nature genetics

A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33

Gloria M Petersen[*1,54], Laufey Amundadottir[2,3,54], Charles S Fuchs[4,5,54], Peter Kraft[6,7,54], Rachael Z Stolzenberg-Solomon[3,54], Kevin B Jacobs[3,8,9], Alan A Arslan[10–12], H Bas Bueno-de-Mesquita[13], Steven Gallinger[14], Myron Gross[15], Kathy Helzlsouer[16], Elizabeth A Holly[17], Eric J Jacobs[18], Alison P Klein[19,20], Andrea LaCroix[21], Donghui Li[22], Margaret T Mandelson[21,23], Sara H Olson[24], Harvey A Risch[25], Wei Zheng[26], Demetrius Albanes[3], William R Bamlet[1], Christine D Berg[27], Marie-Christine Boutron-Ruault[28], Julie E Buring[29,30], Paige M Bracci[17], Federico Canzian[31], Sandra Clipp[32], Michelle Cotterchio[33], Mariza de Andrade[1], Eric J Duell[34], J Michael Gaziano[35,36], Edward L Giovannucci[5,6,37], Michael Goggins[38], Göran Hallmans[39], Susan E Hankinson[5,6], Manal Hassan[22], Barbara Howard[40], David J Hunter[5,6], Amy Hutchinson[3,8], Mazda Jenab[34], Rudolf Kaaks[31], Charles Kooperberg[21], Vittorio Krogh[41], Robert C Kurtz[42], Shannon M Lynch[43], Robert R McWilliams[1], Julie B Mendelsohn[3], Dominique S Michaud[6,44], Hemang Parikh[2,3], Alpa V Patel[18], Petra H M Peeters[44,45], Aleksandar Rajkovic[46], Elio Riboli[44], Laudina Rodriguez[47], Daniela Seminara[43], Xiao-Ou Shu[26], Gilles Thomas[3,48], Anne Tjønneland[49], Geoffrey S Tobias[3], Dimitrios Trichopoulos[6,50], Stephen K Van Den Eeden[51], Jarmo Virtamo[52], Jean Wactawski-Wende[53], Zhaoming Wang[3,8], Brian M Wolpin[4,5], Herbert Yu[25], Kai Yu[3], Anne Zeleniuch-Jacquotte[11,12], Joseph F Fraumeni Jr[3], Robert N Hoover[3,54], Patricia Hartge[3,54] & Stephen J Chanock[2,3,54]

© 2010 Nature America, Inc. All rights reserved.

We conducted a genome-wide association study of pancreatic cancer in 3,851 affected individuals (cases) and 3,934 unaffected controls drawn from 12 prospective cohort studies and 8 case-control studies. Based on a logistic regression model for genotype trend effect that was adjusted for study, age, sex, self-described ancestry and five principal components, we identified eight SNPs that map to three loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Two correlated SNPs, rs9543325 ($P = 3.27 \times 10^{-11}$, per-allele odds ratio (OR) 1.26, 95% CI 1.18–1.35) and rs9564966 ($P = 5.86 \times 10^{-8}$, per-allele OR 1.21, 95% CI 1.13–1.30), map to a nongenic region on chromosome 13q22.1. Five SNPs on 1q32.1 map to *NR5A2*, and the strongest signal was at rs3790844 ($P = 2.45 \times 10^{-10}$, per-allele OR 0.77, 95% CI 0.71–0.84). A single SNP, rs401681 ($P = 3.66 \times 10^{-7}$, per-allele OR 1.19, 95% CI 1.11–1.27), maps to the *CLPTM1L-TERT* locus on 5p15.33, which is associated with multiple cancers. Our study has identified common susceptibility loci for pancreatic cancer that warrant follow-up studies.

Pancreatic cancer is one of the most lethal cancers, with mortality rates approaching its incidence rates[1]. Established risk factors for pancreatic cancer include diabetes, an elevated body-mass index, current or recent smoking and family history of pancreatic cancer[2]. However, only a small fraction of familial aggregation of pancreatic cancer can be explained by previously identified, highly penetrant mutations in *BRCA2*, *CDKN2A* (also known as *p16*), *STK11* (also known as *LKB*), *APC*, *BRCA1*, *PRSS1* and *SPINK*[2,3]. Truncating mutations and deletions in *PALB2* have also recently been shown to be involved in familial pancreatic cancer[4,5].

We recently reported common risk variants for pancreatic cancer that map to the first intron of the *ABO* gene on chromosome 9q34.2 based on a genome-wide association study (GWAS) of 1,896 individuals diagnosed with pancreatic cancer and 1,939 controls[6]. Individuals were drawn from 12 prospective cohort studies (from the Pancreatic Cancer Cohort Consortium) and 1 hospital-based case-control study, the Mayo Clinic Molecular Epidemiology of Pancreatic Cancer Study (Online Methods)[6]. In the first scan, we genotyped approximately 550,000 SNPs and followed up the most significant SNPs that had been found in eight case-control studies (Online Methods)[6].

To identify additional loci, we conducted a second GWAS in which we genotyped approximately 620,000 SNPs in an additional 1,955 cases and 1,995 controls drawn from the same eight case-control studies used to replicate the initial GWAS finding on chromosome 9q34.2. After quality control analysis of genotypes, we combined the datasets, resulting in 551,766 SNPs available for analysis (using Illumina

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

Received 14 October 2009; accepted 17 December 2009; published online 24 January 2010; doi:10.1038/ng.522

Table 1 Association of SNPs on chromosomes 13q22.1, 1q32.1 and 5p15.33 with the risk for pancreatic cancer

Marker[a], alleles[b], chr[c], location[c] and gene[d]	Subset[e]	Rank	MAF[f]		Subjects		χ^{2} [h]	P value[g]	Allelic OR (95% CI)	Genotype OR_{Het} (95% CI)	Genotype OR_{Hom} (95% CI)
			Control	Case	Control	Case					
rs9543325 (T,C)	Cohort	140	0.367	0.416	1,459	1,397	13.55	2.32×10^{-4}	1.23 (1.10–1.37)	1.23 (1.05–1.45)	1.48 (1.18–1.87)
13q22.1 (72814629)	Case-control	3	0.366	0.426	2,182	2,133	31.42	2.08×10^{-8}	1.28 (1.18–1.40)	1.23 (1.08–1.41)	1.68 (1.40–2.02)
None	Combined	1	0.367	0.422	3,641	3,530	44.01	3.27×10^{-11}	1.26 (1.18–1.35)	1.23 (1.11–1.36)	1.61 (1.40–1.86)
rs9564966 (G,A)	Cohort	3,333	0.328	0.364	1,458	1,396	7.54	6.03×10^{-3}	1.17 (1.05–1.31)	1.22 (1.04–1.42)	1.30 (1.02–1.66)
13q22.1 (72794222)	Case-control	9	0.325	0.376	2,179	2,135	23.22	1.44×10^{-6}	1.25 (1.14–1.36)	1.20 (1.06–1.37)	1.60 (1.32–1.95)
None	Combined	6	0.326	0.371	3,637	3,531	29.41	5.86×10^{-8}	1.21 (1.13–1.30)	1.21 (1.09–1.34)	1.48 (1.27–1.72)
Rs3790844 (T,C)	Cohort	821	0.250	0.216	1,459	1,397	10.2	1.40×10^{-3}	0.82 (0.72–0.92)	0.79 (0.68–0.93)	0.72 (0.52–1.00)
1q32.1 (198274055)	Case-control	2	0.239	0.189	2,182	2,135	31.55	1.95×10^{-8}	0.74 (0.67–0.82)	0.72 (0.64–0.82)	0.58 (0.44–0.78)
NR5A2	Combined	2	0.244	0.200	3,641	3,532	40.07	2.45×10^{-10}	0.77 (0.71–0.84)	0.75 (0.68–0.83)	0.64 (0.52–0.79)
rs10919791 (G,A)	Cohort	2,051	0.237	0.205	1,438	1,370	8.42	3.71×10^{-3}	0.83 (0.73–0.94)	0.82 (0.69–0.96)	0.72 (0.51–1.01)
1q32.1 (198231791)	Case-control	1	0.224	0.174	2,177	2,129	31.82	1.69×10^{-8}	0.74 (0.66–0.82)	0.72 (0.63–0.82)	0.57 (0.42–0.78)
NR5A2	Combined	3	0.229	0.186	3,615	3,499	38.2	6.37×10^{-10}	0.77 (0.71–0.84)	0.76 (0.68–0.84)	0.63 (0.50–0.79)
rs3790843 (G,A)	Cohort	781	0.314	0.276	1,459	1,394	10.29	1.34×10^{-3}	0.83 (0.74–0.93)	0.84 (0.71–0.98)	0.69 (0.52–0.90)
1q32.1 (198277447)	Case-control	6	0.297	0.249	2,182	2,134	23.83	1.05×10^{-6}	0.79 (0.72–0.87)	0.77 (0.68–0.87)	0.64 (0.51–0.81)
NR5A2	Combined	4	0.304	0.260	3,641	3,528	33.62	6.69×10^{-9}	0.81 (0.75–0.87)	0.79 (0.72–0.88)	0.66 (0.55–0.79)
rs12029406 (C,T)	Cohort	7,624	0.436	0.404	1,458	1,395	6.06	1.39×10^{-2}	0.88 (0.79–0.97)	0.87 (0.74–1.03)	0.77 (0.62–0.96)
1q32.1 (198172451)	Case-control	8	0.415	0.363	2,182	2,135	23.4	1.32×10^{-6}	0.81 (0.74–0.88)	0.82 (0.72–0.94)	0.64 (0.54–0.77)
NR5A2	Combined	7	0.423	0.379	3,640	3,530	28.31	1.04×10^{-7}	0.83 (0.78–0.89)	0.84 (0.76–0.93)	0.69 (0.60–0.80)
rs4465241 (C,T)	Cohort	970	0.159	0.189	1,459	1,397	9.86	1.69×10^{-3}	1.25 (1.09–1.43)	1.22 (1.03–1.44)	1.69 (1.10–2.59)
1q32.1 (198230245)	Case-control	76	0.155	0.185	2,182	2,134	15.4	8.69×10^{-5}	1.26 (1.12–1.41)	1.24 (1.08–1.42)	1.70 (1.18–2.47)
NR5A2	Combined	9	0.157	0.187	3,641	3,531	25.35	4.79×10^{-7}	1.25 (1.14–1.37)	1.23 (1.11–1.37)	1.68 (1.27–2.23)
rs401681 (C,T)	Cohort	92,235	0.462	0.480	1,459	1,397	1.89	1.70×10^{-1}	1.08 (0.97–1.19)	1.10 (0.92–1.30)	1.15 (0.93–1.42)
5p15.33 (1375087)	Case-control	4	0.437	0.497	2,183	2,135	30.24	3.81×10^{-8}	1.27 (1.17–1.39)	1.28 (1.11–1.48)	1.62 (1.36–1.93)
CLPTM1L	Combined	8	0.447	0.490	3,642	3,532	25.86	3.66×10^{-7}	1.19 (1.11–1.27)	1.20 (1.07–1.34)	1.41 (1.23–1.61)

The results from the unconditional logistic regression of the genotypes generated in a total of 3,851 individuals with pancreatic cancer and 3,934 controls. The analysis was adjusted for age in 10-year categories, sex, study, arm, ancestry and five principal components of population stratification. The SNPs on chromosome 13q22.1 are within a 600-kb intergenic region between *KLF5* and *KLF12*.

[a]NCBI dbSNP identifier. [b]Major allele, minor allele. [c]Chromosome and NCBI Human genome Build 36 location. [d]Gene neighborhood within 20 kb upstream and 10 kb downstream of SNP. [e]Subset: Cohort: cohort studies, Case-control: case-control studies, Combined: all studies. [f]Minor allele frequency. [g]1 d.f. score test. OR, odds ratio; Het, heterozygous; Hom, homozygous for minor allele. CI, 95% confidence interval.

© 2010 Nature America, Inc. All rights reserved.

HumanHap550 and Human 610-Quad chips) in 3,851 individuals with pancreatic cancer and 3,934 controls (Online Methods). A logistic regression model was fit for genotype trend effects (1 degree of freedom (d.f.)) adjusted for study, age, sex, self-described ancestry and five principal components of population stratification. The quantile-quantile plot showed little evidence for inflation of the test statistics as compared to the expected distribution ($\lambda = 1.013$), which excludes the likelihood of substantial hidden population substructure or differential genotype calling between cases and controls (**Supplementary Fig. 1**). A Manhattan plot displays the results of the combined GWAS (**Supplementary Fig. 2a**) and the results from the case-control studies including the full Mayo dataset (**Supplementary Fig. 2b**). Our combined analysis identified three new genomic regions on chromosomes 13q22.1, 1q32.1 and 5p15.33 associated with pancreatic cancer risk that were below the threshold for genome-wide significance ($P < 5 \times 10^{-7}$), as shown in **Table 1** and **Figure 1** (ref. 7). Two different haplotype analyses that involve different test statistics were conducted for each of the three regions: a regularized regression[8] and a sequential haplotype scan[9] (Online Methods). Haplotype analysis across each of the three regions did not identify new or independent markers, thus indicating that the current tag SNPs probably implicate single loci in each region (**Supplementary Fig. 3**).

For the locus on 13q22.1, we observed two highly significant SNPs that ranked number 1 and 6 (most significant and sixth most significant) in the combined analysis: rs9543325 ($P = 3.27 \times 10^{-11}$, per-allele OR 1.26, 95% CI 1.18–1.35; unconstrained heterozygote OR (OR_{het}) 1.23, 95% CI 1.11–1.36 and homozygous OR (OR_{hom}) 1.61, 95% CI 1.40–1.86) and rs9564966 ($P = 5.86 \times 10^{-8}$, per-allele OR 1.21, 95% CI 1.13–1.30; unconstrained OR_{het} 1.21, 95% CI 1.09–1.34 and OR_{hom} 1.48, 95% CI 1.27–1.72). These SNPs, which are 20 kb apart, are highly correlated ($r^2 = 0.82$ in 3,650 study controls of European ancestry and $r^2 = 0.85$ in the HapMap CEU population). SNP rs9564966 was no longer nominally significant after adjusting for rs9543325 ($P = 0.47$), suggesting that the two SNPs mark a single signal in the approximately 600-kb nongenic region between two genes in the family of kruppel-like transcription factors, *KLF5* and *KLF12*, that regulate cell growth and transformation[10,11]. This segment of chromosome 13 is frequently deleted in a spectrum of cancers, including pancreatic cancer[12,13], and may harbor a breast cancer susceptibility locus, as indicated by linkage analysis in families with breast cancer that are negative for mutations in *BRCA1* and *BRCA2* genes[14].

Five highly significant SNPs (ranked 2, 3, 4, 7 and 9 in significance in the combined analysis; $P \leq 5 \times 10^{-7}$) map to a region of chromosome 1q32.1 that harbors *NR5A2* (encoding nuclear receptor subfamily 5, group A, member 2). The SNPs are distributed across a 105-kb genomic region that includes the 5′ end of *NR5A2* and extends to 91 kb upstream of the gene. The two most significant SNPs in this region map to the first intron of *NR5A2* (rs3790844, P = 2.45 × 10^{-10}, per-allele OR 0.77, 95% CI 0.71–0.84; unconstrained OR_{het} 0.75, 95% CI 0.68–0.83 and unconstrained OR_{hom} 0.64, 95% CI 0.52–0.79) and are approximately 32 kb upstream of the gene (rs10919791, P = 6.37 × 10^{-10}; per allele OR 0.77, 95% CI 0.71–0.84; unconstrained OR_{het} 0.76, 95% CI 0.68–0.84 and unconstrained OR_{hom} 0.63, 95% CI 0.50–0.79). The linkage disequilibrium (LD) between these two SNPs is high, with r^2 = 0.81 in study controls and r^2 = 0.71 in the HapMap CEU. In this region, there were three additional SNPs, rs3790843, rs12029406 and rs4465241, that were highly significant ($P < 5 \times 10^{-7}$). Of these three SNPs, the most telomeric one, rs3790843, is highly correlated with rs3790844 and rs10919791 (r^2 = 0.59 and 0.72 in Pancreatic Cancer Cohort Consortium (PanScan) European controls). The two SNPs centromeric to rs3790844 and rs10919791 are less strongly correlated (r^2 = 0.05–0.38 in PanScan European controls). In an analysis adjusted for the most highly associated SNP, rs3790844, three of the other four SNPs, rs10919791, rs3790843 and rs12029406, were no longer nominally significant ($P > 0.05$), whereas the significance of the association with rs4465241 (which had the lowest LD) decreased by several orders of magnitude after adjustment (P = 0.004). Together, these findings suggest that these five SNPs mark a single common allele, but further fine-mapping will be needed to confirm this.

NR5A2 encodes a nuclear receptor of the fushi tarazu (Ftz-F1) subfamily that is predominantly expressed in the exocrine gland of

© 2010 Nature America, Inc. All rights reserved.

Figure 1 Association results, recombination and linkage disequilibrium plots for 13q22.1, 1q32.1 and 5p15.33. Association results are shown in the top panel for all cohort studies (blue squares), case-control studies (green squares) and all studies combined (red diamonds). Overlaid on the association panel for each locus is a plot of recombination rates (cM/Mb) across the region from CEU study controls. (**a**) The LD plot shows a region of chromosome 13q22.1 marked by the SNPs rs9543325 and rs9564966 and bounded by SNPs between 13q22.1:72,721,214 and 13q22.1:72,854,007. These SNPs are within a 600-kb intergenic region between *KLF5* and *KLF12*. (**b**) The LD plot shows a region of chromosome 1q32.1 marked by five SNPs, rs3790844, rs10919791, rs3790843, rs12029406 and rs4465241, and bounded by SNPs between 1q32.1:198,125,014 and 1q32.1:198,317,613. Note that rs3790844 and rs3790843 are located in the first intron of *NR5A2*, shown above the LD plot. (**c**) The LD plot shows a region of chromosome 5p15.33 marked by rs401681 and bounded by SNPs between 5p15.33:1,296,475 and 5p15.33:1,476,905. rs401681 is located in the 13th intron of *CLPTM1L*, shown above the LD plot and 27 kb from the *TERT* gene. For all panels, LD (r^2) is depicted for SNPs with minor allele frequency (MAF) > 5% using PanScan controls of European background (n = 3,650 unrelated individuals). Locations are from NCBI Genome Build 36.

the pancreas, liver, intestine and ovaries in adults. The widespread expression of *NR5A2* in early embryos and the early lethality of *Nr5a2*-knockout mice implies a critical role for this gene in development[15]. NR5A2 plays a role in cholesterol and bile-acid homeostasis, steroidogenesis and cell proliferation (for review, see ref. 16). Evidence for its involvement in cell transformation stems from the fact that NR5A2 interacts with β-catenin to activate expression of cell cycle genes, whereas haploinsufficiency of *NR5A2* attenuates intestinal tumor formation in the $Apc^{Min/+}$ tumor model[17].

The third locus identified is marked by rs401681 ($P = 3.66 \times 10^{-7}$, per-allele OR 1.19, 95% CI 1.11–1.27; unconstrained OR_{het} 1.20, 95% CI 1.07–1.34 and unconstrained OR_{hom} 1.41, 95% CI 1.23–1.61), which maps to chromosome 5p15.33. It resides in intron 13 of *CLPTM1L* (encoding cleft lip and palate transmembrane 1-like), which is part of the *CLPTM1L-TERT* locus that includes *TERT* (encoding telomerase reverse transcriptase), which is only 23 kb away from *CLPTM1L*. Both genes have been implicated in carcinogenesis: *CLPTM1L* is upregulated in cisplatin-resistant cell lines and may play a role in apopotosis[18], whereas *TERT* encodes the catalytic subunit of telomerase, which is essential for maintaining telomere ends. When overexpressed in normal cells, TERT can lead to prolonged cell lifespan and transformation[19,20]. Although telomerase activity cannot be detected in most normal tissues, it is seen in approximately 90% of human cancers[21]. This region of chromosome 5p15.33 has been identified in GWAS of a number of different cancers, including brain tumors, lung cancer, basal cell carcinoma, melanoma and now pancreatic cancer[22–26]. In a recent analysis of lung cancer in smokers, the signal on chromosome 5p15.33 has been shown to be strongly associated with the adenocarcinoma histology subtype[27]. Moreover, another variant in this region, rs402710, that is in LD with our strongest signal, rs401681, has been suggested to be associated with levels of smoking-related bulky aromatic DNA adducts; this is relevant for pancreatic cancer because this cancer is also associated with tobacco use[28]. Germline mutations have been shown to contribute to the development of acute myelogenous leukemia, whereas mutations in *TERT* account for a proportion of individuals with an inherited bone marrow–failure syndrome that is prone to hematologic malignancies[29–31]. SNPs in the *CLPTM1L-TERT* region, including rs401681, are also associated with additional cancers, namely bladder and prostate cancer[22–24]. Notably, the C allele of rs401681 is associated with an increased risk of lung, prostate and bladder cancers, as well as with basal cell carcinoma[22–25], whereas the T allele is associated with increased risk of pancreatic cancer (shown in this study) and melanoma[25]. Lastly, a highly suggestive SNP in this region that did not meet genome-wide significance, rs4635969 (ranked 12th in the combined analysis, $P = 1.05 \times 10^{-6}$), is located between *CLPTM1L* and *TERT* ($r^2 = 0.26$ in 3,650 study controls and $r^2 = 0.36$ in the HapMap CEU population).

It is notable that the estimated ORs for the variants meeting genome-wide significance on chromosomes 13q22, 1q32 and 5p15 were consistent when restricted to data from either the case-control studies or the cohort studies[6]. This similarity of estimated effect size between the two study designs was also observed for rs505922 in the *ABO* locus in our previous report[6]. This consistency of effect supports a role for loci at 13q22.1, 1q32.1, 5p15.33 and *ABO* in risk for pancreatic cancer, whereas the divergent results for *SHH* (reported earlier in ref. 6) on chromosome 7q36 indicate the need for further investigation of the potential influence of study sampling design on detection of risk regions using the GWAS strategy.

GWAS have emerged as a powerful, hypothesis-independent approach to identify common alleles that influence disease risk. Our results show that pancreatic cancer is similar to other complex diseases in that multiple common disease alleles with small effects influence disease risk. Our study has good power to detect common alleles with large effects (over 90% power to detect a per-allele relative risk of 1.4 or greater for an allele with 10% frequency at the $\alpha = 5 \times 10^{-7}$ level) but less power to detect smaller effect sizes. Thus, although it is unlikely that there are common alleles with large effects on most of sporadic pancreatic cancer risk, it is likely that additional susceptibility alleles with moderate to small effects exist. The list of susceptibility alleles should lengthen as further GWAS are performed for pancreatic cancer to catalog the variants with estimated risks below 1.3. Additional studies are needed to assess the clinical utility of risk stratification that combines genetic markers with epidemiologic risk factors already established for pancreatic cancer, namely adiposity, smoking, diabetes and family history.

Our combined analysis of 3,851 individuals with pancreatic cancer and 3,934 controls has yielded three new genomic regions associated with the risk of pancreatic cancer. Two of these regions harbor candidate genes, and the third locus, on chromosome 13q22.1, maps to a large nongenic region analogous to the 8q24 region; however, though the 8q24 region is associated with risk of multiple cancers, including prostate, breast, colorectal and bladder cancers, the locus on chromosome 13q22.1 appears to be specific for pancreatic cancer. The *CPTM1L-TERT* region on chromosome 5p15.33 has been implicated in a disease spectrum that also includes lung cancer, brain tumors, acute myelogenous leukemia, bone marrow failure syndromes and pulmonary fibrosis. The fine-mapping of signals in the three regions identified by our GWAS should guide selection of the optimal variants for functional studies investigating the biological mechanism underpinning pancreatic carcinogenesis. These results, in turn, should help to inform new preventive, diagnostic and/or therapeutic approaches designed to lessen the burden of this highly fatal disease.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS
The authors gratefully acknowledge the energy and contribution of our late colleague Sheila Bingham. Additional acknowledgments are in the **Supplementary Note**.

AUTHOR CONTRIBUTIONS
G.M.P., L.A., C.S.F., P.K., R.Z.S.-S., K.B.J., S.M.L., J.B.M., G.S.T., R.N.H., P.H. and S.J.C. organized and designed the study. L.A., A.H., K.B.J., G.T. and S.J.C. supervised genotyping of samples. L.A., P.K., R.Z.S.-S., C.S.F., K.B.J., C.K., H.P., Z.W., K.Y., R.N.H., P.H. and S.J.C. contributed to the design and execution of statistical analysis. L.A., G.M.P., P.K., R.Z.S.-S., R.N.H., P.H. and S.J.C. wrote the first draft of the manuscript. G.M.P., C.S.F., R.Z.S.-S., A.A.A., H.B.B., S.G., M.G., K.H., E.A.H., E.J.J., A.P.K., A.L., D.L., M.T.M., S.H.O., H.A.R., W.Z., D.A., W.R.B., C.D.B., M.-C.B.-R., J.E.B., P.M.B., F.C., S.C., M.C., M.deA., E.J.D., J.M.G., E.L.G., M.G., G.H., S.E.H., M.H., B.H., D.J.H., M.J., R.K., V.K., R.C.K., R.R.M., D.S.M., A.V.P., P.H.M.P., A.R., E.R., L.R., X.-O.S., A.T., D.T., S.K.V.D.E., J.V., J.W.-W., B.M.W., H.Y., A.Z.-J. and J.F.F.Jr. conducted the epidemiologic studies and contributed samples to the PanScan GWAS and/or replication. All authors contributed to the writing of the manuscript.

COMPETING INTERESTS STATEMENT
The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Ferlay, J., Bray, F., Pisani, P. & Parkin, D.M. *GLOBOCAN 2002: Cancer Incidence, Mortality and Prevalence Worldwide. IARC CancerBase* vol. 5 (IARCPress, Lyon, 2004).

© 2010 Nature America, Inc. All rights reserved.

2. Anderson, K.E.Mack, T.M. & Silverman, D. Cancer of the pancreas. in *Cancer Epidemiology and Prevention* (ed. Schottenfeld, D. & Fraumeni, J.J.) 721–762 (Oxford University Press, New York, (2006)).
3. Shi, C., Hruban, R.H. & Klein, A.P. Familial pancreatic cancer. *Arch. Pathol. Lab. Med.* **133**, 365–374 (2009).
4. Jones, S. *et al.* Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. *Science* **324**, 217 (2009).
5. Tischkowitz, M.D. *et al.* Analysis of the gene coding for the BRCA2-interacting protein PALB2 in familial and sporadic pancreatic cancer. *Gastroenterology* **137**, 1183–1186 (2009).
6. 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).
7. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
8. Li, Y., Sung, W.K. & Liu, J.J. Association mapping via regularized regression analysis of single-nucleotide-polymorphism haplotypes in variable-sized sliding windows. *Am. J. Hum. Genet.* **80**, 705–715 (2007).
9. Yu, Z. & Schaid, D.J. Sequential haplotype scan methods for association analysis. *Genet. Epidemiol.* **31**, 553–564 (2007).
10. Dong, J.T. & Chen, C. Essential role of KLF5 transcription factor in cell proliferation and differentiation and its implications for human diseases. *Cell. Mol. Life Sci.* **66**, 2691–2706 (2009).
11. Nakamura, Y. *et al.* Kruppel-like factor 12 plays a significant role in poorly differentiated gastric cancer progression. *Int. J. Cancer* **125**, 1859–1867 (2009).
12. Chen, C. *et al.* Defining a common region of deletion at 13q21 in human cancers. *Genes Chromosom. Cancer* **31**, 333–344 (2001).
13. Baudis, M. & Cleary, M.L. Progenetix.net: an online repository for molecular cytogenetic aberration data. *Bioinformatics* **17**, 1228–1229 (2001).
14. Kainu, T. *et al.* Somatic deletions in hereditary breast cancers implicate 13q21 as a putative novel breast cancer susceptibility locus. *Proc. Natl. Acad. Sci. USA* **97**, 9603–9608 (2000).
15. Paré, J.F. *et al.* The fetoprotein transcription factor (FTF) gene is essential to embryogenesis and cholesterol homeostasis and is regulated by a DR4 element. *J. Biol. Chem.* **279**, 21206–21216 (2004).
16. Lee, Y.K. & Moore, D.D. Liver receptor homolog-1, an emerging metabolic modulator. *Front. Biosci.* **13**, 5950–5958 (2008).
17. Botrugno, O.A. *et al.* Synergy between LRH-1 and beta-catenin induces G1 cyclin-mediated cell proliferation. *Mol. Cell* **15**, 499–509 (2004).
18. Yamamoto, K., Okamoto, A., Isonishi, S., Ochiai, K. & Ohtake, Y. A novel gene, CRR9, which was up-regulated in CDDP-resistant ovarian tumor cell line, was associated with apoptosis. *Biochem. Biophys. Res. Commun.* **280**, 1148–1154 (2001).
19. Bodnar, A.G. *et al.* Extension of life-span by introduction of telomerase into normal human cells. *Science* **279**, 349–352 (1998).
20. Hahn, W.C. *et al.* Creation of human tumor cells with defined genetic elements. *Nature* **400**, 464–468 (1999).
21. Kim, N.W. *et al.* Specific association of human telomerase activity with immortal cells and cancer. *Science* **266**, 2011–2015 (1994).
22. Wang, Y. *et al.* Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nat. Genet.* **40**, 1407–1409 (2008).
23. McKay, J.D. *et al.* Lung cancer susceptibility locus at 5p15.33. *Nat. Genet.* **40**, 1404–1406 (2008).
24. Rafnar, T. *et al.* Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nat. Genet.* **41**, 221–227 (2009).
25. Stacey, S.N. *et al.* New common variants affecting susceptibility to basal cell carcinoma. *Nat. Genet.* **41**, 909–914 (2009).
26. Shete, S. *et al.* Genome-wide association study identifies five susceptibility loci for glioma. *Nat. Genet.* **41**, 899–904 (2009).
27. 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).
28. Zienolddiny, S. *et al.* The TERT-CLPTM1L lung cancer susceptibility variant associates with higher DNA adduct formation in the lung. *Carcinogenesis* **30**, 1368–1371 (2009).
29. Calado, R.T. *et al.* Constitutional hypomorphic telomerase mutations in patients with acute myeloid leukemia. *Proc. Natl. Acad. Sci. USA* **106**, 1187–1192 (2009).
30. Savage, S.A. & Alter, B.P. Dyskeratosis congenita. *Hematol. Oncol. Clin. North Am.* **23**, 215–231 (2009).
31. Yamaguchi, H. *et al.* Mutations in TERT, the gene for telomerase reverse transcriptase, in aplastic anemia. *N. Engl. J. Med.* **352**, 1413–1424 (2005).

[1]Department of Health Sciences Research, College of Medicine, Mayo Clinic, Rochester, Minnesota, USA. [2]Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA. [3]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department Health and Human Services, Bethesda, Maryland, USA. [4]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [5]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA. [6]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. [7]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA. [8]Core Genotyping Facility, SAIC-Frederick Inc., National Cancer Institute-Frederick, Frederick, Maryland, USA. [9]Bioinformed Consulting Services, Gaithersburg, 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]National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands and Department of Gastroenterology and Hepatology, University Medical Centre Utrecht, Utrecht, The Netherlands. [14]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada. [15]Department of Laboratory Medicine and Pathology, School of Medicine, University of Minnesota, Minneapolis, Minnesota, USA. [16]Prevention and Research Center, Mercy Medical Center, Baltimore, Maryland, USA. [17]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California, USA. [18]Department of Epidemiology, American Cancer Society, Atlanta, Georgia, USA. [19]Department of Oncology, the Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [20]Department of Epidemiology, Bloomberg School of Public Health, The Sol Goldman Pancreatic Research Center, Johns Hopkins Medical Institutions, Baltimore, Maryland, USA. [21]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. [22]Department of Gastrointestinal Medical Oncology, University of Texas M.D. Anderson Cancer Center, Houston, Texas, USA. [23]Group Health Center for Health Studies, Seattle, Washington, USA. [24]Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, New York, USA. [25]Yale University School of Public Health, New Haven, Connecticut, USA. [26]Department of Medicine and Vanderbilt-Ingram Cancer Center, Vanderbilt University, Nashville, Tennessee, USA. [27]Division of Cancer Prevention, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA. [28]Inserm, Paris-Sud University, Institut Gustave-Roussy, Villejuif, France. [29]Divisions of Preventive Medicine and Aging, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. [30]Department of Ambulatory Care and Prevention, Harvard Medical School, Boston, Massachusetts, USA. [31]German Cancer Research Center (DKFZ), Heidelberg, Germany. [32]Johns Hopkins Bloomberg School of Public Health, George W. Comstock Center for Public Health Research and Prevention, Hagerstown, Maryland, USA. [33]Cancer Care Ontario and Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada. [34]Catalan Institute of Oncology (ICO), Barcelona, Spain. [35]Physicians' Health Study, Divisions of Aging, Cardiovascular Disease, and Preventive Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. [36]Massachusetts Veterans Epidemiology Research and Information Center, Veterans Affairs Boston Healthcare System, Boston, Massachusetts, USA. [37]Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA. [38]Departments of Oncology, Pathology and Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [39]Department of Public Health and Clinical Medicine, Nutritional Research, Umeå University, Umeå, Sweden. [40]MedStar Research Institute, Georgetown University, Hyattsville, Maryland, USA. [41]Nutritional Epidemiology Unit, Fondazione Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Istituto Nazionale dei Tumoridi Milano, Milan, Italy. [42]Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York, USA. [43]Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, Maryland, USA. [44]Division of Epidemiology, Public Health and Primary Care, Imperial College London, London, UK. [45]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands. [46]Department of Obstetrics, Gynecology and Reproductive Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [47]Public Health and Participation Directorate, Health and Health Care Services Council, Asturias, Spain. [48]Synergie-Lyon-Cancer, Inserm, Centre Leon Berard, Lyon, Cedex, France. [49]Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, Denmark. [50]Bureau of Epidemiologic Research, Academy of Athens, Athens, Greece. [51]Division of Research, Kaiser Permanente, Northern California Region, Oakland, California, USA. [52]Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland. [53]Department of Social and Preventive Medicine, University at Buffalo, State University of New York, Buffalo, New York, USA. [54]These authors contributed equally to this work. Correspondence should be addressed to S.J.C. (chanocks@mail.nih.gov).

© 2010 Nature America, Inc. All rights reserved.

ONLINE METHODS

Study participants. Participants were drawn from 12 cohort studies and 8 case-control studies[6]. The cohort studies are in the Pancreatic Cancer Cohort Consortium GWAS (PanScan1), part of the National Cancer Institute–sponsored Cohort Consortium. The case-control studies are part of the Pancreatic Cancer Case-Control Consortium (PanC4). The cohort studies include the American Cancer Society Cancer Prevention Study-II (CPS-II)[32]; the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (ATBC)[33]; European Prospective Investigation into Cancer and Nutrition Study (EPIC, which comprises cohorts from Denmark, France, Germany, Great Britain, Greece, Italy, The Netherlands, Spain and Sweden)[34]; Give us a Clue to Cancer and Heart Disease Study (CLUE II)[35]; Health Professionals Follow-up Study (HPFS)[36]; Nurses' Health Study (NHS)[36]; New York University Women's Health Study (NYUWHS)[37], Physicians' Health Study I (PHS I)[36]; Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO)[38]; Shanghai Men's and Women's Health Study (SMWHS); Women's Health Initiative (WHI)[39]; and the Women's Health Study (WHS)[40] (**Supplementary Table 1**). The case-control studies include eight case-control studies from the PanC4 consortium, comprising those from the University of Toronto[41], University of California San Francisco[42], Johns Hopkins University, MD Anderson Cancer Center[43], PACIFIC Study of Group Health and Northern California Kaiser Permanente, Memorial Sloan-Kettering Cancer Center[44] and Yale University[45] and additional cases and controls from the Mayo Clinic Molecular Epidemiology of Pancreatic Cancer Study[46] (**Supplementary Table 2**). 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 from the study.

Each participating study obtained informed consent from study participants and approval from its institutional review board (IRB) for this study and obtained IRB certification permitting data sharing in accordance with the NIH Policy for Sharing of Data Obtained in NIH-Supported or -Conducted Genome-Wide Association Studies (GWAS).

Genotyping and quality control. 8,432 DNA samples, 320 from buccal cells and the remainder extracted from blood products, were selected for genotyping based on quality control measures performed at the Core Genotyping Facility of the National Cancer Institute. 368 samples failed quality control due to a sample completion rate cutoff (**Supplementary Table 2**). The remaining 8,064 samples represent 7,824 distinct individuals. A total of 232 DNA samples were genotyped in duplicate and 4 DNA samples were genotyped in triplicate, which provided a total of 244 plated duplicate pairs.

Genotype clusters were estimated with samples assayed in PanScan I with preliminary completion rates greater than 98%. Genotypes for all samples were called using those clusters. PanScan I samples were divided into four quality groups for batch quality control analysis (QCGROUPs) based on genotype calling metrics: ATBC_PANSCAN_550K, EPIC_PANSCAN_550K, SMWHS_PANSCAN_550K and US_PANSCAN_550K. All PanScan II samples were assigned into a single QCGROUP denoted as PANSCAN2_610K.

Assays for 561,466 loci were attempted on the 4,213 DNA samples in PanScan I using the HumanHap550 Infinium II chip, and assays for 620,901 loci were attempted on the 4,219 DNA samples in PanScan II using the Human 610-Quad chip (Illumina). After quality control, 551,766 SNPs were available for the association analysis. Samples with less than 96% or 98% completion (based on QC group) were excluded. SNP assays with locus call rates lower than 90% were excluded. An average discordance rate of 0.031% was observed for the 244 duplicate pairs.

Deviation from fitness for Hardy-Weinberg proportions was tested[47] for each SNP in control samples of estimated European descent (portion of HapMap CEU ancestry >0.85 by STRUCTURE) of each QCGROUP except the Asian study, SMWHS_PANSCAN_550K (**Supplementary Fig. 4**). SNPs with extreme departures from Hardy-Weinberg proportions ($P < 1 \times 10^{-7}$) were excluded from the association analysis.

Additional participants were excluded based on (i) unanticipated inter-study duplicates (n = 24); (ii) completion rates lower than 96% or 98% for the first and second scans, respectively (n = 368 samples corresponding to 343 participants); (iii) unexpected within-study duplicates (n = 1); (iv) participants who did not meet eligibility requirements (n = 8); and (v) abnormal X chromosome heterozygosity values (n = 6). The final participant count for the combined association analysis was 3,851 cases and 3,934 controls (**Supplementary Table 3**).

We estimated the inflation of the test statistic, λ, adjusted to a sample size of 1,000 cases and 1,000 controls as per the method of de Bakker *et al.*[48] using the formula:

$$\lambda_{\text{corrected}} = 1 + (\lambda - 1) \times [n_{\text{case}}^{-1} + n_{\text{cont}}^{-1}] / [2 \times 10^{-3}].$$

The estimated λ was 1.0035. Assessment of population structure of study participants was performed with STRUCTURE[49] by seeding the analysis with genotypes from HapMap (Phase I and II build 26)[50] and estimating individual admixture coefficients assuming fixed origin and allele frequencies of the members of the three HapMap populations and independence of study participants. A set of 12,898 SNPs with low pairwise correlation ($r^2 < 0.004$) were selected for this analysis[51–53]. A total of 594 participants (315 cases and 279 controls) were estimated to have less than 85% HapMap CEU admixture. No participants were excluded based on results from STRUCTURE, but indicator variables were computed as covariates for the association analysis; participants were classified as 'European' if the HapMap CEU admixture portion was >85%, 'Asian' if the HapMap JPT+CHB admixture was >85% and 'other' if no admixture coefficient was greater than 85% (**Supplementary Fig. 5**). African-American ancestry was defined based on self-report.

A principal-component analysis of samples (excluding inferred sibling and half-sibling pairs) was performed with GLU (a procedure similar to EIGENSTRAT[54]). Five principal components were included as quantitative covariates to correct for population substructure[55].

Ten participant pairs were identified as potential relatives based on genotype sharing in excess of theoretical expectations. A set of 4,546 SNPs was selected (with completion rates > 95%, MAF > 0.3 and $r^2 < 0.01$ in the 3 HapMap populations) and used to run PREST[56]. Seven unexpected full-sibling pairs, one unexpected half-sibling pair and two parent-child pairs (12 cases and 8 controls) were identified and excluded from principal component analysis (but were included in the association analysis).

TaqMan genotyping assays (ABI) were optimized for seven of eight SNPs in the three notable regions to validate the Illumina results. One SNP, rs10919791, could not be manufactured. In an analysis of 2,196 samples from three studies, the comparison of the Illumina calls with the TaqMan assays showed an average concordance rate of 98.2% (with a range of 97.0%–99.8%); no shifts from wild type to homozygotes were observed. The Illumina Infinium genotype probe cluster plots for the eight SNPs are shown in **Supplementary Figure 6**.

Association analysis. All association analyses were conducted using logistic regression, adjusted for age (in 10-year categories), sex, study, arm (for WHI, intervention versus observation), ancestry and five principal components of genetic structure. Each SNP genotype was coded as a count of minor alleles, with the exception of X-linked SNPs among men, which were coded as '2' if the participant carried the minor allele and '0' if he carried the major allele[7,57]. The log-linear odds model has near-optimal power across a wide range of alternative hypotheses, with the exception of those involving rare recessive variants[58]. A score test with 1 d.f. was performed on all genetic parameters in each model. A second, unconstrained model was fit to estimate genotype-specific effects.

We analyzed each study separately and conducted two analyses pooling multiple studies: the first included all cohorts (COHORTS) and the second included all case-control studies (CASE-CONTROL). We assessed heterogeneity in genetic effects across study using the Q and I^2 statistics[59].

We constructed haplotypes from the selected SNPs located in the genomic regions of chromosomes 1q32.1, 5p15.33 and 13q22.1 identified in this scan using fastPHASE. Two approaches were used: (i) the variable-sized sliding-window regularized regression approach[8], in which the maximum window size of a sliding window is determined on the basis of local haplotype diversity and sample size (a regularized regression method is used to tackle the problem of multiple degrees of freedom in the haplotype test[8]); and (ii) the sequential haplotype scan method, which searches for combinations of adjacent markers that are jointly associated with disease status[9]. Association of a single marker with disease is first assessed using the Pearson χ^2 test. Markers are added close to the first one in a sequential manner, but only if the

© 2010 Nature America, Inc. All rights reserved.

contribution of the additional marker to the haplotype association with disease is warranted, conditional on current haplotypes, which is tested using a Mantel-Haenszel statistic.

Data analysis and management was performed with GLU (Genotyping Library and Utilities version 1.0), a suite of tools available as an open-source application for management, storage and analysis of GWAS data. Haplotype analysis was performed using R statistical software.

Estimate of recombination hot spots. SequenceLDhot[60], an approximate marginal likelihood method[61], was used to compute likelihood ratio statistics for a set of putative hot spots across a region. We sequentially analyzed subsets of 100 controls of European background (by pooling 5 controls from each study) and used Phasev2.1[62,63] to infer the haplotypes as well as background recombination rates. To obtain robust results, the analysis was repeated with five nonoverlapping sets of 100 pooled controls.

Data access. The CGEMS data portal provides access to individual level data in 7,785 individuals to investigators from certified scientific institutions after approval of their submitted Data Access Request.

URLs. CGEMS, portal: http://cgems.cancer.gov/; CGF, http://cgf.nci.nih.gov/; GLU, http://code.google.com/p/glu-genetics/; EIGENSTRAT, http://genepath.med.harvard.edu/~reich/EIGENSTRAT.htm; Panc4, http://panc4.org; SNP500Cancer, http://snp500cancer.nci.nih.gov/; STRUCTURE, http://pritch.bsd.uchicago.edu/structure.html; Tagzilla, http://tagzilla.nci.nih.gov/; The R Project for Statistical Computing, http://www.r-project.org/.

32. Calle, E.E. *et al.* The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. *Cancer* **94**, 2490–2501 (2002).
33. The ATBC Cancer Prevention Study Group. The alpha-tocopherol, beta-carotene lung cancer prevention study: design, methods, participant characteristics, and compliance. *Ann. Epidemiol.* **4**, 1–10 (1994).
34. Riboli, E. *et al.* European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr.* **5**, 1113–1124 (2002).
35. Gallicchio, L. *et al.* Single nucleotide polymorphisms in inflammation-related genes and mortality in a community-based cohort in Washington County, Maryland. *Am. J. Epidemiol.* **167**, 807–813 (2008).
36. Wolpin, B.M. *et al.* Circulating insulin-like growth factor binding protein-1 and the risk of pancreatic cancer. *Cancer Res.* **67**, 7923–7928 (2007).
37. Zeleniuch-Jacquotte, A. *et al.* Postmenopausal levels of sex hormones and risk of breast carcinoma in situ: results of a prospective study. *Int. J. Cancer* **114**, 323–327 (2005).
38. Hayes, R.B. *et al.* Methods for etiologic and early marker investigations in the PLCO trial. *Mutat. Res.* **592**, 147–154 (2005).
39. Anderson, G.L. *et al.* Implementation of the Women's Health Initiative study design. *Ann. Epidemiol.* **13**, S5–S17 (2003).
40. Rexrode, K.M., Lee, I.M., Cook, N.R., Hennekens, C.H. & Buring, J.E. Baseline characteristics of participants in the Women's Health Study. *J. Womens Health Gend. Based Med.* **9**, 19–27 (2000).
41. Eppel, A., Cotterchio, M. & Gallinger, S. Allergies are associated with reduced pancreas cancer risk: a population-based case-control study in Ontario, Canada. *Int. J. Cancer* **121**, 2241–2245 (2007).
42. Duell, E.J. *et al.* Detecting pathway-based gene-gene and gene-environment interactions in pancreatic cancer. *Cancer Epidemiol. Biomarkers Prev.* **17**, 1470–1479 (2008).
43. Hassan, M.M. *et al.* Risk factors for pancreatic cancer: case-control study. *Am. J. Gastroenterol.* **102**, 2696–2707 (2007).
44. Olson, S.H. *et al.* Allergies, variants in IL-4 and IL-4R alpha genes, and risk of pancreatic cancer. *Cancer Detect. Prev.* **31**, 345–351 (2007).
45. Risch, H.A. Etiology of pancreatic cancer, with a hypothesis concerning the role of *N*-nitroso compounds and excess gastric acidity. *J. Natl. Cancer Inst.* **95**, 948–960 (2003).
46. McWilliams, R.R. *et al.* Polymorphisms in DNA repair genes, smoking, and pancreatic adenocarcinoma risk. *Cancer Res.* **68**, 4928–4935 (2008).
47. Wigginton, J.E., Cutler, D.J. & Abecasis, G.R. A note on exact tests of Hardy-Weinberg equilibrium. *Am. J. Hum. Genet.* **76**, 887–893 (2005).
48. 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).
49. Pritchard, J.K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
50. Frazer, K.A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
51. Thomas, G. *et al.* Multiple loci identified in a genome-wide association study of prostate cancer. *Nat. Genet.* **40**, 310–315 (2008).
52. Hunter, D.J. *et al.* A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.* **39**, 870–874 (2007).
53. Yu, K. *et al.* Population substructure and control selection in genome-wide association studies. *PLoS One* **3**, e2551 (2008).
54. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
55. Patterson, N., Price, A.L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
56. Sun, L., Wilder, K. & McPeek, M.S. Enhanced pedigree error detection. *Hum. Hered.* **54**, 99–110 (2002).
57. Clayton, D. Testing for association on the X chromosome. *Biostatistics* **9**, 593–600 (2008).
58. Lettre, G., Lange, C. & Hirschhorn, J.N. Genetic model testing and statistical power in population-based association studies of quantitative traits. *Genet. Epidemiol.* **31**, 358–362 (2007).
59. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **21**, 1539–1558 (2002).
60. Fearnhead, P. SequenceLDhot: detecting recombination hotspots. *Bioinformatics* **22**, 3061–3066 (2006).
61. 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).
62. Crawford, D.C. *et al.* Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nat. Genet.* **36**, 700–706 (2004).
63. Li, N.P. & Stephens, M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* **165**, 2213–2233 (2003).

© 2010 Nature America, Inc. All rights reserved.