Novel associations for hypothyroidism include known autoimmune risk loci

Nicholas Eriksson[1,*], Joyce Y. Tung[1], Amy K. Kiefer[1], David A. Hinds[1], Uta Francke[1], Joanna L. Mountain[1], Chuong B. Do[1]
1 23andMe, Inc., Mountain View, CA, USA
∗ **E-mail: nick@23andme.com**

Abstract

Hypothyroidism is the most common thyroid disorder, affecting about 5% of the general population. Here we present the first large genome-wide association study of hypothyroidism, in 2,564 cases and 24,448 controls from the customer base of 23andMe, Inc., a personal genetics company. We identify four genome-wide significant associations, two of which are well known to be involved with a large spectrum of autoimmune diseases: rs6679677 near *PTPN22* and rs3184504 in *SH2B3* (p-values $3.5 \cdot 10^{-13}$ and $3.0 \cdot 10^{-11}$, respectively). We also report associations with rs4915077 near *VAV3* (p-value $8.3 \cdot 10^{-11}$), another gene involved in immune function, and rs965513 near *FOXE1* (p-value $3.1 \cdot 10^{-14}$). Of these, the association with *PTPN22* confirms a recent small candidate gene study, and *FOXE1* was previously known to be associated with thyroid-stimulating hormone (TSH) levels. Although *SH2B3* has been previously linked with a number of autoimmune diseases, this is the first report of its association with thyroid disease. The *VAV3* association is novel. These results suggest heterogeneity in the genetic etiology of hypothyroidism, implicating genes involved in both autoimmune disorders and thyroid function. Using a genetic risk profile score based on the top association from each of the four genome-wide significant regions in our study, the relative risk between the highest and lowest deciles of genetic risk is 2.1.

Introduction

Hypothyroidism is characterized by deficiencies of thyroid hormones T3 (triiodothyronine) and T4 (thyroxine). Thyroid hormones are primarily responsible for the regulation of metabolism, but also play a major role in development. Hypothyroidism is often marked by high thyroid-stimulating hormone (TSH) levels, which are typically indicative of impaired thyroid function; however, this is not always the case, e.g., if reduced T3/T4 levels are caused by insufficient generation of TSH by the pituitary gland rather than thyroid dysfunction. While iodine deficiency is the most common cause of hypothyroidism worldwide, most cases in the developed world are due to autoimmune hypothyroidism (e.g., Hashimoto or Ord thyroiditis). In addition, there are rare forms of congenital hypothyroidism with a number of genetic causes [1].

The genetic determinants of TSH levels are partially understood, with several established associations from genome-wide association studies (GWAS) [2–6]. Known associations include rs4704397 in *PDE8B* (phosphodiesterase 8B) [2, 5] and rs10917469 near *CAPZB* (capping protein (actin filament) muscle Z-line, beta) [4]. In a thyroid cancer GWAS, Gudmundsson et al. discovered rs965513 near *FOXE1* (forkhead box E1), also known as *TTF-2* (thyroid transcription factor 2) and rs944289 near *NKX2-1* (NK2 homeobox 1), also known as *TTF-1* (thyroid transcription factor 1) [3], and further demonstrated an association between these two SNPs and TSH levels. A second SNP near *FOXE1* (rs755109) has also been associated with TSH levels in an isolated Pacific Island population [5]. Almost all of these regions contain genes associated with thyroid function: in addition to the thyroid transcription factors, *PDE8B* is primarily expressed in the thyroid [7] and encodes a phosphodiesterase with a high affinity for cAMP, which mediates TSH effects in the thyroid [8].

Aside from TSH levels, Panicker et al. reported a strong association of rs2235544 in *DIO1* (deiodinase, iodothyronine, type I) with free T4/T3 ratio [9]. There have also been some studies of the genetics of

Hashimoto thyroiditis, the predominant cause of hypothyroidism in the United States; in particular, a candidate gene study demonstrated an association between *PTPN22* (protein tyrosine phosphatase, non-receptor type 22 (lymphoid)) and multiple autoimmune diseases, including Hashimoto thyroiditis [10]. Earlier studies have also linked the HLA (human leukocyte antigen), *CTLA4* (cytotoxic T lymphocyte antigen 4) and the 8q23-24 regions to Hashimoto thyroiditis [11–15]. Graves' disease (another autoimmune thyroid disease, characterized by hyperthyroidism) has been studied in several GWAS, with many loci discovered [16–19].

In this paper, we report on the first GWAS of hypothyroidism. We find four variants significantly associated with hypothyroidism. Two are non-synonymous variants in genes associated with many autoimmune diseases (*PTPN22* R620W and *SH2B3* (SH2B adaptor protein 3) R262W). A third is found in an intron of *VAV3* (vav 3 guanine nucleotide exchange factor), a gene plausibly involved in immune function. A fourth is near the thyroid transcription factor *FOXE1*. We also provide some support for the genes *NKX2-1*, *CAPZB*, and *PDE8B* (previously associated with TSH levels) as further risk loci for hypothyroidism.

Results

We performed a GWAS in 2,564 cases and 24,448 controls from the customer base of 23andMe, Inc., a personal genetics company. Briefly, participants responded to questions about their hypothyroidism diagnosis and related thyroid issues using web-based surveys. We classified as cases individuals who had been diagnosed with hypothyroidism, had elevated TSH levels, or were taking thyroid hormone replacement medication. Controls reported no to at least one of the above questions and yes to none of them. Participants reporting hyperthyroidism, thyroid cancer, thyroid removal, or treatment with radioactive iodine for hyperthyroidism were excluded. From the set of all qualified participants, a subset of unrelated individuals of primarily European ancestry was used in our analysis. Details about the cohort can be found in Table 1 and the Methods. All analyses were controlled for age, sex, genotyping platform, and five principal components. Manhattan and quantile-quantile plots are provided in Figures S1 and S2.

Table 2 shows the genome-wide significant SNPs along with a small set of SNPs that have previously been reported to be associated with thyroid hormone levels. The strongest association is with rs965513, with a p-value of $3.1 \cdot 10^{-14}$ and odds ratio (OR) of 0.78, near *FOXE1* (Figure S3). This SNP has been associated with thyroid cancer and TSH levels previously [3, 20]. Homozygous loss-of-function mutations in this gene cause congenital hypothyroidism due to thyroid dysgenesis and other developmental abnormalities [21].

The second association is with rs6679677, with a p-value of $3.5 \cdot 10^{-13}$ and OR of 1.45 near *PTPN22* (Figure S4). The SNP rs6679677 is in LD ($r^2 \approx 0.9$) with rs2476601 (the missense mutation R620W in *PTPN22*, protein tyrosine phosphatase, non-receptor type 22). This mutation has previously been associated with Hashimoto thyroiditis in a relatively small candidate gene study [10]. *PTPN22* also has well established associations with multiple autoimmune conditions [22], including type 1 diabetes, rheumatoid arthritis, systemic lupus erythematosus, juvenile idiopathic arthritis [23], Graves' disease [24], systemic sclerosis [25], and alopecia areata [26].

The next association, which is novel, is with rs4915077, in an intron of *VAV3*. This SNP has a p-value of $8.3 \cdot 10^{-11}$ and OR of 1.4. See Figure 1 for SNPs in this region.

The final genome-wide significant association presented in Table 2 is a novel association with rs3184504, a missense mutation (R262W) in *SH2B3*, with a p-value of $3.0 \cdot 10^{-10}$ and an OR of 0.82 (Figure S5). This SNP has not previously been associated with thyroid disease; however, it has been associated with a number of autoimmune diseases, including type 1 diabetes [27], celiac disease [28], rheumatoid arthritis [29], and multiple sclerosis [30], as well as to hypertension and myocardial infarction [31]. The T allele of rs3184504 is the variant associated with increased risk in our data and corresponds to the W allele in

the protein, which is also the risk variant in previous studies.

Table 2 also includes attempted replications of several other SNPs previously associated with TSH levels. In the *NKX2-1* region, we find only suggestive evidence for the SNP reported in [3] (rs944289, with a p-value of 0.096); however, a nearby SNP (rs378836) shows a stronger signal (p-value of $3.2 \cdot 10^{-5}$). These SNPs are in partial LD ($r^2 = 0.3$), suggesting that they may represent the same signal (Figure S6). Near *CAPZB*, the SNP that was originally reported, rs10917469, was not present on our genotyping platforms, but we did genotype rs10799824, which is in LD ($r^2 = 1$) with rs10917469. rs10799824 shows only suggestive evidence here (p-value of 0.066), but a second SNP in this region, rs1472565 (which is not in LD with rs10799824 and rs10917469) does show a signal, with a p-value of $7.9 \cdot 10^{-6}$, Figure S7. We replicated the association with rs4704397 near *PDE8B* at a modest level of significance (p-value of 0.021) and failed to replicate both the second association found near *FOXE1* in a non-European population, rs755109 (p-value of 0.73) (see Figure S8) and rs2235544 in *DIO1* (p-value of 0.76).

All SNPs with p-values under 10^{-4} are shown in Table S1. The most notable of these 131 suggestive SNPs is perhaps rs9277535 in the HLA region, with a p-value of $4 \cdot 10^{-5}$. The HLA has been associated with many autoimmune conditions (cf. [32]).

Discussion

We have presented two novel associations with hypothyroidism: the non-synonymous change R262W in *SH2B3*, and the SNP rs4915077 near *VAV3*. *SH2B3* has been associated with a host of autoimmune diseases and thus is a good candidate for hypothyroidism. *VAV3* has not yet been associated with autoimmune disease or thyroid function. However, it has been proposed as the candidate gene in the Idd18.1 region linked with type 1 diabetes in mouse [33], and the Vav1/Vav2/Vav3 family is necessary for adaptive immune function in mouse [34]. *VAV3* is also expressed in the thyroid [35] and is down-regulated in a subset of thyroid tumors [36].

We have found one more association with a mutation well known to be involved in autoimmune disease: R620W in *PTPN22*. This association has been observed before in a small candidate gene study of Hashimoto thyroiditis [10]. That study found, in a sample of 194 cases and 2064 controls, an OR of 1.77 (1.31–2.40) for rs2476601/R620W. We observe an OR for this SNP of 1.42 (1.29–1.57), which is within their confidence interval, despite slightly different phenotypes.

While many autoimmune diseases share genetic risk factors in common [32], there is evidence that these diseases form separate clusters based on genetics [37]. For example, the 620W allele of *PTPN22* has a protective effect in Crohn's disease [38] but is the risk allele for type 1 diabetes and others [39]. The associations with *SH2B3* and the 620W allele of *PTPN22* thus suggest that hypothyroidism falls into the same cluster as type 1 diabetes, for example.

We have also attempted to replicate several SNPs associated with thyroid hormone levels. The only one that replicated strongly was the association between thyroid cancer and TSH levels with the thyroid transcription factor *FOXE1* [3]. We find suggestive (but not genome-wide significant) evidence for the genes *NKX2-1*, *CAPZB* and *PDE8B* previously reported in TSH studies. This suggests that while autoimmune loci may play a predominant role in the genetics of hypothyroidism, there may be smaller effects due to genes more directly related to thyroid function and hormone levels.

Due to the nature of online self-report, we have not gathered clinical measures such as TSH levels or hypothyroidism symptoms. Hypothyroidism is thought to be generally under-diagnosed (but could even be over-diagnosed in a health-conscious group of participants), thus there could be some misclassification among participants. About 7.5% of females and 2.8% of males are estimated to have elevated TSH levels and 2-3% of people are estimated to have overt symptoms of hypothyroidism [40]. This is roughly in the same range as the 9.5% prevalence of hypothyroidism (defined as either elevated TSH levels or hypothyroidism diagnosis) in our cohort. In addition, given the replication of loci associated with TSH, this misclassification does not appear to be a large problem here. Ultimately, the loci discovered here

should be further evaluated in a more deeply phenotyped hypothyroidism population, to determine the interplay between thyroid function genes and immune system genes that leads to this disorder.

Methods

Cohort

All participants in the study were customers of 23andMe, Inc., a personal genetics company, who had been genotyped as part of the 23andMe Personal Genome Service®. Individuals included in the cohort were selected for being of primarily European ancestry, as determined through an analysis of local ancestry via comparison to the three HapMap 2 populations, using an unpublished method substantially similar to [41]. This subset was selected as it is the largest relatively unstructured set within the 23andMe customer base.

A maximal set of unrelated individuals was chosen for the analysis using a segmental identity-by-descent (IBD) estimation algorithm (as used in [42]). Individuals were defined as related if they shared more than 700 cM IBD, including both regions where the two individuals share either one or both genomic segments identical-by-descent. This level of relatedness (roughly 20% of the genome) corresponds approximately to the minimal expected sharing between first-cousins in an outbred population.

This study was conducted according to the principles expressed in the Declaration of Helsinki. The study protocol and consent form were approved by the external AAHRPP-accredited IRB, Ethical and Independent Review Services (E&I Review). Our consent and privacy statement preclude the sharing of individual-level data without explicit consent. We have, however, shared summary statistics for all SNPs with p-values under 10^{-4} in Table S1.

Genotyping

For the 23andMe cohort, DNA extraction and genotyping were performed on saliva samples by National Genetics Institute (NGI), a CLIA-certified clinical laboratory and subsidiary of Laboratory Corporation of America. Samples were genotyped on one of three different platforms. About 70% of the participants were genotyped on one of two platforms based on the Illumina HumanHap550+ BeadChip (called V1 and V2 in Table 1), which included SNPs from the standard HumanHap550 panel augmented with a custom set of approximately 25,000 SNPs selected by 23andMe. Two slightly different versions of this platform were used, as described in [42]. The remaining participants were genotyped on the Illumina OmniExpress+ Bead Chip. This platform has a base set of 730,000 SNPs. It was augmented by approximately 250,000 SNPs to make it approximately a superset of the HumanHap550+, as well as a custom set of about 30,000 SNPs. This platform was called V3 in Table 1. Every sample that failed to reach a 98.5% call rate for SNPs on the standard platforms was re-analyzed. Individuals whose analyses failed repeatedly were re-contacted by 23andMe customer service to provide additional samples, as is done for all 23andMe customers. See [42] for further details on the genotyping and sample quality controls.

SNPs with a call rate under 95% or minor allele frequency under 0.01 were excluded from analysis. Call rates were calculated on a per-platform basis. Additionally, SNPs with Hardy-Weinberg p-values [43] less than than 10^{-20} were excluded. Altogether, 870,065 SNPs (on the union of the two platforms) were retained with an average call rate of 99.78%.

Phenotyping

Participants were able to fill out web-based questionnaires whenever they logged into their 23andMe accounts. Participants answered some of the following questions:

- Have you ever been diagnosed by a doctor with any of the following thyroid conditions? (asked as part of a medical history questionnaire)
  - Hyperthyroidism
  - Hypothyroidism
- Have you been diagnosed with any of the following? (asked as part of a questionnaire on baldness)
  - Hyperthyroidism (overactive thyroid)
  - Hypothyroidism (underactive thyroid)
- Have you ever been diagnosed with hypothyroidism (underactive thyroid)?
- Do you currently take medication for hypothyroidism (low thyroid hormone levels)?
- Have you ever been told by a doctor that your thyroid stimulating hormone (TSH) levels were elevated, indicating hypothyroidism?
- Have you ever been diagnosed with thyroid cancer?
- Have you ever received radioactive iodine treatment for goiter or hyperthyroidism (overactive thyroid)?
- Have you ever had all or part of your thyroid surgically removed?
- Have you ever been diagnosed by a doctor with any of the following common cancers? (asked as part of a medical history questionnaire)
  - Thyroid cancer

Cases answered yes to hypothyroidism or to elevated TSH levels or to taking medication for hypothyroidism. Controls answered no to at least one of the qualifying questions and yes to none of them. Individuals reporting hyperthyroidism or thyroid cancer or treatment with radioactive iodine or thyroid removal were excluded, as all of these could cause hypothyroidism or could signal Graves' disease.

As customers of 23andMe, all participants had the opportunity to view reports based on their genetic information on over 100 traits and diseases. Among these diseases were reports on thyroid cancer (covering the *FOXE1* and *NKX2-1* SNPs from [3]) and Hashimoto thyroiditis (covering *PTPN22*). It is unlikely that a predicted high or low risk for thyroid cancer would lead to misreport of hypothyroidism. For *PTPN22*, there was no evidence of a difference in ORs for rs2476601 and hypothyroidism for people viewing their results before or after answering survey questions ($p = 0.89$ for interaction).

Statistical analysis

For the association analysis, all p-values were calculated using a likelihood ratio test for the logistic regression model, adjusting for sex, age, genotyping platform, and projections onto the first five principal components of the genotype data matrix. The phenotypic status of each individual was coded as 0 for unaffected individuals and 1 for affected individuals. Genotypes were coded 0, 1, or 2, to indicate the number of minor alleles present for the tested SNP (corresponding to a log-additive model of association). Reported odds ratios for each SNP relative to the minor allele were defined as the exponential of the regression coefficients, and the alleles used throughout refer to the plus strand of NCBI build 36.3 of the human genome. We used a cutoff for genome-wide significance of $5 \cdot 10^{-8}$ (corresponding to a Bonferroni correction assuming 1 million independent tests). We did not attempt to formalize a criterion for replication of previously associated SNPs, as previous GWAS have tested different aspects of thyroid function from the broad definition of hypothyroidism used here.

Acknowledgments

We thank the customers of 23andMe who answered surveys and participated in this research. Thanks to all the employees of 23andMe, who together have made this research possible, especially Kimberly Barnholt, Geoffrey Benton, Arnab Chowdry, Emily Drabant, Michael Macpherson, and Brian Naughton.

References

1. Park SM, Chatterjee VK (2005) Genetics of congenital hypothyroidism. J Med Genet 42: 379–389.

2. Arnaud-Lopez L, Usala G, Ceresini G, Mitchell B, Pilia M, et al. (2008) Phosphodiesterase 8B gene variants are associated with serum TSH levels and thyroid function. Am J Hum Genet 82: 1270–80.

3. Gudmundsson J, Sulem P, Gudbjartsson D, Jonasson J, Sigurdsson A, et al. (2009) Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. Nat Genet 41: 460–4.

4. Panicker V, Wilson SG, Walsh JP, Richards JB, Brown SJ, et al. (2010) A locus on chromosome 1p36 is associated with thyrotropin and thyroid function as identified by genome-wide association study. Am J Hum Genet 87: 430–435.

5. Lowe J, Maller J, Pe'er I, Neale B, Salit J, et al. (2009) Genome-wide association studies in an isolated founder population from the Pacific Island of Kosrae. PLoS Genet 5: e1000365.

6. Hwang S, Yang Q, Meigs J, Pearce E, Fox C (2007) A genome-wide association for kidney function and endocrine-related traits in the NHLBI's Framingham Heart Study. BMC Med Genet 8 Suppl 1: S10.

7. Hayashi M, Matsushima K, Ohashi H, Tsunoda H, Murase S, et al. (1998) Molecular cloning and characterization of human PDE8B, a novel thyroid-specific isozyme of 3',5'-cyclic nucleotide phosphodiesterase. Biochem Biophys Res Commun 250: 751–756.

8. Bender A, Beavo J (2006) Cyclic nucleotide phosphodiesterases: molecular regulation to clinical use. Pharmacol Rev 58: 488–520.

9. Panicker V, Cluett C, Shields B, Murray A, Parnell KS, et al. (2008) A common variation in deiodinase 1 gene DIO1 is associated with the relative levels of free thyroxine and triiodothyronine. J Clin Endocrinol Metab 93: 3075–3081.

10. Criswell L, Pfeiffer K, Lum R, Gonzales B, Novitzke J, et al. (2005) Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. Am J Hum Genet 76: 561–71.

11. Zeitlin AA, Simmonds MJ, Gough SC (2008) Genetic developments in autoimmune thyroid disease: an evolutionary process. Clin Endocrinol (Oxf) 68: 671–682.

12. Simmonds MJ, Gough SC (2011) The search for the genetic contribution to autoimmune thyroid disease: the never ending story? Brief Funct Genomics 10: 77–90.

13. Donner H, Braun J, Seidl C, Rau H, Finke R, et al. (1997) Codon 17 polymorphism of the cytotoxic T lymphocyte antigen 4 gene in Hashimoto's thyroiditis and Addison's disease. J Clin Endocrinol Metab 82: 4130–4132.

14. Petrone A, Giorgi G, Mesturino CA, Capizzi M, Cascino I, et al. (2001) Association of DRB1*04-DQB1*0301 haplotype and lack of association of two polymorphic sites at CTLA-4 gene with Hashimoto's thyroiditis in an Italian population. Thyroid 11: 171–175.

15. Tandon N, Zhang L, Weetman AP (1991) HLA associations with Hashimoto's thyroiditis. Clin Endocrinol (Oxf) 34: 383–386.

16. Tomer Y, Davies T (2003) Searching for the autoimmune thyroid disease susceptibility genes: from gene mapping to gene function. Endocr Rev 24: 694–717.

17. Chen P, Fann C, Chu C, Chang C, Chang S, et al. (2011) Comprehensive genotyping in two homogeneous Graves' disease samples reveals major and novel HLA association alleles. PLoS One 6: e16635.

18. Burton P, Clayton D, Cardon L, Craddock N, Deloukas P, et al. (2007) Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nat Genet 39: 1329–37.

19. Newby P, Pickles O, Mazumdar S, Brand O, Carr-Smith J, et al. (2010) Follow-up of potential novel Graves' disease susceptibility loci, identified in the UK WTCCC genome-wide nonsynonymous SNP study. Eur J Hum Genet 18: 1021–6.

20. Landa I, Ruiz-Llorente S, Montero-Conde C, Inglada-Prez L, Schiavi F, et al. (2009) The variant rs1867277 in FOXE1 gene confers thyroid cancer susceptibility through the recruitment of USF1/USF2 transcription factors. PLoS Genet 5: e1000637.

21. Castanet M, Mallya U, Agostini M, Schoenmakers E, Mitchell C, et al. (2010) Maternal isodisomy for chromosome 9 causing homozygosity for a novel FOXE1 mutation in syndromic congenital hypothyroidism. J Clin Endocrinol Metab 95: 4031–6.

22. Gregersen P, Olsson L (2009) Recent advances in the genetics of autoimmune disease. Annu Rev Immunol 27: 363–91.

23. Lee YH, Rho YH, Choi SJ, Ji JD, Song GG, et al. (2007) The PTPN22 C1858T functional polymorphism and autoimmune diseases–a meta-analysis. Rheumatology (Oxford) 46: 49–56.

24. Velaga M, Wilson V, Jennings C, Owen C, Herington S, et al. (2004) The codon 620 tryptophan allele of the lymphoid tyrosine phosphatase (LYP) gene is a major determinant of Graves' disease. J Clin Endocrinol Metab 89: 5862–5.

25. Gourh P, Tan F, Assassi S, Ahn C, McNearney T, et al. (2006) Association of the PTPN22 R620W polymorphism with anti-topoisomerase I- and anticentromere antibody-positive systemic sclerosis. Arthritis Rheum 54: 3945–3953.

26. Betz RC, Konig K, Flaquer A, Redler S, Eigelshoven S, et al. (2008) The R620W polymorphism in PTPN22 confers general susceptibility for the development of alopecia areata. Br J Dermatol 158: 389–391.

27. Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, et al. (2007) Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat Genet 39: 857–864.

28. Zhernakova A, Elbers C, Ferwerda B, Romanos J, Trynka G, et al. (2010) Evolutionary and functional analysis of celiac risk loci reveals SH2B3 as a protective factor against bacterial infection. Am J Hum Genet 86: 970–7.

29. Coenen MJ, Trynka G, Heskamp S, Franke B, van Diemen CC, et al. (2009) Common and different genetic background for rheumatoid arthritis and coeliac disease. Hum Mol Genet 18: 4195–4203.

30. Alcina A, Vandenbroeck K, Otaegui D, Saiz A, Gonzalez JR, et al. (2010) The autoimmune disease-associated KIF5A, CD226 and SH2B3 gene variants confer susceptibility for multiple sclerosis. Genes Immun 11: 439–445.

31. Newton-Cheh C, Johnson T, Gateva V, Tobin M, Bochud M, et al. (2009) Genome-wide association study identifies eight loci associated with blood pressure. Nat Genet 41: 666–676.

32. Lettre G, Rioux J (2008) Autoimmune diseases: insights from genome-wide association studies. Hum Mol Genet 17: R116–121.

33. Fraser H, Dendrou C, Healy B, Rainbow D, Howlett S, et al. (2010) Nonobese diabetic congenic strain analysis of autoimmune diabetes reveals genetic complexity of the Idd18 locus and identifies Vav3 as a candidate gene. J Immunol 184: 5075–5084.

34. Fujikawa K, Miletic A, Alt F, Faccio R, Brown T, et al. (2003) Vav1/2/3-null mice define an essential role for Vav family proteins in lymphocyte development and activation but a differential requirement in MAPK signaling in T and B cells. J Exp Med 198: 1595–1608.

35. Movilla N, Bustelo XR (1999) Biological and regulatory properties of Vav-3, a new member of the Vav family of oncoproteins. Mol Cell Biol 19: 7870–7885.

36. Giordano TJ, Kuick R, Thomas DG, Misek DE, Vinco M, et al. (2005) Molecular classification of papillary thyroid carcinoma: distinct BRAF, RAS, and RET/PTC mutation-specific gene expression profiles discovered by DNA microarray analysis. Oncogene 24: 6646–6656.

37. Sirota M, Schaub M, Batzoglou S, Robinson W, Butte A (2009) Autoimmune disease classification by inverse association with SNP alleles. PLoS Genet 5: e1000792.

38. Barrett J, Hansoul S, Nicolae D, Cho J, Duerr R, et al. (2008) Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet 40: 955–62.

39. Hakonarson H, Grant S, Bradfield J, Marchand L, Kim C, et al. (2007) A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. Nature 448: 591–4.

40. Tunbridge WM, Evered DC, Hall R, Appleton D, Brewis M, et al. (1977) The spectrum of thyroid disease in a community: the Whickham survey. Clin Endocrinol (Oxf) 7: 481–493.

41. Falush D, Stephens M, Pritchard J (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164: 1567–87.

42. Eriksson N, Macpherson J, Tung J, Hon L, Naughton B, et al. (2010) Web-based, participant-driven studies yield novel genetic associations for common traits. PLoS Genet 6: e1000993.

43. Wigginton JE, Cutler DJ, Abecasis GR (2005) A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet 76: 887–893.

44. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, et al. (2010) LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26: 2336–2337.

Figure Legends

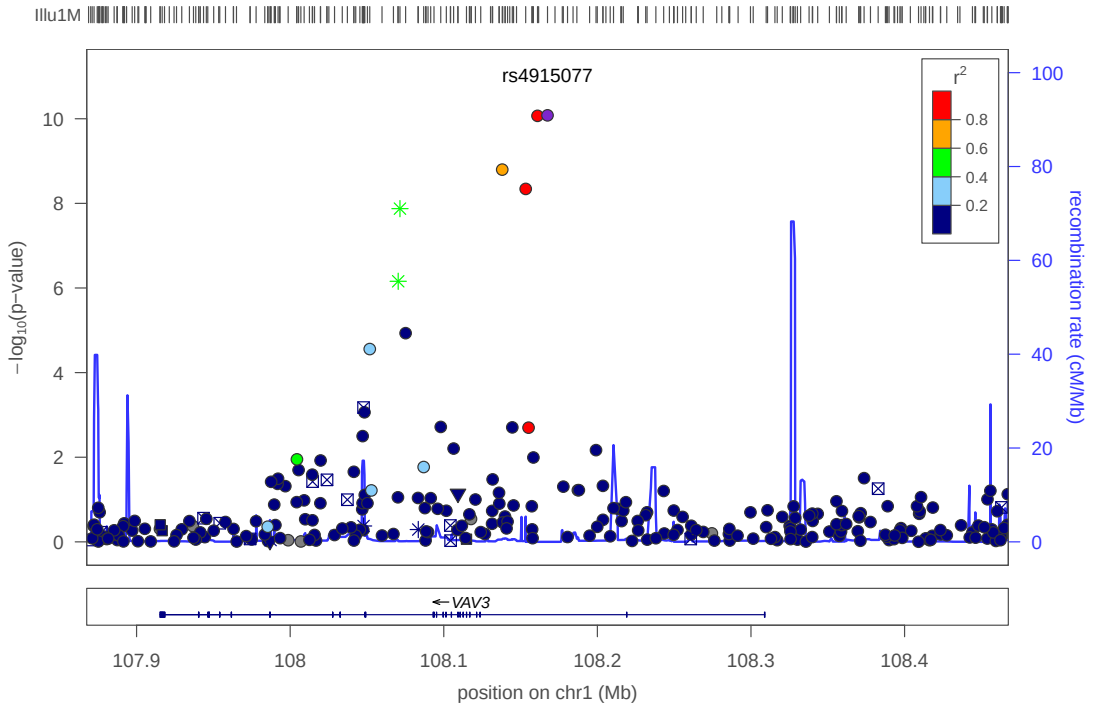

Figure 1. SNPs in the *VAV3* region. In the plot, circles represent unannotated SNPs, upside-down triangles represent non-synonymous variants, and boxes with an "x" are SNPs in regions that are highly conserved across 44 placental mammals. Colors depict the squared correlation (r^2) of each SNP with the most associated SNP (i.e., rs4915077, shown in purple). Gray indicates SNPs for which r^2 information was missing. Plots were produced using the LocusZoom program [44].

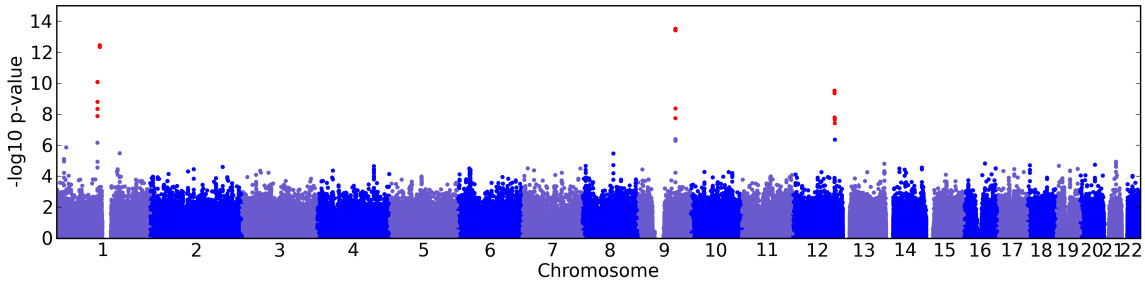

Figure S1. Manhattan plot. Negative log p-values for SNPs by genome position. Genome-wide significant SNPs are shown in red.

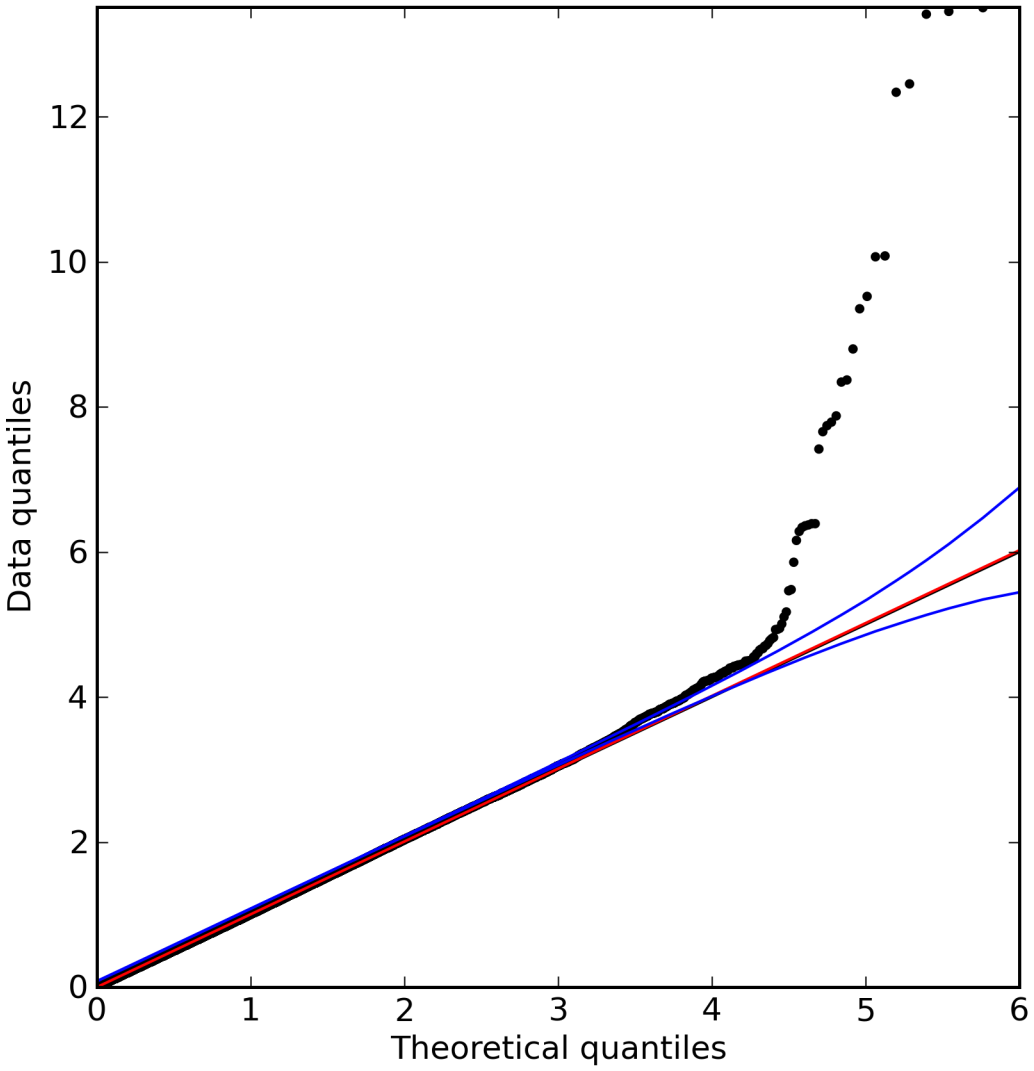

Figure S2. Quantile-quantile plot. Observed p-values versus theoretical p-values under the null hypothesis of no association. The genomic control inflation factor for the study was 0.98 and is indicated by the red line.

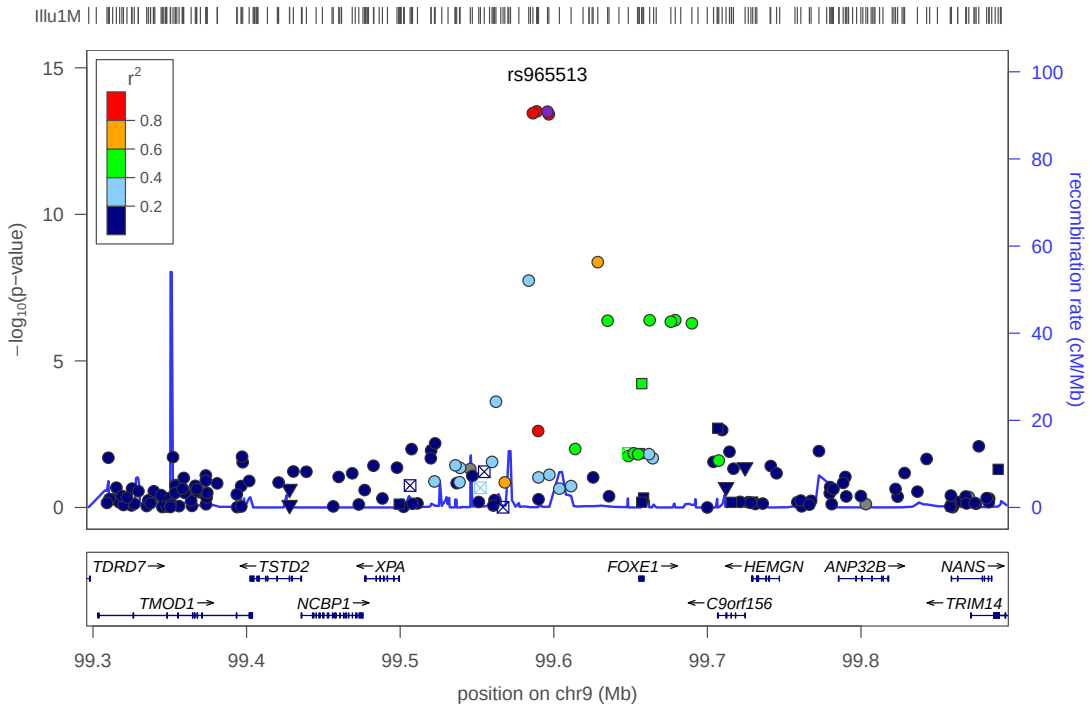

Figure S3. SNPs in the *FOXE1* region. For details, see Figure 1.

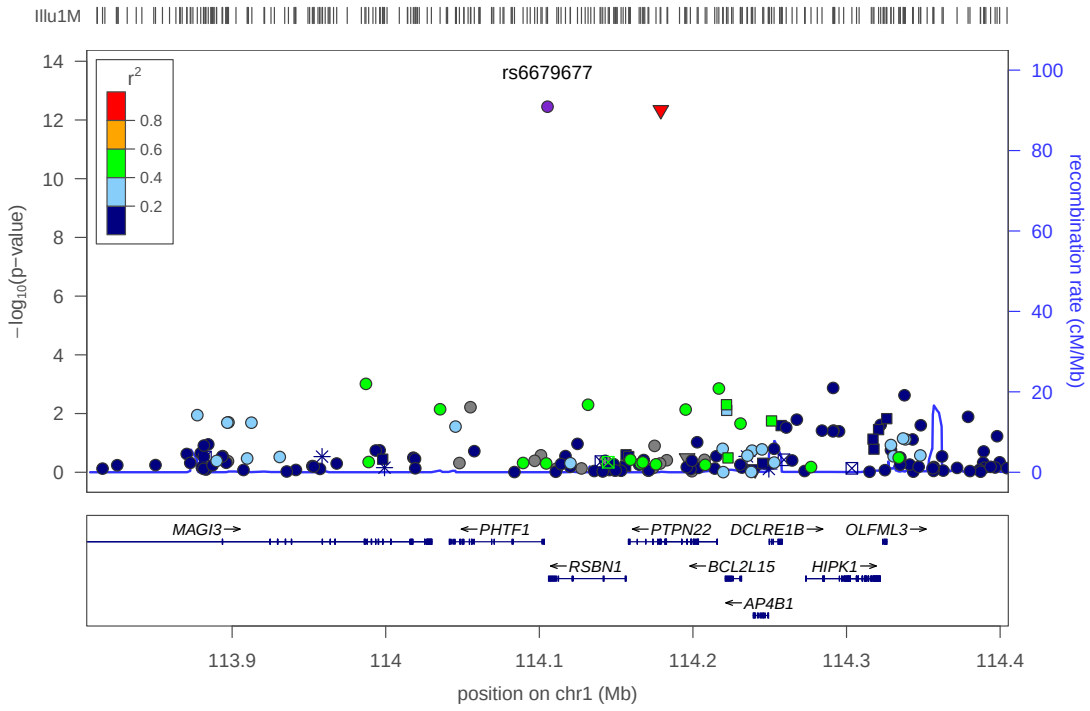

Figure S4. SNPs in the *PTPN22* region. For details, see Figure 1.

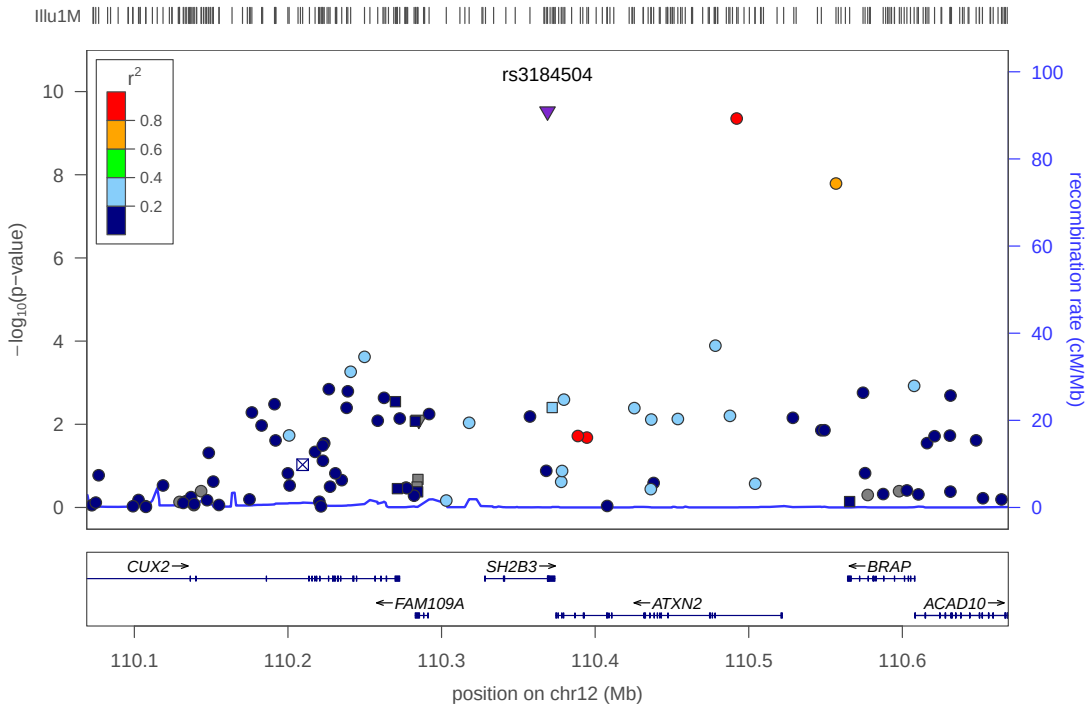

Figure S5. SNPs in the *SH2B3* region. For details, see Figure 1.

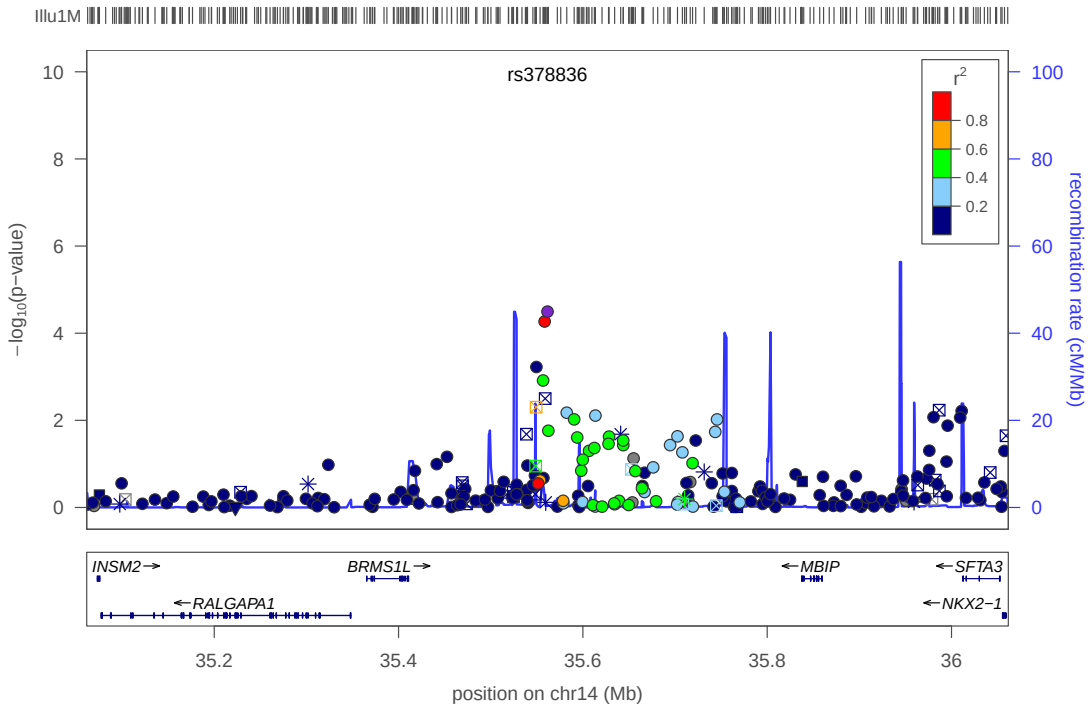

Figure S6. SNPs in the *NKX2-1* region. For details, see Figure 1.

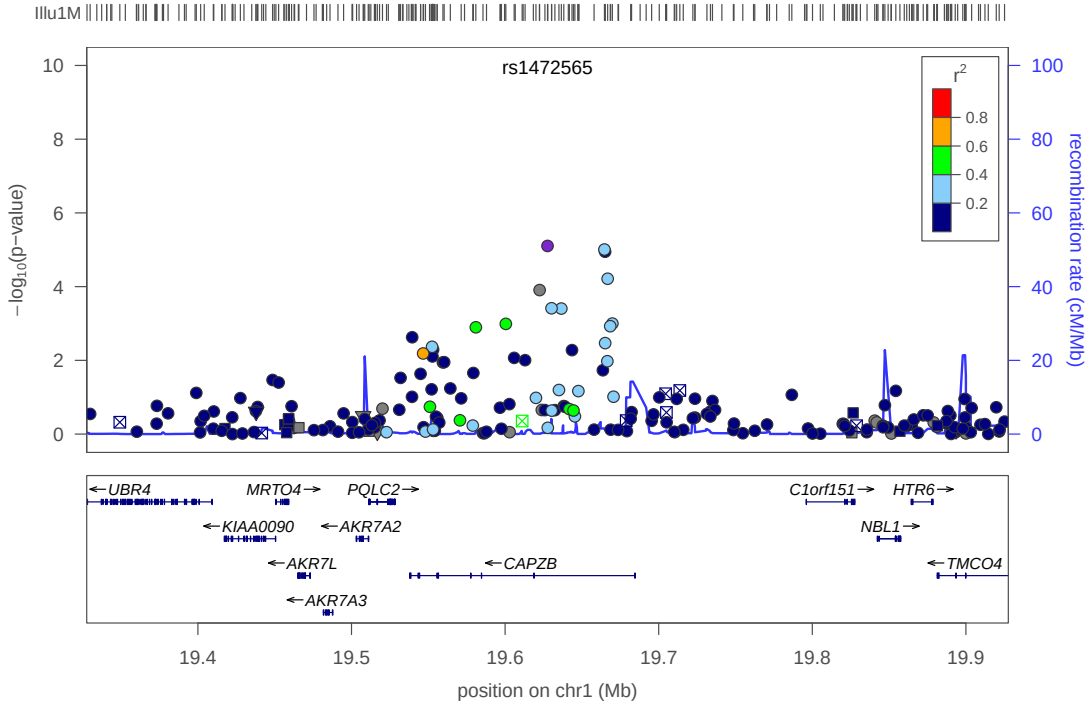

Figure S7. SNPs in the *CAPZB* region. For details, see Figure 1.

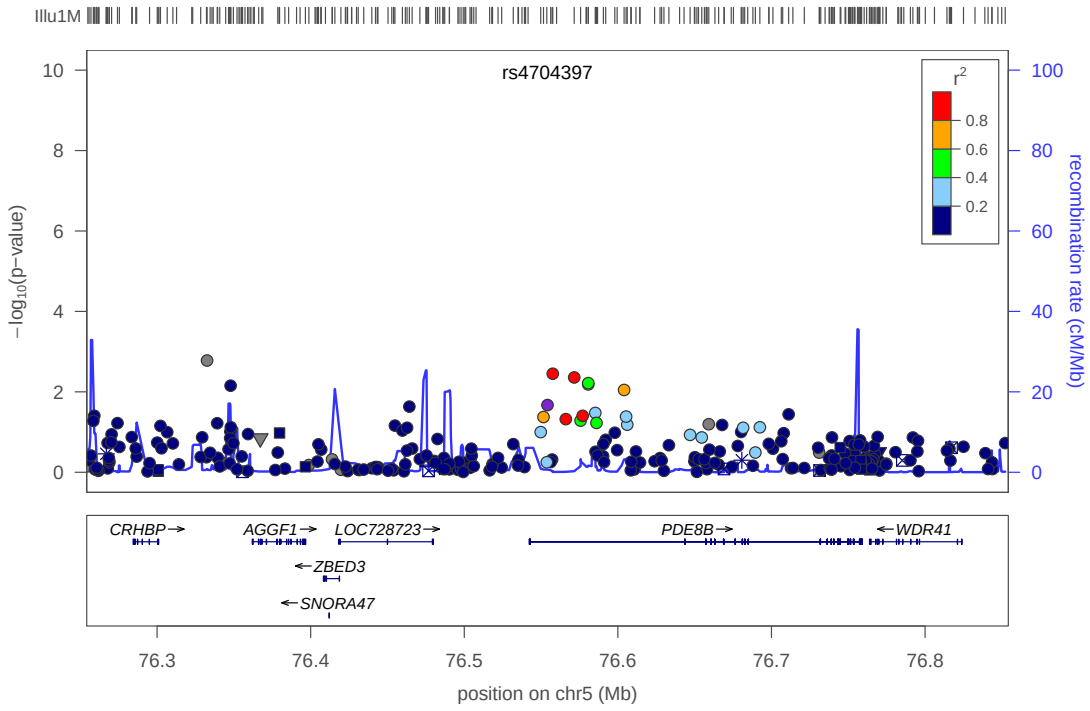

Figure S8. SNPs in the *PDE8B* region. For details, see Figure 1.

Tables

Table 1. Cohort statistics

	Number	Female	Male
Control	24448	8917 (36.5%)	15531 (63.5%)
Case	2564	1891 (73.8%)	673 (26.2%)

	< 45	46–55	56–65	> 65
Control	11850 (48.5%)	3903 (16.0%)	4673 (19.1%)	4022 (16.5%)
Case	580 (22.6%)	419 (16.3%)	801 (31.2%)	764 (29.8%)

	V1	V2	V3
Control	260 (1.1%)	16482 (67.4%)	7706 (31.5%)
Case	27 (1.1%)	1845 (72.0%)	692 (27.0%)

Participants broken down by sex, age, and genotyping platform. V1, V2, and V3 refer to the three platforms used in this study, see Methods.

Table 2. Statistics for genome-wide significant SNPs and selected replications.

SNP	Chr.	Pos.	Region	Alleles	MAF	HWE	p-value	OR
rs965513	9	99595930	*FOXE1*	G/A	0.331	0.23	$3.1 \cdot 10^{-14}$	0.778 (0.73 - 0.83)
rs6679677	1	114105331	*PTPN22*	C/A	0.090	0.74	$3.5 \cdot 10^{-13}$	1.445 (1.31 - 1.59)
rs2476601	1	114179091	*PTPN22*	G/A	0.091	0.66	$4.6 \cdot 10^{-13}$	1.439 (1.31 - 1.58)
rs4915077	1	108167539	*VAV3*	T/C	0.084	0.32	$8.3 \cdot 10^{-11}$	1.397 (1.27 - 1.54)
rs3184504	12	110368991	*SH2B3*	T/C	0.499	0.31	$3 \cdot 10^{-10}$	0.823 (0.77 - 0.87)
rs378836	14	35561627	*NKX2-1*	G/A	0.432	0.96	$3.2 \cdot 10^{-5}$	1.137 (1.07 - 1.21)
rs944289	14	35718997	*NKX2-1*	T/C	0.427	0.61	0.096	1.053 (0.99 - 1.12)
rs1472565	1	19627617	*CAPZB*	T/C	0.476	0.78	$7.9 \cdot 10^{-6}$	1.147 (1.08 - 1.22)
rs10799824	1	19713761	*CAPZB*	G/A	0.153	0.31	0.066	0.853 (0.72 - 1.01)
rs4704397	5	76554198	*PDE8B*	G/A	0.388	0.5	0.021	1.152 (1.02 - 1.30)
rs755109	9	99736024	*FOXE1*	T/C	0.367	0.11	0.73	0.976 (0.86 - 1.10)
rs2235544	1	54148158	*DIO1*	A/C	0.496	0.16	0.76	0.991 (0.93 - 1.05)

All genomic positions are given with respect to NCBI build 36.3. Alleles are listed as major/minor and are specified for the forward strand. Odds ratios are per copy of the minor allele. Two SNPs are listed for *PTPN22*; of these, rs2476601 is the non-synonymous change R620W.

Table S1. SNPs with p-values under 10^{-4}

See Table 2 for details on columns.