepared pancreatic islet cDNA and DNA and reviewed the manuscript. F.P. sorted  $\beta$ -cells, prepared  $\beta$ -cell cDNA, and reviewed the manuscript. R.L.P. cloned the human *G6PC2* promoter. B.P.F., O.C.U., and N.L.C. performed gel retardation and fusion gene analyses. M.V. contributed to the design of the genetic experiments and reviewed the manuscript. O.L. and M. Marre collected data for the DESIR cohort. B.B. collected data for the DESIR cohort and reviewed the manuscript. C.L.-M. collected data for the Haguenau cohort. P.E. collected data for the NFBC 1986. M.-R.J. collected data for the NFBC 1986 and reviewed the manuscript. D.M. collected data for the obese children cohort and reviewed the manuscript. C.D. performed statistical analyses and reviewed the manuscript. J.K.O. performed fusion gene analyses. P.F. was the principle investigator for the genetic studies and reviewed the manuscript. R.M.O'B. was the principle investigator for the gel retardation and fusion gene studies and wrote parts of the manuscript.

The authors thank all the patients and their families for participation in the genetic study; C. Cavalcanti-Proença (CNRS-UMR-8199, France) for statistical analyses and drafting statistical sections of the manuscript; M. Deweinder and F. Allegaert (CNRS-UMR-8199, France) for DNA extraction for part of the cohort studied; and S. Poulain and P. Gallina (CNRS-UMR-8199, France) for the recruitment of the families of obese children in Lille. The authors also thank L. Peltonen (Institute of Molecular Medicine, Finland), A.-L. Hartikainen, and A. Ruokonen (University of Oulu, Finland) for data collection for the NFBC 1986; O. Le Bacquer (CNRS-UMR-8199, France) for advice on expression analyses; S. Del Guerra (University of Pisa, Italy) for preparation of human pancreatic islet DNA and cDNAs for genotype/expression correlations; B. Lukoviac (INSERM U859, France) for RNA extraction from sorted  $\beta$ -cells. The authors also thank R. Stein (Vanderbilt Uni-

versity), S. Efrat (Tel Aviv University), and J-i Miyazaki (Osaka University) for providing the HIT,  $\beta$ TC-3, and Min6 cell lines, respectively; L. Pound (Vanderbilt University) for help with gel retardation assays; and K. Zaret (University of Pennsylvania) for useful comments about *Foxa2*.

#### REFERENCES

- Hutton JC, O'Brien RM. The glucose-6-phosphatase catalytic subunit gene family. *J Biol Chem* 2009;284:29241–29245
- Arden SD, Zahn T, Steegers S, Webb S, Bergman B, O'Brien RM, Hutton JC. Molecular cloning of a pancreatic islet-specific glucose-6-phosphatase catalytic subunit-related protein. *Diabetes* 1999;48:531–542
- Martin CC, Bischof LJ, Bergman B, Hornbuckle LA, Hilliker C, Frigeri C, Wahl D, Svitek CA, Wong R, Goldman JK, Oeser JK, Lepretre F, Froguel P, O'Brien RM, Hutton JC. Cloning and characterization of the human and rat islet-specific glucose-6-phosphatase catalytic subunit-related protein (IGRP) genes. *J Biol Chem* 2001;276:25197–25207
- Wang Y, Martin CC, Oeser JK, Sarkar S, McGuinness OP, Hutton JC, O'Brien RM. Deletion of the gene encoding the islet-specific glucose-6-phosphatase catalytic subunit-related protein autoantigen results in a mild metabolic phenotype. *Diabetologia* 2007;50:774–778
- Hutton JC, Eisenbarth GS. A pancreatic  $\beta$ -cell-specific homolog of glucose-6-phosphatase emerges as a major target of cell-mediated autoimmunity in diabetes. *Proc Natl Acad Sci U S A* 2003;100:8626–8628
- Petrolonis AJ, Yang Q, Tummino PJ, Fish SM, Prack AE, Jain S, Parsons TF, Li P, Dales NA, Ge L, Langston SP, Schuller AG, An WF, Tartaglia LA, Chen H, Hong SB. Enzymatic characterization of the pancreatic islet-specific glucose-6-phosphatase-related protein (IGRP). *J Biol Chem* 2004;279:13976–13983
- Shieh JJ, Pan CJ, Mansfield BC, Chou JY. In islet-specific glucose-6-phosphatase-related protein, the beta cell antigenic sequence that is targeted in diabetes is not responsible for the loss of phosphohydrolase activity. *Diabetologia* 2005;48:1851–1859
- Newgard CB, Lu D, Jensen MV, Schissler J, Boucher A, Burgess S, Sherry AD. Stimulus/secretion coupling factors in glucose-stimulated insulin secretion: insights gained from a multidisciplinary approach. *Diabetes* 2002;51 Suppl 3:S389–393
- Bouatia-Naji N, Rocheleau G, Van Lommel L, Lemaire K, Schuit F, Cavalcanti-Proença C, Marchand M, Hartikainen AL, Sovio U, De Graeve F, Rung J, Vaxillaire M, Tichet J, Marre M, Balkau B, Weill J, Elliott P, Jarvelin MR, Meyre D, Polychronakos C, Dina C, Sladek R, Froguel P. A polymorphism within the *G6PC2* gene is associated with fasting plasma glucose levels. *Science* 2008;320:1085–1088
- Coutinho M, Gerstein HC, Wang Y, Yusuf S. The relationship between glucose and incident cardiovascular events. A meta-regression analysis of published data from 20 studies of 95,783 individuals followed for 12.4 years. *Diabetes Care* 1999;22:233–240
- DECODE Study Group, European Diabetes Epidemiology Group. Is the current definition for diabetes relevant to mortality risk from all causes and cardiovascular and noncardiovascular diseases? *Diabetes Care* 2003;26:688–696
- Chen WM, Erdos MR, Jackson AU, Saxena R, Sanna S, Silver KD, Timpson NJ, Hansen T, Orru M, Grazia Piras M, Bonnycastle LL, Willer CJ, Lyssenko V, Shen H, Kuusisto J, Ebrahim S, Sestu N, Duren WL, Spada MC, Stringham HM, Scott LJ, Olla N, Swift AJ, Najjar S, Mitchell BD, Lawlor DA, Smith GD, Ben-Shlomo Y, Andersen G, Borch-Johnsen K, Jorgensen T, Saramies J, Valle TT, Buchanan TA, Shuldiner AR, Lakatta E, Bergman RN, Uda M, Tuomilehto J, Pedersen O, Cao A, Groop L, Mohlke KL, Laakso M, Schlessinger D, Collins FS, Altshuler D, Abecasis GR, Boehnke M, Scuteri A, Watanabe RM. Variations in the *G6PC2/ABCB11* genomic region are associated with fasting glucose levels. *J Clin Invest* 2008;118:2620–2628
- Prokopenko I, Langenberg C, Florez JC, Saxena R, Soranzo N, Thorleifsson G, Loos RJ, Manning AK, Jackson AU, Aulchenko Y, Potter SC, Erdos MR, Sanna S, Hottenga JJ, Wheeler E, Kaakinen M, Lyssenko V, Chen WM, Ahmadi K, Beckmann JS, Bergman RN, Bochud M, Bonnycastle LL, Buchanan TA, Cao A, Cervino A, Coin L, Collins FS, Crisponi L, de Geus EJ, Dehghan A, Deloukas P, Doney AS, Elliott P, Freimer N, Gateva V, Herder C, Hofman A, Hughes TE, Hunt S, Illig T, Inouye M, Isomaa B, Johnson T, Kong A, Krestyaninova M, Kuusisto J, Laakso M, Lim N, Lindblad U, Lindgren CM, McCann OT, Mohlke KL, Morris AD, Naitza S, Orru M, Palmer CN, Pouta A, Randall J, Rathmann W, Saramies J, Scheet P, Scott LJ, Scuteri A, Sharp S, Sijbrands E, Smit JH, Song K, Steinthorsdottir V, Stringham HM, Tuomi T, Tuomilehto J, Uitterlinden AG, Voight BF, Waterworth D, Wichmann HE, Willemssen G, Witteman JC, Yuan X, Zhao JH, Zeggini E, Schlessinger D, Sandhu M, Boomsma DI, Uda M, Spector TD,

- Penninx BW, Altshuler D, Vollenweider P, Jarvelin MR, Lakatta E, Waeber G, Fox CS, Peltonen L, Groop LC, Mooser V, Cupples LA, Thorsteinsdottir U, Boehnke M, Barroso I, Van Duijn C, Dupuis J, Watanabe RM, Stefansson K, McCarthy MI, Wareham NJ, Meigs JB, Abecasis GR. Variants in MTNR1B influence fasting glucose levels. *Nat Genet* 2008;41:77–81
14. Balkau B. [An epidemiologic survey from a network of French Health Examination Centres, (D.E.S.I.R.): epidemiologic data on the insulin resistance syndrome]. *Rev Epidemiol Sante Publique* 1996;44:373–375
  15. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007;445:881–885
  16. Jarvelin MR, Elliott P, Kleinschmidt I, Martuzzi M, Grundy C, Hartikainen AL, Rantakallio P. Ecological and individual predictors of birthweight in a northern Finland birth cohort 1986. *Paediatr Perinat Epidemiol* 1997;11:298–312
  17. Meyre D, Lecoecur C, Delplanque J, Francke S, Vatin V, Durand E, Weill J, Dina C, Froguel P. A genome-wide scan for childhood obesity-associated traits in French families shows significant linkage on chromosome 6q22.31-q23.2. *Diabetes* 2004;53:803–811
  18. Rolland-Cachera MF, Cole TJ, Sempe M, Tichet J, Rossignol C, Charraud A. Body mass index variations: centiles from birth to 87 years. *Eur J Clin Nutr* 1991;45:13–21
  19. Jaquet D, Collin D, Levy-Marchal C, Czernichow P. Adult height distribution in subjects born small for gestational age. *Horm Res* 2004;62:92–96
  20. Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* 1997;20:1183–1197
  21. Lukowiak B, Vandewalle B, Riachy R, Kerr-Conte J, Gmyr V, Belaich S, Lefebvre J, Pattou F. Identification and purification of functional human  $\beta$ -cells by a new specific zinc-fluorescent probe. *J Histochem Cytochem* 2001;49:519–528
  22. Ebert DH, Bischof LJ, Streeper RS, Chapman SC, Svitek CA, Goldman JK, Mathews CE, Leiter EH, Hutton JC, O'Brien RM. Structure and promoter activity of an islet-specific glucose-6-phosphatase catalytic subunit-related gene. *Diabetes* 1999;48:543–551
  23. Martin C, Flemming B, Wang Y, Oeser J, O'Brien R. Foxa2 and MafA regulate islet-specific glucose-6-phosphatase catalytic subunit-related protein (IGRP/G6PC2) gene expression. *J Mol Endocrinol* 2008;41:315–328
  24. Martin CC, Oeser JK, O'Brien RM. Differential regulation of islet-specific glucose-6-phosphatase catalytic subunit-related protein gene transcription by Pax-6 and Pdx-1. *J Biol Chem* 2004;279:34277–34289
  25. Martin CC, Oeser JK, Svitek CA, Hunter SI, Hutton JC, O'Brien RM. Identification and characterization of a human cDNA and gene encoding a ubiquitously expressed glucose-6-phosphatase catalytic subunit-related protein. *J Mol Endocrinol* 2002;29:205–222
  26. Sambrook J, Fritsch EF and Maniatis T. *Molecular Cloning: A Laboratory Manual*. Plainview, New York, Cold Spring Harbor Laboratory Press, 1989
  27. O'Brien RM, Lucas PC, Yamasaki T, Noisin EL, Granner DK. Potential convergence of insulin and cAMP signal transduction systems at the phosphoenolpyruvate carboxykinase (PEPCK) gene promoter through CCAAT/enhancer binding protein (C/EBP). *J Biol Chem* 1994;269:30419–30428
  28. Tregouet DA, Tiret L. Cox proportional hazards survival regression in haplotype-based association analysis using the Stochastic-EM algorithm. *Eur J Hum Genet* 2004;12:971–974
  29. Stieger B, Meier Y, Meier PJ. The bile salt export pump. *Pflugers Arch* 2007;453:611–620
  30. Quandt K, Frech K, Karas H, Wingender E, Werner T. MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acid Res* 1995;23:4878–4884
  31. Mantovani R. A survey of 178 NF-Y binding CCAAT boxes. *Nucleic Acid Res* 1998;26:1135–1143
  32. Mantovani R. The molecular biology of the CCAAT-binding factor NF-Y. *Gene* 1999;239:15–27
  33. Onuma H, Vander Kooi BT, Boustead JN, Oeser JK, O'Brien RM. Correlation between FOXO1a (FKHR) and FOXO3a (FKHRL1) binding and the inhibition of basal glucose-6-phosphatase catalytic subunit gene transcription by insulin. *Mol Endocrinol* 2006;20:2831–2847
  34. Limbird LE. *Cell Surface Receptors: A Short Course on Theory and Methods*, 3rd Edition, Springer, 2004
  35. Ceribelli M, Dolfini D, Merico D, Gatta R, Vigano AM, Pavesi G, Mantovani R. The histone-like NF-Y is a bifunctional transcription factor. *Mol Cell Biol* 2008;28:2047–2058
  36. Friedman JR, Kaestner KH. The Foxa family of transcription factors in development and metabolism. *Cell Mol Life Sci* 2006;63:2317–2328
  37. O'Brien RM, Noisin EL, Suwanichkul A, Yamasaki T, Lucas PC, Wang JC, Powell DR, Granner DK. Hepatic nuclear factor 3- and hormone-regulated expression of the phosphoenolpyruvate carboxykinase and insulin-like growth factor-binding protein 1 genes. *Mol Cell Biol* 1995;15:1747–1758
  38. Bouatia-Naji N, Bonnefond A, Cavalcanti-Proenca C, Sparso T, Holmkvist J, Marchand M, Delplanque J, Lobbens S, Rocheleau G, Durand E, De Graeve F, Chevre JC, Borch-Johnsen K, Hartikainen AL, Ruokonen A, Tichet J, Marre M, Weill J, Heude B, Tauber M, Lemaire K, Schuit F, Elliott P, Jorgensen T, Charpentier G, Hadjadj S, Cauchi S, Vaxillaire M, Sladek R, Visvikis-Siest S, Balkau B, Levy-Marchal C, Pattou F, Meyre D, Blakemore AI, Jarvelin MR, Walley AJ, Hansen T, Dina C, Pedersen O, Froguel P. A variant near MTNR1B is associated with increased fasting plasma glucose levels and type 2 diabetes risk. *Nat Genet* 2009;41:89–94
  39. Hu C, Zhang R, Wang C, Ma X, Wang C, Fang Q, Bao Y, Xiang K, Jia W. A genetic variant of G6PC2 is associated with type 2 diabetes and fasting plasma glucose level in the Chinese population. *Diabetologia* 2008
  40. Pare G, Chasman DI, Parker AN, Nathan DM, Miletich JP, Zee RY, Ridker PM. Novel association of HK1 with glycated hemoglobin in a non-diabetic population: a genome-wide evaluation of 14,618 participants in the Women's Genome Health Study. *PLoS Genet* 2008;4:e1000312
  41. Reiling E, van 't Riet E, Groenewoud MJ, Welschen LM, van Hove EC, Nijpels G, Maassen JA, Dekker JM, t Hart LM. Combined effects of single-nucleotide polymorphisms in GCK, GCKR, G6PC2 and MTNR1B on fasting plasma glucose and type 2 diabetes risk. *Diabetologia* 2009;52:1866–1870
  42. Takeuchi F, Katsuya T, Chakrewarthy S, Yamamoto K, Fujioka A, Serizawa M, Fujisawa T, Nakashima E, Ohnaka K, Ikegami H, Sugiyama T, Nabika T, Kasturiratne A, Yamaguchi S, Kono S, Takayanagi R, Yamori Y, Kobayashi S, Oghihara T, de Silva A, Wickremasinghe R, Kato N. Common variants at the GCK, GCKR, G6PC2-ABCB11 and MTNR1B loci are associated with fasting glucose in two Asian populations. *Diabetologia* 53:299–308
  43. Dos Santos C, Bougneres P, Fradin D. An SNP in a methylatable Foxa2 binding site of the G6PC2 promoter is associated with insulin secretion in vivo and increased promoter activity in vitro. *Diabetes* 2009;58:489–492
  44. Miranda TB, Jones PA. DNA methylation: the nuts and bolts of repression. *J Cell Physiol* 2007;213:384–390
  45. Pontecorvo G, De Felice B, Carfagna M. Novel methylation at GpC dinucleotide in the fish *Sparus aurata* genome. *Mol Biol Rep* 2000;27:225–230
  46. Overdier DG, Porcella A, Costa RH. The DNA-binding specificity of the hepatocyte nuclear factor 3/forkhead domain is influenced by amino-acid residues adjacent to the recognition helix. *Mol Cell Biol* 1994;14:2755–2766
  47. Smith CL, Hager GL. Transcriptional regulation of mammalian genes in vivo. A tale of two templates. *J Biol Chem* 1997;272:27493–27496
  48. Wang Y, Flemming BP, Martin CC, Allen SR, Walters J, Oeser JK, Hutton JC, O'Brien RM. Long-range enhancers are required to maintain expression of the autoantigen islet-specific glucose-6-phosphatase catalytic subunit-related protein in adult mouse islets in vivo. *Diabetes* 2008;57:133–141
  49. Zaret KS, Watts J, Xu J, Wandzioch E, Smale ST, Sekiya T. Pioneer factors, genetic competence, and inductive signaling: programming liver and pancreas progenitors from the endoderm. *Cold Spring Harb Symp Quant Biol* 2008;73:119–126