Original Article

Common Variants in Maturity-Onset Diabetes of the Young Genes Contribute to Risk of Type 2 Diabetes in Finns

Lori L. Bonnycastle,[1] Cristen J. Willer,[2] Karen N. Conneely,[2] Anne U. Jackson,[2] Cecily P. Burrill,[1] Richard M. Watanabe,[3] Peter S. Chines,[1] Narisu Narisu,[1] Laura J. Scott,[2] Sareena T. Enloe,[1] Amy J. Swift,[1] William L. Duren,[2] Heather M. Stringham,[2] Michael R. Erdos,[1] Nancy L. Riebow,[1] Thomas A. Buchanan,[4] Timo T. Valle,[5] Jaakko Tuomilehto,[5,6,7] Richard N. Bergman,[8] Karen L. Mohlke,[9] Michael Boehnke,[2] and Francis S. Collins[1]

Prior reports have suggested that variants in the genes for maturity-onset diabetes of the young (MODY) may confer susceptibility to type 2 diabetes, but results have been conflicting and coverage of the MODY genes has been incomplete. To complement our previous studies of *HNF4A*, we examined the other five known MODY genes for association with type 2 diabetes in Finnish individuals. For each of the five genes, we selected *1*) nonredundant single nucleotide polymorphisms (SNPs) ($r^2 < 0.8$ with other SNPs) from the HapMap database or another linkage disequilibrium map, *2*) SNPs with previously reported type 2 diabetes association, and *3*) nonsynonymous coding SNPs. We tested 128 SNPs for association with type 2 diabetes in 786 index cases from type 2 diabetic families and 619 normal glucose-tolerant control subjects. We followed up 35 of the most significant SNPs by genotyping them on another 384 case subjects and 366 control subjects from Finland. We also supplemented our previous *HNF4A* results by genotyping 12 SNPs on additional Finnish samples. After correcting for testing multiple correlated SNPs within a gene, we find evidence of type 2 diabetes association with SNPs in five of the six known MODY genes: *GCK*, *HNF1A*, *HNF1B*, *NEUROD1*, and *HNF4A*. Our data suggest that common variants in several MODY genes play a modest role in type 2 diabetes susceptibility. *Diabetes* 55: 2534–2540, 2006

From the [1]Genome Technology Branch, National Genome Research Institute, Bethesda, Maryland; the [2]Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan; the [3]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California; the [4]Department of Medicine, Division of Endocrinology, Keck School of Medicine, University of Southern California, Los Angeles, California; the [5]Diabetes and Genetic Epidemiology Unit, Department of Epidemiology and Health Promotion, National Public Health Institute, Helsinki, Finland; the [6]Department of Public Health, University of Helsinki, Helsinki, Finland; the [7]South Ostrobothnia Central Hospital, Seinäjoki, Finland; the [8]Department of Physiology and Biophysics, Keck School of Medicine, University of Southern California, Los Angeles, California; and the [9]Department of Genetics, University of North Carolina, Chapel Hill, North Carolina.

Address correspondence and reprint requests to Francis S. Collins, MD, PhD, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892-2152. E-mail: francisc@mail.nih.gov.

Received for publication 7 February 2006 and accepted in revised form 15 June 2006.

L.L.B. and C.J.W. contributed equally to this work.

Additional information for this article can be found in an online appendix at http://diabetes.diabetesjournals.org.

FUSION, Finland-U.S. Investigation of NIDDM Genetics Study; LD, linkage disequilibrium; MODY, maturity-onset diabetes of the young; SNP, single nucleotide polymorphism.

DOI: 10.2337/db06-0178

© 2006 by the American Diabetes Association.

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.

The common form of type 2 diabetes results from a complex interaction between genetic background, environment, and health behavior, leading to two primary abnormalities: insulin resistance and insufficient insulin secretion (1). Identifying susceptibility genes for type 2 diabetes has been difficult, as it has been for most multifactorial diseases. In contrast, there has been considerable success in identifying genes involved with an autosomal dominant form of type 2 diabetes, maturity-onset diabetes of the young (MODY). This monogenic disease accounts for 1–5% of all type 2 diabetes cases and is characterized by high penetrance, early age at onset (usually before 25 years), impaired insulin secretion, and mild to severe clinical manifestations (2). Variants in six genes responsible for MODY have been identified (2). These genes are hepatic nuclear factor 4α (*HNF4A*–MODY 1), glucokinase (*GCK*–MODY 2), transcription factor 1 (*HNF1A/TCF1*–MODY 3), insulin promoter factor 1 (*IPF1/PDX1*–MODY 4), transcription factor 2 (*HNF1B/TCF2*–MODY 5), and neurogenic differentiation 1 (*NEUROD1*–MODY 6). Most MODY genes appear to cause disease as a result of haploinsufficiency of the protein product in pancreatic β-cells (3). With the exception of *GCK*, the known MODY genes are transcription factors that directly or indirectly influence expression of insulin and other proteins involved with glucose metabolism and/or β-cell development (4,5).

Genes with rare variants that cause severe phenotypes or monogenic disease are excellent candidates for variants that predispose to related multifactorial diseases (5). Monogenic disorders with diabetes and diabetes-related phenotypes have been associated with mutations in *PPARG* (6) and *KCNJ11* (7). Additionally, common nonsynonymous variations in these genes are two of the most widely replicated type 2 diabetes susceptibility alleles (8). Single nucleotide polymorphism (SNP) variants or haplotypes in MODY genes have been reported to be associated with common type 2 diabetes, but in most instances, these variants have had low frequencies and/or were population specific (9–17). Recently, we and others independently identified the association of common polymorphisms in the P2 promoter region of

TABLE 1
Clinical characteristics of samples

	Stage 1			Stage 2	
	FUSION case subjects	FUSION control subjects	Finrisk control subjects	Finrisk case subjects	Finrisk control subjects
Samples (n)	786	377*	242	384	366
Male:female (n)	437:349	156:221	169:73	232:152	214:152
Age of diagnosis (years)	51.0 (12.0)	NA	NA	59.0 (12.0)	NA
Age at examination (years)	64.3 (10.1)	69.8 (5.9)	64.6 (9.9)	60.5 (12.2)	60.5 (11.3)
BMI (kg/m^2)	29.4 (6.1)	27.1 (5.4)	26.8 (4.3)	30.6 (6.2)	26.6 (4.5)
Fasting glucose (mmol/l)	9.7 (4.9)	5.0 (0.6)	5.6 (0.5)	7.3 (1.3)	5.6 (0.5)
Fasting insulin (pmol/l)	96.0 (84.0)	60.0 (42.0)	42.0 (28.8)	79.2 (61.8)	39.0 (27.6)

Data are median (interquartile range). *225 FUSION elderly control subjects and 152 spouse control subjects. NA, not applicable.

HNF4A with risk of common type 2 diabetes in individuals of Finnish (18) and Ashkenazi Jewish origin (19). Association of these variants has subsequently been reported in U.K., Amish, and Danish samples (20–22) but not in European Americans (14,23), French (24), Polish, Canadian, and Scandinavian samples (23).

The most previous type 2 diabetes studies of MODY genes have had limited gene coverage, focusing primarily on promoter regions, exons, and splice junctions, and most have genotyped only one or a few variants. Such restricted evaluations may have failed to identify disease-predisposing variants located in noncoding regulatory regions, particularly those at some distance from the coding exons. Recent improvements in the understanding of human genetic variation, including the dramatic growth of the dbSNP database (www.ncbi.nlm.nih.gov/projects/SNP/), the HapMap (www.hapmap.org) project (25), and the development of highly cost-efficient genotyping technologies, have made a more thorough analysis of these candidate genes possible. Therefore, to complement and extend our ongoing analysis of *HNF4A*, we performed a comprehensive study to assess type 2 diabetes association of variants in five other MODY genes: *GCK*, *HNF1A*, *IPF1*, *HNF1B*, and *NEUROD1*. We also genotyped 12 SNPs in *HNF4A* in additional samples to further assess the evidence for association in this gene. After correcting for testing of multiple correlated SNPs within a gene, we found evidence of type 2 diabetes association with SNPs in five of the six known MODY genes: *GCK*, *HNF1A*, *HNF1B*, *NEUROD1*, and *HNF4A*.

RESEARCH DESIGN AND METHODS

Independent samples of Finnish individuals from the Finland-U.S. Investigation of NIDDM Genetics (FUSION) study and the Finrisk 2002 (Finrisk) study were used in this study. FUSION subject recruitment and phenotyping have been described previously (26,27). For the current study, we selected 786 affected individuals, one each from 786 FUSION affected sibling pair families, as the stage 1 case sample. FUSION control subjects consisted of 225 elderly subjects and 152 normal glucose-tolerant spouses of type 2 diabetic subjects (Table 1). The remainder of the stage 1 control sample consisted of 242 normal glucose-tolerant individuals (Table 1) from Finrisk 2002, a Finnish population-based national risk factor survey study (28). For stage 2, an additional 384 type 2 diabetic individuals and 366 normal glucose-tolerant individuals were selected from the Finrisk study. These case and control subjects were approximately frequency matched for age, sex, and province of birth (Table 1).

Informed consent was obtained from each study participant, and the institutional review board of participating centers approved the study protocol.

SNP selection. We selected a total of 154 SNPs in the six known MODY genes. A total of 142 of the 154 SNPs were from genomic regions spanning the five MODY genes: *GCK*, *HNF1A*, *IPF1*, *HNF1B*, and *NEUROD1*. We selected nonredundant SNPs ($r^2 < 0.8$ with other SNPs) from the CEU HapMap database (December 2004 releases: www.hapmap.org) and from additional private linkage disequilibrium (LD) maps created using 96 chromosomes from CEPH pedigrees for *GCK*, *IPF1*, *HNF1B*, and *NEUROD1*, provided by D. Altshuler and W. Winckler (D. Altshuler, W. Winckler, personal communication). We also selected *1*) SNPs previously reported to be associated with type 2 diabetes (PubMed: through February 2005), *2*) nonsynonymous coding SNPs from dbSNP, and *3*) *HNF1A* SNPs that showed association ($P < 0.05$) with type 2 diabetes in the initial stage of the Winckler et al. study (29). The average gap size between SNPs on the LD maps within 20 kb of *GCK*, *IPF1*, *NEUROD1*, *HNF1A*, and *HNF1B* were 1.36, 2.11, 4.90, 2.97, and 0.97 kb, respectively. Of the 62 *HNF4A* SNPs genotyped in the FUSION samples, we selected for further analysis 12 SNPs with evidence of association with type 2 diabetes in a subset of the stage 1 samples, 7 of which had been previously reported (15,18).

Genotyping. We carried out SNP genotyping on two platforms: the Sequenom homogeneous MassEXTEND assay (84% of genotypes) and the Illumina Golden Gate assay (16% of genotypes). We observed 51 discrepancies among 54,911 duplicate genotype pairs for the Sequenom system and 66 discrepancies among 115,637 duplicate genotypes for the Illumina system, yielding genotype reproducibility rates of 99.91 and 99.94%, respectively. We obtained call rates ≥90% for 146 of the 154 SNPs and excluded the other 8 SNPs from further analysis. These excluded low-success SNPs were *GCK* rs3217944; *HNF1A* rs2464195 and rs1169305; *IPF1* rs3812861; and *HNF1B* rs1805035, rs3110645, rs916894, and rs1800575. Of the 146 SNPs, we also excluded 4 with an allele frequency estimate <0.005 (*IPF1* Asp76Asn; *NEUROD1* rs1348123, rs8192556, and rs11558286) and 2 that were not consistent with Hardy-Weinberg expectations ($P < 0.001$) (*HNF1B* rs9914818 and rs757211). Consequently, we performed association analyses for 128 SNPs in *GCK*, *HNF1A*, *IPF1*, *HNF1B*, and *NEUROD1* and 12 SNPs in *HNF4A* (Fig. 1 and Table 2).

Staged sample genotyping. We used a two-stage strategy in which we genotyped all SNPs on the stage 1 samples and then followed up the most strongly associated SNPs ($P \leq 0.10$) by genotyping them on the stage 2 sample. We excluded 16 SNPs from further analysis after stage 1 genotyping because they were in strong LD ($r^2 > 0.9$) with another SNP with a higher genotyping success rate. For 35 SNPs, we analyzed genotypes for all samples, for a total of 1,170 case and 985 control subjects. This strategy of joint analysis has greater power than a replication study in which only the results for the second sample are considered (30). Using this design, we have 78% power to detect multiplicative SNP-type 2 diabetes association with odds ratios (ORs) of ≥1.3, at minor allele frequencies ≥10%. Furthermore, this design has 99% of the power of the equivalent single-stage design in which all samples are genotyped for all SNPs. The number of SNPs successfully genotyped and tested for each stage in each gene is shown in Table 2.

Statistical analysis of data. We estimated pairwise LD measures for all SNPs using the LDmax program (31). We tested disease-marker association and estimated ORs using logistic regression under three genetic models: dominant, recessive, and multiplicative. To assess the significance of our results, we calculated P values (designated P_{SNP}) that took into account the testing of three correlated genetic models and of multiple correlated SNPs within a gene region (designated P_{GENE}). To do so, we computed P_{SNP} or P_{GENE} as the probability of observing a P value at least as small as the minimum P value from the three models and/or the multiple SNPs, given the correlation between the genetic models or SNPs, assuming no disease-SNP correlation. We calculated these no-association probabilities using the known asymptotic distributions of the test statistics based on population estimates of the correlation parameters (K.N.C., M.B., personal communication). Our method is similar to those proposed by Lin (2005) (32) and Seaman and Muller-Myhsok (2005) (33) but allows much more rapid estimation of P values. We verified P_{SNP} values for our key results by permutation testing and obtained near-identical results (data not shown). For the *HNF4A* P_{GENE} values, we corrected for all 62 SNPs genotyped in

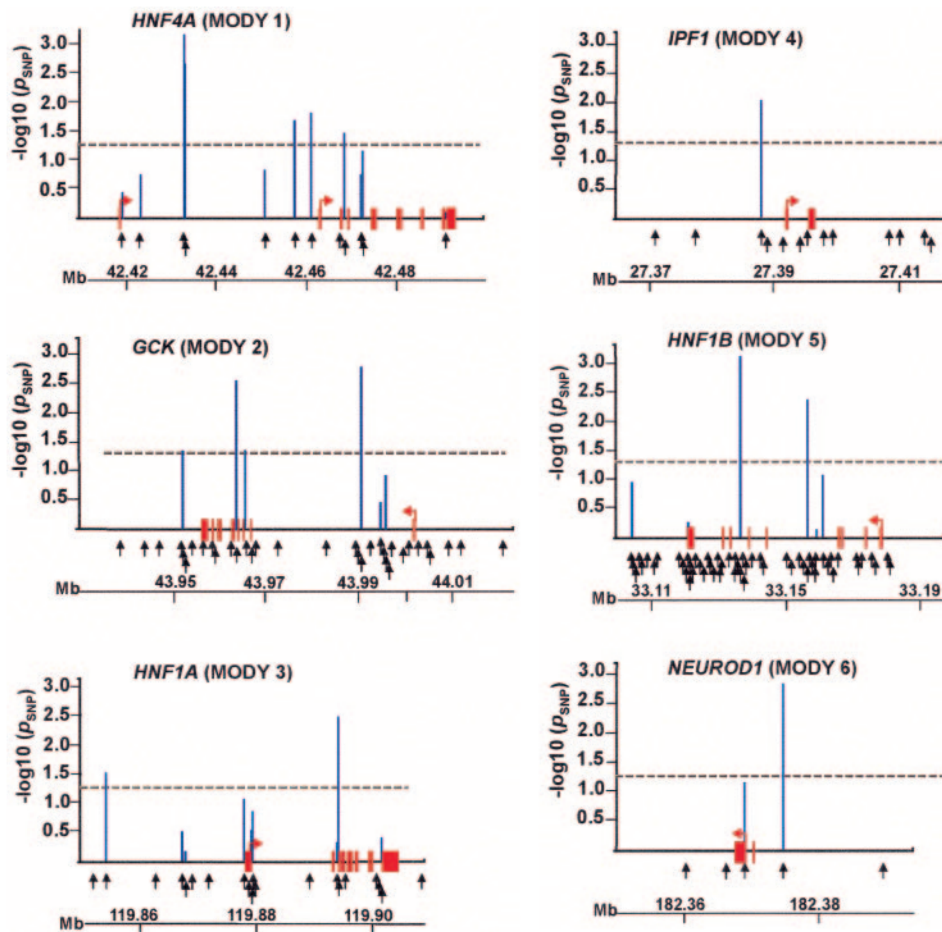

FIG. 1. Association of MODY gene SNPs with type 2 diabetes. The negative log10 P_{SNP} values for the SNPs with both stage 1 + 2 data are plotted in reference to the predicted gene structure (red boxes represent exons; red arrows represent the ATG initiation codons). The dashed line marks the threshold of significant association (P_{SNP} = 0.05). Black arrows below the *x*-axis indicate SNPs successfully genotyped in stage 1. *x*-Axis scale is shown in reference to genome position (hg17) in Mb.

this gene in any subset of our samples and not only the 12 described in this study. We tested for association with the age of diagnosis in case subjects by performing linear regression with sex and BMI as covariates. For each SNP, we tested each of three genetic models, and corrected for the models tested in a similar manner to that used for P_{SNP} for type 2 diabetes association.

We determined the independence of significant association signals ($P_{SNP} <$ 0.05) for two SNPs in the same gene by including the more significant SNP as a covariate in logistic regression and reassessing the evidence for association with the SNP of interest.

To assess association of haplotypes with type 2 diabetes, we first divided genes into sets of SNPs so that all typed SNPs were included in exactly one SNP set. Each gene contained between 1 (for *NEUROD1*) and 12 (for *HNF1B*) sets, and each set contained between 2 and 11 SNPs. We identified regions of moderate D′ between SNP pairs and either extended these SNP sets to incorporate single SNPs between sets or added SNPs that were not previously included in a set but were consecutive in position (Haploview [34]). We then estimated haplotype frequencies separately in case and control subjects for all SNPs within each set and permuted case and control status to obtain a significance level that accounts for the multiple haplotypes examined within each set (FAMHAP software [35]). All haplotypes were used to estimate the significance. However, for simplicity, only haplotypes with frequencies >1% in case or control subjects are shown in Table S3 (online appendix [available at http://diabetes.diabetesjournals.org]).

We assessed interactions with stage 1 data between all SNPs with a $P_{SNP} <$ 0.05 (stage 1 + 2) paired with all other SNPs. A likelihood ratio test where case or control status was predicted using logistic regression was performed to compare two models: the null model, consisting of single-SNP multiplicative and dominant effects, and the full model, which also included terms for multiplicative by multiplicative, multiplicative by dominant, dominant by multiplicative, and dominant by dominant interaction effects. Permuting case and control labels confirmed the appropriateness of the asymptotically derived *P* values.

RESULTS

We evaluated 140 SNPs representing common variation or likely functional changes in and around *GCK*, *HNF1A*, *IPF1*, *HNF1B*, *NEUROD1*, and *HNF4A* in 1,405 stage 1 samples (Tables 1 and 2). We found 39 SNPs with $P_{SNP} \leq$ 0.1 in stage 1. We genotyped 35 SNPs with pairwise $r^2 <$ 0.9 in stage 2 samples. The type 2 diabetes association results from the joint analysis of 2,155 individuals, 1,170 case and 985 control subjects, are shown in Table 3 (stage 1 + 2) and Fig. 1. Genotype counts, ORs, and *P* values for each genetic model for stage 1 and stages 1 + 2 are available in the online apendix Tables S1 and S2, respectively. In the stage 1 + 2 analysis, we found 15 SNPs associated with disease status at significance level $P_{SNP} <$ 0.05 after correcting for three disease models. Because multiple SNPs within a gene were found to be associated with disease, we tested for the independence of these association signals. Eight SNPs showed significant evidence for association ($P_{SNP} <$ 0.05) that was independent of any other SNP: one SNP each in *GCK*, *IPF1*, *NEUROD1*, and *HNF4A* and two each in *HNF1A* and *HNF1B*. After correcting for the total number of SNPs genotyped in each gene and accounting for the LD structure between SNPs,

TABLE 2
Number of SNPs tested for type 2 diabetes association and number of significant SNPs in stage 1 or in stages 1 + 2

Gene	MODY gene	Stage		Stages 1 + 2	
		1	2	$P_{SNP} < 0.05$	$P_{GENE} < 0.05$
GCK	MODY2	34	6	4	1
HNF1A	MODY3	20	9	2	1
IPF1	MODY4	13	1	1	0
HNF1B	MODY5	56	6	2	1
NEUROD1	MODY6	5	2	1	1
Total		128	24	10	4
HNF4A	MODY1	12	11	5	1
Final total		140	35	15	5

P_{SNP}, P value after accounting for testing of three models; P_{GENE}, P value after accounting for testing of multiple correlated SNPs within a gene.

five of the eight independently associated SNPs remained significant ($P_{GENE} < 0.05$). These five SNPs are *GCK* rs2244164, *NEUROD1* rs3916026, *HNF1A* rs2071190, *HNF1B* rs1008284, and *HNF4A* rs6103716. When we carried out disease-haplotype association analysis for a total of 30 SNP sets in the six genes, we observed 2 *HNF1B* SNP sets significant at $P < 0.05$, but these were not significant after correcting for multiple testing (online appendix Table S3).

Considering the early age of onset observed in MODY patients, we tested SNPs for association with age of diagnosis for common type 2 diabetes in the FUSION diabetic individuals. We evaluated 123 SNPs ($r^2 < 0.9$ with other

TABLE 3
Stage 1 + 2 results for SNPs associated with type 2 diabetes in stage 1 ($P_{SNP} < 0.1$)

Gene	SNP ID	Position (kb)*	Alleles (minor/major)	Minor allele frequency†	Best model	Best model OR (CI)	P_{SNP}	P_{GENE}
GCK	rs2284769	+6.33	G/C	0.152	R	1.72 (0.95–3.09)	0.13	
	rs2284773	+7.26	A/G	0.121	M	1.14 (0.95–1.36)	0.31	
	rs2244164	+11.43	C/T	0.470	M	0.81 (0.72–0.92)	0.0018	0.039
	rs12534623	+36.35	G/A	0.346	R	1.36 (1.05–1.76)	0.043‡	
	rs2268573	+38.08	G/T	0.484	R	1.39 (1.14–1.69)	0.0027‡	
	rs882020	+49.81	T/C	0.161	R	1.87 (1.08–3.24)	0.045‡	
HNF1A	rs2701175	−24.90	C/A	0.420	R	1.34 (1.06–1.68)	0.028	0.22
	rs1920792	−11.99	C/T	0.476	D	0.87 (0.72–1.05)	0.28	
	GE117884_349	−11.3	A/G	0.350	D	0.93 (0.78–1.11)	0.67	
	GE117881_360	−1.28	G/A	0.456	R	1.26 (1.01–1.55)	0.076	
	rs1169289	+0.05	G/C	0.471	R	1.17 (0.95–1.44)	0.28	
	rs1169288	+0.08	C/A	0.384	R	1.26 (0.98–1.60)	0.14	
	rs1169300	+14.65	A/G	0.312	R	1.18 (0.88–1.58)	0.47	
	rs2071190	+14.70	A/T	0.209	R	2.08 (1.30–3.31)	0.0032	0.035
	rs735396	+22.27	C/T	0.360	R	1.18 (0.91–1.52)	0.37	
IPF1	rs2297316	−4.06	A/G	0.407	D	0.77 (0.64–0.92)	0.0091	0.074
HNF1B	rs2107131	+18.19	A/G	0.416	M	0.88 (0.78–0.99)	0.085	
	rs916895	+20.65	T/C	0.063	M	1.06 (0.82–1.36)	0.66	
	rs12450628	+22.46	T/C	0.279	R	1.63 (1.20–2.23)	0.0038	0.14
	rs1008284	+42.42	A/G	0.260	R	0.53 (0.37–0.75)	0.00080	0.035
	rs1058166	+57.89	C/T	0.054	M	0.93 (0.71–1.21)	0.58	
	rs3094503	+74.47	C/A	0.217	D	0.84 (0.70–1.00)	0.11	
NEUROD1	rs3916026	−5.42	C/G	0.417	D	0.73 (0.61–0.87)	0.0014	0.0059
	rs1801262	+0.13	T/C	0.337	M	1.15 (1.01–1.30)	0.069	
HNF4A	rs2144908	+1.27	A/G	0.194	M	1.11 (0.95–1.29)	0.34	
	rs6031552	+5.35	A/C	0.213	R	0.68 (0.44–1.05)	0.17	
	rs6103716	+15.19	C/A	0.345	M	1.26 (1.11–1.44)	0.00067	0.028
	rs6031558	+15.20	C/G	0.314	M	0.80 (0.70–0.91)	0.0021§	
	rs6103723	+33.09	A/C	0.026	M	0.73 (0.49–1.11)	0.14	
	rs2425637	+39.60	T/G	0.482	M	1.18 (1.04–1.33)	0.018§	
	rs2425640	+43.59	A/G	0.408	M	0.84 (0.74–0.95)	0.015§	
	rs3212183	+50.69	C/T	0.391	D	1.26 (1.05–1.51)	0.030§	
	rs3212191	+54.34	C/T	0.146	D	1.18 (0.98–1.44)	0.17	
	rs1885088	+54.60	A/G	0.117	D	1.26 (1.03–1.54)	0.056	
	rs3818247	+73.04	T/G	0.341	R	0.91 (0.70–1.18)	0.72	

*Position relative to ATG of longest isoform; GE117881_360 and GE117884_349 are not in dbSNP. †Minor allele frequencies calculated relative to control subjects; P_{SNP}, P value after accounting for testing of three models: R = recessive, D = dominant, M = multiplicative; P_{GENE}, P value after accounting for testing of correlated SNPs within a gene and only reported for SNPs with $P_{SNP} < 0.05$ and with association signal not accounted for by another SNP. ‡Type 2 diabetes association is accounted for by rs2244164. §Type 2 diabetes association is accounted for by rs6103716.

SNPs) in the MODY genes and found 8 SNPs significantly associated with age of diagnosis with $P_{SNP} < 0.05$ and 2 SNPs with $P_{SNP} < 0.01$ (data not shown). This was not different from expectations under the null hypothesis of no association (6.1 at $P_{SNP} < 0.05$ and 1.2 at $P_{SNP} < 0.01$). Among the 15 SNPs genotyped on the combined samples that showed association to type 2 diabetes before correcting for testing of multiple correlated SNPs ($P_{SNP} < 0.05$), only 1 SNP, *HNF4A* rs2425640, was associated with age of diagnosis in FUSION type 2 diabetic individuals ($P_{SNP} = 0.042$).

We examined stage 1 data for interactions between 1,723 SNP pairs, with each pair including at least 1 of the 15 SNPs that were significantly associated with type 2 diabetes after stages 1 + 2 ($P_{SNP} < 0.05$) and all SNPs genotyped on stage 1 samples. We observed 18 pairs with significant evidence for interaction ($P < 0.01$) and 91 pairs with $P < 0.05$. This is not significantly greater than expectations (17.2 at $P < 0.01$ and 86.2 at $P < 0.05$). The two pairs of SNPs that showed the strongest evidence for interaction were *NEUROD1* rs3916026 with *HNF1B* rs2074429 ($P = 0.00018$) and *HNF1B* rs3094508 with *HNF4A* rs2425640 ($P = 0.00024$).

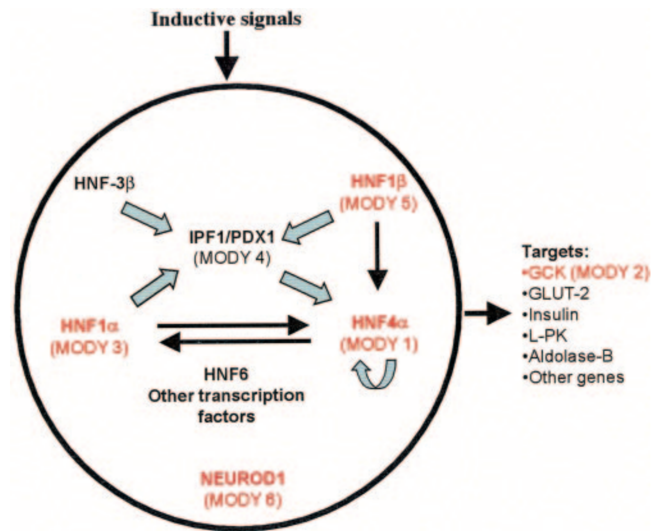

FIG. 2. Transcription factor network in the pancreatic β-cell. Modified from Kulkarni and Kahn (40). Variants in the genes noted in red have shown association with type 2 diabetes in this study. GCK, glucokinase; L-PK, liver pyruvate kinase.

DISCUSSION

We evaluated SNPs in six MODY genes (*GCK*, *HNF1A*, *IPF*, *HNF1B*, *NEUROD1*, and *HNF4A*) and observed eight independent SNPs, at least one per gene, associated with increased risk for type 2 diabetes ($P_{SNP} < 0.05$). After accounting for testing of multiple correlated SNPs within a gene, five of the eight SNPs remained significant ($P_{GENE} <$ 0.05), one per gene in *GCK*, *HNF1A*, *HNF1B*, *NEUROD1*, and *HNF4A*. A subset of these SNPs may play a role in type 2 diabetes susceptibility or be in LD with such variants.

The associated SNPs are all located in noncoding sequences with no obvious functional effect. Theoretically, these SNPs could be in LD with causal SNPs in the coding regions of these genes. However, reference to the data now available from HapMap Phase II (http://www.hapmap.org) does not reveal any previously overlooked common coding variants of these genes. Given this fact, and the observation that the risk alleles have only a modest effect on type 2 diabetes risk, we hypothesize that the type 2 diabetes risk is mediated by noncoding variants exerting subtle effects on gene expression.

There have been several reports of *GCK*, *HNF1A*, *IPF1*, *NEUROD1*, and *HNF4A* gene variants segregating in late-onset type 2 diabetic families or exhibiting association with type 2 diabetes in case and control samples (9–12,16,17,20–22,36,37). In this study, we successfully genotyped the following previously associated SNPs: *GCK* G(−30)A (36), *IPF1* Asp76Asn (10,11), and *NEUROD1* Thr45Ala (37). We found marginal evidence of association with the *NEUROD1* Thr45Ala SNP (two-sided $P_{SNP} = 0.067$, one-sided $P_{SNP} =$ 0.038) with the same risk allele as observed in the original population (37). However, a meta-analysis of 14 case-control studies of *NEUROD1* found no evidence for an overall effect of *NEUROD1* Thr45Ala on type 2 diabetes risk (38). Two in-depth studies of type 2 diabetes and common genetic variation in *HNF1A* in Caucasian samples were published during the course of our study (29,39). We examined several SNPs in common with these studies, and after correcting for multiple testing, only one SNP (rs2071190, $P_{SNP} = 0.0032$) exhibited significant association with type 2 diabetes. The same risk allele of SNP rs2071190 was significantly associated with type 2 diabetes ($P = 0.04$) in the initial case-control sample group of Weedon et al. (39).

Further investigation of *HNF4A* has identified new regions of association in Finns. Two of the recently tested SNPs (rs6103716 and rs6031558) located +15 kb from the *HNF4A* P2 promoter exhibited a stronger association with type 2 diabetes in the combined FUSION and Finrisk samples than previously tested SNPs. If these are true positive findings, the two SNPs may be tagging a single susceptibility variant or haplotype, as the two SNPs are in moderate LD ($r^2 = 0.23$). SNP rs6031558 has been evaluated in the Pima Native-American tribe, and although there was a modest increased risk for type 2 diabetes (multiplicative $P = 0.04$), the risk allele was opposite to our study (16). To our knowledge, SNP rs6103716 has not been tested in other populations, so further evaluation of this SNP will help assess its importance in determining type 2 diabetes susceptibility. Additionally, several SNPs that previously showed evidence for type 2 diabetes association in the FUSION samples (18) were no longer associated after we included data from the Finrisk samples. Variation in the strength and direction of type 2 diabetes association is also observed in multiple studies that have tested these variants (14,20–24). There is evidence of a pancreatic islet–specific network of transcription factors that interacts predominantly at the P2 promoter of *HNF4A* (3,4). As shown in Fig. 2, this *HNF4A* network is likely connected to other networks, involving a variety of transcription factors, including MODY genes *HNF1A*, *HNF1B*, and *IPF1* (3,40). We tested for MODY gene-gene interactions but did not observe an excess of significant associations. Our lack of positive findings may reflect the complexities associated with testing interactions or the lack of statistically detectable interactions between these genes.

Distinguishing MODY from common type 2 diabetes is difficult, and we have attempted to avoid ascertainment of MODY individuals by recruiting FUSION individuals between the ages of 35 and 60 years and Finrisk individuals between the ages of 45 and 74 years. There is a large number of MODY mutations; the vast majority of MODY patients have mutations in *GCK* or *HNF1A*, whereas mutations in the other four genes account for <8% of MODY patients in the U.K. population (41). It is unlikely that undiscovered MODY patients in our sample would carry a MODY-causing allele frequent enough to result in

our evidence for type 2 diabetes association in these genes, particularly in *NEUROD1*, *HNF1B*, and *HNF4A*, which are responsible for so few cases of MODY.

The four SNPs in *GCK*, *HNF1A*, *HNF1B*, and *NEUROD1* with significant type 2 diabetes associations in our study were not significantly associated in a concurrent study of *GCK*, *IPF1*, *HNF1B*, and *NEUROD1* in a large cohort of Swedish, Finnish, Canadian, and Polish type 2 diabetic individuals (D. Altshuler, W. Winckler, personal communication) or in a previous study of *HNF1A* in the same cohort of individuals (29). A number of factors could explain these differences in results, particularly given the likely small effect size of each variant. These include the impact of population and disease heterogeneity, the fact that all our FUSION cases are familial (having an affected sibling), while half the stage 1 cases studied by Winckler are not, and differences in the clinical characteristics of the samples. In contrast to our study, the Winckler et al. study (29) case and control samples (the nonfamilial samples) were matched for BMI, which might obscure any genes contributing to obesity. Also possible is that our significant associations are false-positives or that their lack of concordance with our results is due to their false negatives.

There are several well-known examples of common variants in genes involved in Mendelian diseases that act as susceptibility factors in related complex diseases. As mentioned earlier, this includes *PPARG* and *KCNJ11* in type 2 diabetes (6–8). Truncating mutations in the *APC* gene are responsible for familial adenomatous polyposis, but a missense variant Ile1307Lys confers an increased risk for multifactorial colorectal tumors in individuals of Ashkenazi Jewish origin (42). Mutations in *CFTR* are the cause of the autosomal recessive disorder cystic fibrosis, but milder variants have been linked with idiopathic pancreatitis and male infertility in individuals with no other symptoms of typical cystic fibrosis (43). To date, there are six genes identified to be involved with Mendelian forms of Parkinson's disease (44,45): *SCNA*, *PRKN*, *UCH-L1*, *DJ-1*, *PINK1*, and *LRRK2*. Allelic variants in *SCNA*, *PRKN*, and *LRRK2* have also now been associated with increased risk for common sporadic Parkinson's disease, and a coding variant in *UCH-L1* confers lower risk for this more common late-onset form of disease (44,45). The monogenic disease osteogenesis imperfecta usually arises from mutations in the genes encoding the collagen α1 (*COL1A1*) or α2 (*COL1A2*) chains (46), but a regulatory polymorphism of an Sp1 binding site of the *COL1A1* gene has been associated with low bone density (47) and idiopathic osteoporosis (48).

There is increasing evidence supporting a role for regulatory variants in complex diseases (49,50). Using an LD-based SNP selection and two-stage genotyping approach, we performed a comprehensive evaluation of both coding and noncoding regions surrounding each of the known MODY genes. This strategy differs from most previous studies of MODY genes, and we have observed significant type 2 diabetes association ($P_{GENE} < 0.05$) with noncoding variants, most of which have not been tested in other populations. Replication of these results in additional and larger study samples is crucial in determining the real significance of the associated SNPs. Furthermore, if these are true positive associations, evaluation of these SNPs in populations with different haplotype structures may help identify the disease predisposing SNPs.

Our two-stage genotyping and joint analysis strategy was cost-efficient and enabled us to evaluate more SNPs and samples while retaining the power of a one-stage strategy (30). The inclusion of 65 and 35% of samples in stage 1 and 2, respectively, was not necessarily the optimal design but was a reflection of sample availability during the course of the study.

Although we observed type 2 diabetes association with SNPs in several of the MODY genes, a potential weakness of our study, common to all association studies, is a lack of power to detect rare susceptibility variants or variants with small genetic effect sizes. Another consideration is false negatives or false-positives as a result of population heterogeneity within our sample group. To address this, we compared the genotype counts for the FUSION-ascertained control subjects with the Finrisk-ascertained control subjects. We found no excess of SNPs with significantly different genotype counts for either stage 1 ($P = 0.11$) or stage 2 ($P = 0.51$).

Our data support the hypothesis that mutations in genes involved in MODY play a role in predisposing to type 2 diabetes by implicating SNPs in five of the six known MODY genes. However, the observed effect for each of these variants is small and does not explain a substantial proportion of the hereditary component to type 2 diabetes. The availability of genome-wide association data and the expanding repertoire of genomics tools have the potential to facilitate efforts to identify the major common variants that, individually or in combination, affect susceptibility to this common familial disease.

ACKNOWLEDGMENTS

FUSION was funded by intramural funds from the National Human Genome Research Institute project 1Z01HG000024-11 and by National Institutes of Health (NIH) Grants DK62370 to M.B and DK72193 to K.L.M. J.T. was partially supported by the Academy of Finland (38387 and 46558). K.L.M. was supported by a Burroughs Wellcome Career Award in the Biomedical Sciences. R.N.B. was supported by NIH Grants DK27619 and DK29867. R.M.W. was supported by the American Diabetes Association.

We are indebted to the Finnish citizens who generously participated in the FUSION study. We thank D. Altshuler and W. Winckler for generously providing their LD maps and for helpful discussions. We also thank the Center for Inherited Disease Research (CIDR), Hanyi Zhuang, and Maurine Tong for their assistance with genotyping and other FUSION investigators for their technical assistance.

REFERENCES

1. Lazar MA: How obesity causes diabetes: not a tall tale. *Science* 307:373–375, 2005
2. Fajans SS, Bell GI, Polonsky KS: Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. *N Engl J Med* 345:971–980, 2001
3. Servitja JM, Ferrer J: Transcriptional networks controlling pancreatic development and beta cell function. *Diabetologia* 47:597–613, 2004
4. Yamagata K: Regulation of pancreatic beta-cell function by the HNF transcription network: lessons from maturity-onset diabetes of the young (MODY). *Endocr J* 50:491–499, 2003
5. Malecki MT: Genetics of type 2 diabetes mellitus. *Diabetes Res Clin Pract* 68 (Suppl. 1):S10–S21, 2005
6. Barroso I, Gurnell M, Crowley VE, Agostini M, Schwabe JW, Soos MA, Maslen GL, Williams TD, Lewis H, Schafer AJ, Chatterjee VK, O'Rahilly S: Dominant negative mutations in human PPARgamma associated with severe insulin resistance, diabetes mellitus and hypertension. *Nature* 402:880–883, 1999
7. Gloyn AL, Pearson ER, Antcliff JF, Proks P, Bruining GJ, Slingerland AS, Howard N, Srinivasan S, Silva JM, Molnes J, Edghill EL, Frayling TM,

Temple IK, Mackay D, Shield JP, Sumnik Z, van Rhijn A, Wales JK, Clark P, Gorman S, Aisenberg J, Ellard S, Njolstad PR, Ashcroft FM, Hattersley AT: Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes. *N Engl J Med* 350:1838–1849, 2004
8. Parikh H, Groop L: Candidate genes for type 2 diabetes. *Rev Endocr Metab Disord* 5:151–176, 2004
9. Hegele RA, Cao H, Harris SB, Hanley AJ, Zinman B: The hepatic nuclear factor-1alpha G319S variant is associated with early-onset type 2 diabetes in Canadian Oji-Cree. *J Clin Endocrinol Metab* 84:1077–1082, 1999
10. Hani EH, Stoffers DA, Chevre JC, Durand E, Stanojevic V, Dina C, Habener JF, Froguel P: Defective mutations in the insulin promoter factor-1 (IPF-1) gene in late-onset type 2 diabetes mellitus. *J Clin Invest* 104:R41–R48, 1999
11. Macfarlane WM, Frayling TM, Ellard S, Evans JC, Allen LI, Bulman MP, Ayres S, Shepherd M, Clark P, Millward A, Demaine A, Wilkin T, Docherty K, Hattersley AT: Missense mutations in the insulin promoter factor-1 gene predispose to type 2 diabetes. *J Clin Invest* 104:R33–R39, 1999
12. Malecki MT, Jhala US, Antonellis A, Fields L, Doria A, Orban T, Saad M, Warram JH, Montminy M, Krolewski AS: Mutations in NEUROD1 are associated with the development of type 2 diabetes mellitus. *Nat Genet* 23:323–328, 1999
13. Barroso I, Luan J, Middelberg RP, Harding AH, Franks PW, Jakes RW, Clayton D, Schafer AJ, O'Rahilly S, Wareham NJ: Candidate gene association study in type 2 diabetes indicates a role for genes involved in beta-cell function as well as insulin action. *PLoS Biol* 1:E20, 2003
14. Bagwell AM, Bento JL, Mychaleckyj JC, Freedman BI, Langefeld CD, Bowden DW: Genetic analysis of HNF4A polymorphisms in Caucasian-American type 2 diabetes. *Diabetes* 54:1185–1190, 2005
15. Mohlke KL, Boehnke M: The role of HNF4A variants in the risk of type 2 diabetes. *Curr Diab Rep* 5:149–156, 2005
16. Muller YL, Infante AM, Hanson RL, Love-Gregory L, Knowler W, Bogardus C, Baier LJ: Variants in hepatocyte nuclear factor 4α are modestly associated with type 2 diabetes in Pima Indians. *Diabetes* 54:3035–3039, 2005
17. Karim MA, Wang X, Hale TC, Elbein SC: Insulin promoter factor 1 variation is associated with type 2 diabetes in African Americans. *BMC Med Genet* 6:37, 2005
18. Silander K, Mohlke KL, Scott LJ, Peck EC, Hollstein P, Skol AD, Jackson AU, Deloukas P, Hunt S, Stavrides G, Chines PS, Erdos MR, Narisu N, Conneely KN, Li C, Fingerlin TE, Dhanjal SK, Valle TT, Bergman RN, Tuomilehto J, Watanabe RM, Boehnke M, Collins FS: Genetic variation near the hepatocyte nuclear factor-4 α gene predicts susceptibility to type 2 diabetes. *Diabetes* 53:1141–1149, 2004
19. Love-Gregory LD, Wasson J, Ma J, Jin CH, Glaser B, Suarez BK, Permutt MA: A common polymorphism in the upstream promoter region of the hepatocyte nuclear factor-4 α gene on chromosome 20q is associated with type 2 diabetes and appears to contribute to the evidence for linkage in an Ashkenazi Jewish population. *Diabetes* 53:1134–1140, 2004
20. Weedon MN, Owen KR, Shields B, Hitman G, Walker M, McCarthy MI, Love-Gregory LD, Permutt MA, Hattersley AT, Frayling TM: Common variants of the hepatocyte nuclear factor-4α P2 promoter are associated with type 2 diabetes in the U.K. population. *Diabetes* 53:3002–3006, 2004
21. Damcott CM, Hoppman N, Ott SH, Reinhart LJ, Wang J, Pollin TI, O'Connell JR, Mitchell BD, Shuldiner AR: Polymorphisms in both promoters of hepatocyte nuclear factor 4-α are associated with type 2 diabetes in the Amish. *Diabetes* 53:3337–3341, 2004
22. Hansen SK, Rose CS, Glumer C, Drivsholm T, Borch-Johnsen K, Jorgensen T, Pedersen O, Hansen T: Variation near the hepatocyte nuclear factor (HNF)-4α gene associates with type 2 diabetes in the Danish population. *Diabetologia* 48:452–458, 2005
23. Winckler W, Graham RR, de Bakker PI, Sun M, Almgren P, Tuomi T, Gaudet D, Hudson TJ, Ardlie KG, Daly MJ, Hirschhorn JN, Groop L, Altshuler D: Association testing of variants in the hepatocyte nuclear factor 4α gene with risk of type 2 diabetes in 7,883 people. *Diabetes* 54:886–892, 2005
24. Vaxillaire M, Dina C, Lobbens S, Dechaume A, Vasseur-Delannoy V, Helbecque N, Charpentier G, Froguel P: Effect of common polymorphisms in the HNF4alpha promoter on susceptibility to type 2 diabetes in the French Caucasian population. *Diabetologia* 48:440–444, 2005
25. Consortium TIH: A haplotype map of the human genome. *Nature* 437: 1299–1320, 2005
26. Valle T, Tuomilehto J, Bergman RN, Ghosh S, Hauser ER, Eriksson J, Nylund SJ, Kohtamaki K, Toivanen L, Vidgren G, Tuomilehto-Wolf E, Ehnholm C, Blaschak J, Langefeld CD, Watanabe RM, Magnuson V, Ally DS, Hagopian WA, Ross E, Buchanan TA, Collins F, Boehnke M: Mapping genes for NIDDM: design of the Finland-United States Investigation of NIDDM Genetics (FUSION) study. *Diabetes Care* 21:949–958, 1998
27. Silander K, Scott LJ, Valle TT, Mohlke KL, Stringham HM, Wiles KR, Duren WL, Doheny KF, Pugh EW, Chines P, Narisu N, White PP, Fingerlin TE, Jackson AU, Li C, Ghosh S, Magnuson VL, Colby K, Erdos MR, Hill JE, Hollstein P, Humphreys KM, Kasad RA, Lambert J, Lazaridis KN, Lin G, Morales-Mena A, Patzkowski K, Pfahl C, Porter R, Rha D, Segal L, Suh YD, Tovar J, Unni A, Welch C, Douglas JA, Epstein MP, Hauser ER, Hagopian W, Buchanan TA, Watanabe RM, Bergman RN, Tuomilehto J, Collins FS, Boehnke M: A large set of Finnish affected sibling pair families with type 2 diabetes suggests susceptibility loci on chromosomes 6, 11, and 14. *Diabetes* 53:821–829, 2004
28. Saaristo T, Peltonen M, Lindström J, Saarikoski L, Eriksson J, Tuomilehto J: Cross-sectional evaluation of the Finnish Diabetes Risk Score: a tool to identify undetected type 2 diabetes, abnormal glucose tolerance and metabolic syndrome. *Diabetes Vasc Dis Res* 2:67–72, 2005
29. Winckler W, Burtt NP, Holmkvist J, Cervin C, de Bakker PI, Sun M, Almgren P, Tuomi T, Gaudet D, Hudson TJ, Ardlie KG, Daly MJ, Hirschhorn JN, Altshuler D, Groop L: Association of common variation in the HNF1α gene region with risk of type 2 diabetes. *Diabetes* 54:2336–2342, 2005
30. Skol AD, Scott LJ, Abecasis GR, Boehnke M: Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 38:209–213, 2006
31. Abecasis GR, Cookson WO: GOLD: graphical overview of linkage disequilibrium. *Bioinformatics* 16:182–183, 2000
32. Lin DY: An efficient Monte Carlo approach to assessing statistical significance in genomic studies. *Bioinformatics* 21:781–787, 2005
33. Seaman SR, Muller-Myhsok B: Rapid simulation of P values for product methods and multiple-testing adjustment in association studies. *Am J Hum Genet* 76:399–408, 2005
34. Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265, 2005
35. Becker T, Knapp M: A powerful strategy to account for multiple testing in the context of haplotype analysis. *Am J Hum Genet* 75:561–570, 2004
36. Marz W, Nauck M, Hoffmann MM, Nagel D, Boehm BO, Koenig W, Rothenbacher D, Winkelmann BR: G(-30)A polymorphism in the pancreatic promoter of the glucokinase gene associated with angiographic coronary artery disease and type 2 diabetes mellitus. *Circulation* 109: 2844–2849, 2004
37. Ye L, Xu Y, Zhu Y, Fan Y, Deng H, Zhang J: Association of polymorphism in neurogenic differentiation factor 1 gene with type 2 diabetes. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* 19:484–487, 2002
38. Kavvoura FK, Ioannidis JP: Ala45Thr polymorphism of the NEUROD1 gene and diabetes susceptibility: a meta-analysis. *Hum Genet* 116:192–199, 2005
39. Weedon MN, Owen KR, Shields B, Hitman G, Walker M, McCarthy MI, Hattersley AT, Frayling TM: A large-scale association analysis of common variation of the HNF1α gene with type 2 diabetes in the U.K. Caucasian population. *Diabetes* 54:2487–2491, 2005
40. Kulkarni RN, Kahn CR: Molecular biology: HNFs: linking the liver and pancreatic islets in diabetes. *Science* 303:1311–1312, 2004
41. Frayling TM, Evans JC, Bulman MP, Pearson E, Allen L, Owen K, Bingham C, Hannemann M, Shepherd M, Ellard S, Hattersley AT: β-Cell genes and diabetes: molecular and clinical characterization of mutations in transcription factors. *Diabetes* 50 (Suppl. 1):S94–S100, 2001
42. Laken SJ, Petersen GM, Gruber SB, Oddoux C, Ostrer H, Giardiello FM, Hamilton SR, Hampel H, Markowitz A, Klimstra D, Jhanwar S, Winawer S, Offit K, Luce MC, Kinzler KW, Vogelstein B: Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC. *Nat Genet* 17:79–83, 1997
43. Boyle MP: Nonclassic cystic fibrosis and CFTR-related diseases. *Curr Opin Pulm Med* 9:498–503, 2003
44. Kruger R: Genes in familial parkinsonism and their role in sporadic Parkinson's disease. *J Neurol* 251 (Suppl. 6):VI/2–6, 2004
45. Skipper L, Li Y, Bonnard C, Pavanni R, Yih Y, Chua E, Sung WK, Tan L, Wong MC, Tan EK, Liu J: Comprehensive evaluation of common genetic variation within LRRK2 reveals evidence for association with sporadic Parkinson's disease. 14:3549–3556, 2005
46. Dalgleish R: The human type I collagen mutation database. *Nucleic Acid Res* 25:181–187, 1997
47. Grant SF, Reid DM, Blake G, Herd R, Fogelman I, Ralston SH: Reduced bone density and osteoporosis associated with a polymorphic Sp1 binding site in the collagen type I alpha 1 gene. *Nat Genet* 14:203–205, 1996
48. Peris P, Alvarez L, Oriola J, Guanabens N, Monegal A, de Osaba MJ, Jo J, Pons F, Ballesta AM, Munoz-Gomez J: Collagen type Ialpha1 gene polymorphism in idiopathic osteoporosis in men. *Rheumatology (Oxford)* 39:1222–1225, 2000
49. Hegele RA: Small genetic effects in complex diseases: a review of regulatory sequence variants in dyslipoproteinemia and atherosclerosis. *Clin Biochem* 30:183–188, 1997
50. Knight JC: Regulatory polymorphisms underlying complex disease traits. *J Mol Med* 83:97–109, 2005