METHODOLOGY

A more rapid approach to systematically assessing published associations of genetic polymorphisms and disease risk: type 2 diabetes as a test case

Alex H Cho[1]
Xiaolei Jiang[2]
Devin M Mann[3]
Kensaku Kawamoto[4]
Timothy J Robinson[5]
Nancy Wang[6]
Jeanette J McCarthy[2]
Mark Woodward[7]
Geoffrey S Ginsburg[1,2]

[1]Center for Personalized Medicine and Department of Medicine, Duke University, Durham, NC, [2]Institute for Genome Sciences and Policy, Duke University, Durham, NC, [3]Section of Preventive Medicine and Epidemiology, Department of Medicine, Boston University School of Medicine, Boston, MA, [4]Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, [5]Medical College of Virginia, Richmond, VA, [6]School of Medicine, University of North Carolina-Chapel Hill, Chapel Hill, NC, USA; [7]George Institute for Global Health and University of Sydney, Australia

Correspondence: Alex Cho
DUMC Box 3382, Durham,
NC 27708, USA
Tel + 1 919 314 7361
Email alex.cho@duke.edu

Background: Comparative effectiveness research and research in genomic medicine are not orthogonal pursuits. Both require a robust evidence base, and each stands to benefit from applying the methods of the other. There is an exponentially growing literature reporting associations between single nucleotide polymorphisms (SNPs) and increased risk for diseases such as type 2 diabetes. Literature-based meta-analysis is an important method of assessing the validity of published gene-disease associations, but a traditional emphasis on exhaustiveness makes it difficult to study multiple polymorphisms efficiently. Here we describe a novel two-step search method for broadly yet systematically reviewing the literature to identify the "most-studied" gene-disease associations, thereby selecting those with a high possibility of replication on which to conduct abbreviated, simultaneous meta-analyses. This method was then applied to identify and evaluate the validity of SNPs reported to be associated with increased type 2 diabetes risk, to demonstrate proof of principle.

Methods: A two-step MEDLINE search (1950 to present) was conducted in September 2007 for published genetic association data related to SNPs associated with risk of type 2 diabetes. The top 10 "most-studied" genes were selected for focused searches and final inclusion/exclusion determinations. To demonstrate the ability to efficiently update this two-step search for additions to the literature, an update of the second-step search was conducted 9 months later. Abstracted data were sorted based on study design, risk model, and specific SNPs. Meta-analyses were performed for individual SNPs, with separate analyses done for case-control and prospective studies, and were compared with the results of more recent genome-wide association studies.

Results: The first-step search found 1116 articles covering 108 different genes. The top ten "most-studied" genes were: *ABCC8* (or *SUR1*), *ACE*, *CAPN10*, *KCNJ11* (or *Kir6.2*), *HNF1* alpha, *HNF4* alpha, *IL-6*, *PGC-1* alpha, *PPAR* gamma 2, and *TCF7L2*. The second-step search found a total of 658 articles, yielding 124 articles for initial data abstraction and analysis. We also demonstrated the ability to update this search as newer studies appeared, using the same method almost a year later to find an additional 107 articles (77 were ultimately excluded), bringing the number of included studies to 154. From these studies, data on 90 different DNA variants within the ten genes were abstracted. Simultaneous meta-analyses found that higher-risk alleles for SNPs rs7903146 and rs12255372 in *TCF7L2*, rs1801282 in *PPAR* gamma 2, rs5219 in *KCNJ11*, rs3792267 in *CAPN10*, rs2144909 in *HNF4* alpha, and rs1800795 in *IL-6* appeared to be associated with increased type 2 diabetes risk. These findings were generally highly concordant with the results of traditional literature-based meta-analyses performed for individual genes.

Conclusions: The methodology described in this manuscript represents a reasonable approach to more rapidly identifying and evaluating frequently studied genetic-risk markers for diseases such as type 2 diabetes. Comparison with results of traditional meta-analyses suggests that

1

these gains in efficiency do not necessarily come at the price of reduced accuracy. Given the quickening pace of discovery of such markers, more efficient, unbiased, and readily updatable methods for systematically assessing and re-assessing a changing literature could prove valuable. Good methods for evidence evaluation are also important to the potential application of genetic markers to comparative effectiveness research, and vice versa.

Background

It has been a decade since the "first draft" of the human genome was completed and published.[1] Genetic and genomic information have the potential to help tailor risk assessment, diagnosis, and even therapy to individual biological characteristics, helping to enable the practice of what is often termed "genomic medicine" or "personalized medicine." Comparative effectiveness research (CER) is sometimes thought of as conventionally valuing only the average benefit of a given therapy or diagnostic to the average patient, over entire populations. However, as Sox and Greenfield assert, CER also extends to "research designed to identify the clinical characteristics that predict which intervention would be most successful in an individual patient."[2] Thus, CER and research in genomic medicine are not orthogonal pursuits. As Khoury et al note, both require a robust evidence base to drive "knowledge-driven practice," and each stands to benefit from applying the methods of the other. CER is needed to rigorously compare outcomes for genome-based care versus traditional approaches, and genetic and other biomarkers can help identify the subgroups shown to benefit most from proven therapies.[3]

Type 2 diabetes mellitus represents a highly relevant and important example of a genetically complex chronic disease that has been a focus of both genomic investigation and CER. Close to 20 million people currently suffer from diabetes in the USA, and over a million new cases are diagnosed each year. These numbers have surged in recent years owing to a corresponding increase in obesity.[4] On the other hand, known risk factors for type 2 diabetes such as obesity are modifiable, and studies such as the Diabetes Prevention Program suggest that diabetes onset can be delayed or prevented if lifestyle changes are adopted, even among individuals at higher genetic risk.[5]

Three waves of discovery research into genetic markers associated with type 2 diabetes have been described.[6] The first wave was characterized by family-based multigenerational linkage studies searching for genetic loci associated with inherited forms of diabetes. The second utilized "candidate gene approaches" to look at the association of single or small numbers of genes with diabetes, generally in relatively small case-control studies.

Following completion of the HapMap project – a "sequel" to the Human Genome Project that cataloged sites of common DNA variation (called single-nucleotide polymorphisms, or SNPs) across different populations – a third wave of discovery emerged, driven by large genome-wide association studies (GWAS) using large cohorts of unrelated diabetic and nondiabetic individuals. These GWAS, made possible by newer DNA genotyping technologies that can test for hundreds of thousands of SNPs simultaneously and economically, rapidly accelerated the search for gene-disease connections in type 2 diabetes, increasing the number of currently known genetic loci to as many as 40.[6]

However, a number of issues have been raised regarding the clinical utility of positive findings of genetic risks for complex chronic diseases such as type 2 diabetes.[7] While biologically interesting, the increases in risk noted for common DNA variants are generally small. In addition to clear differences in methods between older and newer studies, there has also been great variation in how results have been reported, further confusing matters. Positive chance findings are not uncommon; thus, the clinical validity of these associations (on which their utility ultimately depends) rests on their replication in multiple studies.[8,9]

One traditional approach to evidence evaluation is to conduct meta-analyses that combine reported results to examine trends across studies. Virtues of such analyses include their efficiency versus conducting de novo trials or even observational studies, their ability to combine data across studies, and rigor. However, there are inherent limitations in using exhaustive searches for individual markers to evaluate associations between multiple genetic variants and risk for complex chronic diseases.

First, these methods require one to select a priori a particular SNP-disease association of interest. But, given the sheer volume of markers that have been and could be studied, deciding which markers ought to be looked at more closely itself introduces a detection or attention bias of its own.

Second, one is expected to conduct an exhaustive search for all published articles that examine this relationship, as well as any unpublished data that one might gain access to. The rationale for this practice is due to real concerns

regarding publication bias with respect to negative results. In the case of studies of genetic risk for complex chronic diseases, due to the modest effect sizes generally seen and very large numbers of subjects required, findings may tend to be falsely negative. Thus, by insisting on being exhaustive up front, one may risk biasing reviews towards the inconclusive results of numerous smaller, underpowered studies; that is, commit a Type II, or beta error, failing to see an association that is real.[10]

Finally, conducting formal meta-analyses of every genetic variant that has ever been studied for a given disease or condition is simply not practical. As chip-based studies continue in different populations, the number of new, validated, and contested associations is likely to continue to grow at a rapid rate.[7] Complicating matters is the fact that individual SNPs cannot easily be searched for in scientific literature databases such as MEDLINE, which presently only index content at the level of individual genes.

At the same time, one would not want to rely on the results of single studies reporting positive gene-disease associations, haphazardly selected. There is thus a need for new strategies to search this literature that are "agnostic" to the identity of the particular markers one is looking for, relatively efficient in terms of capturing the range of new markers for which primary marker-disease association data are being reported, and systematic in their approach.

In this paper we present a two-step, nondirective and updatable search strategy for identifying which genetic polymorphisms have been "most-studied" for an association with risk of type 2 diabetes. Our objective was to identify associations that had been replicated multiple times in a way that was comprehensive yet efficient, broad in coverage (ie, covering many markers at once, or "multiplex"), and capable of maintaining pace with the emerging literature. The results of the studies of individual SNPs found by this strategy were then combined to arrive at point risk estimates for each SNP, and compared with estimates reported in the literature from traditional meta-analyses.

Methods
Study selection

A two-step search of MEDLINE (1950 to present) using PubMed was first conducted in September 2007 for published genetic association data related to SNPs associated with risk of type 2 diabetes. First, an initial comprehensive search was performed to capture the breadth of genes that had been studied in this manner, using the five key search terms plus limits and specific exclusions outlined in Table 1. These were sorted by publication date, to permit the addition of new studies in the future. Articles from the European journal *Diabetologia* were added using a separate search, because of an observed lag between time of publication and indexing in MEDLINE by medical subject headings (MeSH).

Two readers independently reviewed all publication titles for rapid inclusion/exclusion in this initial tally of the number of studies for each gene symbol. Gene symbol was used because MEDLINE indexes by gene and not by individual SNPs. By protocol, abstracts or the article texts themselves were not consulted during this initial, screening phase in a deliberate effort to increase efficiency. Ranking genes in order of frequency of study, the top ten "most-studied" genes were selected for focused searches and final inclusion/ exclusion determinations were made in a second step (see Results, below).

The second step involved conducting individual, focused searches for each of these top ten genes. All listed synonyms from Entrez Gene were used, along with the terms from the initial search. (For example, listed synonyms "ppar gamma," "PPAR gamma," and "peroxisome proliferator-activated receptor gamma" were linked with an "AND"

Table 1 Terms used for Step 1 PubMed search

Description	Search string
Main search	"Diabetes mellitus, type 2/genetics"[MAJR] AND ("genetic markers"[MeSH Terms] OR "polymorphism, genetic"[MeSH Terms] OR "genetic predisposition to disease"[MeSH Terms] OR "epidemiologic studies"[MeSH Terms])
Limits	AND (hasabstract[text] AND "humans"[MeSH Terms] AND English[lang] NOT ("adolescent"[MeSH Terms] AND "child"[MeSH Terms] AND "infant"[MeSH Terms]))
Specific exclusions	NOT review[PT] NOT ("haplotypes"[TIAB] OR ("messenger rna"[Text Word] OR "rna, messenger"[MeSH Terms] OR mRNA[Text Word]) OR (("mice"[TIAB] NOT Medline[SB]) OR "mice"[MeSH Terms] OR mouse[Text Word]) OR ("mitochondrial dna"[Text Word] OR "dna, mitochondrial"[MeSH Terms] OR mtDNA[Text Word]))
Diabetologia add-on	OR (Diabetologia[All Fields] AND ("genetic polymorphism"[Text Word] OR "polymorphism, genetic"[MeSH Terms] OR (SNP[All Fields] OR polymorphism[Text Word])))

statement to the search string from Table 1 to search for *PPAR* gamma 2-related articles that examined its association with type 2 diabetes.) After sorting by publication date, two readers again independently reviewed all titles and abstracts to decide which articles should be obtained for reading based on predetermined excluded categories. Disagreements were adjudicated by a third reader. All searches were reviewed by an expert in library science.

Articles were excluded if they (1) studied complications or comorbidities of diabetes; (2) were "functional" in nature (ie, studied physiology or prediabetic phenotypes); (3) took place in animals or were in vitro; (4) focused on therapies for diabetes; (5) did not study one of the top ten genes; (6) studied conditions other than sporadic type 2 diabetes (including maturity-onset diabetes of the young or gestational diabetes mellitus); (7) presented no new primary data (ie, reviews or meta-analyses); (8) presented interim data that were included in subsequent publications; (9) were "exploratory" in nature (ie, presented data associated with loci but not individual polymorphisms); (10) were not at least a case-control study; (11) did not examine genetic risk; or (12) only presented data for risk associated with more than one genetic marker but not individual markers (ie, haplotypes).

More recent GWAS that had been missed owing to their not having been indexed in MEDLINE by gene were also added (located by substituting the MeSH term "genome-wide association study" for gene symbol).

Article reading and coding

All articles were pulled for more detailed review. Final inclusion/exclusion determinations prior to data abstraction and coding were made, during which process additional articles were excluded based on the same categories as above, again subject to a two-reader review process.

Data were then abstracted in consultation with a genetic epidemiologist, including study design and other characteristics, genetic risk model(s) used, and results. Genetic risk models were categorized as "allelic," in which the unit of analysis was the individual allele; or "genotypic," additive/multiplicative/dominant/recessive, in which the unit of analysis was the individual subject and his/her genotype, with some assumption made about the risk conferred by having zero, one, or two higher-risk forms of the allele in question. Results for case-control studies were recorded as odds ratios (ORs) for having one or two higher-risk alleles, plus confidence intervals (CIs) and *P*-values where available; and for prospective cohort or randomized controlled trials as relative risks or hazard ratios.

Data for all SNPs that had been reported to be significantly associated with type 2 diabetes risk in at least one study were coded. When a series of publications analyzed the same or overlapping data (eg, from the same cohort at different points in time), the latest publication that reported data from the largest or most recent study was the one included. Data were abstracted from both the text of the published article as well as any available supplemental data. All entries were reviewed a second time for completeness and accuracy through a separate quality control step.

Demonstration of ability to readily update search

To demonstrate the ability to efficiently update this two-step search for additions to the literature, an update of the second-step search was conducted 9 months later, in June 2008.

Data sorting and final SNP selection

Data abstracted from the included studies were sorted based on study design (case control, prospective cohort, randomized controlled trial), risk model (allelic, etc) as well as by the specific SNPs. Data for SNPs that were in high linkage disequilibrium with one another (ie, >0.98) were combined under a single SNP label. For instance, studies examining rs5215 and rs5219 in *KCNJ11* (also referred to as E23K in the literature) were grouped under "rs5219" (and will be referred to as such for the purposes of this paper). Of note, the polymorphism identified in the *ACE* gene is an insertion/deletion variant, not an SNP. Only polymorphisms with three or more risk estimates were included in a final dataset for meta-analysis.

For genes in which multiple SNPs had been studied, we did meta-analysis for the one-two most-studied SNPs within these genes. Of note, the polymorphism identified in the *ACE* gene is an insertion/deletion (I/D) variant, not an SNP. And in the special case of *HNF4* alpha, in which more than two dozen SNPs had been studied in this gene alone, we selected six of the most-studied ones, three of which were clustered in the P2 promoter region, and three of which were distributed throughout the coding region.

Meta-analysis

For the purposes of meta-analysis, data from included studies were used if complete risk estimates and confidence intervals were reported. Results from studies using an allelic model of risk were grouped with one-copy "genotypic" results in a single case-control study category. When studies presented both allelic and "genotypic" risk results, the latter were used.

Separate meta-analyses were performed for case-control and prospective studies. Individual risk estimates were pooled using random effects meta-analysis with inverse variance weighting in Stata, version 10 (College Station, TX, USA). It was assumed a priori that there would not be a fixed difference between one included study and the rest of the studies in a given meta-analysis, hence the use of random effects. The I-squared statistic was used to estimate the percentage of variability across studies that is attributable to heterogeneity, and tested for deviation from zero.[11] Odds ratios reported are for higher-risk alleles, regardless of prevalence.

Results

The yield from the two-step search strategy is depicted in Figure 1A and 1B. The first-step search found 1903 articles, yielding 1116 articles covering 108 different genes. The top ten "most-studied" genes, from this first step of the two-step search strategy, were (in alphabetical order): *ABCC8* (or *SUR1*), *ACE*, *CAPN10*, *KCNJ11* (or *Kir6.2*), *HNF1* alpha, *HNF4* alpha, *IL-6*, *PGC-1* alpha, *PPAR* gamma 2, and *TCF7L2*. Cohen's kappa statistic for interrater agreement for article inclusion/exclusion was 0.73.

The second-step search – a focused search for studies of each of these ten genes – found a total of 658 articles. Review of titles and abstracts yielded 135 included articles. Figure 2 depicts the distribution and categories of those excluded. The addition of genome-wide association studies brought the total number of articles pulled for more detailed review to 139. After reading and prior to data abstraction and coding, final inclusion/exclusion determinations excluded 15 additional articles based on the same categories, bringing the total number of articles read and coded to 124.

The update conducted 9 months later found an additional 107 articles; 74 were excluded immediately, and 33 were selected for full-text review. Of these, 30 were ultimately included, bringing the updated total of studies included in the final dataset to 154, of which 134 had complete data for analysis, covering 90 separate SNP–diabetes risk associations.

The results of meta-analyses for SNPs from eight of the ten genes are reported, which met the criteria for numbers of studies and completeness of data described earlier; these are listed in Table 2.[5,12–85,86] Meta-analyses of the case-control studies found that higher-risk alleles for SNPs rs7903146 (OR = 1.42; 95% CI 1.36–1.48) and rs12255372 in *TCF7L2* (OR = 1.40; 95% CI 1.32–1.50), rs1801282 in *PPAR* gamma 2 (OR = 1.16; 95% CI 1.02–1.32), rs5219 in *KCNJ11* (OR = 1.16; 95% CI 1.12–1.20), rs3792267 in *CAPN10* (OR = 1.09; 95% CI 1.03–1.16), rs2144909 in *HNF4*
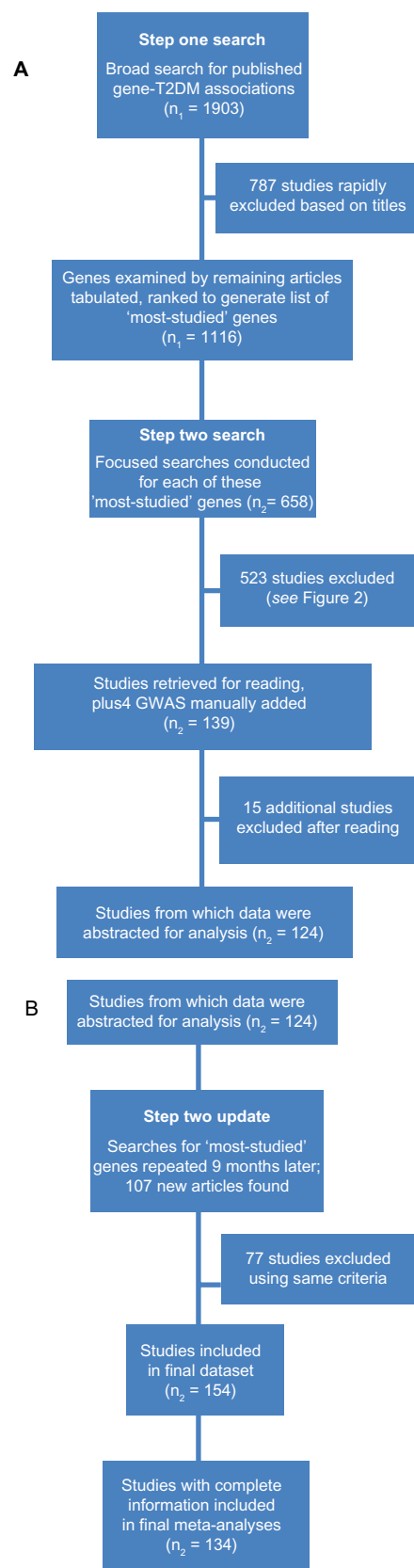

Figure 1 (**A** and **B**) Two-step study selection process.
Notes: n_1 = number of articles yielded at each stage of initial broad search; n_2 = number of articles yielded by subsequent, focused searches for "most-studied" genes.
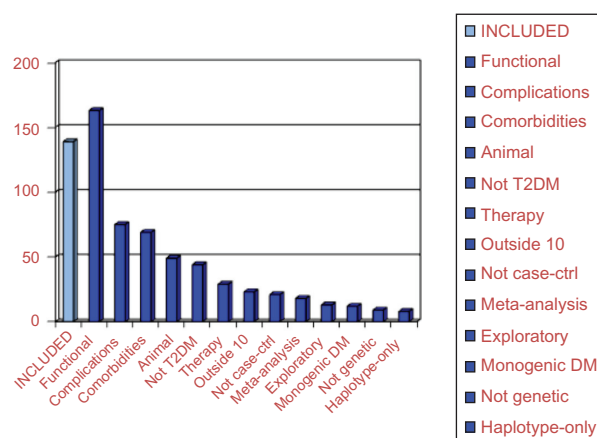**Abbreviation:** GWAS, genome-wide association study.

Figure 2 Distribution and categories of excluded articles.

alpha (OR = 1.13; 95% CI 1.00–1.27), and rs1800795 in *IL-6* (OR = 1.45; 95% CI 1.12–1.88) were each associated with increased type 2 diabetes risk in a statistically significant manner across multiple studies. The *ACE* insertion/deletion (I/D) polymorphism and rs8192678 in *PGC-1* alpha were not found to be significantly associated with risk for type 2 diabetes. Only two SNPs had three or more prospective cohort studies examining progression to diabetes to warrant meta-analysis: rs5219 in *KCNJ11* (OR = 0.96; 95% CI 0.63–1.45) and rs7903146 in *TCF7L2* (OR = 1.24; 95% CI 1.01–1.50). Three representative Forest plots are shown in Figures 3–5. (Also see Supplemental Figures 1–13 for additional plots.)

These results were then compared to the results of other, more traditional meta-analyses that studied individual genes and their association with type 2 diabetes risk. A 2004 meta-analysis of *CAPN10* and rs3792267 (also referred to as

Table 2 SNPs included in meta-analyses

Gene	Polymorphisms[a]
ACE	I/D
CAPN10	rs2975760 (SNP-44)
	rs3792267 (SNP-43)
HNF4 alpha[b]	rs1884614
	rs1885088
	rs2144908
	rs2425637
	rs3818247
IL-6	rs1800795 (C-174G)
KCNJ11	rs5219 (E23K)
PGC-1 alpha	rs8192678 (Gly482Ser)
PPAR gamma 2	rs1801282
TCF7L2	rs7903146
	rs12255372

Notes: [a]Other common names appearing in the literature are in parentheses; [b]a sixth *HNF4* alpha SNP, rs4810424, did not have three studies providing complete data; [c]linkage disequilibrium of rs5219 and rs5215 > 0.99.
Abbreviations: I/D, insertion/deletion; SNPs, single-nucleotide polymorphisms.

UCSNP-43 or SNP-43), which pooled data from eleven studies, found increased odds of type 2 diabetes of 1.19 (95% CI 1.07–1.33).[17] A meta-analysis of *KCNJ11* from 2005, which combined primary data from two case-control studies with data from eight previously published studies, found rs5219 (or E23K) had an OR of 1.12 (95% CI 1.01–1.23).[87] In 2007, Ludovico et al published a meta-analysis of 41 *PPAR* gamma 2 studies, which found the higher-risk form of rs1801282 had an OR of 1.23 (95% CI 1.14–1.33).[88] This finding was replicated by Gouda et al in their 2010 Human Genome Epidemiology (HuGE) review and meta-analysis; across 66 studies the OR for the higher-risk allele was 1.16 (95% CI 1.11–1.23).[89]

Another HuGE review, this one from 2009, examined 35 studies of *TCF7L2* and T2DM risk and found that the higher-risk allele for rs7903146 conferred an OR of 1.41 (95% CI 1.34–1.48); and for rs12255372 an OR of 1.36 (95% CI 1.29–1.43).[90] A meta-analysis published in 2010 by Sookoian et al appeared to validate an association with T2DM risk for two haplotype blocks and three single SNPs in the *HNF4* alpha gene.[91] Finally, another meta-analysis from 2010 looked at 24 studies of the *ACE* I/D polymorphism, concluding that there was a significant association with type 2 diabetes risk, with an OR of 1.14 (95% CI 1.04–1.24).[92] Table 3 summarizes the comparison between our results and those of these other studies.

Discussion

A reproducible, updatable search strategy was developed to systematically and efficiently find the "most-studied" genetic polymorphisms associated with risk of a genetically complex chronic disease, and proof of principle demonstrated in the case of type 2 diabetes. The intent was not to be exhaustive, but to relatively rapidly establish what genes have been most frequently studied in terms of association with disease risk – and thus have the greatest possibility of having been replicated – and then to determine whether or not these associations had in fact been validated multiple times. Six of the ten genes looked at were found to contain at least one SNP meeting this standard: *TCF7L2*, *PPAR* gamma *2*, *KCNJ11*, *CAPN10*, *HNF4* alpha, and *IL-6*. Each SNP conferred an increase in odds of type 2 diabetes that was between 1.0 and 1.5 (1.09–1.42).

In addition, when compared to the results of separately conducted, more traditional meta-analyses for individual genes, the findings from our abbreviated, simultaneous method were generally highly concordant, suggesting that these gains in efficiency do not necessarily come at the price of reduced accuracy. The two exceptions were the *ACE* I/D

TCF7L2 rs7903146

	Odds ratio (95% CI)	% weight
Allelic		
Chang[58]	0.81 (0.50, 1.31)	0.72
Elbein (African-American)[55]	1.05 (0.80, 1.37)	1.96
Saadi[57]	1.16 (0.81, 1.64)	1.26
Rees[56]	1.31 (1.11, 1.56)	3.65
Scott[76]	1.34 (1.21, 1.49)	5.70
Saxena[77]	1.40 (1.30, 1.50)	6.94
Miyake[61]	1.48 (1.20, 1.84)	2.74
Weedon[34]	1.48 (1.36, 1.60)	6.57
Cauchi (Austria)[78]	1.52 (1.29, 1.78)	3.88
Cauchi (Morocco)[78]	1.56 (1.22, 1.89)	2.65
Cauchi[59]	1.69 (1.55, 1.83)	6.50
Salonen[60]	1.71 (1.41, 2.08)	3.10
Elbein (White)[55]	1.72 (1.25, 2.37)	1.48
Subtotal (I-squared = 65.5%, P = 0.001)	1.44 (1.34, 1.55)	47.15
1-copy		
Guo[63]	1.04 (0.82, 1.32)	2.35
Parra[65]	1.23 (0.85, 1.78)	1.16
Ng[79]	1.27 (0.71, 2.29)	0.50
Hayashi[73]	1.31 (1.00, 1.72)	1.93
Marzi[64]	1.33 (1.09, 1.62)	3.03
Scott[76]	1.33 (1.14, 1.56)	3.99
Groves[67]	1.35 (1.19, 1.53)	4.94
Kimber[68]	1.36 (1.21, 1.52)	5.34
van Vliet-Ostaptchouk[69]	1.37 (1.08, 1.73)	2.39
DGI[36]	1.38 (1.31, 1.46)	7.56
Lewis[80]	1.39 (1.21, 1.60)	4.49
Chandak[70]	1.39 (1.08, 1.78)	2.19
Bodhini[66]	1.44 (1.18, 1.76)	2.99
Mayans[72]	1.49 (1.21, 1.83)	2.86
Sladek[81]	1.65 (1.46, 1.84)	5.29
Dahlgren[74]	1.88 (1.32, 2.67)	1.26
Ren[62]	1.98 (1.13, 3.38)	0.57
Subtotal (I-squared = 22.8%, P = 0.190)	1.40 (1.33, 1.46)	52.85
Overall (I-squared = 51.1%, P = 0.001)	1.42 (1.36, 1.48)	100.00

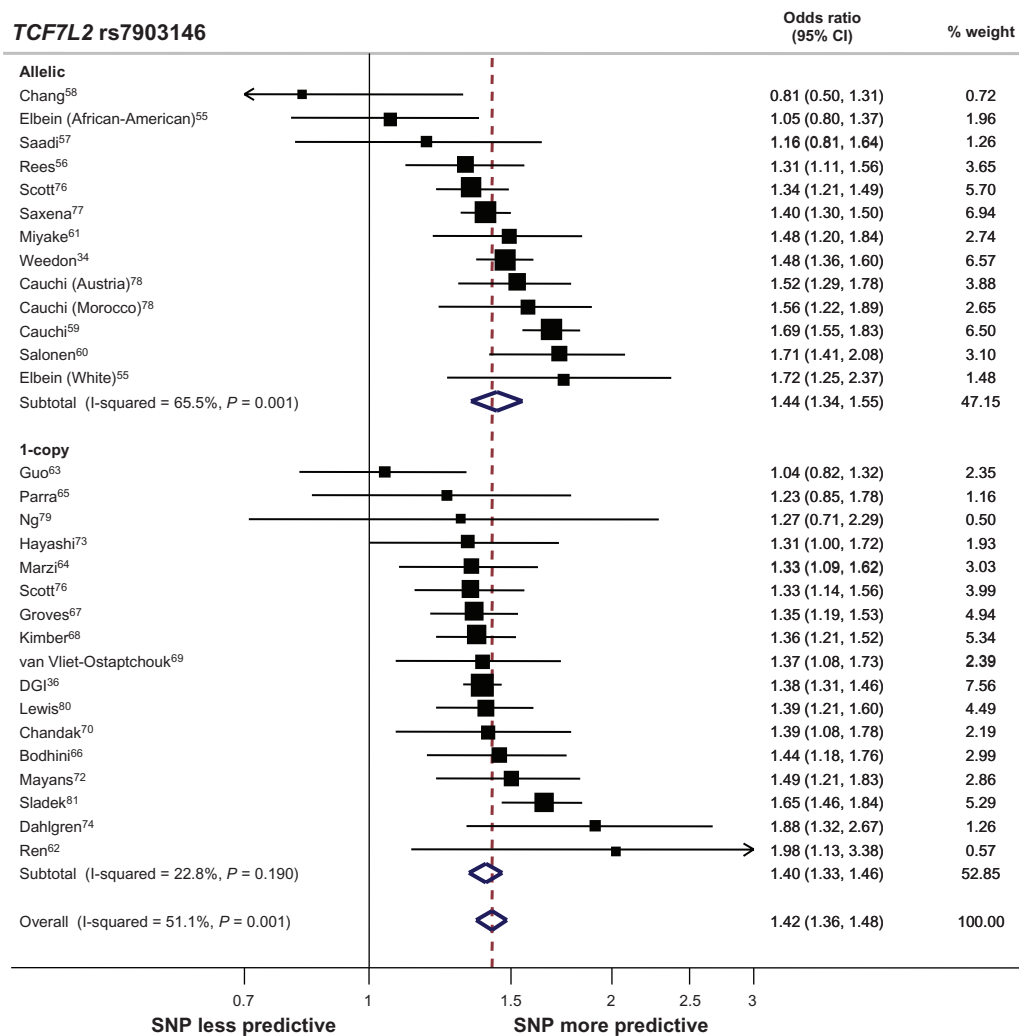**SNP less predictive**   **SNP more predictive**

Figure 3 Forest plot for case control studies of *TCF7L2* rs7903146.
Abbreviations: CI, confidence interval; SNP, single-nucleotide polymorphism.
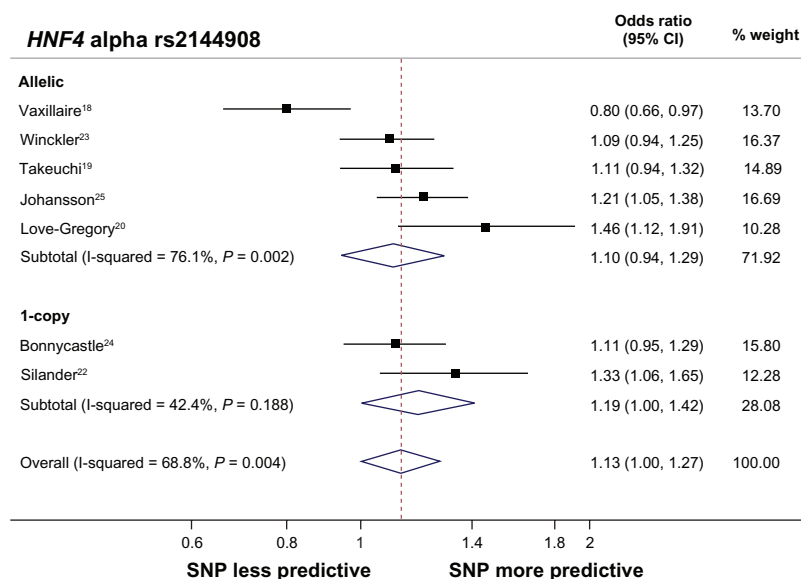
HNF4 alpha rs2144908

	Odds ratio (95% CI)	% weight
Allelic		
Vaxillaire[18]	0.80 (0.66, 0.97)	13.70
Winckler[23]	1.09 (0.94, 1.25)	16.37
Takeuchi[19]	1.11 (0.94, 1.32)	14.89
Johansson[25]	1.21 (1.05, 1.38)	16.69
Love-Gregory[20]	1.46 (1.12, 1.91)	10.28
Subtotal (I-squared = 76.1%, P = 0.002)	1.10 (0.94, 1.29)	71.92
1-copy		
Bonnycastle[24]	1.11 (0.95, 1.29)	15.80
Silander[22]	1.33 (1.06, 1.65)	12.28
Subtotal (I-squared = 42.4%, P = 0.188)	1.19 (1.00, 1.42)	28.08
Overall (I-squared = 68.8%, P = 0.004)	1.13 (1.00, 1.27)	100.00

SNP less predictive   **SNP more predictive**

Figure 4 Forest plot for case-control studies of *HNF4* alpha rs2144908.
Abbreviations: CI, confidence interval; SNP, single-nucleotide polymorphism.
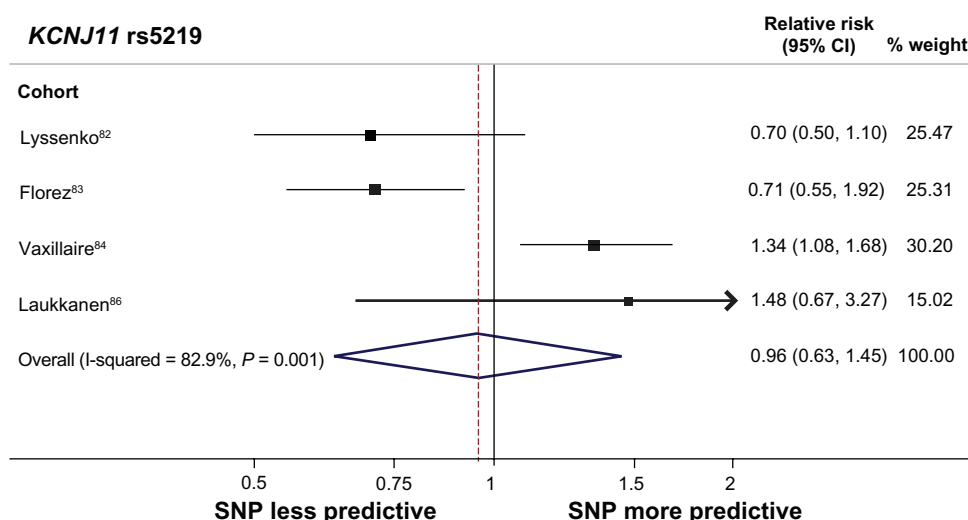
KCNJ11 rs5219

Figure 5 Forest plot for prospective cohort studies of *KCNJ11* rs5219.
Abbreviations: CI, confidence interval; SNP, single-nucleotide polymorphism.

polymorphism and the *HNF4* alpha gene. In the case of *ACE* I/D, only three studies ended up having usable data. In the latter instance, despite the fact that three of four point estimates did not fall within the 95% confidence intervals from a more traditional meta-analysis, there was still quite a bit of overlap between the confidence intervals arrived at via these two methods. Moreover, the use of haplotype blocks by the comparison meta-analysis – defined as comprising two

or more SNPs in linkage disequilibrium (LD) >0.8 – may not make these results directly comparable.[91]

Advantages versus traditional meta-analysis

A strength of this two-step search strategy is the nondirective identification of a list of genes that have been studied multiple times in terms of a possible association with increased risk

Table 3 Comparison of new method with results from traditional meta-analyses

Genetic polymorphism	OR from new method (95% CI)	Number of studies	OR from traditional meta-analysis (95% CI)	Number of studies	New point estimate falls within 95% CI of traditional result?
ACE					
I/D	0.96 (0.73–1.26)	3	1.14 (1.04–1.24)	24	No
CAPN10					
rs3792267	1.09 (1.03–1.16)	3	1.19 (1.07–1.33)	11	Yes
HNF4 alpha					
rs1884614[a]	1.13 (0.96–1.34)	5			
rs2144908[a]	1.13 (1.00–1.27)	7	0.91 (0.87–0.95)	17	No
rs2425637[b]	1.10 (0.98–1.22)	5	1.02 (0.96–1.08)	17	No
rs1885088[c]	1.07 (0.93–1.24)	5	0.97 (0.89–1.05)	17	No
rs3818247	1.00 (0.90–1.10)	6	1.03 (0.98–1.08)	17	Yes
KCNJ11					
rs5219[d]	1.16 (1.12–1.20)	14	1.12 (1.01–1.23)	10	Yes
PPAR gamma 2					
rs1801282	1.16 (1.02–1.32)	15	1.23 (1.14–1.33)[e]	41	Yes
			1.16 (1.11–1.23)[f]	66	Yes
TCF7L2					
rs7903146	1.42 (1.36–1.48)	30	1.41 (1.34–1.48)	35	Yes
rs12255372	1.40 (1.32–1.50)	23	1.36 (1.29–1.43)	35	Yes

Notes: [a]rs1884614 and rs2144908 were included with one other SNP in a haplotype block by Sookoian et al.[91] The OR presented here for the purposes of comparison is the random-effects result for the entire haplotype block; [b]rs2425637 was included with one other SNP in a haplotype block by Sookoian et al.[91] OR for comparison is for this entire block; [c]rs1885088 was included with two other SNPs in a haplotype block by Sookoian et al.[91] OR for comparison is for this entire block; [d]linkage disequilibrium of rs5219 and rs5215 > 0.99; [e]from meta-analysis by Ludovico et al.[88] Because the ORs reported by Ludovico et al were for minor allele, which is the lower-risk one, the OR presented here for comparison was calculated by inverting, ie, = 1/x where x is the OR from that meta-analysis. [f]from meta-analysis by Gouda et al.[89] The same calculation was performed as for the Ludovico et al comparison.[88]
Abbreviations: OR, odds ratio; CI, confidence interval; I/D, insertion/deletion; SNPs, single-nucleotide polymorphisms.

for a common chronic disease, in this case type 2 diabetes. Such an approach is less prone to a form of attention bias that results from the a priori selection of particular candidate genes for further study – mirroring the genome-wide approach to genetic epidemiology that has been made possible by advances in genotyping technology. Each gene on this list has crossed a "threshold" in terms of having been studied on multiple occasions relative to its peers, which is important because it selects for genes that have a high possibility of demonstrated reproducibility, required for any clinically valid biomarker. In addition, this approach allows for evaluation of new gene-disease associations, as other loci not originally included are studied with increasing frequency. Over time, the result is an organic, evolving, and systematically curated list of promising candidate genes. And for SNPs within these genes for which this modified meta-analysis has already been performed, any new results can be incorporated into the previous analyses.

Another strength is the relative efficiency of this method over standard ones, in terms of time and effort spent on a per-variant basis. Using this method, we were able to systematically evaluate 90 polymorphisms across ten genes in a single two-step search. A PubMed search in August 2011 using MeSH terms "diabetes mellitus, type 2" and "polymorphism, genetic," limited by article type to include only meta-analyses, found 15 published meta-analyses of 31 polymorphisms in 13 genes associated with type 2 diabetes risk. The earliest of these studies date back to 1998, since which time hundreds of gene-diabetes associations have been reported in the peer-reviewed literature.[7]

This paucity of secondary meta-analyses of published genetic associations – for a condition that has been among the most studied in this regard – may be due to there being a limited number of type 2 diabetes polymorphisms that have been studied multiple times, but probably also reflects the effort required to complete a traditional meta-analysis. In the method described in this paper, the goal of the searches was not to be exhaustive but to be representative, focusing on the importance of replication as the sine qua non of validity.[93]

Comparison with genome-wide association studies

Some might argue that GWAS will render meta-analyses of the published literature obsolete. However, the considerable costs associated with conducting genome-wide scans across populations of adequate size and consistently characterized phenotypes make it likely there will be limits to the extent to which new GWAS are undertaken solely for the sake

of replication.[4] The continuing need for assessments of replicability, moreover, is illustrated by the example of a recent review, which lists 39 common DNA variants as being associated with type 2 diabetes risk. Thirty-four were identified in GWAS, and a third of these in single studies.[6]

One option for confirming validity would be to use already collected raw data pooled from multiple studies to conduct "mega"-analyses for any gene-disease association of interest. This has been the strategy of groups such as the Diabetes Genetics Replication and Meta-analysis (DIAGRAM) consortium.[94,95] While there is a move towards making these data publicly available and recorded in standard ways – through consortia and other means (including for type 2 diabetes), and emerging tools such as field synopses to help synthesize data from diverse sources – such "mega"-analyses are still somewhat of a futuristic prospect for most diseases.[96]

Another example of an alternative to direct replication has been described, in which an electronic medical record linked to genetic information was used to examine previously reported genetic associations for a number of diseases – a form of GWAS using a naturally occurring cohort.[97] But to date the routine banking of biologic specimens from well-annotated clinical populations is not yet widespread outside of large, independently funded cohort studies (eg, the eMERGE network).[98]

When none of the above approaches is possible, efficient and systematic methods of surveying the broader published literature represent yet another cost-effective strategy to distinguish true signals – particularly when non-replication has been observed in one or more studies – from the proverbial noise. A combination of methods will thus continue to be useful.[4,93]

Limitations

General limitations of the two-step method presented here include that it cannot be as exhaustive as separate meta-analyses of individual polymorphisms. Supplemental data not reported in the body of the published article but made available by the publishing journal (eg, online) were examined where available – especially important in the case of GWAS – but data from unpublished studies were not specifically sought out for the reasons mentioned earlier. These simultaneously conducted "multiplex" reviews also cannot provide as nuanced a view of data quality, populations studied, or other considerations related to study design as when the entire body of available literature on a single association is examined. Such considerations include gene–gene

and gene-environment interactions. In addition, the adequacy of standard statistical methods for addressing heterogeneity among studies becomes a concern if meta-analyses are done over a small number of studies, which is more likely in the method described here.[99]

Some precision in the risk estimates themselves may also be sacrificed. On the other hand, Lin and Zeng have argued that meta-analysis of summary results is statistically as efficient as analysis of individual participant data.[10] The true significance of this limitation is further lessened given that the increases in risk noted from individual markers are generally small (ORs < 1.5). It is certainly reassuring that our results did not appear to differ significantly from those of more traditional meta-analyses, except where clear and knowable issues related to the limitations mentioned above were present.

The inherent limits of indexing in PubMed may also be an issue, as it was for some journals and GWAS not indexed by gene symbol. A related issue is the "sensitivity" of the two-step "multiplex" search method. Although there was no apparent systematic bias in the kinds of studies that were captured, from 13% (three of 24, for *ACE* I/D) to 86% (30 of 35, for rs7903146 of *TCF7L2*) of the articles cited in traditional meta-analyses (where they existed) were included in our abbreviated meta-analyses. In the case of *ACE* I/D, this discrepancy related to data quality and the absence of sufficient quantitative data from older studies excluded from our analysis.

Concerning the specific application of this method reported here, ie, to evaluate common DNA variants associated with type 2 diabetes risk, the use of an older literature dataset (including articles published from 1996–2008) is clearly another limitation. Our analysis therefore may not be fully reflective of the types of studies published since then. However, the period covered is historically and methodologically important for spanning the three major waves of genetic association discovery described earlier, particularly as methods for locating rarer variants and causal variants remain in development.[11,99]

Improvements to our method could include inserting a step, prior to the start of coding, that involves simply counting (and ranking) the number of times specific polymorphisms within these genes have been studied, to help prioritize data-abstraction efforts, as well as automating steps where possible using machine learning and natural language-processing approaches. We could also consider requiring the inclusion of a higher minimum number of studies with usable data for a given association, prior to performing a meta-analysis. At the

same time, particularly with respect to GWAS, study number should probably be balanced against study size.

Future directions include additional testing in other common chronic diseases for which a large number of SNPs have been reported to confer increased risk. One example with CER implications was an analysis of the predictive power of a panel of 101 SNPs and a diagnosis of cardiovascular disease.[100] In this study, the selection criteria for markers included in the panel were boiled down to one: a single published finding of a significant association with either a diagnosis of CVD or a surrogate outcome (such as cholesterol), with a P value $<10^{-5}$. Some critics pointed to the probability that variants which had not been adequately validated were included in the panel, with predictable consequences in terms of disease-risk prediction.[101] Yet, a PubMed search found only one published meta-analysis of genetic polymorphisms associated with increased cardiovascular-disease risk, that of SNPs located in the oft-studied 9p21 locus.[102] The new method presented here could be applied to more systematically examine, for instance, which of the included SNPs had been studied – and validated – on multiple occasions. Though still labor intensive, the time required would be less than that required to conduct traditional meta-analyses of all 101 SNPs.

Conclusions

In summary, the methodology described in this paper represents a reasonable approach to more rapidly identifying and evaluating frequently studied genetic risk markers for diseases such as type 2 diabetes. Comparison with results of traditional meta-analyses suggests that these gains in efficiency do not necessarily come at the price of reduced accuracy. Given the quickening pace of discovery of such markers, more efficient, unbiased, and readily updatable methods for systematically assessing and reassessing a changing literature could prove valuable, having the potential to increase the capacity to determine whether published associations are, in fact, valid. Such a tool can also contribute to CER around the question of genome-based versus non-genome-based care.

Acknowledgments

assisted with data abstraction, and helped draft the manuscript. NW assisted with data abstraction and data quality, and participated in data analysis. MW led the statistical analysis, and performed the final meta-analyses. GG participated in study design, advised its conduct, and helped draft the manuscript. All authors read and approved the final manuscript.

Disclosure

There are no conflicts of interest to report.

References

1. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409(6822):860–921.
2. Sox HC, Greenfield S. Comparative effectiveness research: a report from the Institute of Medicine. *Ann Intern Med*. 2009;151(3):203–205.
3. Khoury MJ, Rich EC, Randhawa G, Teutsch SM, Niederhuber J. Comparative effectiveness research and genomic medicine: an evolving partnership for 21st century medicine. *Genet Med*. 2009;11(10):707–711.
4. Kraft P, Zeggini E, Ioannidis JP. Replication in genome-wide association studies. *Stat Sci*. 2009;24(4):561–573.
5. Florez JC, Jablonski KA, Bayley N, et al. TCF7L2 polymorphisms and progression to diabetes in the Diabetes Prevention Program. *N Engl J Med*. 2006;355(3):241–250.
6. McCarthy MI. Genomics, type 2 diabetes, and obesity. *N Engl J Med*. 2010;363(24):2339–2350.
7. Stolerman ES, Florez JC. Genomics of type 2 diabetes mellitus: implications for the clinician. *Nat Rev Endocrinol*. 2009;5(8):429–436.
8. Weedon MN, Schwarz PE, Horikawa Y, et al. Meta-analysis and a large association study confirm a role for calpain-10 variation in type 2 diabetes susceptibility. *Am J Hum Genet*. 2003;73(5):1208–1212.
9. Kraft P. Curses – winner's and otherwise – in genetic epidemiology. *Epidemiology*. 2008;19(5):649–651.
10. Lin DY, Zeng D. Meta-analysis of genome-wide association studies: no efficiency gain in using individual participant data. *Genet Epidemiol*. 2010;34(1):60–66.
11. Wang K, Dickson SP, Stolle CA, Krantz ID, Goldstein DB, Hakonarson H. Interpretation of association signals and identification of causal variants from genome-wide association studies. *Am J Hum Genet*. 2010;86(5):730–742. Epub April 29, 2010.
12. Singh PP, Naz I, Gilmour A, Singh M, Mastana S. Association of APOE (Hha1) and ACE (I/D) gene polymorphisms with type 2 diabetes mellitus in North West India. *Diabetes Res Clin Pract*. 2006;74(1):95–102.
13. Yang M, Qiu CC, Xu Q, Xiang HD. Association of angiotensin converting enzyme gene I/D polymorphism with type 2 diabetes mellitus. *Biomed Environ Sci*. 2006;19(4):323–327.
14. Bengtsson K, Orho-Melander M, Lindblad U, et al. Polymorphism in the angiotensin converting enzyme but not in the angiotensinogen gene is associated with hypertension and type 2 diabetes: the Skaraborg Hypertension and diabetes project. *J Hypertens*. 1999;17(11):1569–1575.
15. Einarsdottir E, Mayans S, Ruikka K, et al. Linkage but not association of calpain-10 to type 2 diabetes replicated in northern Sweden. *Diabetes*. 2006;55(6):1879–1883.
16. Chen Y, Kittles R, Zhou J, et al. Calpain-10 gene polymorphisms and type 2 diabetes in West Africans: the Africa America Diabetes Mellitus (AADM) Study. *Ann Epidemiol*. 2005;15(2):153–159.
17. Song Y, Niu T, Manson JE, Kwiatkowski DJ, Liu S. Are variants in the CAPN10 gene related to risk of type 2 diabetes? A quantitative assessment of population and family-based association studies. *Am J Hum Genet*. Feb 2004;74(2):208–222.
18. Vaxillaire M, Dina C, Lobbens S, et al. Effect of common polymorphisms in the HNF4alpha promoter on susceptibility to type 2 diabetes in the French Caucasian population. *Diabetologia*. 2005;48(3):440–444.
19. Takeuchi F, Yanai K, Inomata H, et al. Search of type 2 diabetes susceptibility gene on chromosome 20q. *Biochem Biophys Res Commun*. 2007;357(4):1100–1106.
20. Love-Gregory LD, Wasson J, Ma J, et al. A common polymorphism in the upstream promoter region of the hepatocyte nuclear factor-4 alpha 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*. 2004;53(5):1134–1140.
21. Hansen SK, Rose CS, Glümer C, et al. Variation near the hepatocyte nuclear factor (HNF)-4alpha gene associates with type 2 diabetes in the Danish population. *Diabetologia*. 2005;48(3):452–458.
22. Silander K, Mohlke KL, Scott LJ, et al. Genetic variation near the hepatocyte nuclear factor-4 alpha gene predicts susceptibility to type 2 diabetes. *Diabetes*. 2004;53(4):1141–1149.
23. Winckler W, Graham RR, de Bakker PI, et al. Association testing of variants in the hepatocyte nuclear factor 4alpha gene with risk of type 2 diabetes in 7,883 people. *Diabetes*. 2005;54(3):886–892.
24. Bonnycastle LL, Willer CJ, Conneely KN, et al. Common variants in maturity-onset diabetes of the young genes contribute to risk of type 2 diabetes in Finns. *Diabetes*. Sep 2006;55(9):2534–2540.
25. Johansson S, Raeder H, Eide SA, et al. Studies in 3,523 Norwegians and meta-analysis in 11,571 subjects indicate that variants in the hepatocyte nuclear factor 4 alpha (HNF4A) P2 region are associated with type 2 diabetes in Scandinavians. *Diabetes*. 2007;56(12):3112–3117.
26. Illig T, Bongardt F, Schöpfer-Wendels A, et al; KORA Study Group. Genetics of type 2 diabetes: impact of interleukin-6 gene variants. *Gesundheitswesen*. 2005;67 Suppl 1:S122–S126.
27. Illig T, Bongardt F, Schöpfer A, et al. Significant association of the interleukin-6 gene polymorphisms C-174G and A-598G with type 2 diabetes. *J Clin Endocrinol Metab*. 2004;89(10):5053–5058.
28. Vozarova B, Fernandez-Real JM, Knowler WC, et al. The interleukin-6 (-174) G/C promoter polymorphism is associated with type-2 diabetes mellitus in Native Americans and Caucasians. *Hum Genet*. 2003;112(4):409–413.
29. Yokoi N, Kanamori M, Horikawa Y, et al. Association studies of variants in the genes involved in pancreatic beta-cell function in type 2 diabetes in Japanese subjects. *Diabetes*. 2006;55(8):2379–2386.
30. Gloyn AL, Weedon MN, Owen KR, et al. Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. *Diabetes*. 2003;52(2):568–572.
31. Hansen SK, Nielsen EM, Ek J, et al. Analysis of separate and combined effects of common variation in KCNJ11 and PPARG on risk of type 2 diabetes. *J Clin Endocrinol Metab*. 2005;90(6):3629–3637.
32. Horikoshi M, Hara K, Ito C, et al. Variations in the HHEX gene are associated with increased risk of type 2 diabetes in the Japanese population. *Diabetologia*. 2007;50(12):2461–2466.
33. Scott LJ, Mohlke KL, Bonnycastle LL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*. June 1, 2007;316(5829):1341–1345.
34. Weedon MN, McCarthy MI, Hitman G, et al. Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. *PLoS Med*. 2006;3(10):e374.
35. Zeggini E, Weedon MN, Lindgren CM, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science*. 2007;316(5829):1336–1341.
36. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research, Saxena R, Voight BF, Lyssenko V, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*. 2007;316(5829):1331–1336.
37. Willer CJ, Bonnycastle LL, Conneely KN, et al. Screening of 134 single nucleotide polymorphisms (SNPs) previously associated with type 2 diabetes replicates association with 12 SNPs in nine genes. *Diabetes*. 2007;56(1):256–264.

38. Omori S, Tanaka Y, Takahashi A, et al. Association of CDKAL1, IGF2BP2, CDKN2A/B, HHEX, SLC30A8, and KCNJ11 with susceptibility to type 2 diabetes in a Japanese population. *Diabetes*. 2008;57(3):791–795.

39. Sakamoto Y, Inoue H, Keshavarz P, et al. SNPs in the KCNJ11-ABCC8 gene locus are associated with type 2 diabetes and blood pressure levels in the Japanese population. *J Hum Genet*. 2007; 52(10):781–793.

40. Koo BK, Cho YM, Park BL, et al. Polymorphisms of KCNJ11 (Kir6.2 gene) are associated with Type 2 diabetes and hypertension in the Korean population. *Diabet Med*. 2007;24(2):178–186.

41. Doi Y, Kubo M, Ninomiya T, et al. Impact of Kir6.2 E23K polymorphism on the development of type 2 diabetes in a general Japanese population: The Hisayama Study. *Diabetes*. 2007;56(11):2829–2833.

42. Nelson TL, Fingerlin TE, Moss L, Barmada MM, Ferrell RE, Norris JM. The peroxisome proliferator-activated receptor gamma coactivator-1 alpha gene (PGC-1alpha) is not associated with type 2 diabetes mellitus or body mass index among Hispanic and non Hispanic Whites from Colorado. *Exp Clin Endocrinol Diabetes*. 2007;115(4):268–275.

43. Lai CQ, Tucker KL, Parnell LD, et al. PPARGC1 A variation associated with DNA damage, diabetes, and cardiovascular diseases: the Boston Puerto Rican Health Study. *Diabetes*. 2008;57(4):809–816.

44. Sun L, Yang Z, Jin F, et al. The Gly482Ser variant of the PPARGC1 gene is associated with Type 2 diabetes mellitus in northern Chinese, especially men. *Diabet Med*. 2006;23(10):1085–1092.

45. Ek J, Andersen G, Urhammer SA, et al. Mutation analysis of peroxisome proliferator-gamma coactivator-1 (PGC-1) and relationships of identified amino acid polymorphisms to Type II diabetes mellitus. *Diabetologia*. 2001;44(12):2220–2226.

46. Bhat A, Koul A, Rai E, Sharma S, Dhar MK, Bamezai RN. PGC-1-alpha Thr394Thr and Gly482Ser variants are significantly associated with T2DM in two North Indian populations: a replicate case-control study. *Hum Genet*. 2007;121(5):609–614.

47. Mori H, Ikegami H, Kawaguchi Y, et al. The Pro12 → Ala substitution in PPAR-gamma is associated with resistance to development of diabetes in the general population: possible involvement in impairment of insulin secretion in individuals with type 2 diabetes. *Diabetes*. 2001;50(4):891–894.

48. Mancini FP, Vaccaro O, Sabatino L, et al. Pro12Ala substitution in the peroxisome proliferator-activated receptor-gamma2 is not associated with type 2 diabetes. *Diabetes*. 1999;48(7):1466–1468.

49. Altshuler D, Hirschhorn JN, Klannemark M, et al. The common PPAR-gamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet*. 2000;26(1):76–80.

50. Ghoussaini M, Meyre D, Lobbens S, et al. Implication of the Pro12Ala polymorphism of the PPAR-gamma 2 gene in type 2 diabetes and obesity in the French population. *BMC Med Genet*. 22 2005;6:11.

51. Hara K, Okada T, Tobe K, et al. The Pro12Ala polymorphism in PPAR gamma2 may confer resistance to type 2 diabetes. *Biochem Biophys Res Commun*. 2000;271(1):212–216.

52. Moon MK, Cho YM, Jung HS, et al. Genetic polymorphisms in peroxisome proliferator-activated receptor gamma are associated with Type 2 diabetes mellitus and obesity in the Korean population. *Diabet Med*. 2005;22(9):1161–1166.

53. Meshkani R, Taghikhani M, Larijani B, et al. Pro12Ala polymorphism of the peroxisome proliferator-activated receptor-gamma2 (PPARgamma-2) gene is associated with greater insulin sensitivity and decreased risk of type 2 diabetes in an Iranian population. *Clin Chem Lab Med*. 2007;45(4):477–482.

54. Soriguer F, Morcillo S, Cardona F, et al. Pro12Ala polymorphism of the PPARG2 gene is associated with type 2 diabetes mellitus and peripheral insulin sensitivity in a population with a high intake of oleic acid. *J Nutr*. 2006;136(9):2325–2330.

55. Elbein SC, Chu WS, Das SK, et al. Transcription factor 7-like 2 polymorphisms and type 2 diabetes, glucose homeostasis traits and gene expression in US participants of European and African descent. *Diabetologia*. 2007;50(8):1621–1630.

56. Rees SD, Bellary S, Britten AC, et al. Common variants of the TCF7L2 gene are associated with increased risk of type 2 diabetes mellitus in a UK-resident South Asian population. *BMC Med Genet*. 2008;9:8.

57. Saadi H, Nagelkerke N, Carruthers SG, et al. Association of TCF7L2 polymorphism with diabetes mellitus, metabolic syndrome, and markers of beta cell function and insulin resistance in a population-based sample of Emirati subjects. *Diabetes Res Clin Pract*. 2008;80(3):392–398.

58. Chang YC, Chang TJ, Jiang YD, et al. Association study of the genetic polymorphisms of the transcription factor 7-like 2 (TCF7L2) gene and type 2 diabetes in the Chinese population. *Diabetes*. 2007;56(10):2631–2637.

59. Cauchi S, Meyre D, Dina C, et al. Transcription factor TCF7L2 genetic study in the French population: expression in human beta-cells and adipose tissue and strong association with type 2 diabetes. *Diabetes*. 2006;55(10):2903–2908.

60. Salonen JT, Uimari P, Aalto JM, et al. Type 2 diabetes whole-genome association study in four populations: the DiaGen consortium. *Am J Hum Genet*. Aug 2007;81(2):338–345.

61. Miyake K, Horikawa Y, Hara K, et al. Association of TCF7L2 polymorphisms with susceptibility to type 2 diabetes in 4,087 Japanese subjects. *J Hum Genet*. 2008;53(2):174–180.

62. Ren Q, Han XY, Wang F, et al. Exon sequencing and association analysis of polymorphisms in TCF7L2 with type 2 diabetes in a Chinese population. *Diabetologia*. 2008;51(7):1146–1152.

63. Guo T, Hanson RL, Traurig M, et al. TCF7L2 is not a major susceptibility gene for type 2 diabetes in Pima Indians: analysis of 3,501 individuals. *Diabetes*. 2007;56(12):3082–3088.

64. Marzi C, Huth C, Kolz M, et al. Variants of the transcription factor 7-like 2 gene (TCF7L2) are strongly associated with type 2 diabetes but not with the metabolic syndrome in the MONICA/KORA surveys. *Horm Metab Res*. 2007;39(1):46–52.

65. Parra EJ, Cameron E, Simmonds L, et al. Association of TCF7L2 polymorphisms with type 2 diabetes in Mexico City. *Clin Genet*. 2007; 71(4):359–366.

66. Bodhini D, Radha V, Dhar M, Narayani N, Mohan V. The rs12255372(G/T) and rs7903146(C/T) polymorphisms of the TCF7L2 gene are associated with type 2 diabetes mellitus in Asian Indians. *Metabolism*. 2007;56(9):1174–1178.

67. Groves CJ, Zeggini E, Minton J, et al. Association analysis of 6,736 UK subjects provides replication and confirms TCF7L2 as a type 2 diabetes susceptibility gene with a substantial effect on individual risk. *Diabetes*. 2006;55(9):2640–2644.

68. Kimber CH, Doney AS, Pearson ER, et al. TCF7L2 in the Go-DARTS study: evidence for a gene dose effect on both diabetes susceptibility and control of glucose levels. *Diabetologia*. 2007;50(6):1186–1191.

69. van Vliet-Ostaptchouk JV, Shiri-Sverdlov R, Zhernakova A, et al. Association of variants of transcription factor 7-like 2 (TCF7L2) with susceptibility to type 2 diabetes in the Dutch Breda cohort. *Diabetologia*. 2007;50(1):59–62.

70. Chandak GR, Janipalli CS, Bhaskar S, et al. Common variants in the TCF7L2 gene are strongly associated with type 2 diabetes mellitus in the Indian population. *Diabetologia*. 2007;50(1):63–67.

71. Zhang C, Qi L, Hunter DJ, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene and the risk of type 2 diabetes in large cohorts of US women and men. *Diabetes*. 2006;55(9):2645–2648.

72. Mayans S, Lackovic K, Lindgren P, et al. TCF7L2 polymorphisms are associated with type 2 diabetes in northern Sweden. *Eur J Hum Genet*. 2007;15(3):342–346.

73. Hayashi T, Iwamoto Y, Kaku K, Hirose H, Maeda S. Replication study for the association of TCF7L2 with susceptibility to type 2 diabetes in a Japanese population. *Diabetologia*. 2007;50(5):980–984.

74. Dahlgren A, Zethelius B, Jensevik K, Syvanen AC, Berne C. ULSAM Cohort. Variants of the TCF7L2 gene are associated with beta cell dysfunction and confer an increased risk of type 2 diabetes mellitus in the ULSAM cohort of Swedish elderly men. *Diabetologia*. 2007;50(9): 1852–1857.

75. Wang J, Kuusisto J, Vanttinen M, et al. Variants of transcription factor 7-like 2 (TCF7L2) gene predict conversion to type 2 diabetes in the Finnish Diabetes Prevention Study and are associated with impaired glucose regulation and impaired insulin secretion. *Diabetologia*. 2007;50(6):1192–1200.

76. Scott LJ, Bonnycastle LL, Willer CJ, et al. Association of transcription factor 7-like 2 (TCF7L2) variants with type 2 diabetes in a Finnish sample. *Diabetes*. 2006;55(9):2649–2653.

77. Saxena R, Gianniny L, Burtt NP, et al. Common single nucleotide polymorphisms in TCF7L2 are reproducibly associated with type 2 diabetes and reduce the insulin response to glucose in nondiabetic individuals. *Diabetes*. 2006;55(10):2890–2895.

78. Cauchi S, El Achhab Y, Choquet H, et al. TCF7L2 is reproducibly associated with type 2 diabetes in various ethnic groups: a global meta-analysis. *J Mol Med*. 2007;85(7):777–782.

79. Ng MC, Tam CH, Lam VK, So WY, Ma RC, Chan JC. Replication and identification of novel variants at TCF7L2 associated with type 2 diabetes in Hong Kong Chinese. *J Clin Endocrinol Metab*. 2007; 92(9):3733–3737.

80. Lewis JP, Palmer ND, Hicks PJ, et al. Association analysis in african americans of European-derived type 2 diabetes single nucleotide polymorphisms from whole-genome association studies. *Diabetes*. 2008; 57(8):2220–2225.

81. Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*. 22 2007; 445(7130):881–885.

82. Lyssenko V, Almgren P, Anevski D, et al. Genetic prediction of future type 2 diabetes. *PLoS Med*. 2005;2(12):e345.

83. Florez JC, Jablonski KA, Kahn SE, et al. Type 2 diabetes-associated missense polymorphisms KCNJ11 E23K and ABCC8 A1369S influence progression to diabetes and response to interventions in the Diabetes Prevention Program. *Diabetes*. 2007;56(2):531–536.

84. Vaxillaire M, Veslot J, Dina C, et al; DESIR Study Group. Impact of common type 2 diabetes risk polymorphisms in the DESIR prospective study. *Diabetes*. 2008;57(1):244–254.

85. Jensen DP, Urhammer SA, Eiberg H, et al. Variation in CAPN10 in relation to type 2 diabetes, obesity and quantitative metabolic traits: studies in 6018 whites. *Mol Genet Metab*. 2006;89(4):360-367.

86. Laukkanen O, Pihlajamaki J, Lindstrom J, et al; Finnish Diabetes Prevention Study Group. Polymorphisms of the SUR1 (ABCC8) and Kir6.2 (KCNJ11) genes predict the conversion from impaired glucose tolerance to type 2 diabetes. The Finnish Diabetes Prevention Study. *J Clin Endocrinol Metab*. 2004;89(12):6286–6290.

87. van Dam RM, Hoebee B, Seidell JC, Schaap MM, de Bruin TW, Feskens EJ. Common variants in the ATP-sensitive K+ channel genes KCNJ11 (Kir6.2) and ABCC8 (SUR1) in relation to glucose intolerance: population-based studies and meta-analyses. *Diabet Med*. 2005;22(5):590–598.

88. Ludovico O, Pellegrini F, Di Paola R, et al. Heterogeneous effect of peroxisome proliferator-activated receptor gamma2 Ala12 variant on type 2 diabetes risk. *Obesity (Silver Spring)*. 2007;15(5):1076–1081.

89. Gouda HN, Sagoo GS, Harding AH, Yates J, Sandhu MS, Higgins JP. The association between the peroxisome proliferator-activated receptor-gamma2 (PPARG2) Pro12Ala gene variant and type 2 diabetes mellitus: a HuGE review and meta-analysis. *Am J Epidemiol*. 2010;171(6):645–655.

90. Tong Y, Lin Y, Zhang Y, Yang J, Liu H, Zhang B. Association between TCF7L2 gene polymorphisms and susceptibility to type 2 diabetes mellitus: a large Human Genome Epidemiology (HuGE) review and meta-analysis. *BMC Med Genet*. 2009;10:15.

91. Sookoian S, Gemma C, Pirola CJ. Influence of hepatocyte nuclear factor 4alpha (HNF4alpha) gene variants on the risk of type 2 diabetes: a meta-analysis in 49,577 individuals. *Mol Genet Metab*. 2010; 99(1):80–89.

92. Zhou JB, Yang JK, Lu JK, An YH. Angiotensin-converting enzyme gene polymorphism is associated with type 2 diabetes: a meta-analysis. *Mol Biol Rep*. 2010;37(1):67–73.

93. Ioannidis JP, Thomas G, Daly MJ. Validating, augmenting and refining genome-wide association signals. *Nat Rev Genet*. 2009;10(5): 318–329.

94. Qi L, Cornelis MC, Kraft P, et al. Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. *Hum Mol Genet*. 2010; 19(13):2706–2715.

95. Nielsen T, Sparso T, Grarup N, et al. Type 2 diabetes risk allele near CENTD2 is associated with decreased glucose-stimulated insulin release. *Diabetologia*. 2011;54(5):1052–1056.

96. Zeggini E, Ioannidis JP. Meta-analysis in genome-wide association studies. *Pharmacogenomics*. 2009;10(2):191–201.

97. Ritchie MD, Denny JC, Crawford DC, et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. *Am J Hum Genet*. 2010;86(4):560–572.

98. McCarty CA, Chisholm RL, Chute CG, et al. eMERGE Team. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med Genomics*. 2011;4:13.

99. Anderson CA, Soranzo N, Zeggini E, Barrett JC. Synthetic associations are unlikely to account for many common disease genome-wide association signals. *PLoS Biol*. 2011;9(1):e1000580.

100. Paynter NP, Chasman DI, Pare G, et al. Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA*. 2010;303(7):631–637.

101. Holm H, Thorleifsson G, Stefansson K. Genetic risk score and cardiovascular events in women. *JAMA*. 2010;303(20):2032; author reply 2032–2033.

102. Palomaki GE, Melillo S, Bradley LA. Association between 9p21 genomic markers and heart disease: a meta-analysis. *JAMA*. 2010;303(7):648–656.

Supplementary materials
Description

The rest of the Forest plots generated for variants with a sufficient number of usable studies (ie, three or greater).

Supplemental Figure 13 is the Forest plot of prospective cohort studies of *TCF7L2* rs7903146.
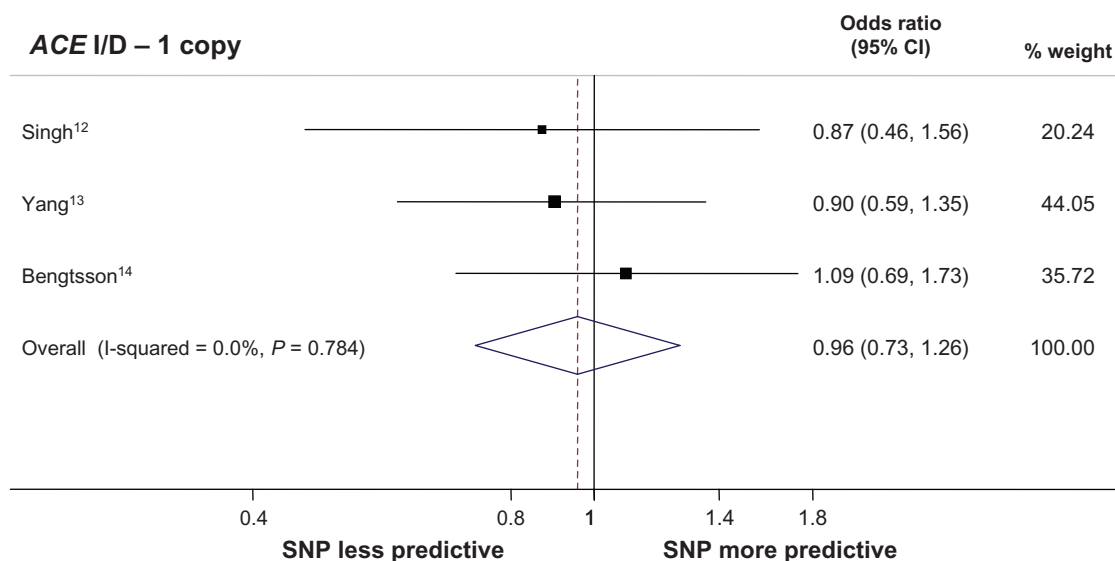
Figure S1 Forest plot for case control studies of *ACE* I/D polymorphism.
Abbreviations: I/D, insertion/deletion; CI, confidence interval; SNP, single-nucleotide polymorphism.

Figure S2 Forest plot for case control studies of *CAPN10* rs2975760.
Abbreviations: CI, confidence interval; SNP, single-nucleotide polymorphism.

CAPN10 rs3792267 – 1 copy

	Odds ratio (95% CI)	% weight
Einarsdottir[15]	1.07 (0.87, 1.30)	8.17
Jensen[85]	1.09 (1.02, 1.16)	79.65
Song[17]	1.12 (0.95, 1.32)	12.18
Overall (I-squared = 0.0%, P = 0.935)	1.09 (1.03, 1.16)	100.00

0.7 0.85 1 1.15 1.3

SNP less predictive   SNP more predictive

Figure S3 Forest plot for case control studies of *CAPN10* rs3792267.
Abbreviations: CI, confidence interval; SNP, single-nucleotide polymorphism.

HNF4 **alpha rs1884614**

	Odds ratio (95% CI)	% weight
Allelic		
Vaxillaire[18]	0.82 (0.67, 0.99)	20.23
Takeuchi[19]	1.10 (0.93, 1.31)	21.59
Love-Gregory[20]	1.45 (1.10, 1.90)	16.01
Subtotal (I-squared = 83.0%, P = 0.003)	1.08 (0.81, 1.45)	57.83
1-copy		
Hansen[21]	1.15 (1.00, 1.31)	23.60
Silander[22]	1.31 (1.04, 1.63)	18.57
Subtotal (I-squared = 0.0%, P = 0.330)	1.19 (1.06, 1.34)	42.17
Overall (I-squared = 73.8%, P = 0.004)	1.13 (0.96, 1.34)	100.00

0.6 0.8 1 1.4 1.8 2

SNP less predictive   SNP more predictive

Figure S4 Forest plot for case control studies of *HNF4* alpha rs1884614.
Abbreviations: CI, confidence interval; SNP, single-nucleotide polymorphism.

HNF4 alpha rs1885088

	Odds ratio (95% CI)	% weight
Allelic		
Love-Gregory[20]	0.85 (0.65, 1.12)	14.74
Winckler[23]	1.05 (0.92, 1.20)	24.92
Subtotal (I-squared = 46.6%, *P* = 0.171)	0.98 (0.80, 1.19)	39.66
1-copy		
Hansen[21]	0.95 (0.84, 1.09)	25.12
Bonnycastle[24]	1.26 (1.03, 1.54)	19.48
Silander[22]	1.37 (1.05, 1.75)	15.74
Subtotal (I-squared = 78.5%, *P* = 0.010)	1.16 (0.91, 1.47)	60.34
Overall (I-squared = 66.4%, *P* = 0.018)	1.07 (0.93, 1.24)	100.00

SNP less predictive   SNP more predictive

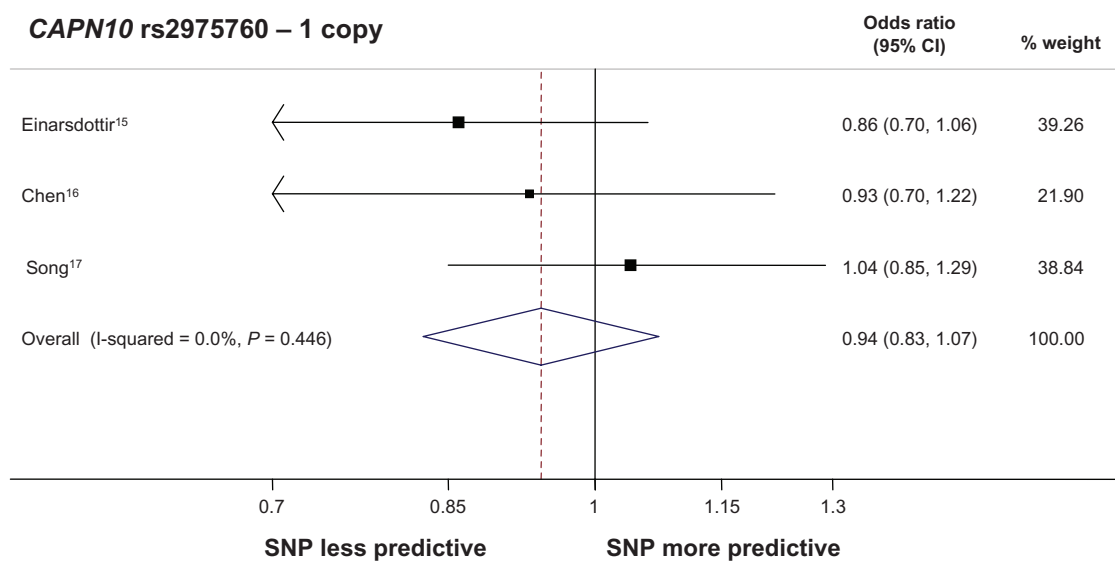**Figure S5** Forest plot for case control studies of *HNF4* alpha rs1885088.
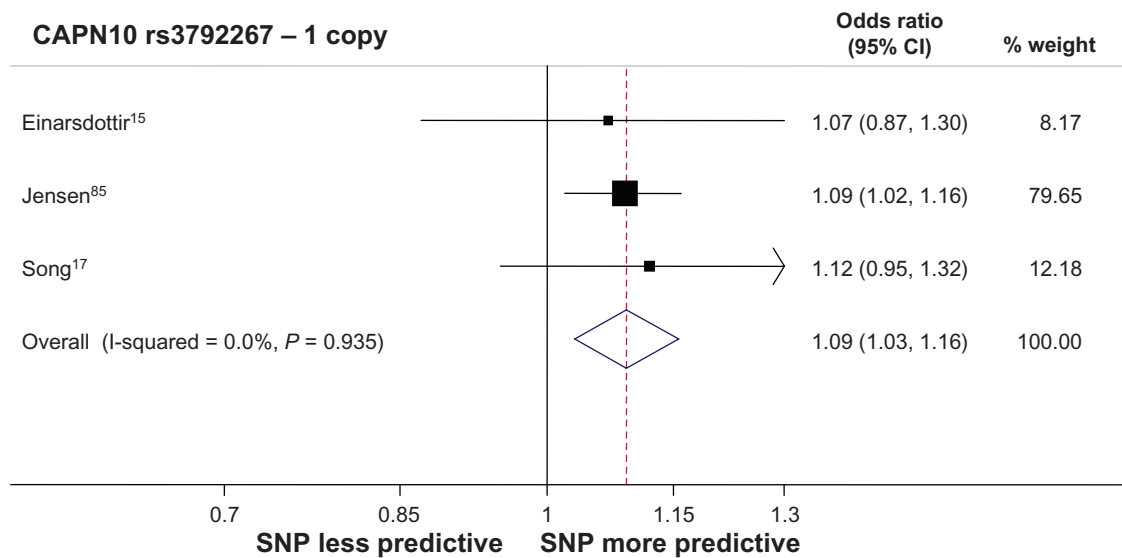**Abbreviations:** CI, confidence interval; SNP, single-nucleotide polymorphism.

HNF4 alpha rs2425637

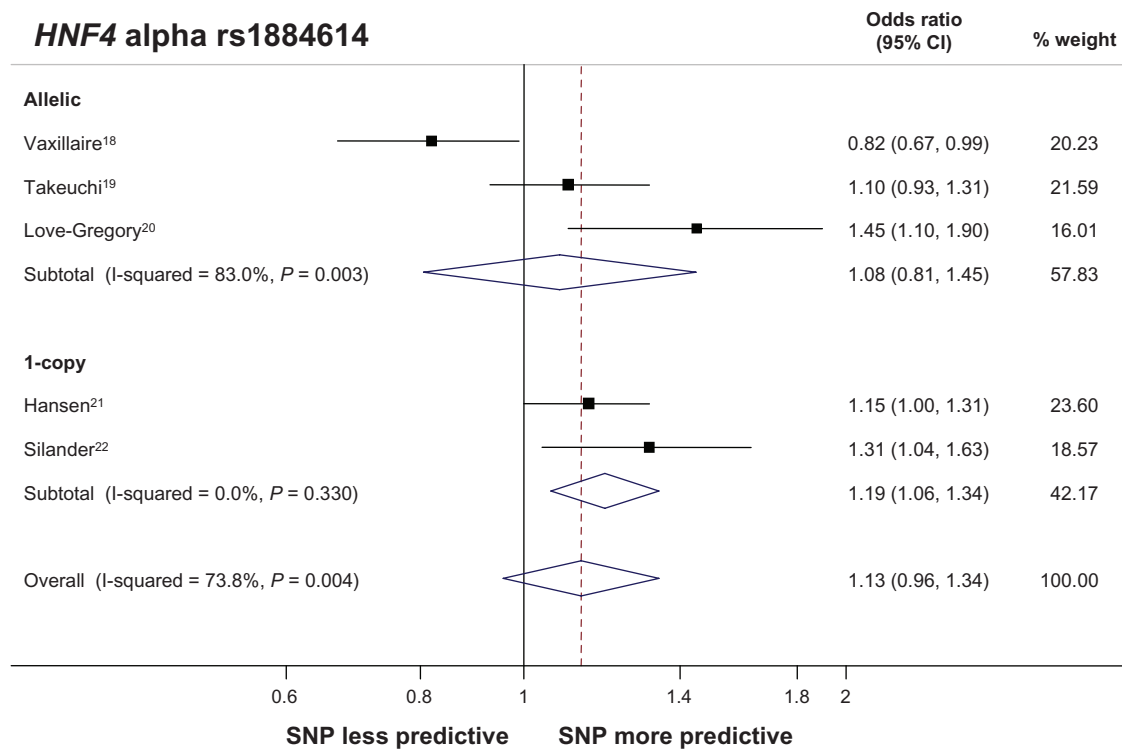|	Odds ratio (95% CI)	% weight
Allelic		
Vaxillaire[18]	1.02 (0.83, 1.26)	15.95
Love-Gregory[20]	1.14 (0.91, 1.43)	14.44
Subtotal (I-squared = 0.0%, *P* = 0.479)	1.07 (0.92, 1.25)	30.39
1-copy		
Hansen[21]	0.94 (0.82, 1.09)	23.58
Bonnycastle[24]	1.18 (1.04, 1.33)	26.36
Silander[22]	1.23 (1.04, 1.47)	19.67
Subtotal (I-squared = 73.4%, *P* = 0.023)	1.11 (0.94, 1.30)	69.61
Overall (I-squared = 50.8%, *P* = 0.087)	1.10 (0.98, 1.22)	100.00

SNP less predictive   SNP more predictive

Figure S6 Forest plot for case control studies of *HNF4* alpha rs2425637.
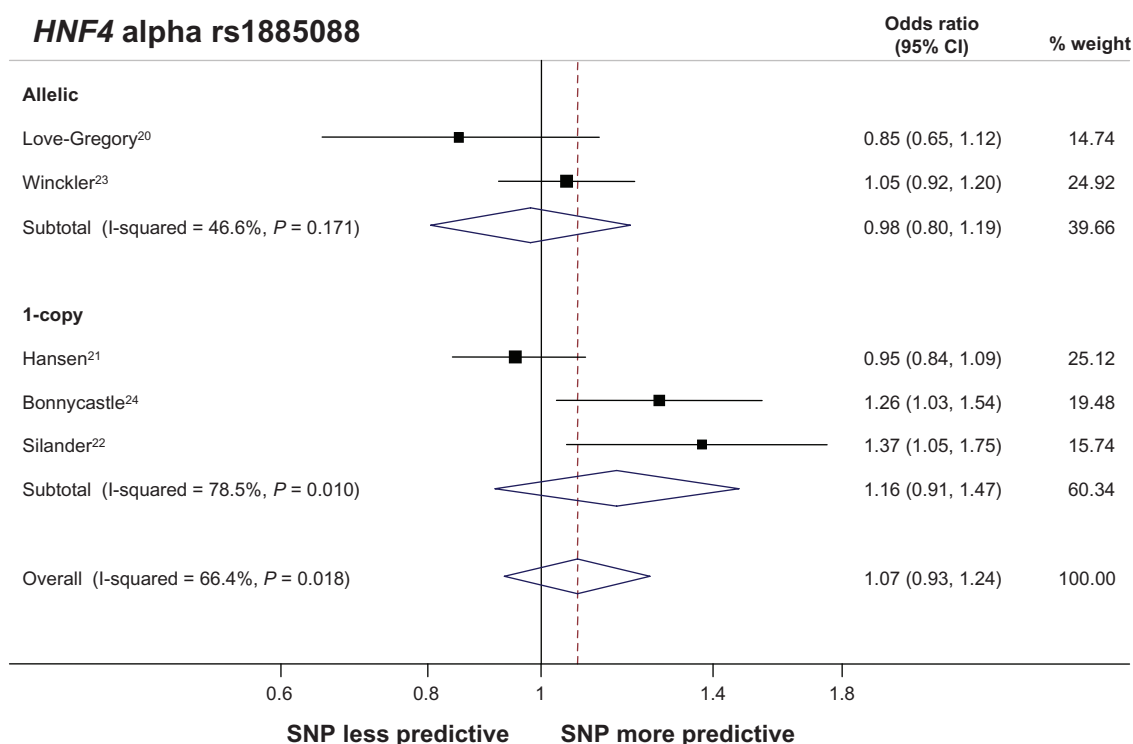**Abbreviations:** CI, confidence interval; SNP, single-nucleotide polymorphism.

HNF4 alpha rs3818247

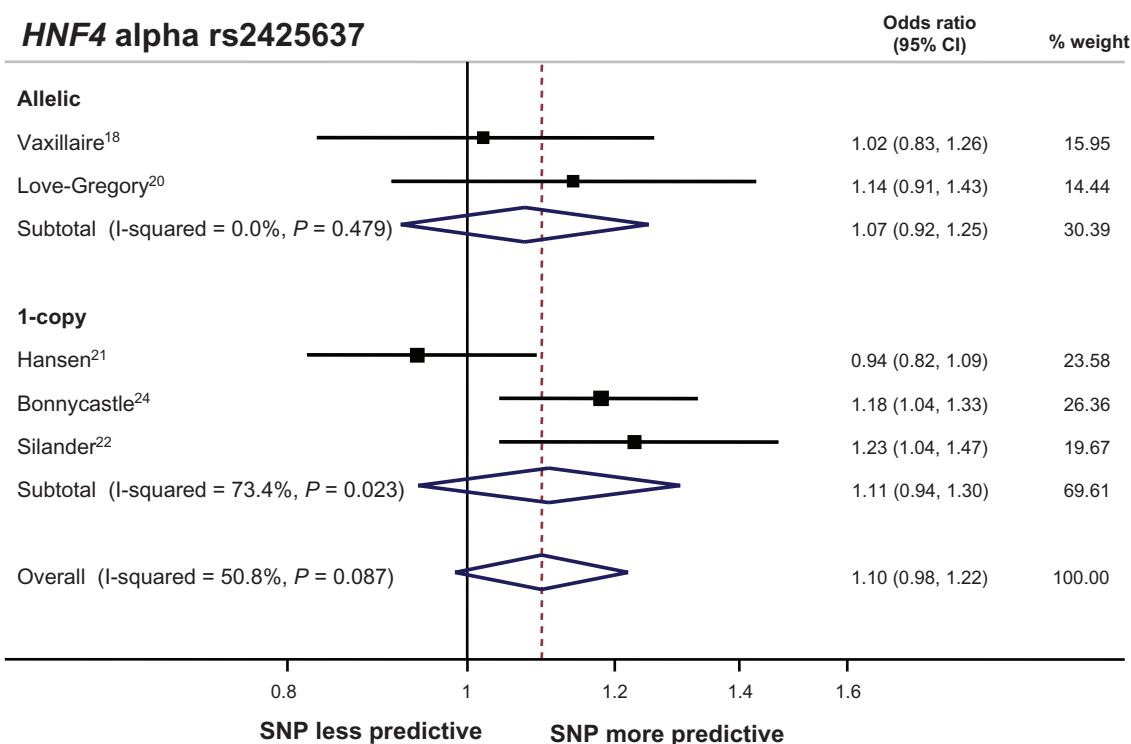|	Odds ratio (95% CI)	% weight
Allelic		
Johansson[25]	0.92 (0.82, 1.03)	20.30
Winckler[23]	0.98 (0.87, 1.10)	20.00
Takeuchi[19]	1.04 (0.88, 1.24)	15.42
Love-Gregory[20]	1.49 (1.15, 1.90)	10.33
Subtotal (I-squared = 75.1%, *P* = 0.007)	1.05 (0.90, 1.22)	66.04
1-copy		
Silander[22]	0.89 (0.75, 1.07)	14.94
Hansen[21]	0.93 (0.82, 1.06)	19.02
Subtotal (I-squared = 0.0%, *P* = 0.694)	0.92 (0.83, 1.02)	33.96
Overall (I-squared = 64.4%, *P* = 0.015)	1.00 (0.90, 1.10)	100.00

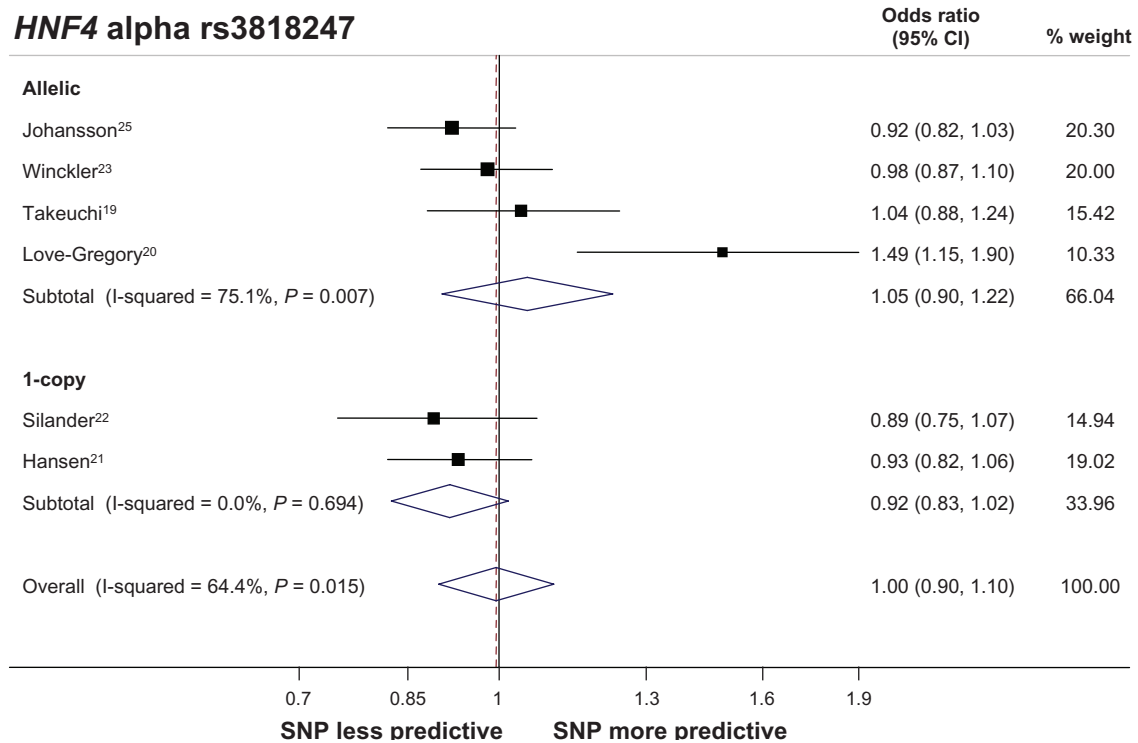SNP less predictive   SNP more predictive

Figure S7 Forest plot for case control studies of *HNF4* alpha rs3818247.
Abbreviations: CI, confidence interval; SNP, single-nucleotide polymorphism.

IL-6 rs1800795 – 1 copy

	Odds ratio (95% CI)	% weight
Illig (2005)[26]	1.20 (0.90, 1.59)	36.33
Illig (2004)[27]	1.51 (1.11, 2.07)	33.41
Vozarova (White)[28]	1.60 (1.12, 2.30)	28.75
Vozarova (Native American)[28]	9.39 (1.17, 75.20)	1.51
Overall (I-squared = 40.9%, *P* = 0.166)	1.45 (1.12, 1.88)	100.00

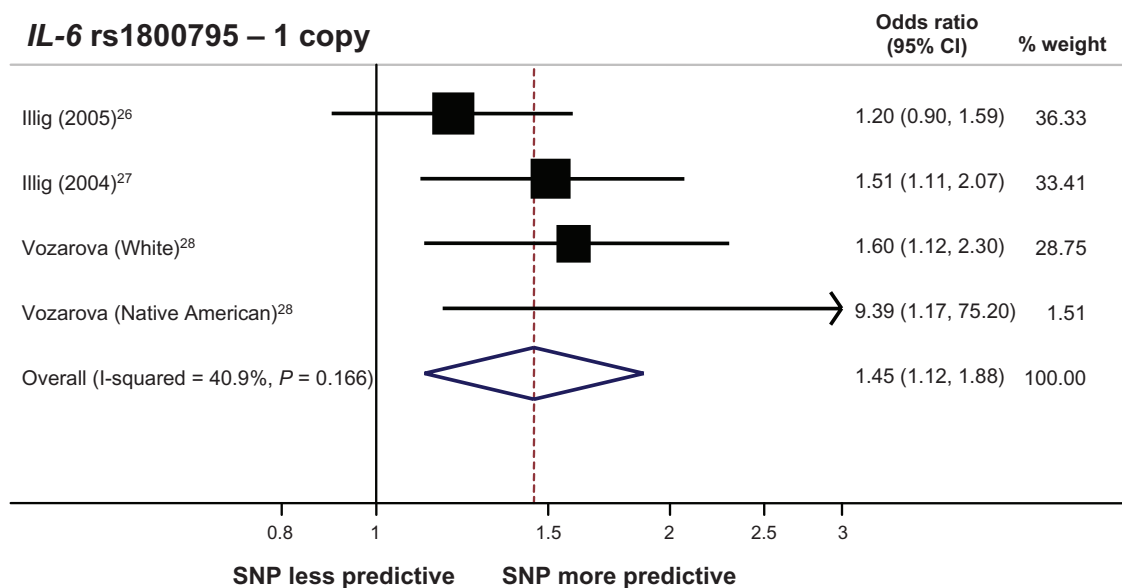SNP less predictive   SNP more predictive

Figure S8 Forest plot for case control studies of *IL-6* rs1800795.
Abbreviations: CI, confidence interval; SNP, single-nucleotide polymorphism.

KCNJ11 rs5219

	Odds ratio (95% CI)	% weight
Allelic		
Yokoi[29]	1.08 (0.97, 1.21)	7.00
Gloyn[30]	1.18 (1.04, 1.34)	5.51
Hansen[31]	1.19 (1.09, 1.30)	10.20
Subtotal (I-squared = 0.0%, P = 0.374)	1.15 (1.09, 1.23)	22.70
1-copy		
Horikoshi[32]	1.03 (0.89, 1.20)	4.08
Scott[33]	1.11 (1.02, 1.21)	10.72
Weedon[34]	1.14 (1.05, 1.23)	12.07
Zeggini[35]	1.15 (1.05, 1.25)	10.38
DGI[36]	1.15 (1.09, 1.21)	21.16
Hansen[31]	1.20 (1.04, 1.38)	4.52
Willer[37]	1.22 (1.08, 1.38)	5.84
Omori[38]	1.25 (1.08, 1.46)	3.99
Sakamoto[39]	1.32 (1.09, 1.60)	2.56
Koo[40]	1.38 (1.09, 1.75)	1.71
Doi[41]	2.10 (1.16, 3.83)	0.28
Subtotal (I-squared = 24.6%, P = 0.209)	1.16 (1.12, 1.21)	77.30
Overall (I-squared = 14.7%, P = 0.292)	1.16 (1.12, 1.20)	100.00

0.8 1 1.2 1.4 1.6

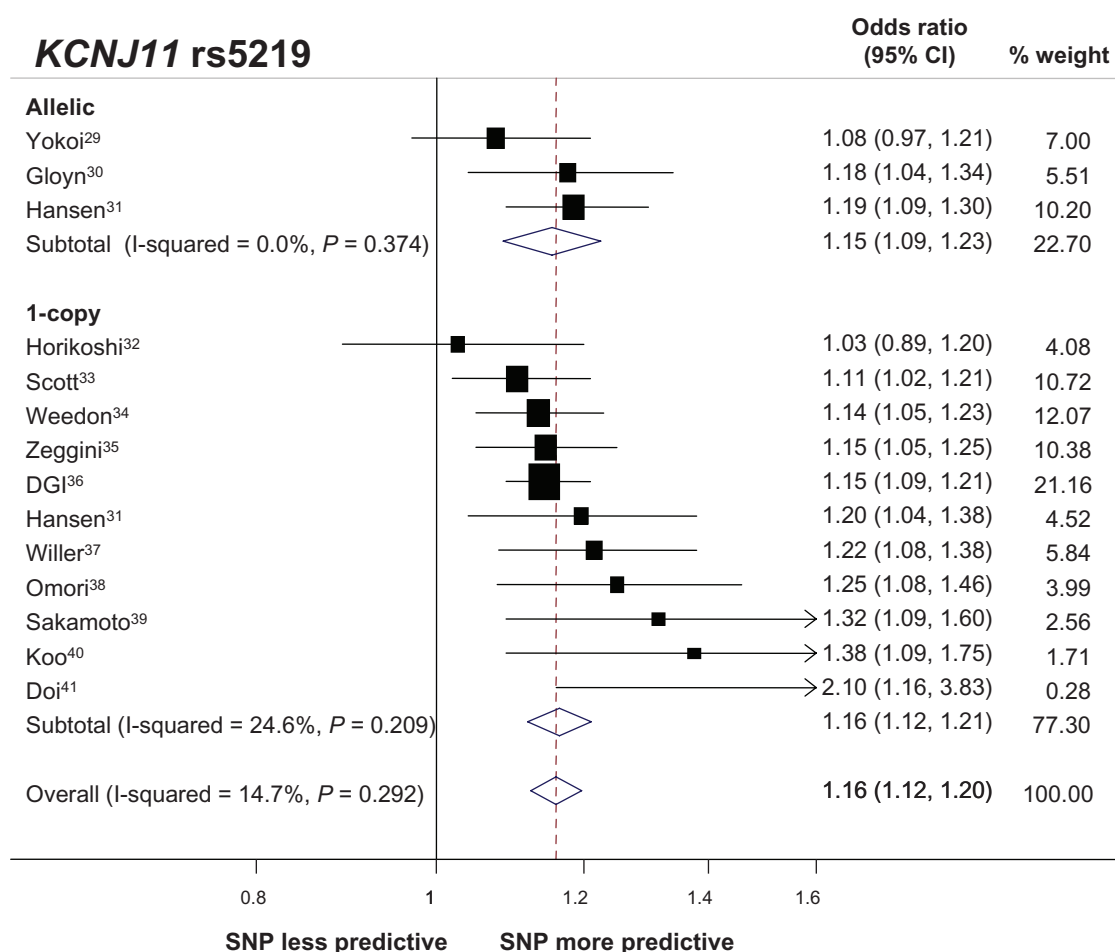**SNP less predictive   SNP more predictive**

Figure S9 Forest plot for case control studies of *KCNJ11* rs5219.
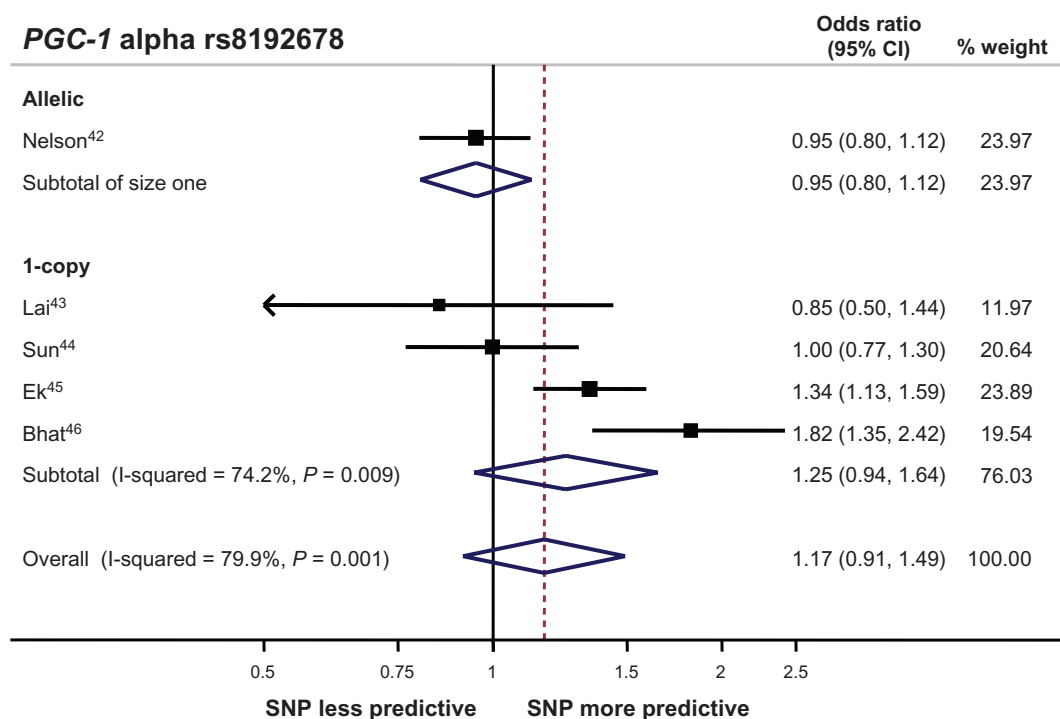**Abbreviations:** CI, confidence interval; SNP, single-nucleotide polymorphism.

PGC-1 alpha rs8192678

	Odds ratio (95% CI)	% weight
Allelic		
Nelson[42]	0.95 (0.80, 1.12)	23.97
Subtotal of size one	0.95 (0.80, 1.12)	23.97
1-copy		
Lai[43]	0.85 (0.50, 1.44)	11.97
Sun[44]	1.00 (0.77, 1.30)	20.64
Ek[45]	1.34 (1.13, 1.59)	23.89
Bhat[46]	1.82 (1.35, 2.42)	19.54
Subtotal (I-squared = 74.2%, P = 0.009)	1.25 (0.94, 1.64)	76.03
Overall (I-squared = 79.9%, P = 0.001)	1.17 (0.91, 1.49)	100.00

0.5 0.75 1 1.5 2 2.5

SNP less predictive   SNP more predictive

Figure S10 Forest plot for case control studies of *PGC-1* alpha rs8192678.
Abbreviations: CI, confidence interval; SNP, single-nucleotide polymorphism.
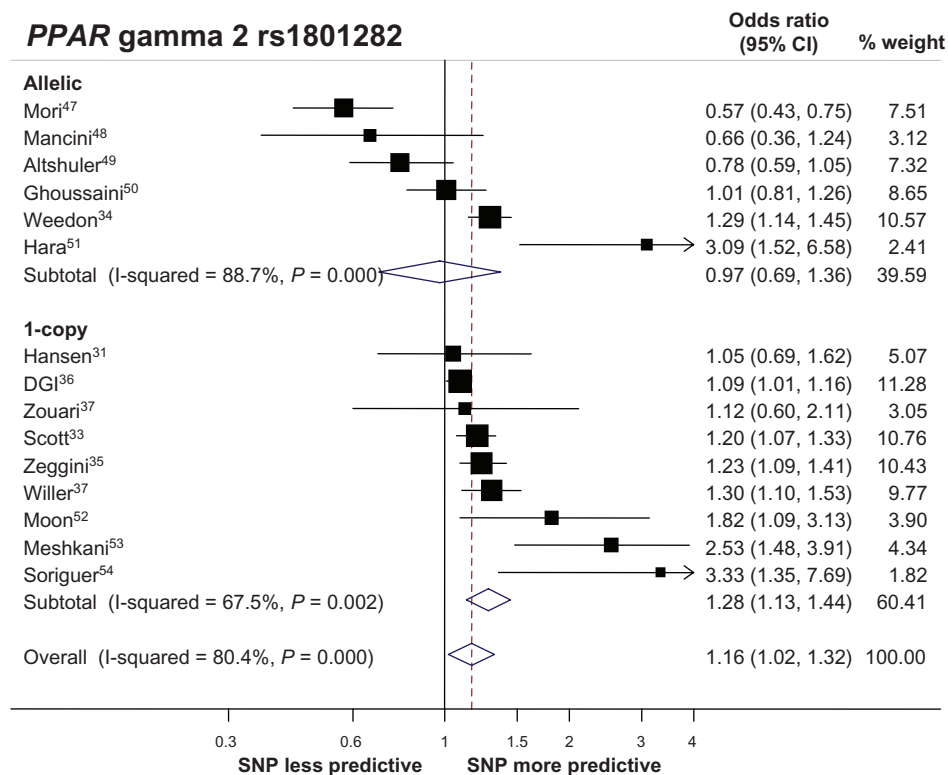
PPAR gamma 2 rs1801282

Figure S11 Forest plot for case control studies of *PPAR* gamma 2 rs1801282.
Abbreviations: CI, confidence interval; SNP, single-nucleotide polymorphism.
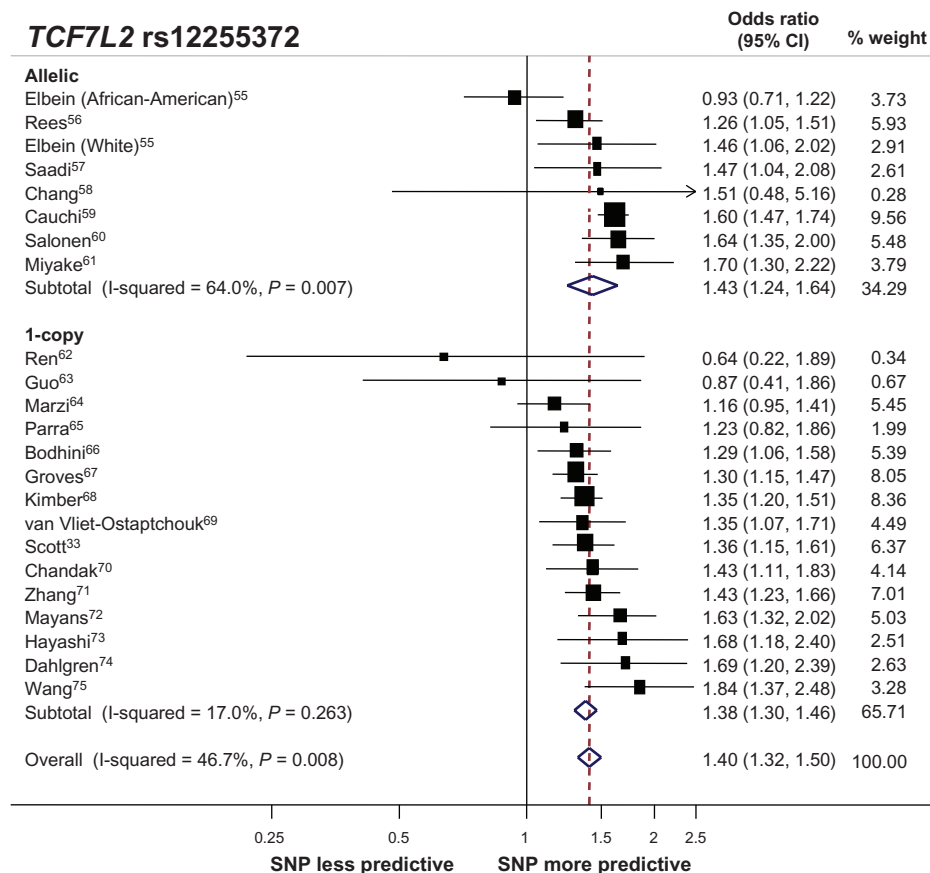
TCF7L2 rs12255372

Figure S12 Forest plot for case control studies of *TCF7L2* rs12255372.
Abbreviations: CI, confidence interval; SNP, single-nucleotide polymorphism.
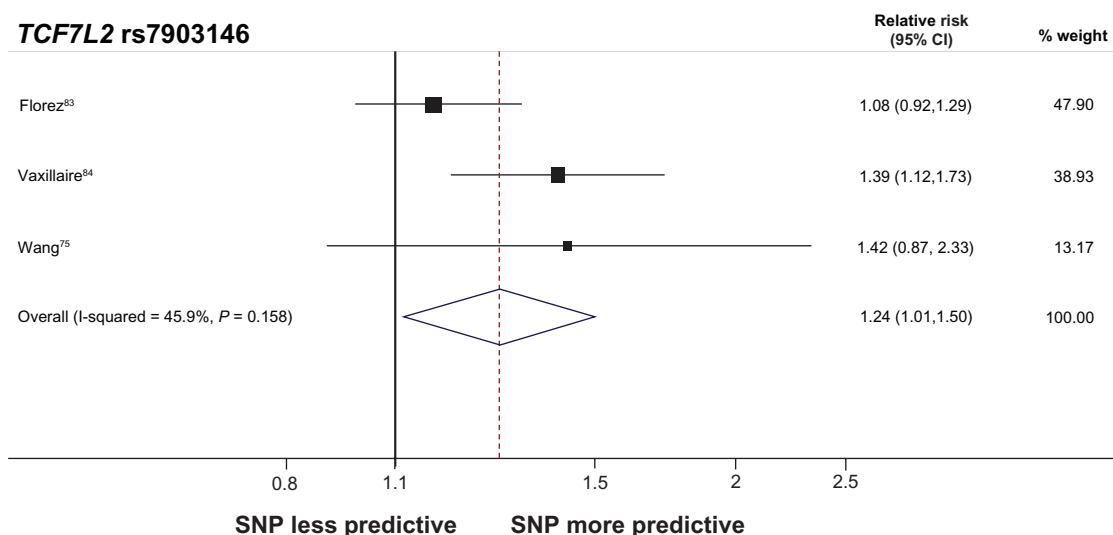
Figure S13 Forest plot for prospective cohort studies of *TCF7L2* rs7903146.
Abbreviations: CI, confidence interval; SNP, single-nucleotide polymorphism.