POGLUT1, the putative effector gene driven by rs2293370 in primary biliary cholangitis susceptibility locus chromosome 3q13.33

Yuki Hitomi1, Kazuko Ueno2,3, Yosuke Kawai3, Nao Nishida4, Kaname Kojima4,5, Minae Kawashima3, Yoshihiro Aiba6, Hitomi Nakamura6, Hiroshi Kouno7, Hirotaka Kouno7, Hajime Ohta7, Kazuhiro Sugi8, Toshiki Nikami7, Tsutomu Yamashita9, Shinji Katsushika8, Toshihiko Komeda1, Keisuke Ario1, Atsushi Naganuma10, Masaaki Shimada11, Noboru Hirashima12, Kaname Yoshizawa13, Fujiyo Makita14, Kiyoshi Furuta15, Masahiro Kikuchi16, Noriaki Naeshiro17, Hironao Takahashi18, Yutaka Mano19, Haruhiro Yamashita20, Kouki Matsushita21, Seiji Tsumenatsu22, Iwao Yabuchi23, Hideo Nishimura24, Yusuke Shimada25, Kazuhiko Yamauchi26, Tatsujii Komatsu27, Rie Sugimoto28, Hironori Sakai29, Eiji Mita30, Masaharu Koda1, Yoko Nakamura1, Hiroshi Kamitsukasa31, Takeaki Sato32, Makoto Nakamuta1, Naohiko Masaki33, Hajime Takikawa34, Atsushi Tanaka35, Hiromasa Ohira36, Mikio Zeniya37, Masanori Abe38, Shuichi Kaneko39, Masao Honda40, Kuniaki Ara41, Teruko Arinaga-Hino42, Etsuko Hashimoto43, Makiko Taniyai44, Takeji Umemura45, Satoru Joshi46, Kazuhiro Nakao47, Tatsuki Ichikawa48, Hidetaka Shibata49, Akinobu Takaki50, Satoshi Yamagita51, Masataka Seike52, Shotaro Sakisaka53, Yasuaki Takeyama15, Masaru Harada54, Michio Senju55, Osamu Yokosuka56, Tatsuo Kanda57, Yoshiyuki Ueno58, Hiroshi Ebinuma59, Takashi Himoto60, Kazumoto Murata61, Shinji Shimoda62, Shinya Nagao46, Seigo Abiru63, Atsumasa Komori64, Kiyoshi Migitaka65, Masahiro Ito66, Hiroshi Yatahashi67, Yoshihiko Maehara68, Shinji Uemoto29, Norihiro Kokudo30, Masao Nagasaki2,3,31, Katsushi Tokunaga1 & Minoru Nakamura6,7,27,32

Primary biliary cholangitis (PBC) is a chronic and cholestatic autoimmune liver disease caused by the destruction of intrahepatic small bile ducts. Our previous genome-wide association study (GWAS) identified six susceptibility loci for PBC. Here, in order to further elucidate the genetic architecture of PBC, a GWAS was performed on an additional independent sample set, then a genome-wide meta-analysis was performed based on a whole-genome single nucleotide polymorphism (SNP) imputation analysis of a total of 4,045 Japanese individuals (2,060 cases and 1,985 healthy controls). A susceptibility locus on chromosome 3q13.33 (including ARHGAP31, TMEM39A, POGLUT1, TIMMDC1, and CD80) was previously identified both in the European and Chinese populations and was replicated in the Japanese population (OR = 0.7241, P = 3.5 × 10⁻⁴). Subsequent in silico and in vitro functional analyses identified rs2293370, previously reported as the top-hit SNP in this locus in the European population, as the primary functional SNP. Moreover, e-QTL analysis indicated that the effector gene of rs2293370 was Protein O-Glucosyltransferase 1 (POGLUT1) (P = 3.4 × 10⁻⁵). This is the first study to demonstrate that POGLUT1 and not CD80 is the effector gene regulated by the primary functional SNP rs2293370, and that increased expression of POGLUT1 might be involved in the pathogenesis of PBC.

1Department of Human Genetics, Graduate School of Medicine, the University of Tokyo, Tokyo, Japan. 2Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan. 3Graduate School of Medicine, Tohoku University, Sendai, Japan. 4The Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Ichikawa, Japan. 5Japan Science and Technology Agency (JST), Tokyo, Japan. 6Clinical Research Center, National Hospital Organization (NHO) Nagasaki Medical Center, Omura, Japan. 7Headquarters of
Primary biliary cholangitis (PBC) is a chronic and progressive cholestatic liver disease characterized by chronic non-suppurative destructive cholangitis (CNSDC), ductopenia, interface hepatitis, fibrosis, and biliary cirrhosis. The destruction of small bile ducts is considered to be mediated by autoimmune responses against biliary epithelial cells (BEC), including CD4+ T cells, CD8+ T cells, B cells, and natural killer (NK) cells. The higher monozygotic/dizygotic (MZ/DZ) ratio and the higher estimated relative sibling risk (Rs) in PBC patients as compared to unaffected individuals indicates the involvement of strong genetic factors in the development of PBC. Previous genome-wide association studies (GWASs), ImmunoChip analyses, and subsequent meta-analyses in populations of European descent identified human leukocyte antigen (HLA) and 30 non-HLA susceptibility regions (nearest candidate genes from the top-hit SNPs in each locus: IL12RB2/DEND1D/C11orf53, YPEL5/LBH, IL1R1/L1/L1R2, STAT4/NAB1, SLC19A3/CCL20, PCL2, TIMMD1C/TMEM39A, IL-12A/IL12A-AS1/IRQ1C/SCIIP1, DGKG, NFKB1/MANBA, IL7R/CAPSL, NUDT12/C2orf30, IL12B, OLG13/TNEAP3, ELM01, IRF5/TNP03, RPS6KA4, DDX6/CXCR5, TNFRSF1A, ATXN2/BRAP, DLEU1/BCMS, RAD51B, EXOC3L4, RN12/CLEC16A, IRF8/FOXF1, ZPP2P2/GSDMB/IKZF3, MAPT, TYK2, SPIB, and SYNGR1/PDGFBR/RPL3) in PBC. Additionally, Asian-specific susceptibility regions for PBC, including CD58, CD28/CTLA4, IL21-AS1, TNFSF15/TNFSF8, IL16, IL21R, CSNK2N2/CCDC113, and AATID3A, were reported in the Japanese and Chinese populations by means of GWAS and subsequent genome-wide meta-analysis with genome-wide SNP imputation (already identified PBC susceptibility loci including this study are shown in Table 1). Thus, the evidence reported to date indicates presence of shared and non-shared genetic susceptibility profiles behind the pathogenesis of PBC in European and Asian populations.
 locus in European and Chinese populations; consequently, this genome-wide meta-analysis involved replicating chromosome 3q13.33 in the Japanese population. Here, we show in silico and in vitro functional analyses and identify the effector gene and the primary functional SNP in the PBC susceptibility locus chromosome 3q13.33.

| Chromosome | Location | Mapped gene(s) | Top-hit SNP | MAFa | ORb | P | Population | Ref. |
|------------|----------|----------------|-------------|-------|-----|---|-----------|-----|
| 1 | 1p13.1 | CD58 | rs2308747 | 0.39 | 1.29 | 2.9E-12 | Chinese | 3 |
| 1p31.3 | IL12RB2/IL2/ENDD1B/C1orf53 | rs72678531 | 0.16 | 1.61 | 2.8E-38 | European | 4 |
| 2 | 2p23.1 | YPEL3/LBH | rs4952108 | 0.19 | 1.28 | 5.0E-08 | European | 18 |
| 2q12.1 | IL1RL1/IL1RL2 | rs12712133 | 0.44 | 1.14 | 5.0E-09 | European | 18 |
| 2q32.2 | STAT4/NAIB | rs3024921, etc. | 0.05 | 0.72 | 9.0E-25 | European | 12 |
| 3 | 3p24.3 | PLCL2 | rs1372072 | 0.32 | 1.31 | 4.0E-14 | Chinese | 17 |
| 3q13.33 | TIMMDC1/TMEM39A | rs2293370 | 0.16 | 1.39 | 7.0E-16 | European | 12 |
| 3q25.33 | IL-12A/IL12A-AS1/IQCJ/SCHIP1 | rs2366643, etc. | 0.38 | 0.62 | 3.0E-35 | European | 12 |
| 4 | 4p16.3 | DGKQ | rs11724804 | 0.44 | 1.22 | 9.0E-12 | European | 14 |
| 4q24 | NFKB1/MANBA | rs7665090 | 0.49 | 1.26 | 8.0E-14 | European | 12 |
| 4q27 | IL21-AS1 | rs925550 | 0.35 | 1.31 | 4.0E-13 | Chinese | 17 |
| 5 | 5p13.2 | IL7R/CAPS | rs6871748 | 0.28 | 1.30 | 2.0E-13 | European | 12 |
| 5q21.1 | IL12B | rs2546890 | 0.50 | 1.15 | 1E-10 | European | 14 |
| 5q21.1 | IL12B | rs2546890 | 0.50 | 1.15 | 1E-10 | European | 14 |
| 5q21.1 | IL12B | rs2546890 | 0.50 | 1.15 | 1E-10 | European | 14 |
| 5q33.3 | IL12B | rs2546890 | 0.50 | 1.15 | 1E-10 | European | 14 |
| 6 | 6q23.3 | OLG3/TNFAIP3 | rs6933404 | 0.17 | 1.18 | 4.0E-08 | European | 14 |
| 7 | 7p14.1 | ELMO1 | rs6974491 | 0.16 | 1.24 | 4.0E-08 | European | 14 |
| 7q32.1 | IRF5/TNPO3 | rs10488631 | 0.10 | 1.59 | 5.0E-23 | European | 14 |
| 9 | 9q32 | TNFSF15/TNFSF8 | rs4979467 | 0.37 | 1.53 | 1E-29 | Chinese | 17 |
| 11 | 11q13.1 | RPS6KA4 | rs538147 | 0.39 | 1.23 | 2.0E-10 | European | 10 |
| 11q23.3 | DDX6/CXCR5 | rs80965107 | 0.19 | 1.39 | 7.0E-16 | European | 12 |
| 12 | 12p13.31 | TNFRSF1A | rs1800693 | 0.43 | 1.17 | 4.0E-09 | Chinese | 17 |
| 12q24.12 | ATXN2/BRAP | rs9591325 | 0.05 | 1.63 | 1E-10 | European | 14 |
| 13 | 13q14.2 | DLEU1/BCMS | rs34536443 | 0.03 | 1.95 | 1E-12 | European | 12 |
| 14 | 14q24.1 | RAD51B | rs91263 | 0.33 | 1.29 | 2.0E-11 | European | 12 |
| 14q32.3 | EXOC3L4 | rs2297067 | 0.22 | 1.39 | 6.0E-19 | European | 14 |
| 15 | 15q25.1 | IL16 | rs11556218 | 0.19 | 1.29 | 9.0E-09 | Chinese | 17 |
| 16 | 16p12.1 | IL21R | rs2189521 | 0.30 | 1.37 | 4.0E-09 | Chinese | 17 |
| 16p12.1 | IL21R | rs2189521 | 0.30 | 1.37 | 4.0E-09 | Chinese | 17 |
| 16q21 | RMI2/CLEC16A | rs1646019, etc. | 0.30 | 1.38 | 2.0E-23 | European | 12 |
| 17 | 17q21.1 | ZPBP2/GSDMB/1KZF3 | rs8067378 | 0.47 | 1.26 | 6.0E-14 | European | 12 |
| 17 | 17q21.3 | MAPT | rs17564829 | 0.23 | 1.25 | 2.0E-09 | European | 12 |
| 19 | 19p13.2 | TYK2 | rs34536443 | 0.03 | 1.95 | 1E-12 | European | 12 |
| 22 | 22q13.1 | SYNGR1/PDGFBR/RPL3 | rs2267407 | 0.24 | 1.29 | 1E-13 | European | 12 |
| 22 | 22q13.1 | SYNGR1/PDGFBR/RPL3 | rs137603 | 0.13 | 1.37 | 3.0E-08 | Chinese | 17 |

Table 1. Non-HLA PBC susceptibility loci in European, Chinese, and Japanese populations. aMAF (minor allele frequency) in 1000 genome project in each population. bOR (odds ratio).
Results

GWAS and genome-wide meta-analysis. We genotyped an independent set of 1,148 samples (668 PBC cases and 480 healthy controls) using the Affymetrix Japonica V1 Array 24. Thirty-four samples were excluded by Dish QC (\(<0.82\)) or overall call rate for a total of 20,000 SNPs (\(<0.97\)). A further 13 samples were excluded because of cryptic relatives. A further 13 samples were located far from the JPT cluster drawn using the first and second components after PCA and were removed from further analysis (Supplementary Fig. 1A). We re-genotype called about 2,897 samples (1,392 PBC cases and 1,505 healthy controls) collected in the previous study16. Eighteen samples were excluded by Dish QC (\(<0.82\)) or overall call rate for a total of 20,000 SNPs (\(<0.97\)) and 15 samples were excluded because of cryptic relatives. Seventeen samples were located far from the HapMap JPT cluster drawn using the first and second components after PCA and were removed from further analysis (Supplementary Fig. 1B).

A quantile-quantile plot of the distribution of test statistics for the comparison of allele frequencies in the PBC cases and healthy controls provided an inflation factor lambda value of 1.097 for all tested SNPs for the 1,148 entries in the current dataset and a value 1.061 for the 2,897 entries in the previous dataset (Supplementary Fig. 2). Genotype imputation and the association study were separately performed for the two datasets. The process of data cleaning and meta-analysis is summarized in Supplementary Fig. 3.

Figure 1 shows a genome-wide view of the single-point association data based on allele frequencies after meta-analysis. The loci HLA, TNFSF15, IL7R, NFKB1/MANBA, and chromosome 17q12-21 showed significant associations with PBC, as reported in the previous GWAS performed on a Japanese population 16. In addition to these regions, meta-analysis to combine the two datasets showed a significant association in chromosome 3q13.33 (Top hit SNP: rs57271503, \(OR = 0.7241, P = 3.5 \times 10^{-9}\), Fig. 2), although this locus appears as evidence of no association with PBC from studies using each platform (Japonica and ASI, Supplementary Fig. 4).

Identification of rs2293370 as the primary functional SNP in chromosome 3q13.33. Among the 29 SNPs whose \(P\) values were less than \(1.0 \times 10^{-6}\) upon genome-wide meta-analysis, SNPs located in the 3′-untranslated region (UTR) and synonymous substitutions were selected as potential candidates for primary functional variation in the chromosome 3q13.33 region (Table 2 and Fig. 2). Five of the 29 SNPs [rs57271503 and rs3830649 in the 3′UTR of CD80, rs2305249 in exon 11 of ARHGAP31 (P567P), rs1131265 in exon 3 of TIMMDC1 (V146V), and rs3732421 in the 3′UTR of TMEM39A] are located in the 3′UTR or synonymous substitutions but are unrelated to their own gene expression as determined by e-QTL analysis 25 (Supplementary Fig. 5) and thus were excluded from further analysis.

Two of the remaining 24 SNPs had RegulomeDB scores higher than 3 and these scores were supported by their location in DNase hyper-sensitivity clusters and the binding of transcription factor. Consequently, these two SNPs were selected as potential candidates (Table 2; rs2293370 in intron 2 of TIMMDC1 and rs56008261 in intron 8 of ARHGAP31). Both SNPs were located in DNase I hyper-sensitivity clusters and in H3K27Ac markers in at least one cell type identified by the UCSC, genome browser 26.

We performed electrophoretic mobility shift assays (EMSAs) to evaluate the effect of candidate SNPs that potentially regulate the binding affinity of transcription factors. A difference in mobility shift between the major allele and the minor allele was detected for rs2293370 in HepG2 (Fig. 3A) and Jurkat (Supplementary Fig. 6A) cells. The shifted band was abrogated by incubation with a \(200\times\) concentration of a non-labeled probe (competitor probe) (Fig. 3B and Supplementary Fig. 6B). In contrast, there was no difference in mobility shift for rs56008261 between the major allele and the minor allele (Fig. 3A and Supplementary Fig. 6A).
Additionally, in order to deny the possibility of the existence of other variations with independent genetic contributions in chromosome 3q13.33, conditional logistic regression analysis was performed. When the rs2293370 was conditioned on, significant associations of other SNPs in chromosome 3q13.33 totally disappeared (Supplementary Fig. 7). These results indicated that rs2293370 is the primary functional SNP in chromosome 3q13.33.

**Molecular features of disease susceptibility influenced by rs2293370.** We performed luciferase reporter assays in HepG2 and Jurkat cells to determine the differences in transcription efficiency between the C (major allele, PBC-susceptibility) and T (minor allele, PBC-protective) alleles of rs2293370. Concordant with the result of EMSA, the luciferase activity of cells 24 h after transfection with a reporter construct containing the C allele of rs2293370 was significantly higher than that of cells containing the T allele for both cell lines (Fig. 3C,D and Supplementary Fig. 6C).

Next, we performed *in silico* prediction of transcription factor binding using the TRANSFAC database to identify the transcription factor responsible for the mobility shift associated with the C allele of rs2293370. The C allele of rs2293370, but not the T allele, was predicted to be involved in a binding motif of Runt-related transcription factors (Supplementary Fig. 8). Although the DMRT and Myb families also showed similar patterns, they are not expressed in HepG2 or Jurkat cells. Of the Runt-related transcription factors, Runt-related transcription factor (RUNX1) -1, but not RUNX-2 and RUNX-3, was confirmed to be expressed in both HepG2 and Jurkat cells (Supplementary Fig. 9). Consistent with the *in silico* prediction of transcription factor binding, the mobility shift associated with the C allele of rs2293370 was reduced by pre-incubation with an anti-RUNX1 antibody prior to electrophoresis (Fig. 3E).

These results indicated that the PBC susceptibility allele of rs2293370 enhances transcription via binding with RUNX1.

**The mRNA expression level of POGLUT1 is influenced by rs2293370.** We used the GTEx portal database to assess the influence of rs2293370 on endogenous gene expression by comparing the expression levels of all genes in the human genome for the different genotypes of rs2293370 in every organ whose expression level of *POGLUT1* was above the threshold for detection. Individuals carrying the C allele (i.e., the PBC-susceptible allele) of rs2293370 showed a significantly higher level of expression of *POGLUT1* in several organs (Fig. 4; statistical significance level: \( P < 0.05/47 \) organs = 0.00106). Other genes located in chromosome 3q13.33 (*ARHGAP31*, *TMEM39A*, *POGLUT1*, *TIMMDC1*, and *CD80*) showed no significant association between rs2293370 genotypes and gene expression (Supplementary Fig. 10).

**Discussion**

In the present study we identified chromosome 3q13.33, which includes the genes *ARHGAP31*, *TMEM39A*, *POGLUT1*, *TIMMDC1*, and *CD80*, as a PBC susceptibility locus in the Japanese population by genome-wide meta-analysis based on whole-genome SNP imputation analysis of two distinct data sets of Japanese PBC-GWAS. The role of chromosome 3q13.33 had previously been identified in European and Chinese populations. Consequently, rs2293370, which is located in intron 2 of *TIMMDC1*, was identified as the primary functional SNP for disease susceptibility to PBC in chromosome 3q13.33 by *in silico* and *in vitro* functional analyses. In addition, the disease protective allele of rs2293370 was shown to disrupt a RUNX1 binding site and was associated with significantly decreased *POGLUT1* mRNA expression levels in tissues compared with individuals without this allele.
Table 2. 29 SNPs associated with susceptibility to PBC in the Japanese population in chromosome 3q13.33 by high-density association mapping.

| rs number | SNP location (Chr.3) | GWAS/imputation | Allele 1 | Allele 2 | P value | OR | Regulome DB score | UCSC (Regulatory Motifs) | Location |
|-----------|----------------------|------------------|----------|----------|---------|----|------------------|--------------------------|----------|
| rs57271503 | 119244593            | imputation       | G        | A        | 3.48E-09 | 0.724 | No data | ×                        | CD80 YUTR   |
| rs9855065  | 119130141            | imputation       | G        | A        | 3.57E-09 | 0.725 | No data | ×                        | ARHGAP31 intron 11 |
| rs3830649  | 119246385            | imputation       | G        | del      | 4.66E-09 | 0.727 | No data | ×                        | CD80 YUTR   |
| rs2305249  | 119128398            | imputation       | G        | A        | 5.07E-09 | 0.728 | No data | ×                        | ARHGAP31 Exon 11 (P567P) |
| rs13092998 | 119245044            | GWAS (Japonica)  | G        | T        | 5.45E-09 | 0.728 | 6        | ×                        | CD80 intron 6 |
| rs62264485 | 119237798            | imputation       | C        | A        | 6.00E-09 | 0.728 | 6        | ×                        | TIMMDC1 intron 6 |
| rs2293370  | 119219934            | GWAS (ASI, Japonica) | G    | A        | 6.08E-09 | 0.728 | 3a       | Δ                        | TIMMDC1 intron 2 |
| rs14633138 | 119128634            | imputation       | T        | C        | 6.34E-09 | 0.751 | 6        | ×                        | ARHGAP31 intron 11 |
| rs1132165  | 11922456             | imputation       | G        | C        | 6.57E-09 | 0.729 | No data | ×                        | TIMMDC1 Exon 3 (V146V) |
| rs1463319  | 11912682             | imputation       | A        | G        | 6.70E-09 | 0.749 | No data | ×                        | ARHGAP31 intron 11 |
| rs3732421  | 119150089            | imputation       | A        | G        | 9.93E-09 | 0.732 | 5        | ×                        | TMEM39A 3′UTR |
| rs7650774  | 119205050            | imputation       | T        | C        | 1.01E-08 | 0.731 | No data | ×                        | POGUT1 intron 6 |
| rs12636784 | 119174383            | imputation       | A        | G        | 1.27E-08 | 0.733 | 6        | ×                        | TMEM39A intron 3 |
| rs4687853  | 11913036             | imputation       | A        | G        | 1.47E-08 | 0.754 | 6        | ×                        | ARHGAP31 intron 11 |
| rs9843355  | 119228508            | imputation       | G        | A        | 1.79E-08 | 0.735 | 6        | ×                        | TIMMDC1 intron 4 |
| rs1530687  | 11911451             | imputation       | T        | C        | 6.26E-08 | 0.767 | 5        | Δ                        | ARHGAP31 intron 8 |
| rs9831023  | 119111762            | GWAS (ASI, Japonica) | T    | C        | 6.98E-08 | 0.768 | 5        | ×                        | ARHGAP31 intron 7 |
| rs9840409  | 119116150            | imputation       | G        | A        | 9.34E-08 | 0.753 | 5        | ×                        | ARHGAP31 intron 8 |
| rs100198   | 11913820             | imputation       | A        | C        | 1.34E-07 | 0.771 | 6        | ×                        | ARHGAP31 intron 8 |
| rs11922594 | 119125822            | imputation       | T        | C        | 1.50E-07 | 0.771 | 5        | ×                        | ARHGAP31 intron 10 |
| rs6773050  | 119123814            | GWAS (ASI)       | C        | T        | 1.62E-07 | 0.774 | 6        | ×                        | ARHGAP31 intron 10 |
| rs1249341  | 119122820            | imputation       | T        | C        | 1.90E-07 | 0.758 | No data | ×                        | ARHGAP31 intron 10 |
| rs4279094  | 119114693            | GWAS (ASI)       | A        | G        | 2.12E-07 | 0.774 | 4        | ×                        | ARHGAP31 intron 8 |
| rs9846306  | 119116064            | imputation       | A        | C        | 2.12E-07 | 0.774 | 5        | ×                        | ARHGAP31 intron 8 |
| rs56008261 | 119114927            | GWAS (Japonica)  | T        | C        | 2.61E-07 | 0.776 | 1b       | 0                        | ARHGAP31 intron 8 |
| rs6776377  | 119115556            | imputation       | T        | C        | 2.81E-07 | 0.776 | 5        | ×                        | ARHGAP31 intron 8 |
| rs6787836  | 119115567            | imputation       | A        | G        | 5.85E-07 | 0.782 | 5        | ×                        | ARHGAP31 intron 8 |
| rs11715698 | 11918497            | imputation       | A        | G        | 6.61E-07 | 0.788 | 5        | ×                        | ARHGAP31 intron 9 |

The contribution of POGUT1 to the pathogenesis of PBC has not been reported to date. Endoplasmic reticulum (ER)-localized protein O-glucosyltransferase 1, which is encoded by POGUT1, adds glucose moieties to serine residues of the epidermal growth factor (EGF)-like domains of Notch family proteins. Notch signaling is an evolutionarily conserved cascade that includes four receptors (Notch 1–4) and five ligands (Jagged 1, Jagged 2, Delta-like ligand 1 (DLL1), DLL3 and DLL4). Therefore, it might be possible that genetic polymorphisms affecting the expression levels of POGUT1 influence the Notch signaling pathway by altering Notch glycosylation. The generation and development of diverse blood cell lineages and peripheral immune responses are regulated by this Notch signaling cascade, especially in hematopoiesis during T cell lineage commitment and maturation in the thymus, and during marginal zone B (MZB) cell development in the spleen. Recently, dendritic cell (DC) homeostasis and the development of several lymphocyte subsets belonging to the innate immune system have been reported to be regulated by Notch. Therefore, inappropriate immune responses against self-antigens could occur due to impaired regulation of Notch signaling. In experimental autoimmune encephalomyelitis (EAE) and non-obese diabetic (NOD) mice, which are mouse models for multiple sclerosis and type 1 diabetes (T1D), respectively, disease progression was impeded by the administration of blocking antibodies against Notch receptors or DLL4. Therefore, higher endogenous levels of POGUT1 caused by the PBC-susceptible allele of rs2293370 may induce excessive Notch signaling and inappropriate immune responses against self-antigens. Very importantly, Notch signaling is also involved in the development or formation of intrahepatic bile ducts. Mutations in JAG1 or Notch2 are known causes of Alagille syndrome, an autosomal dominant disease characterized by congenital cholangiopathy with jaundice and bile duct paucity. POGUT1 was shown to regulate the number of bile ducts around portal veins in a Jag1-dependent manner using Jag1+/− and POGUT1+/− mice. These results indicate that POGUT1 might be involved in the mechanism of bile duct loss in PBC. However, a limitation of this study was that endogenous POGUT1 expression levels in PBC patients were not examined in this study. Additional studies are warranted to improve our understanding of the relationship between PBC pathogenesis and POGUT1.

A more than 100-kb stretch of the genome is located in chromosome 3q13.33 that includes ARHGAP31, TMEM39A, POGUT1, TIMMDC1, and CD80. In addition to the present study on the Japanese population, this
locus, as represented by top-hit SNP rs2293370, has been reported as a susceptibility region to PBC in European and Chinese populations\textsuperscript{10,14,17}. The present study identified rs2293370 as the primary functional SNP by \textit{in silico} and \textit{in vitro} functional analysis. Therefore, \textit{POGLUT1} is likely the effector gene for susceptibility to PBC in not only the Japanese population but also other populations. This locus has also been reported as a susceptibility region for celiac disease, multiple sclerosis, systemic lupus erythematosus, and vitiligo, represented by rs11712165 in \textit{ARHGAP31}, rs2293370 in \textit{TIMMDC1}, and rs6804441 and rs148136154 in \textit{CD80}, respectively\textsuperscript{37–40}. The LDs were not strong between these represented tag-SNPs and rs2293370, which was identified as the primary functional SNP in this study ($r^2 < 0.4$ in 1000 genomes Asian, Supplementary Fig. 11). Regardless, rs2293370, whose effector gene is \textit{POGLUT1}, may operate as a functional SNP in both PBC and multiple sclerosis.

The present study, compared with the GWAS in the European descent, identified four out of 30 non-HLA loci (\textit{IL7R}, \textit{NFKB1/MANBA}, chromosome 17q12-21, and chromosome 3q13.33) as susceptibility loci in the Japanese population. Additionally, even though the $p$-values were below the genome-wide significance level ($P < 5.0 \times 10^{-8}$), the direction of OR was the same between a European descent population and the Japanese population in \textit{MMEL1}, \textit{STAT4}/\textit{STAT1}, \textit{IL12A}, \textit{CXCR5/DDX6}, and \textit{SPIB} (Table 1). These loci are likely candidates as shared susceptibility loci for the pathogenesis of PBC between European and Japanese populations. We are currently working to clarify the shared gene profiles between different populations by conducting SNP imputation and subsequent meta-analysis using GWAS data from an international collaboration involving the UK, Italy, the USA, Canada, China and Japan.

In conclusion, genome-wide meta-analysis together with \textit{in silico}/\textit{in vitro} functional analyses identified the primary functional SNP rs2293370 and the effector gene \textit{POGLUT1}. Chromosome 3q13.33 contains \textit{CD80}, which encodes a well-known co-stimulatory signaling molecule necessary for antigen presentation from HLA class II.
to T cell receptor (TCR). CD80 had been assumed as a candidate effector gene in previous GWASs, whereas our approach identified POGLUT1 as the target transcript for disease susceptibility in this locus comprehensively without stereotypes. Of the PBC susceptibility genes identified in the Japanese population, we previously identified the primary functional SNPs of TNFSF15, PRKCB, and chromosome 17q12-21, and their effector genes16,22,41, as well as chromosome 3q13.33 in the present study. Similar post-GWAS approaches for susceptibility genes are needed to further clarify the molecular mechanisms of disease development.

Materials and Methods

Subjects. DNA samples for GWAS using the Japonica array platform were collected from 1,148 individuals (668 PBC cases and 480 healthy controls). The demographics of the PBC cases and controls are shown in Supplementary Table 1. Written informed consent was obtained from all participants. The protocol of this study was approved by the committee on research ethics and genetically modified organisms of the Graduate School of Medicine, The University of Tokyo, by the ethics committees of Nagasaki Medical Center, and by the ethics committees of Tohoku Medical Megabank Organization, Tohoku University. All experiments were performed in accordance with relevant guidelines and regulations.

Genotyping, quality control, and genotype imputation. Genotyping was performed using the Japonica V1 array (1,148 Japanese individuals in the present study; Affymetrix Japan). Genotype calling was performed with the apt-probeset-genotype program in Affymetrix Power Tools ver. 1.18.2 (Thermo Fisher Scientific Inc., Waltham, MA). Sample quality control was conducted by following the manufacturer’s recommendation (dish QC > 0.82 and sample call rate > 97%). Clustering of each SNP was evaluated by the Ps classification function in the SNPolisher package (version 1.5.2, Thermo Fisher Scientific Inc.). SNPs that were assigned “recommended” by the Ps classification function were used for downstream analyses. SNPs that satisfied the following criteria were used for genotype imputation: a call rate > 99.0%, a Hardy-Weinberg Equilibrium (HWE) p-value > 0.0001, and a minor allele frequency (MAF) > 0.5%. Pre-phasing was conducted with EAGLE v2.3.246.

Figure 4. rs2293370 genotypes are associated with differences in endogenous POGLUT1 expression levels. (A) The relationship between rs2293370 genotype and the endogenous expression of POGLUT1 was compared in all tissues registered in the GTEx database. The effect sizes of the rs2293370 minor allele (T allele; disease protective) in every organ are shown. Statistical significance levels before Bonferroni multiple comparison correction were \( P = 0.00116 \), \( *Pc < 0.05 \), \( **Pc < 0.005 \), and \( ***Pc < 0.0005 \), after Bonferroni multiple comparison correction. (B and C) Box plots showing the association between endogenous POGLUT1 expression and rs2293370 genotypes in (B). Transformed Fibroblast and (C) Whole Blood.
with default settings. Genotype imputation was conducted with IMPUTE4 v1.0 using a phased reference panel of 2,049 Japanese individuals from a prospective, general population cohort study performed by the Tohoku Medical Megabank Organization (ToMMo). These procedures were conducted using default settings. Cryptic relatives were excluded using PRIMUS with default settings. In addition, principal component analysis (PCA) was performed using East Asian samples from the International 1000 Genome Project (104 JPT, 103 CHB, 93 CHS, 91 CDX, and 99 KHV samples) in addition to the case and control samples. PCA identified outliers to be excluded using the Smirnov-Grubbs test with a Bonferroni corrected p-value < 0.05. We had previously analyzed the data (2,897 Japanese individuals) using the Axiom Genome-Wide ASI 1 Array and re-analyzed the data using the above-mentioned procedures.

**Association analysis and meta-analysis.** Association analysis was performed with PLINK (version 1.9) in each dataset (i.e., 2,897 ASI array data and 1,148 Japonica array data). The following options were used for PLINK: a call rate > 97.0%, a HWE p-value > 0.000001, a minor allele frequency (MAF) > 0.1%, and a logistic regression model.

Meta-analysis was performed using PLINK with the meta-analysis option after excluding duplicates between the two datasets. The fixed-effects meta-analysis p-value was used.

**Databases.** The probability that a candidate functional variation might influence transcription regulation was evaluated using the RegulomeDB database (http://www.regulomedb.org/index) and the UCSC genome browser (http://genome.ucsc.edu/index.html). Transcription factor binding was predicted using TRANSFAC Professional (QIAGEN, Valencia CA; http://www.gene-regulation.com/pub/databases.html). Gene expression levels in each cell line and the correlation between the genotypes of candidate SNPs and gene expression were examined using GeneCards (http://www.genecards.org/) and the GTEx portal database (http://gtexportal.org/home/), respectively. P values < 0.05, after adjustment for multiple testing (Bonferroni correction), were regarded as statistically significant.

**Electrophoretic mobility shift assay (EMSA).** EMSA was performed using a LightShift Chemiluminescent EMSA Kit (Thermo-Fisher Scientific) and biotin-labeled double-stranded oligonucleotide probes corresponding to each major and minor allele (Supplementary Table 2), according to the manufacturer's instructions. These oligonucleotide probes (10 fmol/μL) and a nuclear extract (2.5 μg/μL) of HepG2 or Jurkat cells (Nuclear Extract Kit; Active Motif, Carlsbad, CA) were incubated together for 30 min at 25°C.

The super-shift assay was performed by incubating the biotin-labeled probe with the nuclear extracts for 30 min at 25°C, before subsequently incubating these complexes with Anti-RUNX1/AML1 antibody - ChIP Grade (ab23980) (Abcam, Cambridge, UK) for 30 min at 25°C.

Each assay was independently performed three times.

**Luciferase reporter assay.** Amplicons of part of the 2nd intron of TIMMDC1, which contain each allele of rs2293370, were obtained from human genomic DNA using specific PCR primers (Supplementary Table 3) and then subcloned into the luciferase reporter pGL4.23 (luc2/minP) vector (Promega, Madison, WI). pGL4.23 constructs (500 ng) of each allele and 50 ng of the pGL4.74 (hRluc/TK) vector as an internal control were transfected into Jurkat and HepG2 cells using Lipofectamine 3000 (Thermo-Fisher Scientific). The Dual-Luciferase Reporter Assay system (Promega) was used to measure luciferase activity. Differences in relative luciferase activity were compared between the major and minor alleles of each SNP using Student’s t-test. P values < 0.05 were regarded as statistically significant. Each figure shows representative data from experiments performed independently three times. The data in the figures represent the mean ± standard deviation of triplicate assays in a single experiment.

**References**

1. Kaplan, M. M. & Gershwin, M. E. Primary biliary cholangitis. *N. Engl. J. Med.* 353, 1261–1273 (2005).
2. Selmi, C., Bowls, C. L., Gershwin, M. E. & Coppel, R. L. Primary biliary cirrhosis. *Lancet* 377, 1600–1609 (2011).
3. Gershwin, M. E. & Mackay, I. R. The causes of primary biliary cirrhosis: convenient and inconvenient truths. *Hepatology* 47, 737–745 (2008).
4. Shimoda, S. et al. Natural killer cells regulate T cell immune responses in primary biliary cirrhosis. *Hepatology* 62, 1817–1827 (2015).
5. Jones, D. E., Watt, F. E., Metcalf, J. V., Bassendine, M. F. & James, O. F. Familial primary biliary cholangitis reassessed: a geographically based population study. *J. Hepatol.* 30, 402–407 (1999).
6. Selmi, C. et al. Primary biliary cholangitis in monozygotic and dizygotic twins: genetics, epigenetics, and environment. *Gastroenterology* 127, 485–492 (2004).
7. Hirschfeld, G. M. et al. Primary biliary cholangitis associated with HLA, IL12A, and IL12RB2 variants. *N. Engl. J. Med.* 360, 2544–2555 (2009).
8. Hirschfeld, G. M. et al. Variants at IRF5-TNP3, 17q12-21 and MME1 are associated with primary biliary cholangitis. *Nat. Genet.* 42, 655–657 (2010).
9. Liu, X. et al. Genome-wide meta-analyses identify three loci associated with primary biliary cholangitis. *Nat. Genet.* 42, 658–660 (2010).
10. Mells, G. F. et al. Genome-wide association study identifies 12 new susceptibility loci for primary biliary cholangitis. *Nat. Genet.* 43, 329–332 (2011).
11. Hirschfeld, G. M. et al. Association of primary biliary cholangitis with variants in the CLEC16A, SOCS1, SPIB and SIAE immunomodulatory genes. *Gastroenterology* 133, 328–335 (2012).
12. Liu, J. Z. et al. Dense fine-mapping study identifies new susceptibility loci for primary biliary cholangitis. *Nat. Genet.* 44, 1137–1141 (2012).
13. Juran, B. D. et al. Immunochip analyses identify a novel risk locus for primary biliary cirrhosis at 13q14, multiple independent associations at four established risk loci and epistasis between 1p31 and 7q32 risk variants. *Hum. Mol. Genet.* 21, 5209–5221 (2012).
14. Cordell, H. J. et al. International genome-wide meta-analysis identifies new primary biliary cholangitis risk loci and targetable pathogenic pathways. Nat. Commun. 6, 8019 (2015).
15. Nakamura, M. et al. Genome-wide association study identified TNFSF15 and POU2AF1 as susceptibility locus for primary biliary cholangitis in the Japanese population. Am. J. Hum. Genet. 91, 721–728 (2012).
16. Kawashima, M. et al. Genome-wide association study identified PRKCB as a genetic susceptibility locus for primary biliary cholangitis in a Japanese population. Hum. Mol. Genet. 26, 650–659 (2017).
17. Qu, F. et al. A genome-wide association study identifies six novel risk loci for primary biliary cholangitis. Nat. Commun. 8, 14828 (2017).
18. Welter, D. et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. 42, D1001–D1006 (2014).
19. Montgomery, S. B. & Dermitzakis, E. T. From expression QTLs to personalized transcriptomics. Nat. Rev. Genet. 12, 277–282 (2011).
20. Cheung, V. G. et al. Polymorphic cis- and trans-regulation of human gene expression. PLoS Biol. 8, e1000480 (2010).
21. Claussnitzer, M. et al. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. N. Engl. J. Med. 373, 895–907 (2015).
22. Hitomi, Y. et al. Identification of the functional driving variant ORMDL3 and GSDMB expression in human chromosome 17q12-21 in primary biliary cholangitis. Sci. Rep. 7, 29041 (2017).
23. Cantero-Recasens, G., Fandos, C., Rubio-Moscardo, F., Valverde, M. A. & Vicente, R. The asthma-associated ORMDL3 gene product regulates endoplasmic reticulum-mediated calcium signaling and cellular stress. Hum. Mol. Genet. 19, 111–121 (2010).
24. Kawai, Y. et al. Japonica array: improved genotype imputation by designing a population-specific SNP array with 1070 Japanese individuals. J. Hum. Genet. 60, 581–587 (2015).
25. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat. Genet. 45, 580–585 (2013).
26. Boyle, A. P. et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 22, 1790–1797 (2012).
27. Kent, W. J. et al. The human genome browser at UCSC. Genome Res. 12, 996–1006 (2002).
28. Wingender, E., Dietze, P., Karas, H. & Knüppel, R. TRANSFAC: a database on transcription factors and their DNA binding sites. Nucleic Acids Res. 24, 238–241 (1996).
29. Rebhan, M., Chalifa-Caspi, V., Prilusky, J. & Lancet, D. GeneCards: integrating information about genes, proteins and diseases. Trends Genet. 13, 163 (1997).
30. Acar, M. et al. Rumi is a CAP10 domain glycosyltransferase that modifies Notch and is required for Notch signaling. Cell 132, 247–258 (2008).
31. Moloney, D. J. et al. Mammalian Notch1 is modified with two unusual forms of O-linked glycosylation found on epidermal growth factor-like modules. J. Biol. Chem. 275, 9604–9611 (2000).
32. Radilke, F., MacDonald, H. R. & Tacchini-Cottier, F. Regulation of innate and adaptive immunity by Notch. Nat. Rev. Immunol. 13, 427–437 (2013).
33. Li, L. et al. Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1. Nat. Genet. 16, 243–251 (1997).
34. Oda, T. et al. Mutations in the human Jagged1 gene are responsible for Alagille syndrome. Nat. Genet. 16, 235–242 (1997).
35. McDanieli, R. et al. NOTCH2 mutations cause Alagille syndrome, a heterogeneous disorder of the notch signaling pathway. Am. J. Hum. Genet. 79, 169–173 (2006).
36. Thakurdas, S. M. et al. Jagged1 heterozygosity in mice results in a congenital cholangiopathy which is reversed by concomitant deletion of one copy of Poglut (Rumi). Hepatology 63, 550–565 (2016).
37. Dubois, P. C. et al. Multiple common variants for celiac disease influencing immune gene expression. Nat. Genet. 42, 295–302 (2010).
38. International Multiple Sclerosis Genetics Consortium et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature 476, 214–219 (2011).
39. Yang, W. et al. Meta-analysis followed by replication identifies loci in or near CDKN1B, TET3, CD80, DRAM1, and ARID5B as associated with systemic lupus erythematosus in Asians. Am. J. Hum. Genet. 92, 41–51 (2013).
40. Jin, Y. et al. Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants. Nat. Genet. 48, 1418–1424 (2016).
41. Hitomi, Y. et al. Human primary biliary cirrhosis-susceptible allele of rs4979462 enhances TNFSF15 expression by binding NF-1. Hum. Genet. 134, 737–747 (2015).
42. Loh, P. R. et al. Reference-based phasing using the Haplotype Reference Consortium panel. Nat. Genet. 48, 1443–1448 (2016).
43. Bycroft, C. et al. Genome-wide genetic data on ~500,000 UK Biobank participants. bioRxiv, 166298 (2017).
44. Nagasaki, M. et al. Rare variant discovery by deep whole-genome sequencing of 1,070 Japanese individuals. Nat. Commun. 6, 8018 (2015).
45. Yamaguchi-Kabata, Y. et al. iJGVD: an integrative Japanese genome variation database based on whole-genome sequencing. Hum. Genome Var. 2, 15050 (2015).
46. Staples, J. et al. PRIMUS: rapid reconstruction of pedigrees from genome-wide estimates of identity by descent. Am. J. Hum. Genet. 95, 553–564 (2014).

Acknowledgements
We would like to thank all patients and volunteers who enrolled in the study. We also thank Drs Nobuoyoshi Fukushima, Yukio Ohara, Toyokichi Muro, Eiichi Takesaki, Hitoshi Takaki, Tetsuo Yamamoto, Michio Kato, Yuko Nakagaki, Shigeki Hayashi, Koichi Honda, Jinnya Ishida, Yukio Watanabe, Masakazu Kobayashi, Michiaki Koga, Takeo Saoshiro, Michiyasu Yamagata, Keisuke Hirata (Members of PBC Research in the NHO Study Group for Liver Disease in Japan (NHOSLJ)) for collecting clinical data and blood samples, and for obtaining informed consent from PBC cases. We also thank Professors Takafumi Ichida, Hirohito Tsuobuchi, Kazuaki Chayama, Morikazu Onji, Kazuhide Yamamoto, Masashi Mizokami, and Hiromi Ishibashi (Directors and/or Councillors in The Japan Society of Hepatology) for helpful comments and discussion. We are also grateful to Ms. Mayumi Ishii, Takayo Tsuchiura (National Center for Global Health and Medicine), Ms. Natsumi Baba, Ms. Yoshimi Shigemori, Ms. Tomoko Suzuki (The University of Tokyo), for their technical and administrative assistance. This work was supported by Grants-in-Aid for Scientific Research from the Japan Society for the Promotion of Science to Yuki Hitomi (15K19314, 17K06924), Minae Kawashima (#15K19357, 17K09449), and Minoru Nakamura (#23391006, #26293181), a Grant-in-Aid for Clinical Research from the National Hospital Organization to Minoru Nakamura, a grant from the Research Program of Intractable Disease provided by the Ministry of Health, Labor, and Welfare of Japan to Minoru Nakamura, a grant from the Platform Program for the Promotion of Genome Medicine (16 km 0405205h0101) from the Japan Agency for Medical Research and Development to Katsushi Tokunaga and Masao Nagasaki, a grant from the Takeda Foundation to Yuki Hitomi, and a grant from the Uehara Memorial Foundation to Yuki Hitomi.
Author Contributions

Y.H. and K.U. wrote the main manuscript text and made the Tables and Figures. Y. Kawai, N. Nishida, K.K., M. Kawashima, Y.A., M. Nagasaki, M. Nakamura and K.T. contributed to materials of the research and reviewed the manuscript. H. Nakamura, Hiroshi Kouno, Hirotaka Kouno, H. Ota, K.S., T.N., T.Y., S. Katsushima, T. Komeda, K. Ario, A.N., M. Shimada, N.H., K. Yoshizawa, F.M., K.F., M. Kikuchi, N. Naeshiro, H. Takahashi, Y. Mano, H. Yamashita, K. Matsushita, S.T., I.Y., H. Nishimura, Y.S., K. Yamauchi, T. Komatsu, R.S., H. Sakai, E.M., M. Koda, Y.N., H. Kamitsukasa, T.S., M. Nakamuta, N.M., H. Takikawa, A. Tanaka, H.O., M.Z., M.A., S. Kaneko, M. Honda, K. Arai, T.A.H., E.H., M.T., T.U., S.J., K.N., T.I., H. Shibata, A. Takaki, S.Y., M. Seike, S. Sakisaka, Y.T., M. Harada, M. Senju, O.Y., T. Kanda, Y.U., H.E., T.H., K. Murata, S. Shimoda, S.N., S.A., A.K., K. Migita, M.I., H. Yatsuhashi, Y. Maehara, S.U. and N.K. contributed to collecting DNA samples.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-36490-1.

Competing Interests: The authors declare no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019