

Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer

Brian M Wolpin^{1,2,101}, Cosmeri Rizzato^{3,101}, Peter Kraft^{4,5,101}, Charles Kooperberg^{6,101}, Gloria M Petersen^{7,101}, Zhaoming Wang^{8,9}, Alan A Arslan^{10–12}, Laura Beane-Freeman⁸, Paige M Bracci¹³, Julie Buring^{14,15}, Federico Canzian³, Eric J Duell¹⁶, Steven Gallinger¹⁷, Graham G Giles^{18–20}, Gary E Goodman⁶, Phyllis J Goodman²¹, Eric J Jacobs²², Aruna Kamineni²³, Alison P Klein^{24,25}, Laurence N Kolonel²⁶, Matthew H Kulke¹, Donghui Li²⁷, Núria Malats²⁸, Sara H Olson²⁹, Harvey A Risch³⁰, Howard D Sesso^{4,14,15}, Kala Visvanathan³¹, Emily White^{32,33}, Wei Zheng^{34,35}, Christian C Abnet⁸, Demetrius Albanes⁸, Gabriella Andreotti⁸, Melissa A Austin³³, Richard Barfield⁵, Daniela Basso³⁶, Sonja I Berndt⁸, Marie-Christine Boutron-Ruault^{37–39}, Michelle Brotzman⁴⁰, Markus W Büchler⁴¹, H Bas Bueno-de-Mesquita^{42–44}, Peter Bugert⁴⁵, Laurie Burdette^{8,9}, Daniele Campa⁴⁶, Neil E Caporaso⁸, Gabriele Capurso⁴⁷, Charles Chung^{8,9}, Michelle Cotterchio^{48,49}, Eithne Costello⁵⁰, Joanne Elena⁵¹, Niccola Funel⁵², J Michael Gaziano^{14,15,53}, Nathalia A Giese⁴¹, Edward L Giovannucci^{4,54,55}, Michael Goggins^{56–58}, Megan J Gorman¹, Myron Gross⁵⁹, Christopher A Haiman⁶⁰, Manal Hassan²⁷, Kathy J Helzlsouer⁶¹, Brian E Henderson⁶², Elizabeth A Holly¹³, Nan Hu⁸, David J Hunter^{2,63,64}, Federico Innocenti⁶⁵, Mazda Jenab⁶⁶, Rudolf Kaaks⁴⁶, Timothy J Key⁶⁷, Kay-Tee Khaw⁶⁸, Eric A Klein⁶⁹, Manolis Kogevinas^{70–72}, Vittorio Krogh⁷³, Juozas Kupcinskis⁷⁴, Robert C Kurtz⁷⁵, Andrea LaCroix⁶, Maria T Landi⁸, Stefano Landi⁷⁶, Loïc Le Marchand⁷⁷, Andrea Mambrini⁷⁸, Satu Mannisto⁷⁹, Roger L Milne^{18,19}, Yusuke Nakamura⁸⁰, Ann L Oberg⁸¹, Kouros Owzar⁸², Alpa V Patel²², Petra H M Peeters^{44,83}, Ulrike Peters⁸⁴, Raffaele Pezzilli⁸⁵, Ada Piepoli⁸⁶, Miquel Porta^{71,87,88}, Francisco X Real^{89,90}, Elio Riboli⁸³, Nathaniel Rothman⁸, Aldo Scarpa⁹¹, Xiao-Ou Shu^{34,35}, Debra T Silverman⁸, Pavel Soucek⁹², Malin Sund⁹³, Renata Talar-Wojnarowska⁹⁴, Philip R Taylor⁸, George E Theodoropoulos⁹⁵, Mark Thornquist⁶, Anne Tjønneland⁹⁶, Geoffrey S Tobias⁸, Dimitrios Trichopoulos^{4,97,98}, Pavel Vodicka⁹⁹, Jean Wactawski-Wende¹⁰⁰, Nicolas Wentzensen⁸, Chen Wu⁴, Herbert Yu⁷⁷, Kai Yu⁸, Anne Zeleniuch-Jacquotte^{11,12}, Robert Hoover⁸, Patricia Hartge^{8,102}, Charles Fuchs^{1,54,102}, Stephen J Chanock^{8,9,102}, Rachael S Stolzenberg-Solomon^{8,102} & Laufey T Amundadottir^{8,102}

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 Union^{1,2}. 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).

A full list of author affiliations appears at the end of the paper.

Received 26 February; accepted 10 July; published online 3 August 2014; doi:10.1038/ng.3052

Table 1 Subject numbers and characteristics of pancreatic cancer cases and controls

	Cases <i>n</i> (%)	Controls <i>n</i> (%)
Number of subjects		
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		
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		
Male	4,107 (53.5)	8,841 (61.4)
Female	3,576 (46.5)	5,556 (38.6)
Age, years		
≤60	1,972 (25.7)	4,577 (31.8)
61–70	2,688 (35.0)	5,906 (41.0)
>70	3,023 (39.3)	3,914 (27.2)
Smoking status^a		
Current or past	2,634 (51.6)	4,541 (51.3)
Never	1,642 (32.2)	3,186 (36.0)
Unknown	831 (16.3)	1,118 (12.6)

^aSmoking 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 I⁵ (1,757 cases and 1,801 controls from 12 cohort studies and 1 case-control study typed on the Illumina HumanHap550 array) and PanScan II⁶ (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, $P = 1.6 \times 10^{-16}$), rs9543325 at 13q22.1 (*KLF5/KLF12*, OR = 1.23, 95% CI 1.18–1.30, $P = 4.3 \times 10^{-14}$), rs10919791 at 1q32.1 (*NR5A2*, OR = 0.79, 95% CI 0.75–0.85, $P = 1.4 \times 10^{-11}$) and rs31490 at 5p15.33 (*CLPTM1L*, OR = 1.20, 95% CI 1.14–1.27, $P = 2.0 \times 10^{-11}$).

We observed two new SNPs below genome-wide significance ($P < 5 \times 10^{-8}$) in the meta-analysis of stages 1 and 2, plus 11 additional promising SNPs ($P < 5 \times 10^{-5}$) 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) cases in stage 1 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	Position ^b	Minor allele ^c	Major allele ^c	Stage	Allelic OR (95% CI)	Minor allele frequency		<i>P</i> ^f
								Controls	Cases	
5p15.33	<i>TERT</i> , <i>MIR4457</i> , <i>CLPTM1L</i>	rs2736098	1,294,086	T	C	Stage 1	0.76 (0.68–0.86)	0.268	0.216	8.22×10^{-6}
						Stage 2	0.82 (0.74–0.90)	0.284	0.259	2.63×10^{-5}
						Replication ^d	0.81 (0.74–0.89)			1.36×10^{-5}
						Combined^e	0.80 (0.76–0.85)			9.78×10^{-14}
7q32.3	<i>LINC-PINT</i>	rs6971499	130,680,521	C	T	Stage 1	0.79 (0.68–0.90)	0.155	0.127	6.58×10^{-4}
						Stage 2	0.81 (0.74–0.90)	0.147	0.124	4.69×10^{-5}
						Replication ^d	0.77 (0.69–0.86)			4.37×10^{-6}
						Combined^e	0.79 (0.74–0.84)			2.98×10^{-12}
16q23.1	<i>BCAR1</i> , <i>CTRB1</i> , <i>CTRB2</i>	rs7190458	75,263,661	A	G	Stage 1	1.61 (1.32–1.96)	0.042	0.065	4.14×10^{-6}
						Stage 2	1.47 (1.20–1.82)	0.039	0.049	2.17×10^{-4}
						Replication ^d	1.33 (1.10–1.61)			3.14×10^{-3}
						Combined^e	1.46 (1.30–1.65)			1.13×10^{-10}
13q12.2	<i>PDX1</i>	rs9581943	28,493,997	A	G	Stage 1	1.23 (1.12–1.35)	0.397	0.441	1.34×10^{-5}
						Stage 2	1.12 (1.05–1.20)	0.406	0.434	5.51×10^{-4}
						Replication ^d	1.11 (1.03–1.20)			4.80×10^{-3}
						Combined^e	1.15 (1.10–1.20)			2.35×10^{-9}
22q12.1	<i>ZNRF3</i>	rs16986825	29,300,306	T	C	Stage 1	1.25 (1.10–1.42)	0.150	0.184	4.96×10^{-4}
						Stage 2	1.15 (1.05–1.26)	0.149	0.168	2.11×10^{-3}
						Replication ^d	1.18 (1.08–1.30)			5.13×10^{-4}
						Combined^e	1.18 (1.12–1.25)			1.18×10^{-8}
8q24.21	<i>MIR1208</i> , <i>PVT1</i>	rs1561927	129,568,078	C	T	Stage 1	0.88 (0.78–0.97)	0.269	0.251	1.59×10^{-2}
						Stage 2	0.86 (0.80–0.93)	0.279	0.250	1.11×10^{-4}
						Replication ^d	0.89 (0.82–0.97)			6.44×10^{-3}
						Combined^e	0.87 (0.83–0.92)			1.30×10^{-7}

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), rs6971499 (7,435/13,289), rs7190458 (7,412/13,291), rs9581943 (7,415/13,286), rs16986825 (7,413/13,196) and rs1561927 (7,486/13,274). ^f1-degree-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 consortium¹⁰ and (iii) cases enrolled in CALGB 80303, a US cooperative group clinical trial¹¹ (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*, $P = 9.8 \times 10^{-14}$), rs6971499 at 7q32.3 (*LINC-PINT*, $P = 3.0 \times 10^{-12}$), rs7190458 at 16q23.1 (*BCAR1/CTRB1/CTRB2*, $P = 1.1 \times 10^{-10}$), rs9581943 at 13q12.2 (*PDX1*, $P = 2.4 \times 10^{-9}$), rs16986825 at 22q12.1 (*ZNRF3*, $P = 1.2 \times 10^{-8}$) (Table 2 and Fig. 1) and rs4962153 at 9q34.2 (*ADAMTS13*, $P = 1.5 \times 10^{-8}$). 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, $P = 1.3 \times 10^{-7}$) 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 muskulin 1) and *KLF14* (Supplementary Table 7 and Fig. 1). Muskulin 1 is an intracellular protein that mediates cell responses to the extracellular matrix, in particular influencing cell adhesion and cytoskeleton organization¹². *KLF14* is a member of the Kruppel-like family of transcription factors, which have been implicated as tumor suppressors, including in mutant *KRAS*-driven tumors¹³. *KLF14* has also been identified as a regulator of several metabolic phenotypes, including type 2 diabetes¹⁴. 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, *CTRB1* (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^{15,16}.

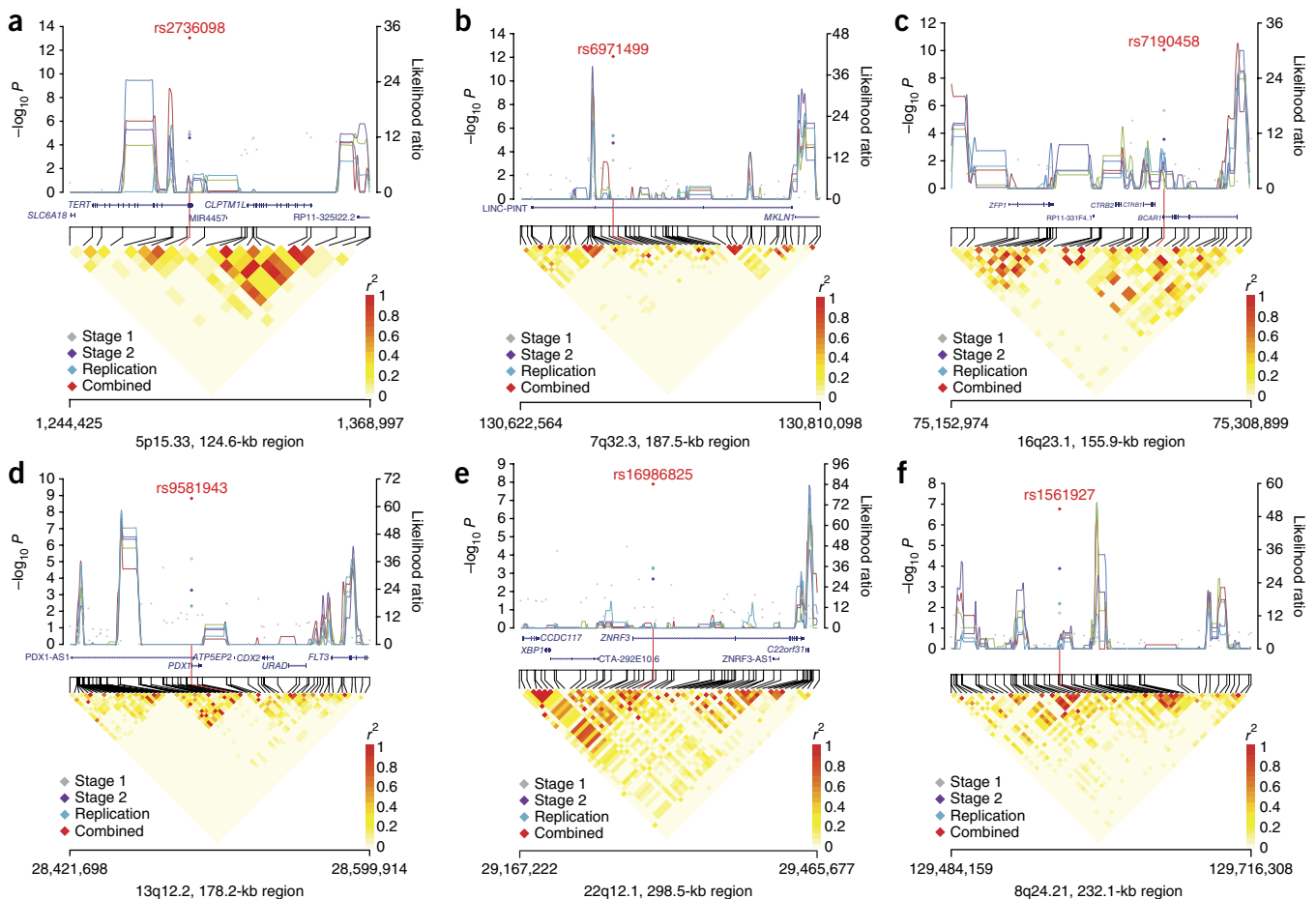


Figure 1 Association results, recombination hotspots and LD plots for new pancreatic cancer susceptibility regions and one suggestive region. (a–f) Results in the newly identified susceptibility regions (a–e) and the suggestive region (f). Top, association results of GWAS data from stage 1 (grey diamonds), stage 2 (purple diamonds) and the replication (blue diamonds), as well as the combined data from stages 1–3 (red diamonds), plotted against $-\log_{10} P$ values (left y axis). Overlaid are likelihood ratio statistics (right y axis) estimating putative recombination hotspots across the 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).

Table 3 Conditional analyses of SNPs at chromosomes 5p15.33 and 9q34.2

New SNP	Chr.	Position ^a	Gene	OR ^b (95% CI)	P ^b	Conditional OR ^c (95% CI)	Established SNP	r ^{2, d}	OR ^e (95% CI)	P ^e	Conditional OR ^f (95% CI)	Conditional P ^f
rs2736098	5p15.33	1,294,086	TERT, CLPTMIL	0.80 (0.74–0.86)	1.42 × 10 ⁻⁹	0.88 (0.80–0.95)	rs401681/rs31490	0.22	0.83 (0.79–0.88)	1.97 × 10 ⁻¹¹	0.86 (0.81–0.92)	7.55 × 10 ⁻⁶
rs4962153	9q34.2	136,323,754	ADAMTS13, ABO	1.20 (1.10–1.30)	1.97 × 10 ⁻⁵	1.05 (0.96–1.16)	rs505902/rs687289	0.17	1.27 (1.20–1.35)	1.64 × 10 ⁻¹⁶	1.25 (1.18–1.33)	6.28 × 10 ⁻¹²

^aPosition of the SNP in NCBI genome build 37 (Hg19). ^bPer-allele ORs for the minor allele and P values for the new SNP from the unconditional meta-analysis of stage 1 and stage 2. ^cPer-allele ORs for the minor allele and P values for the new SNP from the conditional meta-analysis of stage 1 and stage 2. ^dLD values between the new SNP and the established SNP at the locus based on 1000G CEU data. ^ePer-allele ORs for the minor allele and P values for the established SNP from the unconditional meta-analysis of stage 1 and stage 2. ^fPer-allele ORs for the minor allele and P values for the established SNP from the conditional meta-analysis of stage 1 and stage 2.

Chr., chromosome.

The chymotrypsinogens are members of a family of serine proteases that are secreted by the pancreas into the gastrointestinal tract¹⁷. Mutations in the related genes *PRSS1* (encoding trypsin 1) and *CTRC* have been associated with hereditary pancreatitis¹⁸, a known risk factor for pancreatic cancer¹⁹. In addition, a susceptibility locus for types 1 and 2 diabetes^{20,21} 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 function²² and influences the expression of *CTRB1* and *CTRB2* in pancreatic tissue²³.

At chromosome 13q12.2, the newly identified SNP, rs9581943, is approximately 200 bp upstream of *PDX1* (encoding pancreatic and duodenal homeobox1 protein 1) 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 differentiation of the exocrine pancreas and regulates beta-cell function in the mature pancreas^{24,25}. Mutations in *PDX1* have been linked to agenesis of the pancreas²⁴ and maturity onset diabetes of the young²⁶, a dominantly inherited disorder of nonautoimmune diabetes. Furthermore, *PDX1* has been implicated in glucose-dependent regulation of insulin gene transcription²⁷, and GWAS have identified a SNP (rs2293941, $r^2 = 0.20$ in the 1000G CEU data) at the *PDX1* locus that is associated with fasting glucose levels²⁸.

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 pathway²⁹. Additionally, *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 damage³⁰. Alterations in *CHEK2* have been implicated in susceptibility to several cancer types³¹.

We performed conditional analyses to assess whether the newly identified SNPs at 5p15.33 (*CLPTMIL* and *TERT*) and 9q34.2 (*ABO* and *ADAMTS13*) were independent from those identified previously. After conditioning on the reported SNP within intron 13 of *CLPTMIL*, 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)³², 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^{33,34}. *CLPTMIL* 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^{35,36}. Variants across the *TERT* and *CLPTMIL* region have been associated previously with risk of multiple cancers. Furthermore, independent signals within *TERT* and *CLPTMIL* have been identified for bladder cancer³⁷, chronic lymphocytic leukemia³⁸ and lung cancer^{37,39}, and fine-mapping studies have identified at least four independent signals across the *TERT* and *CLPTMIL* region that are associated with cancer^{40,41}. 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^{37,40,41}. 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 cancer⁴⁰. Although further characterization of this region will be necessary, the new SNP in exon 2 of *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 *ABO*, the SNP rs4962153 in *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 *PVT1* and *LINC00977* (Supplementary Table 7 and Fig. 1). 8q24.21 is known to contain multiple cancer susceptibility loci that span over 2 Mb^{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)⁴⁴, and the closest genes are centromeric to rs1561927: *MIR1208* (406 kb), *PVT1* (455 kb) and *MYC* (814 kb). Several 8q24.21 risk loci have been shown to interact with *MYC* or *PVT1* promoters through long-range chromosomal interaction, and allele-specific effects on the expression of both genes have been reported^{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^{5,6} (Table 2). We also assessed previously published pancreatic cancer risk loci from individuals of Chinese⁷ and Japanese⁸ 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 HaploReg⁴⁷ (Supplementary Table 12). We also evaluated expression quantitative trait locus (eQTL) effects^{48–50} (Supplementary Table 12). We noted *cis*-eQTLs on chromosome 16q23.1 in peripheral blood (*CFDP1*), chromosome 13q12.2 in skin and liver (*POMP*), chromosome 22q12.1 in liver and peripheral blood (*CCDC117*) and peripheral blood (*XBPI*) and chromosome 8q24.21 in adipose tissue (*PVT1*). *XBPI* at chromosome 22q12.1 regulates pancreatic beta-cell function with effects on systemic glucose control⁵¹ and modulates acinar cell homeostasis²⁵. In a gene set enrichment analysis⁵² 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 ($P = 3.3 \times 10^{-4}$). Understanding the functional consequences of pancreatic cancer susceptibility variants will require further laboratory investigation.

In a linear-mixed model analysis⁵³ (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 ≤ 6 risk alleles had an OR of 0.55 (95% CI 0.44–0.68) and those with ≥ 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](#).

Note: Any Supplementary Information and Source Data files 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 K07 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. organized and designed the study. B.M.W., C.R., F.C., L.B., R.S.S.-S. and L.T.A. conducted and supervised the genotyping of samples. B.M.W., C.R., P.K., C.K., Z.W., R.B., R.S.S.-S. and L.T.A. contributed to the design and execution of statistical analyses. B.M.W., R.S.S.-S. and L.T.A. wrote the first draft of the manuscript. B.M.W., C.R., P.K., C.K., G.M.P., A.A.A., L.B.-F., P.M.B., J.B., F.C., E.J.D., S.G., G.G.G., G.E.G., P.J.G., E.J.J., A.K., A.P.K., L.N.K., M.H.K., D.L., N.M., S.H.O., H.A.R., H.D.S., K.V., E.W., W.Z., C.C.A., D.A., G.A., M.A.A., D.B., S.I.B., M.-C.B.-R., M.B., M.W.B., H.B.B.-d.-M., P.B., D.C., N.E.C., G.C., C.C., M.C., E.C., J.E., N.F., J.M.G., N.A.G., E.L.G., M. Goggins, M.J.G., M. Gross, C.A.H., M.H., K.J.H., B.E.H., E.A.H., N.H., D.J.H., F.I., M.J., R.K., T.J.K., K.-T.K., E.A.K., M.K., V.K., J.K., R.C.K., A.L., M.T.L., S.L., L.L.M., A.M., S.M., R.L.M., Y.N., A.L.O., K.O., A.V.P., P.H.M.P., U.P., R.P., A.P., M.P., F.X.R., E.R., N.R., A.S., X.-O.S., D.T.S., P.S., M.S., R.T.-W., P.R.T., G.E.T., M.T., A.T., G.S.T., D.T., P.V., J.W.-W., N.W., C.W., H.Y., K.Y., A.Z.-J., R.H., P.H., C.F., S.J.C., R.S.S.-S. and L.T.A. conducted the epidemiological studies and contributed samples to the GWAS and/or follow-up genotyping. All authors contributed to 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>.

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

11. Innocenti, F. *et al.* A genome-wide association study of overall survival in pancreatic cancer patients treated with gemcitabine in CALGB 80303. *Clin. Cancer Res.* **18**, 577–584 (2012).
12. Adams, J.C., Seed, B. & Lawler, J. Muskulin, a novel intracellular mediator of cell adhesive and cytoskeletal responses to thrombospondin-1. *EMBO J.* **17**, 4964–4974 (1998).
13. Fernandez-Zapico, M.E. *et al.* A functional family-wide screening of SP/KLF proteins identifies a subset of suppressors of KRAS-mediated cell growth. *Biochem. J.* **435**, 529–537 (2011).
14. Small, K.S. *et al.* Identification of an imprinted master *trans* regulator at the KLF14 locus related to multiple metabolic phenotypes. *Nat. Genet.* **43**, 561–564 (2011).
15. Barrett, A., Pellet-Many, C., Zachary, I.C., Evans, I.M. & Frankel, P. p130Cas: a key signalling node in health and disease. *Cell. Signal.* **25**, 766–777 (2013).
16. Cabodi, S., del Pilar Camacho-Leal, M., Di Stefano, P. & Defilippi, P. Integrin signalling adaptors: not only figurants in the cancer story. *Nat. Rev. Cancer* **10**, 858–870 (2010).
17. Whitcomb, D.C. & Lowe, M.E. Human pancreatic digestive enzymes. *Dig. Dis. Sci.* **52**, 1–17 (2007).
18. Whitcomb, D.C. Genetic aspects of pancreatitis. *Annu. Rev. Med.* **61**, 413–424 (2010).
19. Chen, J.M. & Ferec, C. Chronic pancreatitis: genetics and pathogenesis. *Annu. Rev. Genomics Hum. Genet.* **10**, 63–87 (2009).
20. Barrett, J.C. *et al.* Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat. Genet.* **41**, 703–707 (2009).
21. Morris, A.P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).
22. Harder, M.N. *et al.* Type 2 diabetes risk alleles near BCAR1 and in ANK1 associate with decreased beta-cell function whereas risk alleles near ANKRD55 and GRB14 associate with decreased insulin sensitivity in the Danish Inter99 cohort. *J. Clin. Endocrinol. Metab.* **98**, E801–E806 (2013).
23. Hart, L.M. *et al.* The CTRB1/2 locus affects diabetes susceptibility and treatment via the incretin pathway. *Diabetes* **62**, 3275–3281 (2013).
24. Stoffers, D.A., Zinkin, N.T., Stanojevic, V., Clarke, W.L. & Habener, J.F. Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence. *Nat. Genet.* **15**, 106–110 (1997).
25. MacDonald, R.J., Swift, G.H. & Real, F.X. Transcriptional control of acinar development and homeostasis. *Prog. Mol. Biol. Transl. Sci.* **97**, 1–40 (2010).
26. Vaxillaire, M., Bonnefont, A. & Froguel, P. The lessons of early-onset monogenic diabetes for the understanding of diabetes pathogenesis. *Best Pract. Res. Clin. Endocrinol. Metab.* **26**, 171–187 (2012).
27. Ohlsson, H., Karlsson, K. & Edlund, T. IPF1, a homeodomain-containing transactivator of the insulin gene. *EMBO J.* **12**, 4251–4259 (1993).
28. Manning, A.K. *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat. Genet.* **44**, 659–669 (2012).
29. Hao, H.X. *et al.* ZNRF3 promotes Wnt receptor turnover in an R-spondin-sensitive manner. *Nature* **485**, 195–200 (2012).
30. Antoni, L., Sodha, N., Collins, I. & Garrett, M.D. CHK2 kinase: cancer susceptibility and cancer therapy—two sides of the same coin? *Nat. Rev. Cancer* **7**, 925–936 (2007).
31. Gronwald, J. *et al.* Cancer risks in first-degree relatives of CHEK2 mutation carriers: effects of mutation type and cancer site in proband. *Br. J. Cancer* **100**, 1508–1512 (2009).
32. Fearnhead, P. SequenceLDhot: detecting recombination hotspots. *Bioinformatics* **22**, 3061–3066 (2006).
33. Ballew, B.J. & Savage, S.A. Updates on the biology and management of dyskeratosis congenita and related telomere biology disorders. *Expert Rev. Hematol.* **6**, 327–337 (2013).
34. Horn, S. *et al.* TERT promoter mutations in familial and sporadic melanoma. *Science* **339**, 959–961 (2013).
35. James, M.A., Vikis, H.G., Tate, E., Rymaszewski, A.L. & You, M. CRR9/CLPTM1L regulates cell survival signaling and is required for Ras transformation and lung tumorigenesis. *Cancer Res.* **74**, 1116–1127 (2014).
36. Jia, J. *et al.* CLPTM1L promotes growth and enhances aneuploidy in pancreatic cancer cells. *Cancer Res.* **74**, 2785–2795 (2014).
37. Rafnar, T. *et al.* Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nat. Genet.* **41**, 221–227 (2009).
38. Speedy, H.E. *et al.* A genome-wide association study identifies multiple susceptibility loci for chronic lymphocytic leukemia. *Nat. Genet.* **46**, 56–60 (2014).
39. McKay, J.D. *et al.* Lung cancer susceptibility locus at 5p15.33. *Nat. Genet.* **40**, 1404–1406 (2008).
40. Bojesen, S.E. *et al.* Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat. Genet.* **45**, 371–384 (2013).
41. Kote-Jarai, Z. *et al.* Fine-mapping identifies multiple prostate cancer risk loci at 5p15, one of which associates with TERT expression. *Hum. Mol. Genet.* **22**, 2520–2528 (2013).
42. Grisanzio, C. & Freedman, M.L. Chromosome 8q24-associated cancers and MYC. *Genes Cancer* **1**, 555–559 (2010).
43. Huppi, K., Pitt, J.J., Wahlberg, B.M. & Caplen, N.J. The 8q24 gene desert: an oasis of non-coding transcriptional activity. *Front. Genet.* **3**, 69 (2012).
44. Pharoah, P.D. *et al.* GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. *Nat. Genet.* **45**, 362–370 (2013).
45. Ahmadiyeh, N. *et al.* 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. *Proc. Natl. Acad. Sci. USA* **107**, 9742–9746 (2010).
46. Meyer, K.B. *et al.* A functional variant at a prostate cancer predisposition locus at 8q24 is associated with PVT1 expression. *PLoS Genet.* **7**, e1002165 (2011).
47. Ward, L.D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934 (2012).
48. Grundberg, E. *et al.* Mapping *cis*- and *trans*-regulatory effects across multiple tissues in twins. *Nat. Genet.* **44**, 1084–1089 (2012).
49. Yang, T.P. *et al.* Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics* **26**, 2474–2476 (2010).
50. Westra, H.J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
51. Lee, A.H., Heidtman, K., Hotamisligil, G.S. & Glimcher, L.H. Dual and opposing roles of the unfolded protein response regulated by IRE1 α and XBP1 in proinsulin processing and insulin secretion. *Proc. Natl. Acad. Sci. USA* **108**, 8885–8890 (2011).
52. Tabas-Madrid, D., Nogales-Cadenas, R. & Pascual-Montano, A. GeneCodis3: a non-redundant and modular enrichment analysis tool for functional genomics. *Nucleic Acids Res.* **40**, W478–W483 (2012).
53. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).

¹Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. ²Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. ³Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁴Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. ⁵Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA. ⁶Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. ⁷Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA. ⁸Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA. ⁹Cancer Genomics Research Laboratory, National Cancer Institute, Division of Cancer Epidemiology and Genetics, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, Maryland, USA. ¹⁰Department of Obstetrics and Gynecology, New York University School of Medicine, New York, New York, USA. ¹¹Department of Environmental Medicine, New York University School of Medicine, New York, New York, USA. ¹²New York University Cancer Institute, New York, New York, USA. ¹³Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, California, USA. ¹⁴Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. ¹⁵Division of Aging, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. ¹⁶Unit of Nutrition, Environment and Cancer, Cancer Epidemiology Research Program, Catalan Institute of Oncology (ICO), Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain. ¹⁷Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada. ¹⁸Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Victoria, Australia. ¹⁹Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia. ²⁰Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Victoria, Australia. ²¹Southwest Oncology Group Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. ²²Epidemiology Research Program, American Cancer Society, Atlanta, Georgia, USA. ²³Group Health Research Institute, Seattle, Washington, USA. ²⁴Department of Oncology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. ²⁵Department of Epidemiology, Bloomberg School of Public Health, Baltimore, Maryland, USA. ²⁶The Cancer Research Center of Hawaii (retired), Honolulu, Hawaii, USA. ²⁷Department of Gastrointestinal Medical Oncology, University of Texas MD Anderson Cancer Center, Houston, Texas, USA. ²⁸Genetic and Molecular Epidemiology Group, CNIO—Spanish National Cancer Research Centre, Madrid, Spain. ²⁹Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, New York, USA. ³⁰Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, Connecticut, USA. ³¹Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA. ³²Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. ³³Department of Epidemiology, University of Washington, Seattle, Washington, USA. ³⁴Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA. ³⁵Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, Tennessee, USA. ³⁶Department of Laboratory Medicine, University Hospital of Padova, Padua, Italy. ³⁷INSERM, Centre for Research in Epidemiology and Population Health (CESP), Nutrition, Hormones and Women's Health

Team, Villejuif, France. ³⁸University Paris Sud, UMRS 1018, Villejuif, France. ³⁹Institut Gustave Roussy (IGR), Villejuif, France. ⁴⁰Westat, Rockville, Maryland, USA. ⁴¹Department of General Surgery, University Hospital Heidelberg, Heidelberg, Germany. ⁴²National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands. ⁴³Department of Gastroenterology and Hepatology, University Medical Centre Utrecht, Utrecht, the Netherlands. ⁴⁴Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands. ⁴⁵Institute of Transfusion Medicine and Immunology, Medical Faculty Mannheim, Heidelberg University, German Red Cross Blood Service Baden-Württemberg-Hessen, Mannheim, Germany. ⁴⁶Division of Cancer Epidemiology, DKFZ, Heidelberg, Germany. ⁴⁷Digestive and Liver Disease Unit, 'Sapienza' University of Rome, Rome, Italy. ⁴⁸Cancer Care Ontario, University of Toronto, Toronto, Ontario, Canada. ⁴⁹Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada. ⁵⁰National Institute for Health Research Liverpool Pancreas Biomedical Research Unit, University of Liverpool, Liverpool, UK. ⁵¹Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA. ⁵²Department of Surgery, Unit of Experimental Surgical Pathology, University Hospital of Pisa, Pisa, Italy. ⁵³Massachusetts Veteran's Epidemiology, Research and Information Center, Geriatric Research Education and Clinical Center, Veterans Affairs Boston Healthcare System, Boston, Massachusetts, USA. ⁵⁴Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. ⁵⁵Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA. ⁵⁶Department of Pathology, Sidney Kimmel Cancer Center and Johns Hopkins University, Baltimore, Maryland, USA. ⁵⁷Department of Medicine, Sidney Kimmel Cancer Center and Johns Hopkins University, Baltimore, Maryland, USA. ⁵⁸Department of Oncology, Sidney Kimmel Cancer Center and Johns Hopkins University, Baltimore, Maryland, USA. ⁵⁹Laboratory of Medicine and Pathology, University of Minnesota, Minneapolis, Minnesota, USA. ⁶⁰Preventive Medicine, University of Southern California, Los Angeles, California, USA. ⁶¹Prevention and Research Center, Mercy Medical Center, Baltimore, Maryland, USA. ⁶²Cancer Prevention, University of Southern California, Los Angeles, California, USA. ⁶³Harvard School of Public Health, Boston, Massachusetts, USA. ⁶⁴Harvard Medical School, Boston, Massachusetts, USA. ⁶⁵The University of North Carolina Eshelman School of Pharmacy, Center for Pharmacogenomics and Individualized Therapy, Lineberger Comprehensive Cancer Center, School of Medicine, Chapel Hill, North Carolina, USA. ⁶⁶International Agency for Research on Cancer, Lyon, France. ⁶⁷Cancer Epidemiology Unit, University of Oxford, Oxford, UK. ⁶⁸School of Clinical Medicine, University of Cambridge, Cambridge, UK. ⁶⁹Glickman Urological and Kidney Institute, Cleveland Clinic, Cleveland, Ohio, USA. ⁷⁰Centre de Recerca en Epidemiologia Ambiental (CREAL), CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain. ⁷¹Hospital del Mar Institute of Medical Research (IMIM), Barcelona, Spain. ⁷²Department of Nutrition, National School of Public Health, Athens, Greece. ⁷³Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy. ⁷⁴Department of Gastroenterology, Lithuanian University of Health Sciences, Kaunas, Lithuania. ⁷⁵Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York, USA. ⁷⁶Department of Biology, University of Pisa, Pisa, Italy. ⁷⁷Cancer Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii, USA. ⁷⁸Oncology Department, ASL1 Massa Carrara, Massa Carrara, Italy. ⁷⁹Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland. ⁸⁰Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. ⁸¹Alliance Statistics and Data Center, Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA. ⁸²Alliance Statistics and Data Center, Department of Biostatistics and Bioinformatics, Duke Cancer Institute, Duke University Medical Center, Durham, North Carolina, USA. ⁸³Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK. ⁸⁴Department of Epidemiology, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. ⁸⁵Pancreas Unit, Department of Digestive Diseases and Internal Medicine, Sant'Orsola-Malpighi Hospital, Bologna, Italy. ⁸⁶Department of Gastroenterology, Scientific Institute and Regional General Hospital 'Casa Sollievo della Sofferenza', Opera di Padre Pio da Pietrelcina, San Giovanni Rotondo, Italy. ⁸⁷Department of Epidemiology, School of Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain. ⁸⁸CIBERESP, Madrid, Spain. ⁸⁹Epithelial Carcinogenesis Group, CNIO-Spanish National Cancer Research Centre, Madrid, Spain. ⁹⁰Departament de Ciències i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain. ⁹¹ARC-NET: Centre for Applied Research on Cancer, University and Hospital Trust of Verona, Verona, Italy. ⁹²Toxicogenomics Unit, Center for Toxicology and Safety, National Institute of Public Health, Prague, Czech Republic. ⁹³Department of Surgical and Perioperative Sciences, Umeå University, Umeå, Sweden. ⁹⁴Department of Digestive Tract Diseases, Medical University of Łódź, Łódź, Poland. ⁹⁵1st Propaedeutic Surgical Department, Hippocraton University Hospital, Athens, Greece. ⁹⁶Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, Denmark. ⁹⁷Bureau of Epidemiologic Research, Academy of Athens, Athens, Greece. ⁹⁸Hellenic Health Foundation, Athens, Greece. ⁹⁹Department of Molecular Biology of Cancer, Institute of Experimental Medicine, Academy of Sciences of the Czech Republic, Prague, Czech Republic. ¹⁰⁰Department of Social and Preventive Medicine, University at Buffalo, State University of New York, Buffalo, New York, USA. ¹⁰¹These authors contributed equally to this work. ¹⁰²These authors jointly directed this work. Correspondence should be addressed to R.S.S.-S. (rachael.solomon@nih.gov) or L.T.A. (amundadottiri@mail.nih.gov).

ONLINE METHODS

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 I⁵, 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 GenomeStudio 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 $P < 1 \times 10^{-7}$ 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 $\text{Pi-hat} > 0.40$) 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, 1,582 cases of European ancestry were available for analysis (**Supplementary Tables 2 and 3**).

Controls of $\geq 80\%$ 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, as previously described^{9,61-64}. In total, 5,203 controls were included in the analysis (**Supplementary Tables 2 and 3**). A total of 608,202 SNPs with overall completion rate >80% in both cases and controls were advanced to the association analysis. To evaluate population substructure, a principal components analysis was performed using the *struct.pca* module of GLU, version 1.0, which is similar to EIGENSTRAT⁶⁵. Plots of the first six principal components are shown in **Supplementary Figure 4**. The estimated inflation of the test statistic, λ , was 1.02 (ref. 66); a quantile-quantile (QQ) plot is shown in **Supplementary Figure 5**.

Association analysis was performed assuming a log-additive genetic model and adjusting for age, sex, geographic region and six significant eigenvectors

(i.e., eigenvectors that were nominally significant in a baseline risk model adjusting for age, sex and geographic region). Geographic region was defined as REGION_US (United States): AgHealth, CPS-II, DFCI, HPFS, MEC, NHS, NYU-WHS, PHS, PLCO, SELECT, VITAL, WHI; REGION_CNE (central and northern Europe): ATBC, EPIC, PANDoRA-Heidelberg, MCCS (Melbourne); and REGION_SE (southern Europe): SBCS (Spain controls), PANKRAS-II (cases). All data analyses and management were conducted using GLU.

Stage 2: PanScan I and II data. The second stage involved the primary whole genome-scan data from the previously reported PanScan I⁵ and PanScan II⁶ studies. PanScan I and PanScan II were genotyped on 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 data⁹. 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 SMWHS⁶⁸. 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 $< 5 \times 10^{-5}$) 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 plus samples from cases that did not have sufficient DNA for the full GWAS and control subjects previously genotyped at CGR; (ii) PANDoRA: case and control subjects from the PANDoRA pancreatic cancer case-control consortium¹⁰ (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**).

Subjecting 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, $n = 512$ pairs) showed >99.9% genotype concordance. Patients enrolled in CALBG/Alliance 80303 (group iii) were previously genotyped using the Illumina HumanHap550v3 Genotyping BeadChip array¹¹. Control subjects from PLCO previously genotyped at CGR using the Illumina HumanHap550v3 Genotyping BeadChip array⁷⁰ 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 r^2 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 SequenceLDhot³², a program that uses the approximate marginal likelihood method⁷¹ 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 program⁷⁴. 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 ($r^2 < 0.5$) 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 $r^2 > 0.7$ 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 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 publically 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 value⁵².

54. Alavanja, 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. Rothman, 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 E2C2. *Hum. Genet.* **133**, 211–224 (2014).
65. Price, A.L. *et al.* Principal components analysis corrects 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 multipoint 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. Thornquist, M.D. *et al.* Statistical 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 5p15 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. & Nicodemus, 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).