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

Published in final edited form as:
Nat Genet. 2014 September ; 46(9): 994–1000. doi:10.1038/ng.3052.

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

A full list of authors and affiliations appears at the end of the article.

Abstract

We performed a multistage genome-wide association study (GWAS) 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; $P = 3.0\times10^{-12}$), rs7190458 at 16q23.1 (*BCAR1/CTRB1/CTRB2*; OR = 1.46; 95% CI = 1.30–1.65; $P = 1.1\times10^{-10}$), rs9581943 at 13q12.2 (*PDX1*; OR = 1.15; 95% CI = 1.10–1.20; $P = 2.4\times10^{-9}$), and rs16986825 at 22q12.1 (*ZNRF3*; OR = 1.18; 95% CI = 1.12–1.25; $P = 1.2\times10^{-8}$). An independent signal was identified in exon 2 of *TERT* at the established region 5p15.33 (rs2736098; OR = 0.80; 95% CI = 0.76–0.85; $P = 9.8\times10^{-14}$). We also identified a locus at 8q24.21 (rs1561927; $P = 1.3\times10^{-7}$) that approached genome-wide significance located 455 kb telomeric of *PVT1*. Our study has identified multiple new susceptibility alleles for pancreatic cancer 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 months[3]. Rare, moderately- to highly-penetrant mutations account for a small fraction of the familial aggregation of pancreatic cancer[4]. In two previous GWAS called PanScan I[5] and PanScan II[6], we identified common variants at four loci associated with risk of sporadic pancreatic cancer in European populations. Subsequent GWAS demonstrated five distinct susceptibility loci among individuals of Chinese descent[7] and three suggestive loci among individuals of Japanese descent[8].

Correspondence to: Rachael S. Stolzenberg-Solomon; Laufey T. Amundadottir.
*These authors contributed equally to this work
†These authors jointly directed this work

AUTHOR CONTRIBUTIONS

B.M.W., C.R., P.K., C.K., G.M.P, P.H., C.F., S.J.C., R.S.S., and L.T.A. organized and designed the study. B.M.W., C.R., F.C., L.B., R.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., and L.T.A. contributed to the design and execution of statistical analyses. B.M.W., R.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., MC.BR., M.B., M.W.B., H.B.B., P.B., D.C., N.E.C., G.C., M.C., E.C., J.E., N.F., J.M.G., N.A.G., E.L.G., M.G., M.J.G., M.G., 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., KT.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., XO.S., D.T.S., P.S., M.S., R.TW., P.R.T., G.E.T., M.T., A.T., G.S.T., D.T., P.V., J.WW., N.W., C.W., H.Y., K.Y., A.ZJ., R.H., P.H., C.F., S.J.C., R.SS., 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.

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). In stage 1, we newly genotyped 1,582 cases from 13 prospective cohort studies, 2 case series, and 1 case-control study using Illumina OmniExpress Beadchip array. The control population included 5,203 cancer-free individuals previously genotyped using second-generation Illumina SNP microarrays (e.g. OmniExpress, Omni 1M or Omni 2.5M) and drawn from PanScan III prospective cohorts and a Spanish case-control study of bladder cancer. Of newly genotyped cases, 94% passed quality control criteria (Online Methods, 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 I[5] (1,757 cases and 1,801 controls from 12 cohort studies and 1 case-control study typed on Illumina HumanHap550 array) and PanScan II[6] (1,768 cases and 1,841 controls from 8 case-control studies typed with Illumina Human 610-Quad array) studies. To address differences in typed SNPs across the arrays, we utilized the DCEG Imputation Reference Set[9] 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; $P = 1.6\times10^{-16}$), rs9543325 at 13q22.1 (*KLF5/KLF12*, OR = 1.23; 95% CI = 1.18–1.30; $P = 4.3\times10^{-14}$), rs10919791 at 1q32.1 (*NR5A2*, OR = 0.79; 95% CI = 0.75–0.85; $P = 1.4\times10^{-11}$), and rs31490 at 5p15.33 (*CLPTM1L*, OR = 1.20; 95% CI = 1.14–1.27 $P = 2.0\times10^{-11}$).

We observed two new SNPs below genome-wide significance ($P<5\times10^{-8}$) in the meta-analysis of stages 1 and 2, plus 11 additional promising SNPs ($P<5\times10^{-5}$) from distinct regions (Supplementary Table 4). These 13 SNPs were carried forward for replication (stage 3) in 2,576 cases and 5,552 controls, drawn from: (a) cases in stage 1 with DNA quantity insufficient for full GWAS, (b) cases and controls from the PANDoRA consortium[10], and (c) cases enrolled to CALGB 80303, a U.S. cooperative group clinical trial[11] (Supplementary Table 5). Additional control subjects were selected from cancer-free individuals previously genotyped using Illumina HumanHap550 array (Online Methods). Of 13 SNPs advanced to replication, nine SNPs 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*, $P=9.8\times10^{-14}$), rs6971499 at 7q32.3 (*LINC-PINT*, $P=3.0\times10^{-12}$), rs7190458 at 16q23.1 (*BCAR1/CTRB1/CTRB2*, $P=1.1\times10^{-10}$), rs9581943 at 13q12.2 (*PDX1*, $P=2.4\times10^{-9}$), rs16986825 at 22q12.1 (*ZNRF3*, $P=1.2\times10^{-8}$) (Table 2 and Figure 1), and rs4962153 at 9q34.2 (*ADAMTS13*, $P=1.5\times10^{-8}$). 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, $P=1.3\times10^{-7}$) and located in a region previously associated with multiple cancers (Table 2 and Figure 1).

The SNP, rs6971499, at 7q32.3 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 *Muskelin 1 (MKLN1)* and *KLF14* (Supplementary Table 7 and Figure 1). Muskelin is an intracellular protein that mediates cell response to the extracellular matrix, particularly influencing cell adhesion and cytoskeleton organization[12]. 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[13]. KLF14 has also been identified as a regulator of several metabolic phenotypes, including type 2 diabetes[14]. Notably, the previously established susceptibility locus at 13q22.1 is located in an intergenic region between *KLF5* and *KLF12*, two other members of the Kruppel-like family of transcription factors.

The SNP, rs7190458, at 16q23.1 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* (23kb) (Supplementary Table 7 and Figure 1). Aberrant expression of *BCAR1* has been linked with 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]. The chymotrypsinogens are members of a family of serine proteases that are secreted by the pancreas into the gastrointestinal tract[17]. Mutations in the related genes *PRSS1* (*trypsin 1*) and *CTRC* have been associated with hereditary pancreatitis[18], a known risk factor for pancreatic cancer[19]. 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 1000G CEU data). Functional analyses indicate that this variant (rs7202877) leads to impaired pancreatic beta-cell function[22] and influences expression of *CTRB1* and *CTRB2* in pancreatic tissue[23].

At chromosome 13q12.2, the newly identified SNP, rs9581943, is approximately 200bp upstream of *PDX1* (*pancreatic and duodenal homeobox1 protein 1*) and intronic to *PDX1-AS1* (*PDX1 antisense RNA 1*), a recently identified noncoding RNA (Supplementary Table 7 and Figure 1). PDX1 is critical for early pancreatic development, plays a role in differentiation of exocrine pancreas, and regulates beta-cell function in the mature pancreas[24,25]. Mutations in *PDX1* have been linked to agenesis of the pancreas[24] and maturity onset diabetes of the young (MODY)[26], a dominantly inherited disorder of non-autoimmune diabetes. Furthermore, PDX1 has been implicated in glucose-dependent regulation of insulin gene transcription[27], and GWAS have identified a SNP (rs2293941, r^2=0.20 in 1000G CEU data) at the *PDX1* locus associated with fasting glucose levels[28].

The signal at 22q12.1, rs16986825, maps to an intron in *ZNRF3* (*zinc and ring finger 3*) (Supplementary Table 7 and Figure 1), encoding a cell surface transmembrane E3 ubiquitin protein ligase that is a negative regulator of the Wnt signaling pathway[29]. 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 damage[30]. Alterations in *CHEK2* have been implicated in susceptibility to several cancer types[31].

We performed conditional analyses to assess whether the newly identified SNPs at 5p15.33 (*CLPTM1L/TERT*) and 9q34.2 (*ABO/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=2.4×10^{-3}) (Table 3). Two strong recombination hotspots lie between the established and new SNPs in 1000G CEU data (likelihood ratios, LR of 27.1 and 261.0)[32], and the two SNPs are in modest linkage disequilibrium (LD; r^2=0.22 in 1000G CEU data) (Figure 1).

TERT encodes the catalytic subunit of telomerase reverse transcriptase, a component of the ribonucleoprotein complex that maintains 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]. *CLPTM1L* encodes the cleft lip and palate associated transmembrane 1 like protein involved in mediating apoptosis, aneuploidy, cisplatin resistance, and RAS-mediated malignant transformation[35,36]. Variants across the *TERT/CLPTM1L* region have previously been associated with risk of multiple cancers. Furthermore, independent signals within *TERT* and *CLPTM1L* have been identified for bladder cancer[37], CLL[38], and lung cancer[37,39], and fine-mapping studies have identified at least four independent signals across the *TERT*/*CLPTM1L* region associated with cancer[40,41]. The new SNP identified in PanScan III (rs2736098) is located in a region of LD spanning ~4kb 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 and *TERT* promoter activity[37,40,41]. The minor allele of rs2736098 that is associated with a lower risk of pancreatic cancer in PanScan was associated with longer telomeres and lower risk of breast cancer[40]. 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).

A promising risk locus was identified at 8q24.21 (rs1561927; P = 1.3×10^{-7}) in a nongenic region between *PVT1* and *LINC00977* (Supplementary Table 7 and Figure 1). 8q24.21 is known to contain multiple cancer susceptibility loci that span over 2Mb[42,43]. The promising pancreatic cancer SNP is in LD with a SNP associated with ovarian cancer risk (rs10088218, r^2=0.37 in 1000G CEU data, 24kb 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 1000G CEU data).

In stratified analyses, no statistically significant heterogeneity was noted 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

Chinese[7] and Japanese[8] ancestry, noting no loci and one locus, respectively, as nominally statistically significant in PanScan (Supplementary Table 11).

To pursue the first steps towards understanding the functional underpinnings of the newly identified risk alleles, we conducted bioinformatic analyses using HaploReg[47] (Supplementary Table 12). We also evaluated expression quantitative trait locus (eQTL) effects[48–50] (Supplementary Table 12). Cis-eQTLs were noted on chr16q23.1 in peripheral blood (*CFDP1*), chr13q12.2 in skin and liver (*POMP*), chr22q12.1 in liver and peripheral blood (*CCDC117*) and peripheral blood (*XBP1*), and chr8q24.21 in adipose tissue (*PVT1*). *XBP1* at chr22q12.1 regulates pancreatic beta-cell function with effects on systemic glucose control[51] and modulates acinar cell homeostasis[25]. In gene set enrichment analysis[52] of genes within 100 kb of the 10 index SNPs identified in PanScan, the only statistically significant pathway was maturity onset diabetes of the young (P=3.3×10^{-4}). Understanding the functional consequences of pancreatic cancer susceptibility variants will require further laboratory investigation.

In a linear-mixed model analysis[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 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 Figure 2).

In conclusion, our multistage GWAS revealed new loci associated with pancreatic cancer risk, as well as promising loci that merit follow-up. Several of the new loci harbor plausible candidate genes 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.

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). Pancreatic cancer cases included individuals newly identified from nine cohort studies that participated in PanScan I[5], 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], Melbourne Collaborative Cohort Study (MCCS)[55], Multiethnic Cohort Study (MEC)[56], Selenium and Vitamin E Cancer Prevention Trial (SELECT)[57], and Vitamins and Lifestyle Study (VITAL)[58]. The new case-based studies were the Gastrointestinal Cancer Clinic of Dana-Farber Cancer Institute (DFCI-GCC), Spanish Pancreatic Cancer Study PANKRAS-II[59], and PANDoRA-Heidelberg pancreatic cancer case-control study[10]. 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 pancreatic cancer cases with 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 based on the optimized cluster file. Extensive quality-control metrics were applied to the data: SNPs with a call rate <94% or Hardy-Weinberg Proportion p value $<1\times10^{-7}$ were excluded (n=18,765); samples with a call rate <94% (n=78), 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, 1,582 cases of European ancestry were available for analysis (Supplementary Tables 2 and 3).

Controls of ≥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 previously genotyped 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 EIGENSTRAT[65]. Plots of the first six principal components are shown in Supplementary Figure 4. The estimated inflation of the test statistic, λ, was 1.02[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 6 significant eigenvectors (i.e. EVs 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 I[5] and PanScan II[6] 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 (~300K), imputation of missing genotypes was performed using phased haplotypes from the DCEG Reference Set and IMPUTE2[9,67]. The DCEG reference set is well-designed for "filling in" missing genotypes across chip designs in PanScan since it is based on several of the same studies included in PanScan and the imputation accuracy is improved over 1,000 Genomes and HapMap data[9]. Imputed SNPs with low minor allele frequencies (MAF <0.01) or low-quality scores (IMPUTE2 information score <0.3) were removed prior to association analysis. The same quality thresholds as described above for stage 1 were applied for stage 2. 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 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). 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 SMWHS[68]. Quality control and association analysis were performed as described above for European ancestry subjects.

Stage 3: Replication studies

Thirteen SNPs (*P*-value threshold of $<5\times10^{-5}$) 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: (A) CGR: pancreatic cancer case and control subjects from CARET[69] plus samples from cases that did not have sufficient DNA for full GWAS and control subjects previously genotyped at CGR; (B) PANDoRA: case and control subjects from the PANDoRA pancreatic cancer case-control consortium[10] (no overlap with the

PANDoRA-Heidelberg cases genotyped in stage 1); and (C) CALBG/Alliance 80303: cases from a randomized clinical trial of gemcitabine plus placebo versus gemcitabine plus bevacizumab[11], and control subjects previously genotyped at CGR (Supplementary Table 5).

Genotyping for cases in group A was performed using custom TaqMan genotyping assays (Applied Biosystems) at CGR. Genotyping for cases and controls from PANDoRA (group B) 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 on CALGB/Alliance 80303 (group C) were previously genotyped using the Illumina HumanHap550v3 Genotyping BeadChip array[11]. Control subjects from PLCO previously genotyped at CGR using the Illumina HumanHap550v3 Genotyping BeadChip array[70] were used for groups (A) and (C) (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 sample and loci in the replication are listed in Supplementary Table 5B. 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, 2 and 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 OmniExpress array contents and confirmatory TaqMan assays (n= 511 samples from PanScan I and II) yielded an r^2 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 estimation of recombination hotspots between loci in *TERT* and *CLPTM1L* on chr5p15.33, we used the 1000G (version 3) CEU data.

Heritability analysis

To estimate 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 HWE *P*-values >10^{-4}. Non-autosomal 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 via 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 zero to 20 risk alleles, based on their genotypes at the 10 identified loci. Odds ratios were calculated using multivariable-adjusted unconditional logistic regression with meta-analysis to combine data from stages 1 and 2, as 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 10 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 1000 Genomes CEU data, we identified SNPs with r^2 >0.7 with our lead SNP. We used HaploReg v2[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 non-transformed 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*-value[52].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Brian M. Wolpin[1,2,*], Cosmeri Rizzato[3,*], Peter Kraft[4,5,*], Charles Kooperberg[6,*], Gloria M. Petersen[7,*], Zhaoming Wang[8,9], Alan A. Arslan[10,11,12], Laura Beane-Freeman[8], Paige M. Bracci[13], Julie Buring[14,15], Federico Canzian[3], Eric J. Duell[16], Steven Gallinger[17], Graham G. Giles[18,19,20], Gary E. Goodman[6], Phyllis J. Goodman[21], Eric J. Jacobs[22], Aruna Kamineni[23], Alison P. Klein[24,25], Laurence N. Kolonel[26], Matthew H. Kulke[1], Donghui Li[27], Núria Malats[28], Sara H. Olson[29], Harvey A. Risch[30], Howard D. Sesso[4,14,15], Kala Visvanathan[31], Emily White[32,33], Wei Zheng[34,35], Christian C. Abnet[8], Demetrius Albanes[8], Gabriella Andreotti[8], Melissa A. Austin[33], Richard Barfield[5], Daniela Basso[36], Sonja I. Berndt[8], Marie-Christine Boutron-Ruault[37,38,39], Michelle Brotzman[40], Markus W. Büchler[41], H. Bas Bueno-de-Mesquita[42,43,44], Peter Bugert[45], Laurie Burdette[8,9], Daniele Campa[46], Neil E. Caporaso[8], Gabriele Capurso[47], Charles Chung[8,9], Michelle Cotterchio[48,49],

Eithne Costello[50], Joanne Elena[51], Niccola Funel[52], J. Michael Gaziano[14,15,53], Nathalia A. Giese[41], Edward L. Giovannucci[4,54,55], Michael Goggins[56,57,58], Megan J. Gorman[1], Myron Gross[59], Christopher A. Haiman[60], Manal Hassan[27], Kathy J. Helzlsouer[61], Brian E. Henderson[62], Elizabeth A. Holly[13], Nan Hu[8], David J. Hunter[2,63,64], Federico Innocenti[65], Mazda Jenab[66], Rudolf Kaaks[46], Timothy J. Key[67], Kay-Tee Khaw[68], Eric A. Klein[69], Manolis Kogevinas[70,71,72], Vittorio Krogh[73], Juozas Kupcinskas[74], Robert C. Kurtz[75], Andrea LaCroix[6], Maria T. Landi[8], Stefano Landi[76], Loic Le Marchand[77], Andrea Mambrini[78], Satu Mannisto[79], Roger L. Milne[18,19], Yusuke Nakamura[80], Ann L. Oberg[81], Kouros Owzar[82], Alpa V. Patel[22], Petra H. M. Peeters[83,84], Ulrike Peters[85], Raffaele Pezzilli[86], Ada Piepoli[87], Miquel Porta[71,88,89], Francisco X. Real[90,91], Elio Riboli[44], Nathaniel Rothman[8], Aldo Scarpa[92], Xiao-Ou Shu[34,35], Debra T. Silverman[8], Pavel Soucek[93], Malin Sund[94], Renata Talar-Wojnarowska[95], Philip R. Taylor[8], George E. Theodoropoulos[96], Mark Thornquist[6], Anne Tjønneland[97], Geoffrey S. Tobias[8], Dimitrios Trichopoulos[4,98,99], Pavel Vodicka[100], Jean Wactawski-Wende[101], Nicolas Wentzensen[8], Chen Wu[4], Herbert Yu[77], Kai Yu[8], Anne Zeleniuch-Jacquotte[11,12], Robert Hoover[8], Patricia Hartge[8,†], Charles Fuchs[1,54,†], Stephen J. Chanock[8,9,†], Rachael S. Stolzenberg-Solomon[8,†], and Laufey T. Amundadottir[8,†]

Affiliations

[1]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA [2]Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA [3]Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany [4]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA [5]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA [6]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA [7]Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA [8]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA [9]Cancer 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 [10]Department of Obstetrics and Gynecology, New York University School of Medicine, New York, New York, USA [11]Department of Environmental Medicine, New York University School of Medicine, New York, New York, USA [12]New York University Cancer Institute, New York, New York, USA [13]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California, USA [14]Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA [15]Division of Aging, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA [16]Unit of Nutrition, Environment and Cancer, Cancer Epidemiology Research Program, Bellvitge Biomedical Research Institute (IDIBELL), Catalan Institute of Oncology (ICO), Barcelona, Spain [17]Samuel

Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada [18]Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Victoria, Australia [19]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Victoria, Australia [20]Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Victoria, Australia [21]Southwest Oncology Group Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA [22]Epidemiology Research Program, American Cancer Society, Atlanta, Georgia, USA [23]Group Health Research Institute, Seattle, Washington, USA [24]Department of Oncology, the Johns Hopkins University School of Medicine, Baltimore, Maryland, USA [25]Department of Epidemiology, the Bloomberg School of Public Health, Baltimore, Maryland, USA [26]The Cancer Research Center of Hawaii (retired), Honolulu, Hawaii, USA [27]Department of Gastrointestinal Medical Oncology, University of Texas M.D. Anderson Cancer Center, Houston, Texas, USA [28]Genetic and Molecular Epidemiology Group, CNIO-Spanish National Cancer Research Centre, Madrid, Spain [29]Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, New York, USA [30]Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, Connecticut, USA [31]Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA [32]Fred Hutchinson Cancer Research Center, Seattle, Washington, USA [33]Department of Epidemiology, University of Washington, Seattle, Washington, USA [34]Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA [35]Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, Tennessee, USA [36]Department of Laboratory Medicine, University Hospital of Padova, Padua, Italy [37]Inserm, Centre for Research in Epidemiology and Population Health (CESP), U1018, Nutrition, Hormones and Women's Health Team, F-94805, Villejuif, France [38]University Paris Sud, UMRS 1018, F-94805, Villejuif, France [39]IGR, F-94805, Villejuif, France [40]Westat, Rockville, Maryland, USA [41]Department of General Surgery, University Hospital Heidelberg, Heidelberg, Germany [42]National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands [43]Department of Gastroenterology and Hepatology, University Medical Centre Utrecht, Utrecht, The Netherlands [44]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom [45]Institute of Transfusion Medicine and Immunology, Heidelberg University, Medical Faculty Mannheim, German Red Cross Blood Service Baden-Württemberg-Hessen, Mannheim, Germany [46]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany [47]Digestive and Liver Disease Unit, 'Sapienza' University of Rome, Rome, Italy [48]Cancer Care Ontario, University of Toronto, Toronto, Ontario, Canada [49]Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada [50]National Institute for Health Research Liverpool Pancreas Biomedical Research Unit, University of Liverpool, Liverpool, United Kingdom [51]Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA [52]Department of Surgery, Unit of Experimental Surgical

Pathology, University Hospital of Pisa, Pisa, Italy [53]Massachusetts Veteran's Epidemiology, Research, and Information Center, Geriatric Research Education and Clinical Center, Veterans Affairs Boston Healthcare System, Boston, Massachusetts, USA [54]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, and Harvard Medical School, Boston, Massachusetts, USA [55]Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA [56]Department of Pathology, Sidney Kimmel Cancer Center and Johns Hopkins University, Baltimore, Maryland, USA [57]Department of Medicine, Sidney Kimmel Cancer Center and Johns Hopkins University, Baltimore, Maryland, USA [58]Department of Oncology, Sidney Kimmel Cancer Center and Johns Hopkins University, Baltimore, Maryland, USA [59]Laboratory of Medicine and Pathology, University of Minnesota, Minneapolis, Minnesota, USA [60]Preventive Medicine, University of Southern California, Los Angeles, California, USA [61]Prevention and Research Center, Mercy Medical Center, Baltimore, Maryland, USA [62]Cancer Prevention, University of Southern California, Los Angeles, California, USA [63]Harvard School of Public Health, Boston, Massachusetts, USA [64]Harvard Medical School, Boston, Massachusetts, USA [65]The 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 [66]International Agency for Research on Cancer, Lyon, France [67]Cancer Epidemiology Unit, University of Oxford, Oxford, United Kingdom [68]School of Clinical Medicine, University of Cambridge, United Kingdom [69]Glickman Urological and Kidney Institute, Cleveland Clinic, Cleveland, OH, USA [70]Centre de Recerca en Epidemiologia Ambiental (CREAL), CIBER Epidemiología y Salud Pública (CIBERESP), Spain [71]Hospital del Mar Institute of Medical Research (IMIM), Barcelona, Spain [72]National School of Public Health, Athens, Greece [73]Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy [74]Department of Gastroenterology, Lithuanian University of Health Sciences, Kaunas, Lithuania [75]Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York, USA [76]Department of Biology, University of Pisa, Pisa, Italy [77]Cancer Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA [78]Oncology Department, ASL1 Massa Carrara, Massa Carrara, Italy [79]National Institute for Health and Welfare, Department of Chronic Disease Prevention, Helsinki, Finland [80]Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan [81]Alliance Statistics and Data Center, Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA [82]Alliance Statistics and Data Center, Department of Biostatistics and Bioinformatics, Duke Cancer Institute, Duke University Medical Center, Durham, North Carolina, USA [83]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands [84]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom [85]Epidemiology, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA [86]Pancreas Unit, Department of Digestive

Diseases and Internal Medicine, Sant'Orsola-Malpighi Hospital, Bologna, Italy [87]Department of Gastroenterology, Scientific Institute and Regional General Hospital "Casa Sollievo della Sofferenza", Opera di Padre Pio da Pietrelcina, San Giovanni Rotondo, Italy [88]School of Medicine, Universitat Autònoma de Barcelona, Spain [89]CIBER de Epidemiología y Salud Pública (CIBERESP), Spain [90]Epithelial Carcinogenesis Group, CNIO-Spanish National Cancer Research Centre, Madrid, Spain [91]Departament de Ciències i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain [92]ARC-NET: Centre for Applied Research on Cancer, University and Hospital Trust of Verona, Verona, Italy [93]Toxicogenomics Unit, Center for Toxicology and Safety, National Institute of Public Health, Prague, Czech Republic [94]Department of Surgical and Peroperative Sciences, Umeå University, Umeå, Sweden [95]Department of Digestive Tract Diseases, Medical University of Łodz, Łodz, Poland [96]1st Propaideutic Surgical Department, Hippocration University Hospital, Athens, Greece [97]Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, Denmark [98]Bureau of Epidemiologic Research, Academy of Athens, Athens, Greece [99]Hellenic Health Foundation, Athens, Greece [100]Department of Molecular Biology of Cancer, Institute of Experimental Medicine, Academy of Sciences of the Czech Republic, Prague, Czech Republic [101]Department of Social and Preventive Medicine, University at Buffalo, Buffalo, New York, USA

Acknowledgments

This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. Major support for PanScan III sample identification and processing was provided by the Lustgarten Foundation for Pancreatic Cancer Research. Additional support from NIH/NCI K07 CA140790, American Society of Clinical Oncology Conquer Cancer Foundation, Howard Hughes Medical Institute, Lustgarten Foundation, Robert T. and Judith B. Hale Fund for Pancreatic Cancer Research, and Promises for Purple to Dr. Brian Wolpin. A full list of acknowledgments for each participating study is provided in the Supplementary Note.

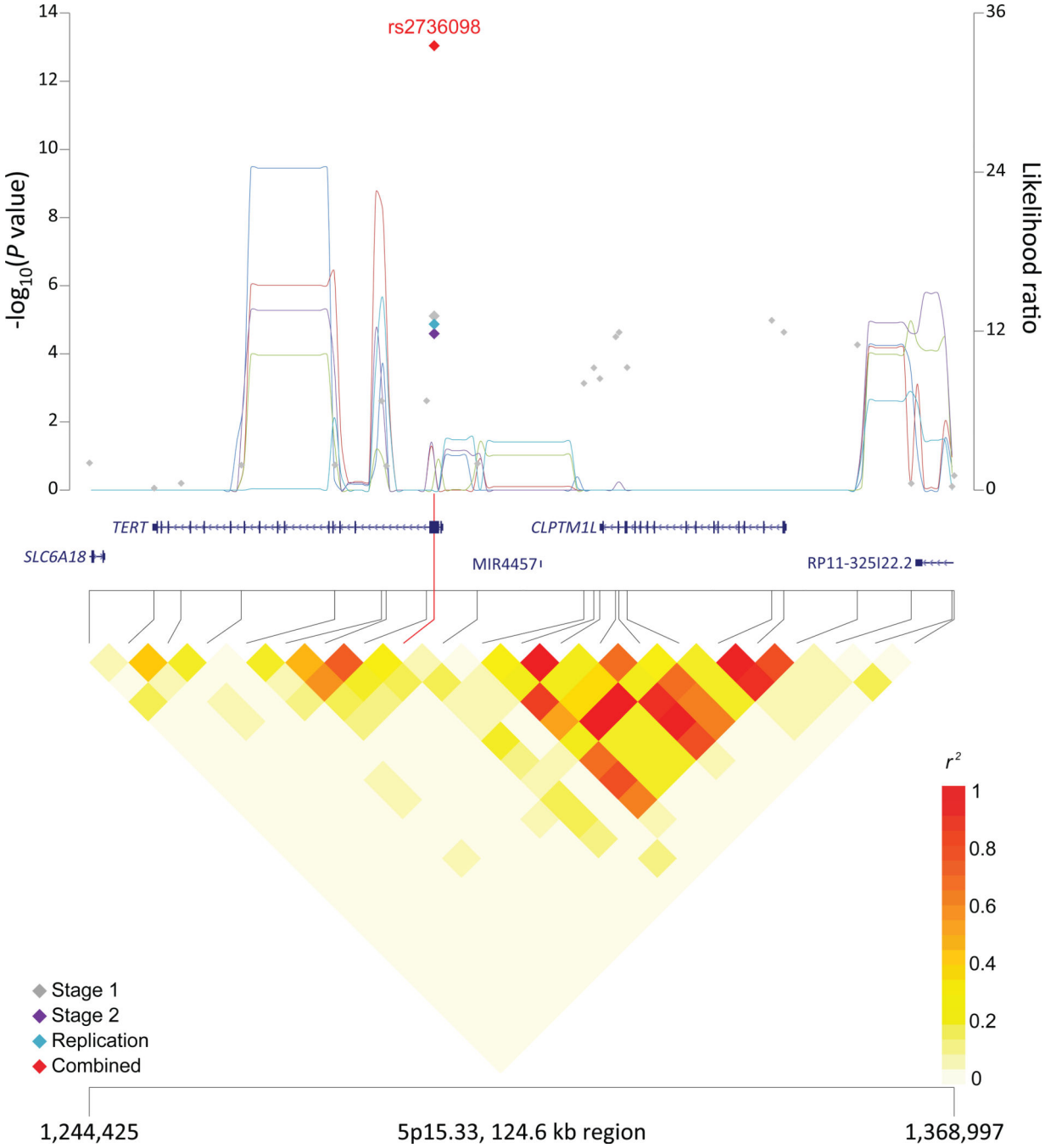
REFERENCES

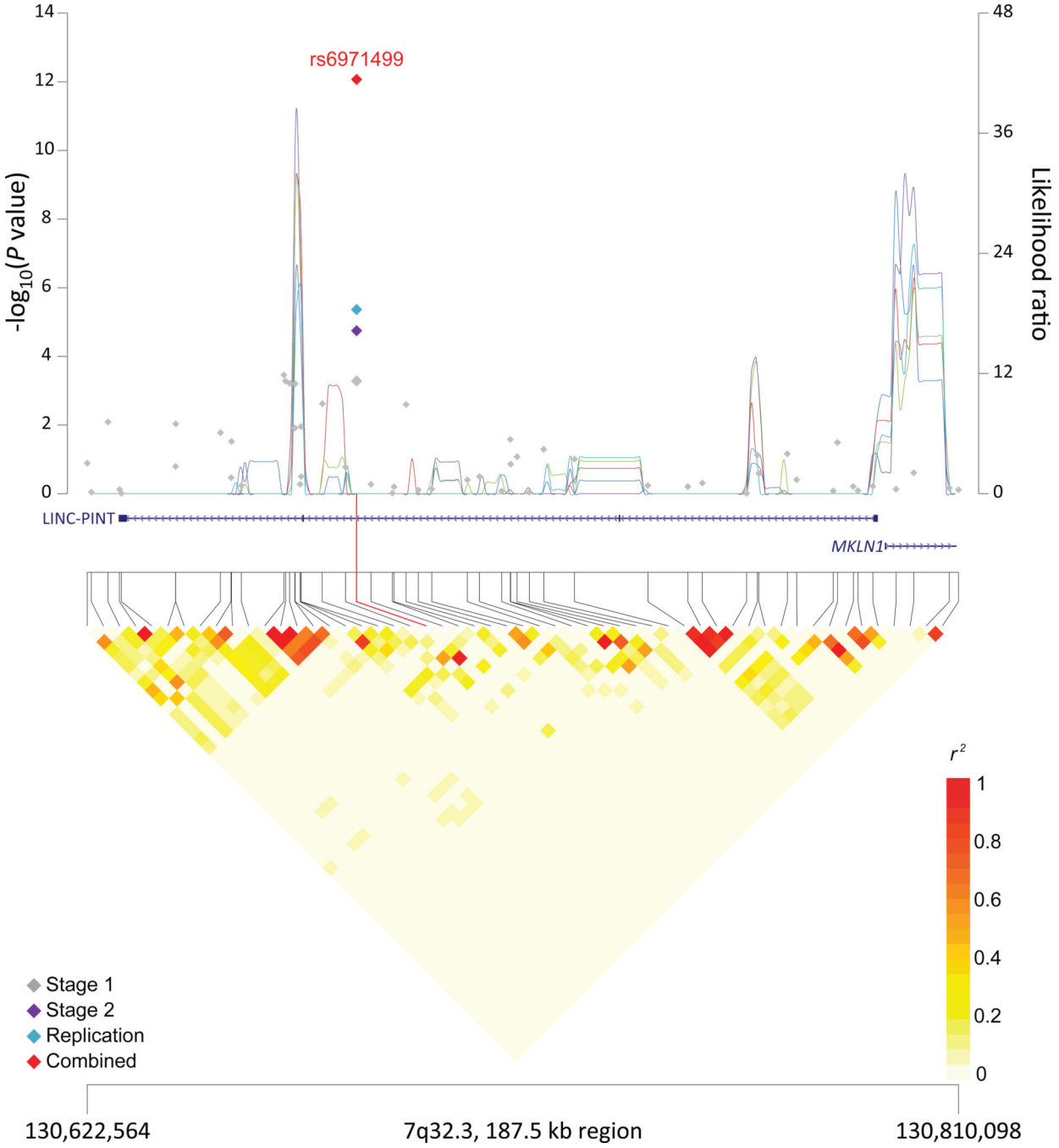
1. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. CA Cancer J Clin. 2014; 64:9–29. [PubMed: 24399786]
2. Ferlay J, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. Eur J Cancer. 2013; 49:1374–1403. [PubMed: 23485231]
3. Hidalgo M. Pancreatic cancer. N Engl J Med. 2010; 362:1605–1617. [PubMed: 20427809]
4. Klein AP. Genetic susceptibility to pancreatic cancer. Mol Carcinog. 2012; 51:14–24. [PubMed: 22162228]
5. Amundadottir L, et al. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. Nat Genet. 2009; 41:986–990. [PubMed: 19648918]
6. Petersen GM, et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Nat Genet. 2010; 42:224–228. [PubMed: 20101243]
7. Wu C, et al. Genome-wide association study identifies five loci associated with susceptibility to pancreatic cancer in Chinese populations. Nat Genet. 2012; 44:62–66. [PubMed: 22158540]

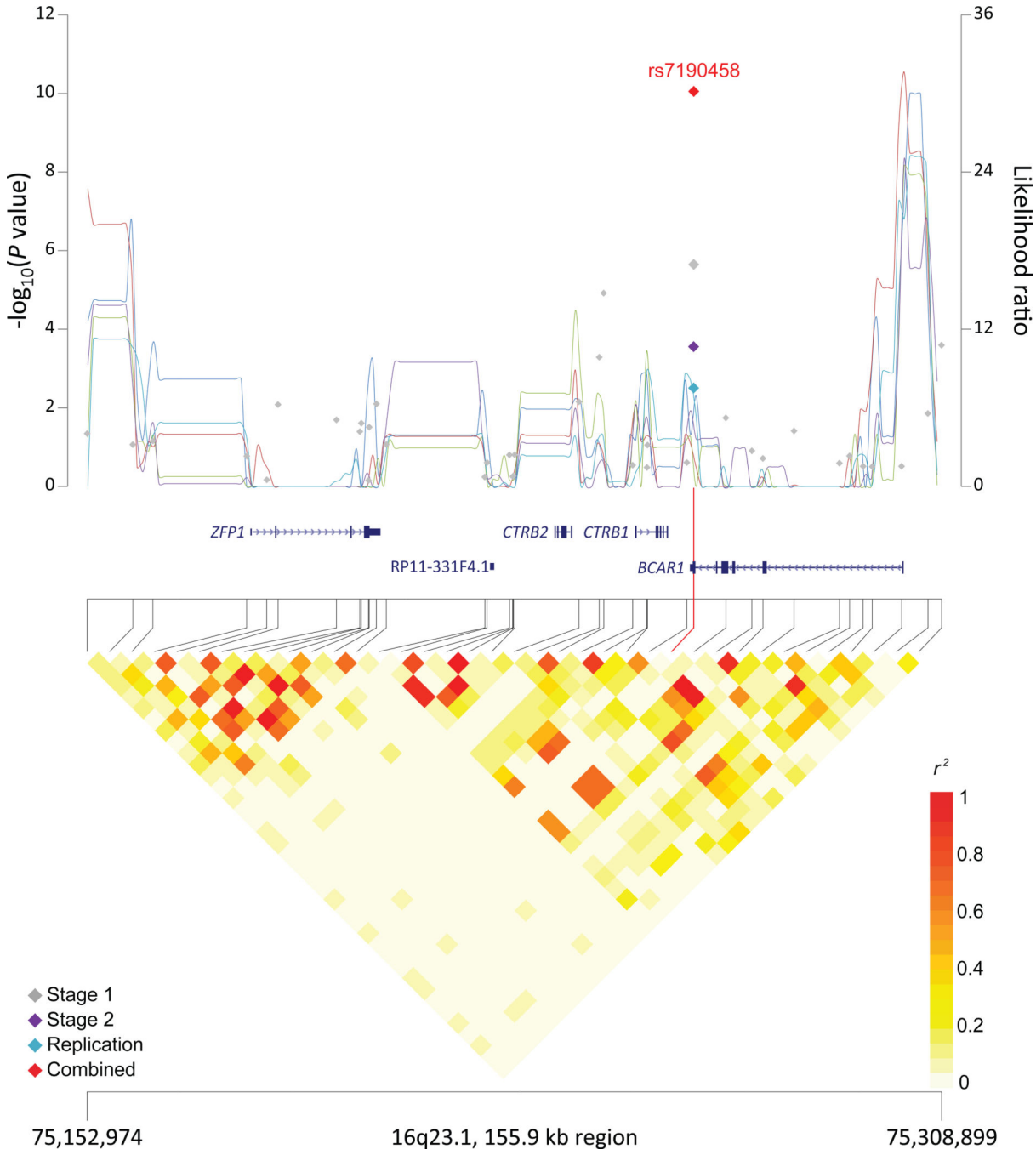
8. Low SK, et al. Genome-wide association study of pancreatic cancer in Japanese population. PLoS ONE. 2010; 5:e11824. [PubMed: 20686608]
9. Wang Z, et al. Improved imputation of common and uncommon SNPs with a new reference set. Nat Genet. 2012; 44:6–7. [PubMed: 22200770]
10. Campa D, et al. Genetic susceptibility to pancreatic cancer and its functional characterisation: the PANcreatic Disease ReseArch (PANDoRA) consortium. Dig Liver Dis. 2013; 45:95–99. [PubMed: 23206934]
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. 2012; 18:577–584. [PubMed: 22142827]
12. Adams JC, Seed B, Lawler J. Muskelin, a novel intracellular mediator of cell adhesive and cytoskeletal responses to thrombospondin-1. EMBO J. 1998; 17:4964–4974. [PubMed: 9724633]
13. Fernandez-Zapico ME, et al. A functional family-wide screening of SP/KLF proteins identifies a subset of suppressors of KRAS-mediated cell growth. Biochem J. 2011; 435:529–537. [PubMed: 21171965]
14. Small KS, et al. Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. Nat Genet. 2011; 43:561–564. [PubMed: 21572415]
15. Barrett A, Pellet-Many C, Zachary IC, Evans IM, Frankel P. p130Cas: a key signalling node in health and disease. Cell Signal. 2013; 25:766–777. [PubMed: 23277200]
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. 2010; 10:858–870. [PubMed: 21102636]
17. Whitcomb DC, Lowe ME. Human pancreatic digestive enzymes. Dig Dis Sci. 2007; 52:1–17. [PubMed: 17205399]
18. Whitcomb DC. Genetic aspects of pancreatitis. Annu Rev Med. 2010; 61:413–424. [PubMed: 20059346]
19. Chen JM, Ferec C. Chronic pancreatitis: genetics and pathogenesis. Annu Rev Genomics Hum Genet. 2009; 10:63–87. [PubMed: 19453252]
20. Barrett JC, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet. 2009; 41:703–707. [PubMed: 19430480]
21. Morris AP, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet. 2012; 44:981–990. [PubMed: 22885922]
22. Harder MN, 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. 2013; 98:E801–E806. [PubMed: 23457408]
23. t Hart LM, et al. The CTRB1/2 Locus Affects Diabetes Susceptibility and Treatment via the Incretin Pathway. Diabetes. 2013; 62:3275–3281. [PubMed: 23674605]
24. Stoffers DA, Zinkin NT, Stanojevic V, Clarke WL, Habener JF. Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence. Nat Genet. 1997; 15:106–110. [PubMed: 8988180]
25. MacDonald RJ, Swift GH, Real FX. Transcriptional control of acinar development and homeostasis. Prog Mol Biol Transl Sci. 2010; 97:1–40. [PubMed: 21074728]
26. Vaxillaire M, Bonnefond A, Froguel P. The lessons of early-onset monogenic diabetes for the understanding of diabetes pathogenesis. Best Pract Res Clin Endocrinol Metab. 2012; 26:171–187. [PubMed: 22498247]
27. Ohlsson H, Karlsson K, Edlund T. IPF1, a homeodomain-containing transactivator of the insulin gene. EMBO J. 1993; 12:4251–4259. [PubMed: 7901001]
28. Manning AK, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet. 2012; 44:659–669. [PubMed: 22581228]
29. Hao HX, et al. ZNRF3 promotes Wnt receptor turnover in an R-spondin-sensitive manner. Nature. 2012; 485:195–200. [PubMed: 22575959]

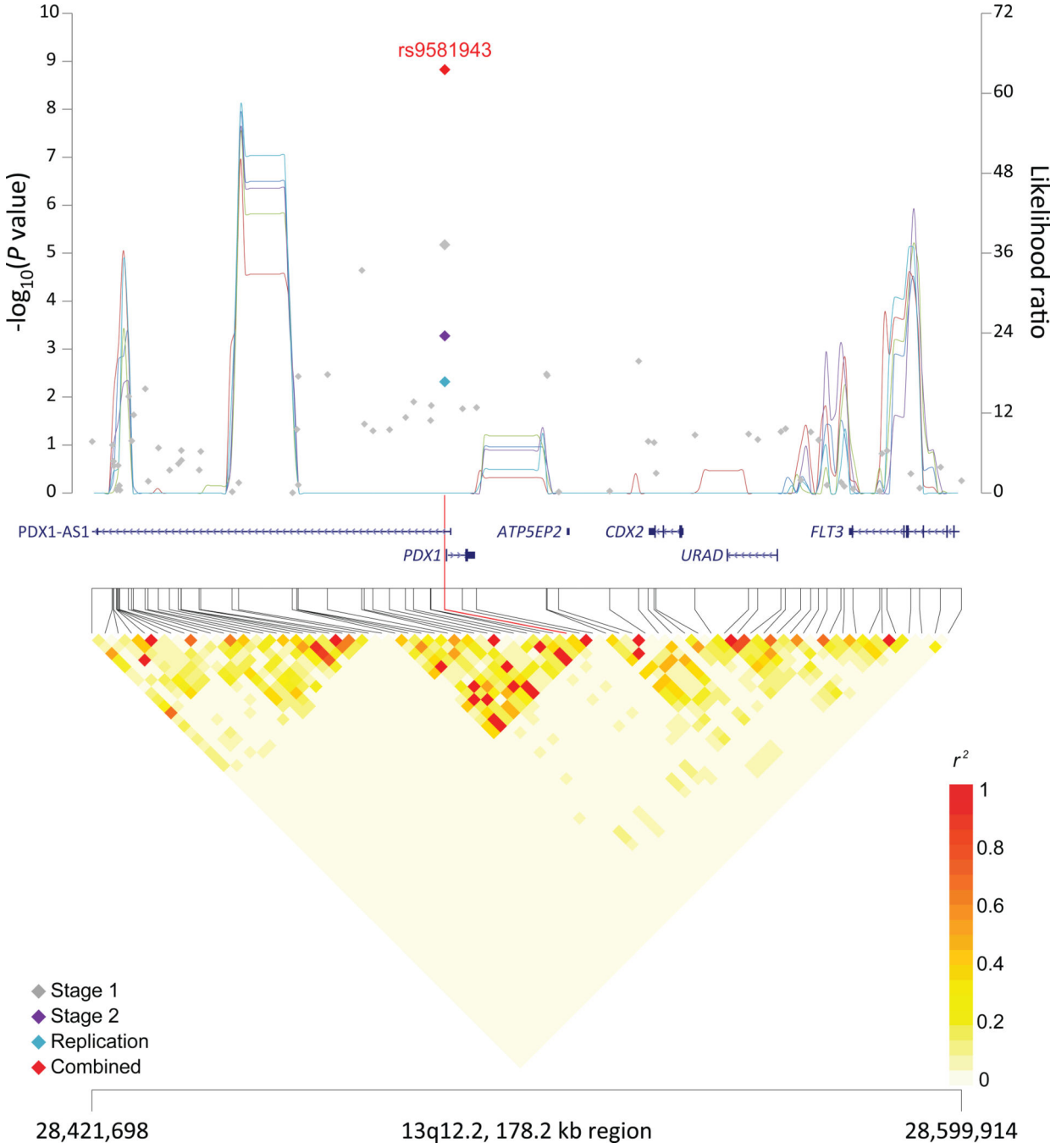
30. Antoni L, Sodha N, Collins I, Garrett MD. CHK2 kinase: cancer susceptibility and cancer therapy - two sides of the same coin? Nat Rev Cancer. 2007; 7:925–936. [PubMed: 18004398]
31. Gronwald J, et al. Cancer risks in first-degree relatives of CHEK2 mutation carriers: effects of mutation type and cancer site in proband. Br J Cancer. 2009; 100:1508–1512. [PubMed: 19401704]
32. Fearnhead P. SequenceLDhot: detecting recombination hotspots. Bioinformatics. 2006; 22:3061–3066. [PubMed: 17060358]
33. Ballew BJ, Savage SA. Updates on the biology and management of dyskeratosis congenita and related telomere biology disorders. Expert Rev Hematol. 2013; 6:327–337. [PubMed: 23782086]
34. Horn S, et al. TERT promoter mutations in familial and sporadic melanoma. Science. 2013; 339:959–961. [PubMed: 23348503]
35. James MA, Vikis HG, Tate E, Rymaszewski AL, You M. CRR9/CLPTM1L Regulates Cell Survival Signaling and is Required for Ras Transformation and Lung Tumorigenesis. Cancer Res. 2013
36. Jia J, et al. CLPTM1L Promotes Growth and Enhances Aneuploidy in Pancreatic Cancer Cells. Cancer Res. 2014; 74:2785–2795. [PubMed: 24648346]
37. Rafnar T, et al. Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. Nat Genet. 2009; 41:221–227. [PubMed: 19151717]
38. Speedy HE, et al. A genome-wide association study identifies multiple susceptibility loci for chronic lymphocytic leukemia. Nat Genet. 2014; 46:56–60. [PubMed: 24292274]
39. McKay JD, et al. Lung cancer susceptibility locus at 5p15.33. Nat Genet. 2008; 40:1404–1406. [PubMed: 18978790]
40. Bojesen SE, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. Nat Genet. 2013; 45:371–384. 384e1–384e2. [PubMed: 23535731]
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. 2013; 22:2520–2528. [PubMed: 23535824]
42. Grisanzio C, Freedman ML. Chromosome 8q24-Associated Cancers and MYC. Genes Cancer. 2010; 1:555–559. [PubMed: 21779458]
43. Huppi K, Pitt JJ, Wahlberg BM, Caplen NJ. The 8q24 gene desert: an oasis of non-coding transcriptional activity. Front Genet. 2012; 3:69. [PubMed: 22558003]
44. Pharoah PD, et al. GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. Nat Genet. 2013; 45:362–370. 370e1–370e2. [PubMed: 23535730]
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 U S A. 2010; 107:9742–9746. [PubMed: 20453196]
46. Meyer KB, et al. A functional variant at a prostate cancer predisposition locus at 8q24 is associated with PVT1 expression. PLoS Genet. 2011; 7:e1002165. [PubMed: 21814516]
47. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012; 40:D930–D934. [PubMed: 22064851]
48. Grundberg E, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat Genet. 2012; 44:1084–1089. [PubMed: 22941192]
49. Yang TP, et al. Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. Bioinformatics. 2010; 26:2474–2476. [PubMed: 20702402]
50. Westra HJ, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. 2013; 45:1238–1243. [PubMed: 24013639]
51. Lee AH, Heidtman K, Hotamisligil GS, Glimcher LH. Dual and opposing roles of the unfolded protein response regulated by IRE1alpha and XBP1 in proinsulin processing and insulin secretion. Proc Natl Acad Sci U S A. 2011; 108:8885–8890. [PubMed: 21555585]
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. 2012; 40:W478–W483. [PubMed: 22573175]

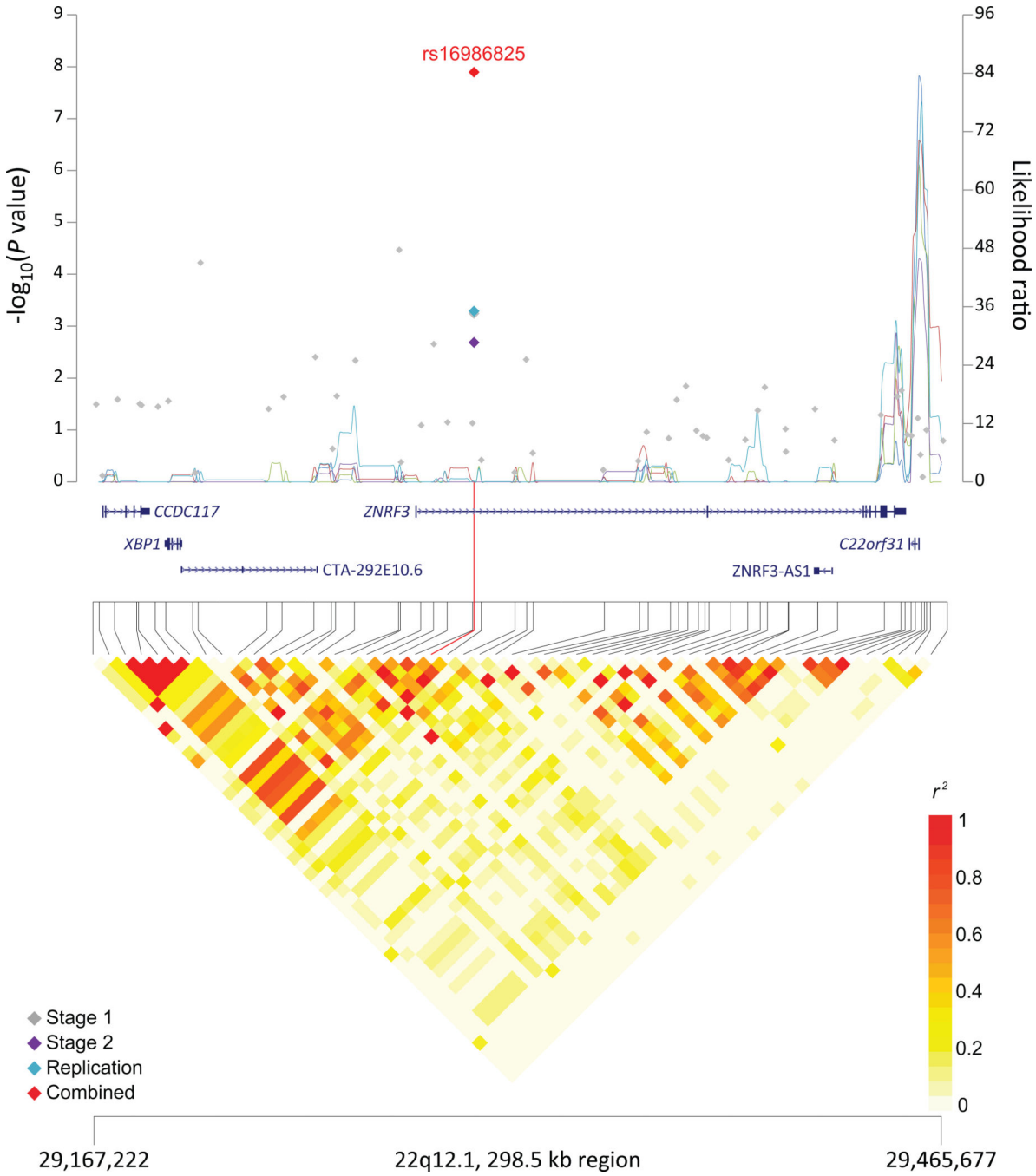
53. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011; 88:76–82. [PubMed: 21167468]
54. Alavanja MC, et al. Cancer and noncancer risk to women in agriculture and pest control: the Agricultural Health Study. J Occup Med. 1994; 36:1247–1250. [PubMed: 7532217]
55. Giles GG, English DR. The Melbourne Collaborative Cohort Study. IARC Sci Publ. 2002; 156:69–70. [PubMed: 12484128]
56. Kolonel LN, et al. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. Am J Epidemiol. 2000; 151:346–357. [PubMed: 10695593]
57. Lippman SM, et al. Designing the Selenium and Vitamin E Cancer Prevention Trial (SELECT). J Natl Cancer Inst. 2005; 97:94–102. [PubMed: 15657339]
58. White E, et al. VITamins And Lifestyle cohort study: study design and characteristics of supplement users. Am J Epidemiol. 2004; 159:83–93. [PubMed: 14693663]
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. 2009; 24:513–521. [PubMed: 19797353]
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. 2006; 15:1348–1354. [PubMed: 16835335]
61. Berndt SI, et al. Genome-wide association study identifies multiple risk loci for chronic lymphocytic leukemia. Nat Genet. 2013; 45:868–876. [PubMed: 23770605]
62. Rothman N, et al. A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. Nat Genet. 2010; 42:978–984. [PubMed: 20972438]
63. Figueroa JD, et al. Genome-wide association study identifies multiple loci associated with bladder cancer risk. Hum Mol Genet. 2013
64. De Vivo I, et al. Genome-wide association study of endometrial cancer in E2C2. Hum Genet. 2014; 133:211–224. [PubMed: 24096698]
65. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–909. [PubMed: 16862161]
66. de Bakker PI, et al. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum Mol Genet. 2008; 17:R122–R128. [PubMed: 18852200]
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. 2007; 39:906–913. [PubMed: 17572673]
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. 2012; 44:1090–1097. [PubMed: 22960999]
69. Thornquist MD, et al. Statistical design and monitoring of the Carotene and Retinol Efficacy Trial (CARET). Control Clin Trials. 1993; 14:308–324. [PubMed: 8365195]
70. Landi MT, 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. 2009; 85:679–691. [PubMed: 19836008]
71. Fearnhead P, Harding RM, Schneider JA, Myers S, Donnelly P. Application of coalescent methods to reveal fine-scale rate variation and recombination hotspots. Genetics. 2004; 167:2067–2081. [PubMed: 15342541]
72. Li N, Stephens M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics. 2003; 165:2213–2233. [PubMed: 14704198]
73. Crawford DC, et al. Evidence for substantial fine-scale variation in recombination rates across the human genome. Nat Genet. 2004; 36:700–706. [PubMed: 15184900]
74. Luna A, Nicodemus KK. snp.plotter: an R-based SNP/haplotype association and linkage disequilibrium plotting package. Bioinformatics. 2007; 23:774–776. [PubMed: 17234637]
75. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. Am J Hum Genet. 2011; 88:294–305. [PubMed: 21376301]











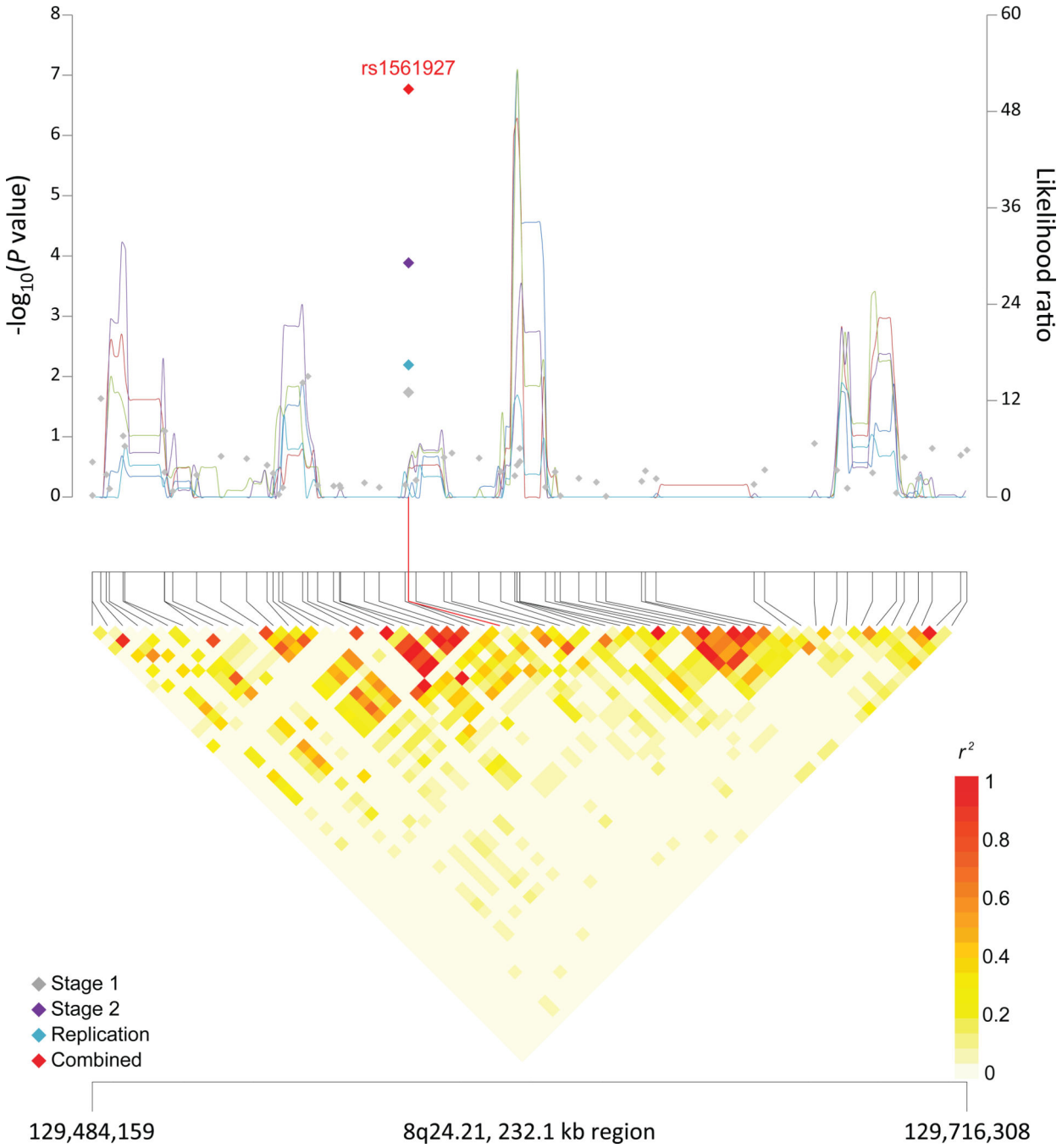


Figure 1. Association results, recombination hotspots and LD plots for new pancreatic cancer susceptibility regions (a–e) and one suggestive region (f)

Top, association results of GWAS data from the stage 1 (gray diamonds), stage 2 (purple diamonds), replication (blue diamonds) and 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 map based on r^2 values from the total control populations for all SNPs included in the GWAS. The data are based on a total number 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 8q24.1 (f).

Table 1

Subject numbers and characteristics of pancreatic cancer cases and controls

	Cases No. (%)	Controls No. (%)
No. 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/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*		
Current / 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)

* Smoking status was available for subjects in Stages 1 and 2.

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		*P*[f]
								Controls	Cases	
5p15.33	*TERT, MIR4457, CLPTM1L*	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	*LINC-PINT*	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	*BCAR1, CTRB1, CTRB2*	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	*PDX1*	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	*ZNRF3*	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	*MIR1208, PVT1*	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 from unconditional logistic regression of the genotypes generated in Stage 1, Stage 2, and replication (total of 7,683 individuals diagnosed with pancreatic cancer and 14,397 controls).

[a] Closest RefSeq gene(s). Genes located within 25 kb of SNP are listed in black in order of closest gene to those further away; closest genes outside this 50 kb window are listed in grey.

[b] Position of SNP in NCBI genome build 37 (Hg19).

[c] Minor and major alleles.

[d] The replication is a meta-analysis of three groups and thus, minor allele frequency (MAF) is not listed.

[e] Number of case and control subjects in joint analysis of Stage 1, Stage 2 and 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), rs1561927 (7,486/13,274).

[f] 1 d.f. 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 Replication. Text in bold indicates the combined meta-analysis results.

Table 3

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

New SNP	Chr	Position[a]	Gene	OR[b]	P[b]	Conditional OR[c]	Conditional P[c]	Established SNP	r^2,[d]	OR[e]	P[e]	Conditonal OR[f]	Conditional P[f]
rs2736098	5p15.33	1,294,086	*TERT, CLPTM1L*	0.80 (0.74–0.86)	1.42×10^{-9}	0.88 (0.80–0.95)	2.44×10^{-3}	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)	0.28	rs505902 / rs687289	0.17	1.27 (1.20–1.35)	1.64×10^{-16}	1.25 (1.18–1.33)	6.28×10^{-12}

[a] Position of SNP in NCBI genome build 37 (Hg19).

[b] Per-allele ORs for the minor allele and *P* values for the new SNP from the unconditional meta-analysis of Stage 1 and Stage 2.

[c] Per-allele ORs for the minor allele and *P* values for the new SNP from the conditional meta-analysis.

[d] r^2 LD values between the new SNP and the established SNP at the locus based on 1000 Genomes Project CEU data.

[e] Per-allele ORs for the minor allele and *P* values for the established SNP from the unconditional meta-analysis of Stage 1 and Stage 2.

[f] Per-allele ORs for the minor allele and *P* values for the established SNP from the conditional meta-analysis.

Chr: chromosome.