We performed a multistage genome-wide association study including 7,683 individuals with pancreatic cancer and 14,397 controls of European descent. Four new loci reached genome-wide significance: rs6971499 at 7q32.3 (LINC-PINT, per-allele odds ratio (OR) = 0.79, 95% confidence interval (CI) 0.74–0.84, \( P = 3.0 \times 10^{-12} \)), rs7190458 at 16q23.1 (BCAR1/CTRB1/CTRB2, OR = 1.46, 95% CI 1.30–1.65, \( P = 1.1 \times 10^{-10} \)), rs9581943 at 13q12.2 (PDX1, OR = 1.15, 95% CI 1.10–1.20, \( P = 2.4 \times 10^{-9} \)) and rs16986825 at 22q12.1 (ZNRF3, OR = 1.18, 95% CI 1.12–1.25, \( P = 1.2 \times 10^{-8} \)). We identified an independent signal in exon 2 of TERT at the established region 5p15.33 (rs2736098, OR = 0.80, 95% CI 0.76–0.85, \( P = 9.8 \times 10^{-14} \)). We also identified a locus at 8q24.21 (rs1561927, \( P = 1.3 \times 10^{-7} \)) that approached genome-wide significance located 455 kb telomeric of PVT1. Our study identified multiple new susceptibility alleles for pancreatic cancer that are worthy of follow-up studies.

Pancreatic cancer is the fourth-leading cause of cancer-related death in the United States and the fifth-leading cause in the European Union1,2. Over 80% of patients have incurable disease at the time of diagnosis, and the majority live for less than 12 months3. Rare, moderately to highly penetrant mutations account for a small fraction of the familial aggregation of pancreatic cancer4. In two previous genome-wide association studies (GWAS) called PanScan I5 and PanScan II6, we identified common variants at four loci that are associated with risk of sporadic pancreatic cancer in European populations. Subsequent GWAS demonstrated five distinct susceptibility loci among individuals of Chinese descent7 and three suggestive loci among individuals of Japanese descent8.

In the current study (designated PanScan III), we performed a multistage GWAS of 7,683 individuals diagnosed with pancreatic cancer and 14,397 control individuals of European descent (Online Methods, Table 1, Supplementary Table 1 and Supplementary Fig. 1).
Table 1: Subject numbers and characteristics of pancreatic cancer cases and controls

| Number of subjects     | Cases n (%) | Controls n (%) |
|------------------------|-------------|----------------|
| Stage 1                | 1,582       | 5,203          |
| Stage 2                | 3,525       | 3,642          |
| Replication            | 2,576       | 5,552          |
| Full study population  | 7,683       | 14,397         |

| Geographic region      | Cases (n) (%) | Controls (n) (%) |
|------------------------|---------------|------------------|
| United States          | 4,387 (57.1)  | 7,962 (55.3)     |
| Central or northern Europe | 2,264 (29.5)  | 3,853 (26.8)     |
| Southern Europe        | 1,032 (13.4)  | 2,582 (17.9)     |

| Sex      | Cases n (%) | Controls (n) (%) |
|----------|-------------|------------------|
| Male     | 4,107 (53.5) | 8,841 (61.4)     |
| Female   | 3,576 (46.5) | 5,556 (38.6)     |

| Age, years | Cases (n) (%) | Controls (n) (%) |
|------------|---------------|------------------|
| <60        | 1,972 (25.7)  | 4,577 (25.7)     |
| 61–70      | 2,688 (35.0)  | 5,906 (35.0)     |
| >70        | 3,023 (40.3)  | 8,360 (53.2)     |

| Smoking statusa | Cases n (%) | Controls (n) (%) |
|-----------------|-------------|------------------|
| Current or past | 2,634 (56.1) | 4,541 (51.3)     |
| Never           | 1,642 (33.2) | 3,186 (36.0)     |
| Unknown         | 831 (16.3)   | 1,118 (12.6)     |

| Smoking status information was available for subjects in stages 1 and 2.

In stage 1, we newly genotyped 1,582 cases from 13 prospective cohort studies, 2 case series and 1 case-control study using the Illumina OmniExpress Beadchip array. The control population included 5,203 cancer-free individuals that were genotyped previously using second-generation Illumina SNP microarrays (for example, OmniExpress, Omni 1M or Omni 2.5M) and were drawn from PanScan III prospective cohorts and a Spanish case-control study of bladder cancer. Of the newly genotyped cases, 94% passed quality-control criteria (Online Methods and Supplementary Tables 2 and 3), and 712,704 SNPs were included with a minimum call rate of 94%. In stage 2, we used the primary whole genome–scan data from the reported PanScan II (1,757 cases and 1,801 controls from 12 cohort studies and 1 case-control study typed on the Illumina HumanHap550 array) and PanScan II (1,768 cases and 1,841 controls from 8 case-control studies typed with the Illumina Human 610-Quad array) studies. To address differences in typed SNPs across the arrays, we used the Division of Cancer Epidemiology and Genetics (DCEG) Imputation Reference Set9 to fill in missing genotypes (Online Methods).

In a meta-analysis of stages 1 and 2, we observed robust associations for the four previously identified loci in individuals of European descent: rs687289 at 9q34.2 (ABO, OR = 1.27, 95% CI 1.20–1.35, P = 1.6 × 10−6), rs9543325 at 13q22.1 (KLFS/KLF12, OR = 1.23, 95% CI 1.18–1.30, P = 4.3 × 10−10), rs10919791 at 1q32.1 (NRAS, OR = 0.79, 95% CI 0.75–0.85, P = 1.4 × 10−11) and rs31490 at 9p13.3 (CLPTM1L, OR = 1.20, 95% CI 1.14–1.27, P = 2.0 × 10−11).

We observed two new SNPs below genome-wide significance (P < 5 × 10−8) in the meta-analysis of stages 1 and 2, plus 11 additional promising SNPs (P < 5 × 10−7) from distinct regions (Supplementary Table 4). We carried these 13 SNPs forward for replication (stage 3) in 2,576 cases and 5,552 controls, drawn from (i) stages 1 and 2 with DNA quantity insufficient for the full GWAS, (ii) cases and controls.

Table 2: Association results for five new pancreatic cancer susceptibility loci and one suggestive locus

| Chr., Nearest gene(s)a SNP Positionb Minor allelec Major alleled Stage | Allelic OR (95% CI) | Minor allele frequency |
|--------------------------------------------------------|----------------------|------------------------|

| 5p15.33 | TERT, MIR4457, CLPTM1L rs2736098 1,294,086 T C Stage 1 | 0.76 (0.68–0.86) | 0.268 |
|---------|--------------------------------------------------|------------------|--------|
|         | Stage 2                                           | 0.82 (0.74–0.90) | 0.284  |
|         | Replication                                       | 0.81 (0.74–0.89) | 0.296  |
|         | Combined                                          | 0.80 (0.76–0.85) | 0.298  |
| 7q32.3  | LINC-PINT rs6971499 130,680,521 C T Stage 1         | 0.79 (0.68–0.90) | 0.155  |
|         | Stage 2                                           | 0.81 (0.74–0.90) | 0.147  |
|         | Replication                                       | 0.77 (0.69–0.86) | 0.147  |
|         | Combined                                          | 0.79 (0.74–0.84) | 0.147  |
| 16q23.1 | BCAR1, CTRB1, CTRB2 rs7190458 75,263,661 A G Stage 1| 1.61 (1.32–1.96) | 0.042  |
|         | Stage 2                                           | 1.47 (1.20–1.82) | 0.039  |
|         | Replication                                       | 1.33 (1.10–1.61) | 0.040  |
|         | Combined                                          | 1.46 (1.30–1.65) | 0.039  |
| 13q12.2 | PDX1 rs9581943 28,493,997 A G Stage 1               | 1.23 (1.12–1.35) | 0.397  |
|         | Stage 2                                           | 1.12 (1.05–1.20) | 0.406  |
|         | Replication                                       | 1.11 (1.03–1.20) | 0.406  |
|         | Combined                                          | 1.15 (1.10–1.20) | 0.406  |
| 22q12.1 | ZNRF3 rs16986825 29,300,306 T C Stage 1             | 1.25 (1.10–1.42) | 0.150  |
|         | Stage 2                                           | 1.15 (1.05–1.26) | 0.149  |
|         | Replication                                       | 1.18 (1.08–1.30) | 0.149  |
|         | Combined                                          | 1.18 (1.12–1.25) | 0.149  |
| 8q24.21 | MIR1208, PVT1 rs1561927 129,568,078 C T Stage 1     | 0.88 (0.78–0.97) | 0.269  |
|         | Stage 2                                           | 0.86 (0.78–0.93) | 0.279  |
|         | Replication                                       | 0.89 (0.82–0.97) | 0.279  |
|         | Combined                                          | 0.87 (0.83–0.92) | 0.279  |

Results are shown from an unconditional logistic regression analysis of the genotypes generated in stage 1, stage 2 and the replication (a total of 7,683 individuals diagnosed with pancreatic cancer and 14,397 controls).

aClosest RefSeq gene(s). Genes located within 25 kb of the given SNP are listed in order of the closest gene to those further away for 5p15.33, 7q32.3, 16q23.1, 13q12.2 and 22q12.1; the closest genes outside this 50 kb window are shown for 8q24.21. bPosition of the SNP in NCBI genome build 37 (Hg19). cMinor and major alleles. dThe replication is a meta-analysis of three groups and, thus, minor allele frequency (MAF) is not listed. eNumber of case and control subjects in the joint analysis of stage 1, stage 2 and the replication: rs2736098 (7,199/13,121), rs7190458 (7,412/13,286), rs9581943 (7,415/13,160) and rs1561927 (7,486/13,274). fDegree-of-freedom score test. Chr., chromosome and band; OR, per-allele OR for the minor allele adjusted for age, sex, geographic region and significant principal components for stage 1; per-allele OR adjusted for age, sex, study, arm and significant principal components for stage 2; per-allele OR adjusted for age, sex and study for the replication. Text in bold indicates results from the combined meta-analysis.
from the PANDoRA consortium and (iii) cases enrolled in CALGB 80303, a US cooperative group clinical trial (Supplementary Table 5). We selected additional control subjects from cancer-free individuals genotyped previously using the Illumina HumanHap550 array (Online Methods). Of the 13 SNPs advanced to replication, 9 were associated with pancreatic cancer risk (P < 0.05) in the replication stage (Supplementary Table 6).

For the complete study of 7,683 cases and 14,397 controls, we applied a fixed-effect meta-analysis to the results from the three stages. Overall, six SNPs had P values below genome-wide significance: rs2736098 at 5p15.33 (a second signal in TERT, P = 9.8 × 10^(-14)), rs6971499 at 7q32.3 (LINC-PINT, P = 3.0 × 10^(-12)), rs7190458 at 16q23.1 (BCAR1/CTRBI/CTRB2, P = 1.1 × 10^(-10)), rs9581943 at 13q12.2 (PDX1, P = 2.4 × 10^(-6)), rs16986825 at 22q12.1 (ZNRF3, P = 1.2 × 10^(-8)) (Table 2 and Fig. 1) and rs4962153 at 9q34.2 (ADAMTS13, P = 1.5 × 10^(-8)). In a subsequent conditional analysis described below, rs4962153 in ADAMTS13 marked the same signal as rs687289 in ABO identified in PanScan I and II. An additional locus at 8q24.21 was close to genome-wide significance (rs1561927, 10^(-7)) and is located in a region that has been associated previously with multiple cancers (Table 2 and Fig. 1).

The SNP at 7q32.3, rs6971499, maps to an intron in LINC-PINT, which is a p53-induced long intergenic non–protein coding RNA located in a 375-kb region between MKLN1 (encoding muskelin 1) and KLF14 (Supplementary Table 7 and Fig. 1). Muskelin 1 is an intracellular protein that mediates cell responses to the extracellular matrix, in particular influencing cell adhesion and cytoskeleton organization. KLF14 is a member of the Kruppel-like family of transcription factors, which have been implicated as tumor suppressors, including in mutant KRAS-driven tumors. KLF14 has also been identified as a regulator of several metabolic phenotypes, including type 2 diabetes. Notably, the previously established susceptibility locus at 13q22.1 is located in an intergenic region between KLF5 and KLF12, which encode two other members of the Kruppel-like family of transcription factors.

The SNP at 16q23.1, rs7190458, is a synonymous SNP residing in the last exon of BCAR1 (also known as P130Cas) and close to two chymotrypsinogen genes, CTRBI (5 kb) and CTRB2 (23 kb) (Supplementary Table 7 and Fig. 1). Aberrant expression of BCAR1 has been linked with the transformation and progression of multiple cancer types, and BCAR1 functions as an adaptor protein that coordinates cell cycle control, cytoskeleton organization and cell migration.

The SNP at 13q12.2, 178.2-kb region, is a second signal within the 13q22.1 region on the basis of five unique sets of 100 randomly selected control samples. Bottom, LD heat maps based on R^2 values from the total control populations for all SNPs included in the GWAS. The data are based on a total of 7,683 individuals with pancreatic cancer and 14,397 controls of European descent. Shown are results for 5p15.33 (a), 7q32.3 (b), 16q23.1 (c), 13q12.2 (d), 22q12.1 (e) and 8q24.21 (f).
The chymotrypsinogens are members of a family of serine proteases that are secreted by the pancreas into the gastrointestinal tract. Mutations in the related genes PRSSI1 (encoding trypsin 1) and CTRC have been associated with hereditary pancreatitis, a known risk factor for pancreatic cancer. In addition, a susceptibility locus for types 1 and 2 diabetes is located 16 kb centromeric to rs7190458 (rs7202877, \(r^2 = 0.32\) in the 1000 Genomes Project [1000G] CEU data). Functional analyses have indicated that this variant (rs7202877) leads to impaired pancreatic beta-cell function and influences the expression of CTRB1 and CTRB2 in pancreatic tissue.

At chromosome 13q12.2, the newly identified SNP, rs9581943, is approximately 200 bp upstream of PDX1 (encoding pancreatic and duodenal homeobox 1 protein) and is intronic to PDX1-AS1 (encoding PDX1 antisense RNA 1), a recently identified noncoding RNA (Supplementary Table 7 and Fig. 1). PDX1 is critical for early pancreatic development, has a role in the differentiation of the exocrine pancreas and regulates beta-cell function in the mature pancreas. Mutations in PDX1 have been linked to agenesis of the pancreas and maturity onset diabetes of the young, a dominantly inherited disorder of nonautoimmune diabetes. Furthermore, PDX1 has been implicated in glucose-dependent regulation of insulin gene transcription, and GWAS have identified a SNP (rs2293941, \(P = 0.20\) in the 1000G CEU data) at the PDX1 locus that is associated with fasting glucose levels.

The signal at 22q12.1, rs16986825, maps to an intron in ZNRF3 (zinc and ring finger 3) (Supplementary Table 7 and Fig. 1), encoding a cell-surface transmembrane E3 ubiquitin protein ligase that is a negative regulator of the Wnt signaling pathway. CHEK2 is located 162 kb centromeric to the marker SNP and encodes a cell cycle-checkpoint kinase that cooperates with P53, BRCA1 and ATM in response to DNA damage. Alterations in CHEK2 have been implicated in susceptibility to several cancer types.

We performed conditional analyses to assess whether the newly identified SNPs at 5p15.33 (CLPTM1L and TERT) and 9q34.2 (ABO and ADAMTS13) were independent from those identified previously. After conditioning on the reported SNP within intron 13 of CLPTM1L, the newly identified synonymous SNP within the second exon of TERT (rs2736098) remained statistically significant (\(P = 2.4 \times 10^{-3}\)) (Table 3). Two strong recombination hotspots lie between the established and new SNPs in the 1000G CEU data (likelihood ratios of 27.1 and 261.0) and the two SNPs are in modest linkage disequilibrium (LD; \(r^2 = 0.22\) in the 1000G CEU data) (Fig. 1).

TERT encodes the catalytic subunit of telomerase reverse transcriptase, a component of the ribonucleoprotein complex that maintains the integrity of chromosome ends. Inherited mutations affecting TERT underlie cases of dyskeratosis congenita, aplastic anemia, acute myeloid leukemia, familial melanoma and pulmonary fibrosis. CLPTM1L encodes the cleft lip and palate associated transmembrane 1-like protein that is involved in mediating apoptosis, aneuploidy, cisplatin resistance and RAS-mediated malignant transformation.

Variants across the TERT and CLPTM1L region have been associated previously with risk of multiple cancers. Furthermore, independent signals within TERT and CLPTM1L have been identified for bladder cancer, chronic lymphocytic leukemia and lung cancer, and fine-mapping studies have identified at least four independent signals across the TERT and CLPTM1L region that are associated with cancer. The new SNP identified in PanScan III (rs2736098) is located in a region of LD spanning ~4 kb from the promoter region to exon 2 of TERT. This SNP and several correlated SNPs have been associated with telomere length in white blood cells, as well as TERT promoter activity. The minor allele of rs2736098 that is associated with...
a lower risk of pancreatic cancer in PanScan is also associated with longer telomeres and lower risk of breast cancer.\textsuperscript{40} Although further characterization of this region will be necessary, the new SNP in exon 2 of \textit{TERT} appears to mark an independent risk locus for pancreatic cancer.

After conditioning on the established SNP at 9q34.2 in the first intron of \textit{ABO}, the SNP rs4962153 in \textit{ADAMTS13} identified in PanScan III was not statistically significant (\(P = 0.28\)), indicating that these two SNPs point to the same susceptibility haplotype (Table 3).

We identified a promising risk locus at 8q24.21 (rs1561927, \(P = 1.3 \times 10^{-7}\)) in a nongenic region between \textit{PVT1} and \textit{LINC00977} (Supplementary Table 7 and Fig. 1). 8q24.21 is known to contain multiple cancer susceptibility loci that span over 2 Mb\textsuperscript{42-43}. The promising pancreatic cancer SNP is in LD with a SNP that is associated with ovarian cancer risk (rs10088218, \(r^2 = 0.37\) in the 1000G CEU data, 24 kb upstream)\textsuperscript{14}, and the closest genes are centromeric to rs1561927: \textit{MIR1208} (406 kb), \textit{PVT1} (455 kb) and \textit{MYC} (814 kb). Several 8q24.21 risk loci have been shown to work with \textit{MYC} or \textit{PVT1} promoters through long-range chromosomal interaction, and allele-specific effects on the expression of both genes have been reported.\textsuperscript{42,45-46} However, these loci are located more than 1 Mb upstream of rs1561927 on 8q24.21 (\(r^2 < 0.03\) in the 1000G CEU data).

In stratified analyses, we noted no statistically significant heterogeneity by geographic region or smoking status (Supplementary Tables 8 and 9). In a preliminary analysis that included 173 cases and 430 controls of Asian ancestry (Supplementary Table 10), we examined the susceptibility loci identified in individuals of European descent\textsuperscript{5,6} (Table 2). We also assessed previously published pancreatic cancer risk loci from individuals of Chinese\textsuperscript{7} and Japanese\textsuperscript{8} ancestry, noting no loci and one locus, respectively, as being nominally statistically significant in PanScan (Supplementary Table 11).

To pursue the first steps toward understanding the functional underpinnings of the newly identified risk alleles, we conducted bioinformatic analyses using HaploReg\textsuperscript{47} (Supplementary Table 12). We also evaluated expression quantitative trait locus (eQTL) effects\textsuperscript{48-50} (Supplementary Table 12). We noted \textit{cis}-eQTLs on chromosome 16q23.1 in peripheral blood (\textit{CFDP1}), chromosome 13q12.2 in skin and liver (\textit{POMP}), chromosome 22q12.1 in liver and peripheral blood (\textit{CCDC117}) and peripheral blood (\textit{XBPI}) and chromosome 8q24.21 in adipose tissue (\textit{PVT1}). \textit{XBPI} at chromosome 22q12.1 regulates pancreatic beta-cell function with effects on systemic glucose control\textsuperscript{51} and modulates acinar cell homeostasis.\textsuperscript{52} In a gene set enrichment analysis\textsuperscript{53} of genes within 100 kb of the ten index SNPs identified in PanScan, the only statistically significant pathway was maturity onset diabetes of the young (\(P < 3 \times 10^{-4}\)). Understanding the functional consequences of pancreatic cancer susceptibility variants will require further laboratory investigation.

In a linear-mixed model analysis\textsuperscript{53} (Online Methods), we estimated that the heritability for pancreatic cancer due to common SNPs present on GWAS arrays was 13\% (95\% CI 4–22\%). Furthermore, we estimated that the nine loci identified in individuals of European ancestry account for approximately 9\% of the total heritability tagged by common SNPs. We also evaluated the cumulative association with pancreatic cancer of risk alleles at susceptibility loci identified in individuals of European descent. Compared to individuals with the most prevalent number of risk alleles in controls (\(n = 10\)), those with \(\leq 6\) risk alleles had an OR of 0.55 (95\% CI 0.44–0.68) and those with \(\geq 14\) risk alleles had an OR of 2.24 (95\% CI 1.80–2.80) for pancreatic cancer (Supplementary Fig. 2).

In conclusion, our multistage GWAS revealed new loci that are associated with pancreatic cancer risk, as well as promising loci that merit follow-up study. Several of the new loci harbor plausible candidate genes that have been implicated in pancreas development, pancreatic beta-cell function and predisposition to diabetes. Further investigation is warranted to understand the biological underpinnings of these common pancreatic cancer susceptibility alleles.

**METHODS**

Methods and any associated references are available in the online version of the paper.

**ACKNOWLEDGMENTS**

This project was funded in whole or in part with federal funds from the National Cancer Institute (NCI), US National Institutes of Health (NIH) under contract number HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the US Department of Health and Human Services, and mention of trade names, commercial products or organizations does not imply endorsement by the US government. Major support for PanScan III sample identification and processing was provided by the Lustgarten Foundation for Pancreatic Cancer Research. Additional support was received from NIH/NCI R01 CA140790, the American Society of Clinical Oncology Conquer Cancer Foundation, the Howard Hughes Medical Institute, the Lustgarten Foundation, the Robert T and Judith B Hale Fund for Pancreatic Cancer Research and Promises for Purple to B.M.W. A full list of acknowledgments for each participating study is provided in the Supplementary Note.

**AUTHOR CONTRIBUTIONS**

B.M.W., C.R., P.K., C.K., G.M.P., P.H., C.F., S.J.C., R.S.S.-S. and L.T.A. conducted the epidemiological studies and/or follow-up genotyping. All authors contributed samples to the GWAS and/or follow-up genotyping. All authors participated in the writing of the manuscript.

**COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Siegel, R., Ma, J., Zou, Z. & Jemal, A. Cancer statistics, 2014. \textit{CA Cancer J. Clin.} \textbf{64}, 9–29 (2014).
2. Ferlay, J. et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. \textit{Eur. J. Cancer} \textbf{49}, 1374–1403 (2013).
3. Hidalgo, M. Pancreatic cancer. \textit{N. Engl. J. Med.} \textbf{362}, 1605–1617 (2010).
4. Abarca, H. A.P. Genetic susceptibility to pancreatic cancer. \textit{Mol. Carcinog.} \textbf{51}, 14–24 (2012).
5. Amundadottir, L. et al. Genome-wide association study identifies variants in the \textit{ABO} locus associated with susceptibility to pancreatic cancer. \textit{Nat. Genet.} \textbf{41}, 986–990 (2009).
6. Petersen, G.M. et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. \textit{Nat. Genet.} \textbf{42}, 224–228 (2010).
7. Wu, C. et al. Genome-wide association study identifies five loci associated with susceptibility to pancreatic cancer in Chinese populations. \textit{Nat. Genet.} \textbf{44}, 62–66 (2012).
8. Low, S.K. et al. Genome-wide association study of pancreatic cancer in Japanese population. \textit{PLoS ONE} \textbf{5}, e11824 (2010).
9. Wang, Z. et al. Improved imputation of common and uncommon SNPs with a new reference set. \textit{Nat. Genet.} \textbf{44}, 6–7 (2012).
10. Dopman, D. et al. Genetic susceptibility to pancreatic cancer and its functional characterisation: the \textit{Pancreatic Disease ReseArch} (PANDoRA) consortium. \textit{Dig. Liver Dis.} \textbf{45}, 95–99 (2013).

© 2014 Nature America, Inc. All rights reserved.
Stage 1: GWAS for PanScan III. We conducted a GWAS of pancreatic cancer using case and control subjects from 17 studies (Supplementary Table 1). Cases with pancreatic cancer included individuals newly identified from nine cohort studies that participated in PanScan II, as well as those from five new cohort studies, two new case series and one new case-control study. The new cohort studies included the Agricultural Health Study (AHS)54, the Melbourne Collaborative Cohort Study (MCCS)55, the Multiethnic Cohort Study (MEC)56, the Selenium and Vitamin E Cancer Prevention Trial (SELECT)57 and the Vitamins and Lifestyle Study (VITAL)58. The new case-based studies were the Gastrointestinal Cancer Clinic of Dana-Farber Cancer Institute (DFCI-GCC) study, the Spanish Pancreatic Cancer Study PANKRAS-II59 and the PANDORA-Heidelberg pancreatic cancer case-control study10. Cases were defined as those individuals having primary adenocarcinoma of the exocrine pancreas (ICD-O-3 code C250-C259). Those with non-exocrine pancreatic tumors (histology types 8150, 8151, 8153, 8155 and 8240) were excluded. Each participating study obtained informed consent from study participants, approval from its institutional review board (IRB) for this study and IRB certification permitting data sharing in accordance with the NIH Policy for Sharing of Data Obtained in NIH-Supported or NIH-Conducted Genome-Wide Association Studies.

All samples from cases with pancreatic cancer that had sufficient DNA (n = 1,894) were genotyped on the Illumina OmniExpress chip at the NCI Cancer Genomic Research Laboratory (CGR) (Supplementary Table 2). Genotypes were called using the Illumina Genomewide studio software. Genotype clusters for new cases were estimated using samples with a completion rate of 98% to optimize accuracy. Genotypes for all samples, including those initially excluded, were subsequently called on the basis of the optimized cluster file. Extensive quality-control metrics were applied to the data: SNPs with a call rate <94% or Hardy-Weinberg proportion P < 1 × 10^-7 were excluded (n = 18,765); samples with a call rate <94% (n = 78) and those with mean heterozygosity <26% or >33% (n = 2) based on autosomal SNPs or gender discordance (>5% heterozygosity based on the X chromosome SNPs for males or <20% heterozygosity based on the X chromosome SNPs for females, n = 5) were excluded. Unexpected duplicates (>99.9% concordance, n = 3) and first-degree relatives (n = 2, on the basis of identity-by-descent sharing with Pi-hat > 0.40) were removed. Quality-control duplicate samples in PanScan III (n = 38 pairs) showed >99.9% genotype concordance. Duplicates with PanScan I or II were removed (>99.9% concordance, n = 21). Ancestry was assessed using the Genotyping Library and Utilities (GLU) struct.admix module. Participants with <80% European ancestry (n = 199) were excluded for the primary analysis of individuals of European ancestry (Supplementary Fig. 3).

After exclusions, 1,382 cases of European ancestry were available for analysis (Supplementary Tables 2 and 3). Controls of 280% European ancestry were drawn from ten of the studies included in PanScan III (Alpha-Tocopherol, Beta-Carotene Lung Cancer Prevention Study (ATBC), American Cancer Society Cancer Prevention Study-II Cohort (CPS-II), European Prospective Investigation into Cancer and Nutrition (EPIC), Health Professionals Follow-Up Study (HPFS), Melbourne Collaborative Cohort Study (MCCS), Multiethnic Cohort Study (MEC), Nurses’ Health Study (NHS), Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO), Spanish Pancreatic Cancer Study (SPCS, PANKRAS-II and Spanish Bladder Cancer SBC/EPICURO studies)60 and Women’s Health Initiative (WHI)). These controls had no history of cancer, were not included in PanScan I or PanScan II and had been genotyped previously at CGR on the Illumina OmniExpress, Omni1M or Omni2.5M arrays with extensive quality-control metrics. Genotyping for cases in group (i) was performed using custom TaqMan genotyping assays (Applied Biosystems) at CGR. Genotyping for cases and controls from PANDORA (group ii) was performed in the same manner but at the German Cancer Research Center (DKFZ) in Heidelberg, Germany. Quality-control duplicate samples in the replication (CGR, n = 20 pairs; PANDORA, n = 912 pairs) showed >99.9% genotype concordance. Patients enrolled in CALBG/Alliance 80303 (group iii) were previously genotyped using the Illumina HumanHap550 Infinium II and the Human 610-Quad chips, respectively, whereas PanScan III was genotyped on the OmniExpress chip. As the number of overlapping SNPs between the three chips is moderate (∼300,000), imputation of missing genotypes was performed using phased haplotypes from the DCEG reference set and IMPUTE2 (refs. 9,67). The DCEG reference set is well designed for filling in missing genotypes across chip designs in PanScan, as it is based on several of the same studies included in PanScan, and the imputation accuracy is improved over 1,000G and HapMap data68. Imputed SNPs with low MAF (<0.01) or low-quality scores (IMPUTE2 information score <0.3) were removed before the association analysis. The same quality thresholds as described above for stage 1 were applied for stage 2. The final numbers of cases and controls included in stage 2 were 1,757 cases and 1,801 controls from PanScan I and 1,768 cases and 1,841 controls from PanScan II.

To combine the data from PanScan I, II and III, meta-analyses were performed using the fixed-effects inverse-variance method based on the β estimates and standard errors. No heterogeneity was observed across stages 1 and 2 for the SNPs identified as GWAS significant or suggestive in the full study (P heterogeneity ≥ 0.11; Supplementary Table 4). A Manhattan plot for the results of the meta-analysis of stage 1 and stage 2 is shown in Supplementary Figure 6.

Association analysis was also performed in 173 cases and 430 controls of Asian ancestry from the Shanghai Men’s and Women’s Health Study (SMWHS) (Supplementary Table 10). This analysis included case and control subjects from stages 1 and 2 of PanScan III and previously genotyped control subjects from SMWHS66. The quality-control methods and association analysis were performed as described above for the European ancestry subjects.

Stage 3: replication studies. Thirteen SNPs (P value threshold of <5 × 10^-5) were taken forward for de novo replication in an additional 2,576 cases and 5,552 controls. The replication samples were analyzed individually as three groups: (i) CGR: pancreatic cancer case and control subjects from CARET69; (ii) CALBG/Alliance 80303: cases from a randomized clinical trial of gemcitabine plus placebo versus gemcitabine plus bevacizumab11 and control subjects previously genotyped at CGR; (iii) PANDORA: case and control subjects from the PANDORA pancreatic cancer case-control consortium12 (no overlap with the PANDORA-Heidelberg cases genotyped in stage 1); and (iii) CALBG/Alliance 80303: cases from a randomized clinical trial of gemcitabine plus placebo versus gemcitabine plus bevacizumab11 and control subjects previously genotyped at CGR (Supplementary Table 5).

Genotyping for cases in group (i) was performed using custom TaqMan genotyping assays (Applied Biosystems) at CGR. Genotyping for cases and controls from PANDORA (group ii) was performed in the same manner but at the German Cancer Research Center (DKFZ) in Heidelberg, Germany. Quality-control duplicate samples in the replication (CGR, n = 20 pairs; PANDORA, n = 912 pairs) showed >99.9% genotype concordance. Patients enrolled in CALBG/Alliance 80303 (group iii) were previously genotyped using the Illumina HumanHap550v3 Genotyping BeadChip array12. Control subjects from PLCO previously genotyped at CGR using the Illumina HumanHap550v3 Genotyping BeadChip array were used for groups (i) and (iii) (Supplementary Table 5) and did not overlap with control subjects included in PanScan I, II or III. CALBG/Alliance 80303 and control genotypes were imputed to OmniExpress SNP content in the same manner as described above for stage 1. Quality-control thresholds and exclusions for samples and loci in the replication are listed in Supplementary Table 5. Association results for the replication studies were adjusted for age, sex and study, and a meta-analysis of the three replication groups was performed using the fixed-effects inverse-variance method based on the β estimates and standard errors.
(Supplementary Table 6). This was followed by a meta-analysis of stages 1 and 2 and the replication for the 13 SNPs using the same fixed-effects inverse-variance method.

Technical validation. A comparison of the genotyping calls from the imputation of PanScan I and II into the OmniExpress array contents and confirmatory TaqMan assays (n = 511 samples from PanScan I and II) yielded r² values of 0.74, 0.96, 0.56, 0.99, 0.98 and 1.00 for rs2736098, rs6971499, rs7190458, rs9581943, rs16986825 and rs1561927, respectively.

Estimate of recombination hotspots. To identify recombination hotspots, we used SequenceLDhot^32, a program that uses the approximate marginal likelihood method^71 and calculates likelihood ratio statistics at a set of possible hotspots. We tested five unique sets of 100 control samples. The PHASE v2.1 program was used to calculate background recombination rates^72,73, and LD heat maps were visualized using the snp.plotter program^74. For an estimation of recombination hotspots between loci in TERT and CLPTM1L on chromosome 5p15.33, we used the 1000G (version 3) CEU data.

Heritability analysis. To estimate the heritability explained by common SNPs present on GWAS arrays on the liability scale (lifetime disease risk of 0.015), we used GCTA^53,75 on a set of LD-pruned SNPs (r² < 0.5) that passed the following stringent quality-control thresholds: MAF > 1%, SNP missing rates <5%, subject missing rate <1% and Hardy-Weinberg equilibrium P values >10⁻⁴. Nonautosomal SNPs and pairs of subjects with genetic relatedness >5% were removed. These analyses were run separately in PanScan I, II and III adjusting for age, sex, study (or geographic region in PanScan III) and the significant principal components in each study. PanScan III analyses were restricted to participating studies that contributed both cases and controls. PanScan I, II and III results were combined in a meta-analysis. We repeated the analyses restricted to the genome-wide significant SNPs in individuals of European ancestry to estimate the proportion of heritability tagged by these nine SNPs.

Further follow-up analyses. We constructed a genetic risk score for pancreatic cancer incorporating the susceptibility loci identified in PanScan I, II and III. For this analysis, subjects could possess 0–20 risk alleles based on their genotypes at the 10 identified loci. ORs were calculated using multivariable-adjusted unconditional logistic regression with meta-analysis to combine data from stages 1 and 2, as was done in the analyses of individual SNPs. Replication samples were not genotyped for the four susceptibility loci identified in PanScan I and II, and therefore these subjects could not be included in the risk score analysis. Subjects with missing genotypes for one or more of the ten SNPs (n = 898) were assigned the most common genotype at that SNP among cases or controls. In sensitivity analyses, results were unchanged if these subjects were excluded. Using 1000G CEU data, we identified SNPs with r² > 0.7 with our lead SNP. We used HaploReg v2 (ref. 47), a tool for exploring noncoding functional annotation using ENCODE data, to evaluate the genome surrounding our identified SNPs (Supplementary Table 12). In addition, we evaluated cis associations between all new and promising SNPs discovered in this study and the expression of nearby genes in skin biopsies, adipose biopsies and nontransformed peripheral blood samples from subjects of European descent from publicly available data sets^48,50 (Supplementary Table 12). Gene set enrichment analysis was also performed for genes in pancreatic cancer risk loci identified in subjects of European descent (in a window of 100 kb centered on the most significant SNP in each locus) based on KEGG (Kyoto Encyclopedia of Genes and Genomes) annotations using GeneCodis3 with reporting of the corrected hypergeometric P value^52.

54. Alavonian, M.C. et al. Cancer and noncancer risk to women in agriculture and pest control: the Agricultural Health Study. J. Occup. Med. 36, 1247–1250 (1994).
55. Giles, G.G. & English, D.R. The Melbourne Collaborative Cohort Study. IARC Sci. Publ. 156, 69–70 (2002).
56. Kolonel, L.N. et al. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. Am. J. Epidemiol. 151, 346–357 (2000).
57. Lippman, S.M. et al. Designing the Selenium and Vitamin E Cancer Prevention Trial (SELECT). J. Natl. Cancer Inst. 97, 94–102 (2005).
58. White, E. et al. ViTamins And Lifestyle cohort study: study design and characteristics of supplement users. Am. J. Epidemiol. 159, 83–93 (2004).
59. Porta, M. et al. In pancreatic ductal adenocarcinoma blood concentrations of some organochlorine compounds and coffee intake are independently associated with KRAS mutations. Mutagenesis 24, 513–521 (2009).
60. Samanic, C. et al. Smoking and bladder cancer in Spain: effects of tobacco type, timing, environmental tobacco smoke, and gender. Cancer Epidemiol. Biomarkers Prev. 15, 1348–1354 (2006).
61. Berndt, S.I. et al. Genome-wide association study identifies multiple risk loci for chronic lymphocytic leukemia. Nat. Genet. 45, 868–876 (2013).
62. Rottmann, N. et al. A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. Nat. Genet. 42, 978–984 (2010).
63. Figueroa, J.D. et al. Genome-wide association study identifies multiple loci associated with bladder cancer risk. Hum. Mol. Genet. 23, 1387–1398 (2014).
64. De Vivo, I. et al. Genome-wide association study of endometrial cancer in EZ2C. Hum. Mutat. 33, 211–224 (2014).
65. Price, A.L. et al. Principal components analysis methods correct for stratification in genome-wide association studies. Nat. Genet. 38, 904–909 (2006).
66. de Bakker, P.I. et al. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum. Mol. Genet. 17, R122–R128 (2008).
67. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multirpoint method for genome-wide association studies by imputation of genotypes. Nat. Genet. 39, 906–913 (2007).
68. Wu, C. et al. Genome-wide association analyses of esophageal squamous cell carcinoma in Chinese identify multiple susceptibility loci and gene-environment interactions. Nat. Genet. 44, 1090–1097 (2012).
69. Thorquist, M.D. et al. Statistic design and monitoring of the Carotene and Retinol Efficacy Trial (CARET). Control. Clin. Trials 14, 308–324 (1993).
70. Landi, M.T. et al. A genome-wide association study of lung cancer identifies a region of chromosome 2q15 associated with risk for adenocarcinoma. Am. J. Hum. Genet. 85, 679–691 (2009).
71. Fearnhead, P., Harding, R.M., Schneider, J.A., Myers, S. & Donnelly, P. Application of coalescent methods to reveal fine-scale rate variation and recombination hotspots. Genetics 167, 2067–2081 (2004).
72. Li, N. & Stephens, M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics 165, 2213–2233 (2003).
73. Crawford, D.C. et al. Evidence for substantial fine-scale variation in recombination rates across the human genome. Nat. Genet. 36, 700–706 (2004).
74. Luna, A. & Nicos, K.K. SNP.plotter: an R-based SNP/haplotype association and linkage disequilibrium plotting package. Bioinformatics 23, 774–776 (2007).
75. Lee, S.H., Wray, N.R., Goddard, M.E. & Visscher, P.M. Estimating missing heritability for disease from genome-wide association studies. Am. J. Hum. Genet. 88, 294–305 (2011).