

Evaluation of Association of *HNF1B* Variants with Diverse Cancers: Collaborative Analysis of Data from 19 Genome-Wide Association Studies

Katherine S. Elliott¹, Eleftheria Zeggini^{1,2}, Mark I. McCarthy^{1,3}, Julius Gudmundsson⁴, Patrick Sulem⁴, Simon N. Stacey⁴, Steinunn Thorlacius⁴, Laufey Amundadottir⁵, Henrik Grönberg⁶, Jianfeng Xu^{7,8}, Valerie Gaborieau⁹, Rosalind A. Eeles¹⁰, David E. Neal¹¹, Jenny L. Donovan¹², Freddie C. Hamdy¹³, Kenneth Muir¹⁴, Shih-Jen Hwang¹⁵, Margaret R. Spitz¹⁶, Brent Zanke^{17,18}, Luis Carvajal-Carmona¹, Kevin M. Brown¹⁹, Australian Melanoma Family Study Investigators²⁰, Nicholas K. Hayward²⁰, Stuart Macgregor²⁰, Ian P. M. Tomlinson¹, Mathieu Lemire¹⁷, Christopher I. Amos¹⁶, Joanne M. Murabito²¹, William B. Isaacs²², Douglas F. Easton²³, Paul Brennan⁹, The PanScan Consortium²⁴, Rosa B. Barkardottir^{24,25}, Daniel F. Gudbjartsson⁴, Thorunn Rafnar⁴, David J. Hunter²⁶, Stephen J. Chanock⁵, Kari Stefansson⁴, John P. A. Ioannidis^{27,28*}

1 Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom, **2** Applied Statistical Genetics, Wellcome Trust Sanger Institute, University of Cambridge, Cambridge, United Kingdom, **3** Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, United Kingdom, **4** deCODE Genetics, Reykjavik, Iceland, **5** Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, United States of America, **6** Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, **7** Center for Cancer Genomics, Wake Forest University School of Medicine, Winston-Salem, North Carolina, United States of America, **8** Center for Human Genomics, Wake Forest University School of Medicine, Winston-Salem, North Carolina, United States of America, **9** Genetic Epidemiology Group, International Agency for Research on Cancer (IARC), Lyon, France, **10** Oncogenetics Team, The Institute of Cancer Research, Sutton, United Kingdom, **11** Department of Oncology, University of Cambridge, Cambridge, United Kingdom, **12** Department of Social Medicine, University of Bristol, Bristol, United Kingdom, **13** Nuffield Department of Surgery, University of Oxford, Oxford, United Kingdom, **14** Health Sciences Research Institute, University of Warwick, Coventry, United Kingdom, **15** Framingham Study, National Heart, Lung, and Blood Institute, Bethesda, Maryland, United States of America, **16** Department of Epidemiology, M. D. Anderson Cancer Center, Houston, Texas, United States of America, **17** Ontario Institute for Cancer Research, MaRS Centre, Toronto, Ontario, Canada, **18** Ottawa Health Research Institute, University of Ottawa, Ottawa, Ontario, Canada, **19** Integrated Cancer Genomics Division, The Translational Genomics Research Institute, Phoenix, Arizona, United States of America, **20** Queensland Institute of Medical Research, Royal Brisbane Hospital, Brisbane, Queensland, Australia, **21** Section of General Internal Medicine, Boston University School of Medicine, Boston, Massachusetts, United States of America, **22** The Brady Urological Institute, Johns Hopkins Medical Institutions, Baltimore, Maryland, United States of America, **23** Cancer Research UK Genetic Epidemiology Unit, University of Cambridge, Cambridge, United Kingdom, **24** Department of Pathology, Landspítali-University Hospital of Iceland, Reykjavik, Iceland, **25** Faculty of Medicine, University of Iceland, Reykjavik, Iceland, **26** Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, United States of America, **27** Department of Hygiene and Epidemiology, University of Ioannina School of Medicine and Biomedical Research Institute, Foundation for Research and Technology-Hellas, Ioannina, Greece, **28** Center for Genetic Epidemiology and Modelling, Tufts University School of Medicine, Boston, Massachusetts, United States of America

Abstract

Background: Genome-wide association studies have found type 2 diabetes-associated variants in the *HNF1B* gene to exhibit reciprocal associations with prostate cancer risk. We aimed to identify whether these variants may have an effect on cancer risk in general versus a specific effect on prostate cancer only.

Methodology/Principal Findings: In a collaborative analysis, we collected data from GWAS of cancer phenotypes for the frequently reported variants of *HNF1B*, rs4430796 and rs7501939, which are in linkage disequilibrium ($r^2 = 0.76$, HapMap CEU). Overall, the analysis included 16 datasets on rs4430796 with 19,640 cancer cases and 21,929 controls; and 21 datasets on rs7501939 with 26,923 cases and 49,085 controls. Malignancies other than prostate cancer included colorectal, breast, lung and pancreatic cancers, and melanoma. Meta-analysis showed large between-dataset heterogeneity that was driven by different effects in prostate cancer and other cancers. The per-T2D-risk-allele odds ratios (95% confidence intervals) for rs4430796 were 0.79 (0.76, 0.83) per G allele for prostate cancer ($p < 10^{-15}$ for both); and 1.03 (0.99, 1.07) for all other cancers. Similarly for rs7501939 the per-T2D-risk-allele odds ratios (95% confidence intervals) were 0.80 (0.77, 0.83) per T allele for prostate cancer ($p < 10^{-15}$ for both); and 1.00 (0.97, 1.04) for all other cancers. No malignancy other than prostate cancer had a nominally statistically significant association.

Conclusions/Significance: The examined *HNF1B* variants have a highly specific effect on prostate cancer risk with no apparent association with any of the other studied cancer types.

Citation: Elliott KS, Zeggini E, McCarthy MI, Gudmundsson J, Sulem P, et al. (2010) Evaluation of Association of *HNF1B* Variants with Diverse Cancers: Collaborative Analysis of Data from 19 Genome-Wide Association Studies. PLoS ONE 5(5): e10858. doi:10.1371/journal.pone.0010858

Editor: Florian Kronenberg, Innsbruck Medical University, Austria

Received: April 8, 2010; **Accepted:** April 28, 2010; **Published:** May 28, 2010

Copyright: © 2010 Elliott et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: JPAI is supported by funding for the Tufts Clinical and Translational Science Institute and the Center for Genetic Epidemiology and Modeling is supported by grant UL1 RR025752 from the National Institutes of Health. KSE is supported by the Wellcome Trust (WT075491/Z/04). EZ is supported by the Wellcome Trust (WT088885/Z/09/Z). NH and SM were funded by the Australian National Health and Medical Research Council (NHMRC) and the National Institutes of Health (grant R01-CA88363). HG is supported by Swedish Cancer Society (Cancerfonden), Swedish Research Council, and Linneus grant "Prediction and prevention of breast cancer and prostate cancer". RE and KM are supported by Cancer Research UK Grant C5047/A8385; C5047/A7357, The Institute of Cancer Research and The Everyman Campaign and The Prostate Cancer Research Foundation. The authors acknowledge support from the NIHR to the Biomedical Research Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust. DEN, JLD and FCH would like to acknowledge the support of the National Cancer Research Institute (NCRI) formed by the Department of Health, the Medical Research Council and Cancer Research UK. The Prostate Testing for Cancer and Treatment (ProtecT) study is funded by the UK National Institute for Health Research Health Technology Assessment Programme (projects 96/20/06, 96/20/99). The NCRI provided funding through ProMPT (Prostate Mechanisms of Progression and Treatment) and this support is gratefully acknowledged. The Cambridge prostate biorepository also received funding from the NIHR Comprehensive Biomedical Research Centre Grant. This support is gratefully acknowledged. The views and opinions expressed therein are those of the authors and are not necessarily those of the funding bodies. deCODE acknowledges grant support from the 6th Framework Program of the EU: contract 018827 (Polygene). The Australian Melanoma Family Study (AMFS) was supported by grants from the NHMRC, Cancer Councils of Queensland, NSW and Victoria, and by National Institutes of Health grant R01-CA83115 to the GenoMEL consortium. The Framingham Heart Study (FHS) The Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine were supported by the National Heart, Lung and Blood Institute's Framingham Heart Study Contract No. N01-HC-25195 and its contract with Affymetrix, Inc for genotyping services (Contract No. N02-HL-6-4278). Analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. A portion of this research was conducted using the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: jioannid@cc.uoi.gr

✉ See acknowledgments for author list.

Introduction

A large number of epidemiological studies have suggested correlations between type 2 diabetes (T2D) and various cancers [1,2,3]. Most evidence suggests an inverse correlation between T2D and prostate cancer [4,5,6] although not all studies agree on this [7]. Several studies also suggest positive correlations between other cancers and T2D [1,2,3]. It is unclear whether these correlations, if true, represent causal relationships and whether they may also reflect some shared genetic background. Recently, with the advent of genome-wide association studies (GWAS), a large number of genetic variants have been identified that confer susceptibility to T2D or specific types of cancer [8]. An interesting observation has been that specific variants in the *HNF1B* gene (formerly *TCF2*) have been demonstrated to be associated both with the risk of prostate cancer [9,10,11] and the risk of T2D [9,12] with the effects being in the opposite direction for these two phenotypes.

HNF1B was previously known to be mutated in individuals with maturity-onset diabetes of the young type 5 (MODY 5) [13], but a biological explanation of the impact of the identified common variation on T2D and prostate cancer risk remains elusive. The identified genetic effects are small in magnitude even for prostate cancer and T2D, representing odds ratios [ORs] per allele in the range of 1.2 [9,11] and 0.9 [9,12], respectively. Therefore, small effects for other cancer types would not be readily detectable, unless very large studies were performed or data were combined from several studies.

A definitive answer on whether *HNF1B* variants modulate also the risk of other malignancies, or show specificity for prostate cancer, requires large sample sizes. Here we present the results of a large collaborative meta-analysis of *HNF1B*, rs4430796 and rs7501939, which have the most consistent associations with both prostate cancer and T2D. Relevant data were collected on the two variants from GWAS on cancer phenotypes in Caucasian populations in order to examine whether they have an effect on cancer risk in general, on few specific cancers, or only on prostate cancer.

Results

Database of contributed information

All the originally contacted investigators of cancer-related GWA studies agreed to participate in this collaborative analysis, with the exception of the investigators of 3 GWA studies [14,15,16] (1 on

breast cancer, 1 on colorectal cancer and 1 on neuroblastoma), 1 of which had no data on the requested variants, as they had used a Affymetrix platform [15]. Investigators who agreed to participate in the collaborative analysis contributed data on 13 datasets for rs4430796 and 19 datasets for rs7501939 [11,17,18,19, 20,21,22,23,24,25,26,27,28,29,30,31,32,33]. For 5 datasets, data were available only for the latter polymorphism either because the polymorphism was not available on the platform used or the SNP failed quality control criteria.

The contributing teams and datasets are shown in Table 1 with data on the number of cases and controls for each polymorphism and for each type of cancer. Datasets from the Framingham cohort contained imputed data for both polymorphisms since an Affymetrix platform had been used, rs4430796 data from the M.D. Anderson Cancer Center was imputed since this SNP had not been directly genotyped, and melanoma data from AMFS and Q-MEGA contained counts from pooling experiments, otherwise all other datasets had direct genotyping on individual participants. Detailed demographic and other characteristics of the study populations can be found in the respective primary publications of these GWA studies [14,15,16,17,18,20,21,22,23,24,25,26,27,28, 29,30,31,32,34].

Overall, the collaborative analysis included data on rs4430796 for 19,640 cancer cases and 21,929 controls; for prostate cancer there were 11,145 cases and 9,650 controls, while for all other cancers there were 8,495 cases and 12,279 controls. The collected data on rs7501939 included 26,923 cases and 49,085 controls; for prostate cancer there were 12,898 cases and 40,371 controls, while for the other cancers there were 14,025 cases and 43,893 controls. Malignancies other than prostate cancer in these datasets included colorectal, breast, lung and pancreatic cancers, and melanoma (Table 1). deCODE contributed data on 4 different cancers and had a common population control group for all 4 of them. The Framingham Heart Study (FHS) contributed data on 4 different cancers and had a common population control group (subjects ≥ 65 years at the last contact who are not nuclear family member of the cancer cases) for all 4 studies with the exception of prostate and breast cancer which used male and female only controls respectively. The common control groups for deCODE and FHS are only counted once in the total sample sizes above.

The meta-analysis of all datasets (Table 2, Figure 1) showed a per T2D risk allele association with both rs4430796 (G allele OR 0.91 [95% CI: 0.88, 0.94] $p = 3 \times 10^{-10}$) and rs7501939 (T allele

Table 1. Characteristics of datasets included in the collaborative meta-analysis.

Study Centre	Cancer	Genotyping platform(s)	rs4430796 #cases	rs4430796 #controls	rs7501939 #cases	rs7501939 #controls
*ARCTIC	colorectal[23]	Sequenom homogenous MassExtend (in house)	1,079	1,089	1,075	1,087
*AMFS	melanoma[17,24]	Illumina 550K (pooled)	490 ^P	427 ^P	490 ^P	427 ^P
Cambridge	breast [33]	Perlegen	387	363	387	363
*CGEMS	prostate[11,25]	Illumina 550K	4,960	5,021	4,869	4,930
*CAPS	prostate[26]	Sequenom (in house)	2,874	1,708	2,865	1,707
*CORGI	colorectal[27]	Illumina 550K	n/a	n/a	900	908
deCODE	breast[28]	Illumina 300K	n/a	n/a	1,815	30,742
deCODE	colorectal[29]	Illumina 300K	n/a	n/a	988	30,742
deCODE	lung[29,30]	Illumina 300K	n/a	n/a	651	30,742
deCODE	prostate[9,31,32]	Illumina 300K	n/a	n/a	1619	30,742
*FHS	breast[34]	Affymetrix 500K and MIPS 50K combined	182 ⁱ	852 ⁱ	182 ⁱ	852 ⁱ
*FHS	colorectal[34]	Affymetrix 500K and MIPS 50K combined	108 ⁱ	1,498 ⁱ	108 ⁱ	1,498 ⁱ
*FHS	lung[34]	Affymetrix 500K and MIPS 50K combined	90 ⁱ	1,498 ⁱ	90 ⁱ	1,498 ⁱ
*FHS	prostate[34]	Affymetrix 500K and MIPS 50K combined	190 ⁱ	646 ⁱ	190 ⁱ	646 ⁱ
*IARC	lung[20,21]	Illumina 300K	641	2,435	1,797	2,378
*JHH	prostate[26]	Sequenom (in house)	1,512	478	1,521	479
*MDACC	lung[22]	Illumina 317K	1,152 ⁱ	1,137 ⁱ	1,152	1,137
*PANSCAN	pancreatic Stage 1[19,48]	Illumina 550K and 610K	1,754	1,796	1,757	1,796
*PANSCAN	pancreatic Stage 2[19,48]	Illumina 550K and 610K	1,748	1,818	1,769	1,841
*Q-MEGA	melanoma[24]	Illumina 550K (pooled)	864 ^P	864 ^P	864 ^P	864 ^P
*UKGPCS	prostate[18]	Illumina 550K	1,609	1,797	1,834	1,867

Unless otherwise indicated all data is from direct genotyping. *ARCTIC (Assessment of Risk for Colorectal Tumors in Canada), AMFS (Australian Melanoma Family Study), CGEMS (Cancer Genetics Markers of Susceptibility), CAPS (Cancer of the Prostate in Sweden), CORGI (Colorectal Tumour Gene Identification), FHS (Framingham Heart Study), IARC (International Agency for Research on Cancer), JHH (Johns Hopkins Hospital), MDACC (M.D. Anderson Cancer Center, Texas), PANSCAN (Pancreatic Cancer Cohort Consortium), Q-MEGA (Queensland study of Melanoma: Environment and Genetic Associations), UKGPCS (UK Genetic Prostate Cancer Study). n/a: no available data; ⁱ: imputed; ^P: pooled.

doi:10.1371/journal.pone.0010858.t001

OR 0.91 [95% CI: 0.88, 0.94] $p = 5 \times 10^{-10}$) according to fixed effects calculations, while by random effects calculations there was nominal significance (OR 0.94 [95% CI: 0.88, 1.00], $p = 0.033$ for rs4430796 and 0.93 [95% CI: 0.86, 1.01], $p = 0.07$ for rs7501939). The reason for this diversity is that there was very large between-study heterogeneity in the effect sizes (I^2 of 82% [95% CI: 73–89%] and 80% [95% CI: 70–86%], respectively, for the two polymorphisms; Q-test p -value < 0.001 for both polymorphisms), and this makes the fixed effects calculations less reliable. Results were qualitatively similar when we increased the variance in deCODE, FHS, and IARC estimates to account for the overlapping control group (not shown).

The heterogeneity was largely driven by the diversity in the effect sizes between prostate cancer and all other cancers. A meta-analysis limited to prostate cancer datasets gave consistent associations with both rs4430796 (OR per copy of T2D risk allele (A) 0.79 [95% CI: 0.76, 0.83], $p < 10^{-15}$ by fixed effects and 0.79 [95% CI: 0.74, 0.84] $p = 10^{-13}$ by random effects), and rs7501939 (OR per copy of T2D risk allele (T) 0.80 [95% CI: 0.77, 0.83] $p < 10^{-15}$ by fixed effects and 0.79 [95% CI: 0.74, 0.85], $p = 2 \times 10^{-11}$ by random effects) (Table 2). There was some residual between-study heterogeneity even within the prostate cancer datasets (I^2 of 42% [95% CI: 0–79%] and 56% [95% CI:

0–82%], respectively, for the two polymorphisms; Q-test p -value 0.037 and 0.14, respectively), although the heterogeneity pertained only to the exact magnitude of the genetic effects and a nominally statistically significant association was seen in each of the datasets except for the Framingham study where the number of prostate cancer cases was more limited.

Conversely, the results for all other cancers suggested no significant association and results were consistent across studies. The summary OR was 1.03 and 1.00 for the two polymorphisms respectively ($p = 0.14$ and 0.81 by fixed effects) and the 95% CIs excluded ORs deviating more than 7% from the null (OR = 1.00) for rs4430796 and more than 4% from the null for rs7501939 (Table 2). The Q-test p -value was 0.99 and 0.45 for the two polymorphisms respectively and random effects estimates were thus identical to fixed effects estimates.

There was also no convincing evidence for an association between either of the two polymorphisms and any of the other cancers (besides prostate cancer), when each cancer type was evaluated separately. Point estimates were in the opposite direction (odds ratio 1.03–1.05) for pancreatic and lung cancer, but were not nominally statistically significant (Table 2). The difference between the prostate cancer and other cancers' effect estimates was beyond chance ($p < 0.05$) for both polymorphisms.

Table 2. Summary of results for association between rs4430796 and rs7501939 and diverse cancer types.

Cancer type	rs4430796	rs4430796	rs4430796	rs7501939	rs7501939	rs7501939
	Studies (cases, controls)	OR (95% CI)	I ² (95% CI)	Studies (cases, controls)	OR (95% CI)	I ² (95% CI)
All cancers	16 (19,640, 21,929)*	0.91 (0.88, 0.94)	82 (73, 89)	21 (26,923, 49,085)*	0.92 (0.90, 0.95)	80 (70, 86)
Prostate	5 (11,145, 9,650)	0.79 (0.76, 0.83)	42 (0, 79)	6 (12,898, 40,371)	0.80 (0.77, 0.83)	56 (0, 82)
All Others	11 (8,495, 12,279)*	1.03 (0.99, 1.07)	0 (0, 60)	15 (14,025, 43,893)*	1.00 (0.97, 1.04)	0 (0, 54)
Breast	2 (569, 1,215)	1.00 (0.84, 1.20)	n/a	3 (2,384, 31,957)	0.97 (0.91, 1.04)	0 (0, 90)
Lung	3 (1,883, 5,070)	1.05 (0.98, 1.13)	0 (0, 90)	4 (3,690, 35,755)	1.03 (0.96, 1.10)	0 (0, 85)
Colorectal	2 (1,187, 2,587)	1.01 (0.90, 1.14)	n/a	4 (3,071, 34,235)	1.01 (0.94, 1.08)	0 (0, 85)
Melanoma	2 (1,354, 1,291)	0.98 (0.87, 1.01)	n/a	2 (1,354, 1,291)	1.01 (0.90, 1.13)	n/a
Pancreatic	2 (3,502, 3,614)	1.04 (0.98, 1.11)	n/a	2 (3,526, 3,637)	1.03 (0.97, 1.10)	n/a

OR: odds ratio, CI: confidence interval, n/a: not applicable (heterogeneity I² confidence intervals are not calculated when there are only 2 studies). Odds ratios are based on fixed effects calculations. When point estimates or confidence intervals differ by over 1% in random effects calculations, random effects results are mentioned in the text. * the common control groups of deCODE and FHS are counted only once.

doi:10.1371/journal.pone.0010858.t002

Discussion

The current collaborative analysis documents that both rs4430796 and rs7501939 have robust support for association with prostate cancer, while we did not observe any convincing evidence for an association of any of the other cancers examined with either polymorphism. When data from all other cancers, excluding prostate cancer, were combined the summary effects had 95% CIs that excluded even subtle associations. Apart from prostate cancer, when other datasets for each individual cancer type was combined, the 95% CIs consistently excluded associations with modest effects. This would suggest that the effects mediated by these polymorphisms are specific to T2D and prostate cancer and they do not involve any other cancer types.

The *HNF1B* gene encodes a transcription factor and it was initially identified as a *MODY* gene[13]. Subsequent studies have suggested that mutations in this gene may also be related to renal disease[35] and chromophobe renal cell carcinoma[36]. No GWAS evaluating kidney cancer were included in our analysis, and no kidney cancer GWAS has been published to-date. The expression profile of the gene shows a tissue-specific pattern. It is essential for embryonic survival and is expressed in the gut, kidney, liver, lung, pancreas, prostate, thymus and genital tract [37,38]. It could be speculated that the lack of association with some cancers studied here may be due to the low or absent expression of this gene in those tissues (for example breast cancer). We did not have data on liver cancer, thymoma or genital tract cancer, but data on lung, pancreatic, and colorectal cancer showed no association, with point estimates very near to the null.

The two variants that we assessed are not necessarily the functional culprits. GWA studies typically derive markers of phenotypes that are probably linked with the functional genetic variation[39]. However, identifying the functional variants is difficult. Even if they could be identified, it is unlikely that substantially large genetic effects for other cancers would exist, if the tagging variants have so consistently null effects. Another caveat is that we only examined populations of Caucasian descent. This reduces the heterogeneity that could be due to different LD patterns in populations of different ancestry. However, it would be worthwhile to investigate the associations of the *HNF1B* variants for T2D, prostate cancer, and other cancers, also in non-Caucasian populations. Preliminary data suggest that both of the examined variants had consistent

associations with T2D in Caucasian, Asian (Hong Kong), and West African ancestry participants[9], while the association of rs4430796 with prostate cancer risk was found to be even stronger in the Japanese than in Caucasian populations[40]. Moreover, it would be useful to dissect associations with specific disease subsets. Even within the analyzed Caucasian-descent populations, we observed some modest between-study heterogeneity in the strength of the association between the *HNF1B* variants and prostate cancer. This may be due to different associations in different sub-phenotypes. For example, some data suggest that the rs4430796 A allele may primarily increase the risk for early-onset (<50 years) prostate cancer rather than later-onset disease[41].

In conclusion, while the two examined *HNF1B* variants conclusively have pleiotropic effects on both T2D and prostate cancer, the pleiotropy apparently does not extend to other cancer types. Genetic effects may offer a way to dissect comorbidity between specific cancers and metabolic phenotypes. Besides *HNF1B*, other gene loci have started appearing where variants are identified that modulate susceptibility to both T2D and some malignancy, e.g. prostate cancer for the *JAZF1* locus gene [11,42] and melanoma for the *CDKN2A* locus [43], although different, unlinked variants are implicated in the susceptibility to the malignancy and T2D, respectively. The elucidation of correlated pleiotropic effects on diverse phenotypes will require very large studies, given the generally subtle effects involved. Collaborative efforts between multiple teams, as in the current study, may offer a suitable approach to answer such questions.

Methods

Eligible GWA investigations and data

We used the NHGRI catalogue of published GWA studies[44], a comprehensive database of GWA investigations to identify GWA studies on cancer phenotypes published as of May 20, 2008. We also performed additional PubMed searches to identify whether any additional GWA studies on cancer phenotypes had been published until then. We focused on solid cancers, excluding hematologic malignancies. Given that these GWAS did not include any studies on pancreatic cancer (of special interest, given the association with T2D), we also identified GWAS on pancreatic cancer that had not been published by that time, so as to ensure their inclusion.

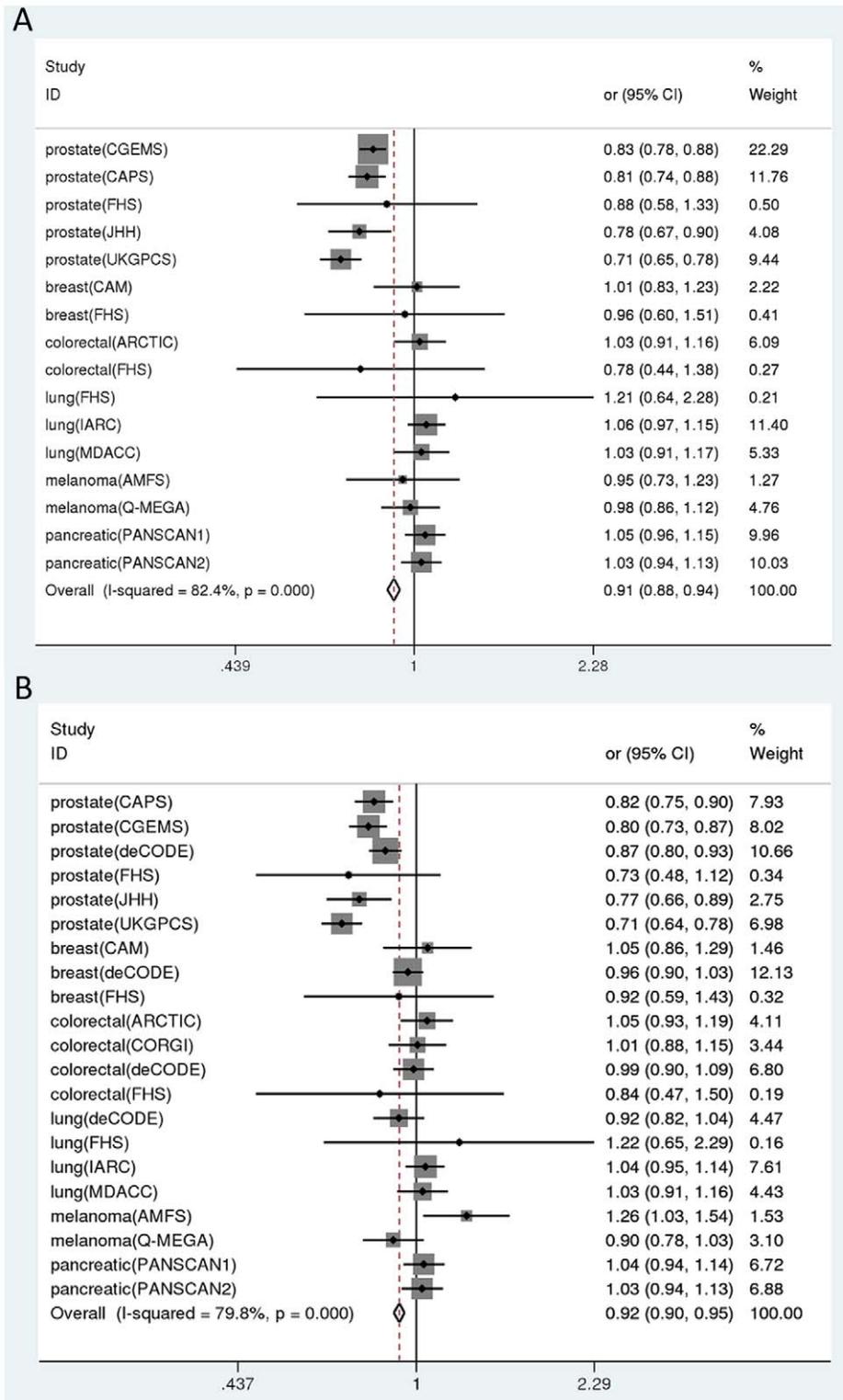


Figure 1. Association of rs4430796 and rs7501939 with diverse cancer types. Panel A shows results for rs4430796 and panel B shows results for rs7501939. Each study is shown by its odds ratio and 95% confidence intervals. Prostate cancer studies appear on the top and other cancer studies follow in alphabetical order. For the abbreviations of the names of the studies see Table 1. The summary diamond at the bottom corresponds to the fixed effects summary. Weight indicates the relative proportion of the total evidence found in each study (the weight is inversely proportional to the variance).

doi:10.1371/journal.pone.0010858.g001

We communicated with the corresponding and principal investigators of all of these studies to request their participation in the collaborative meta-analysis. The investigators of these studies were asked to contribute relevant data on genotype frequencies in cancer cases and non-cancer controls for the *HNF1B* variants, rs4430796 and rs7501939. The risk alleles for prostate cancer are A and C for rs4430796 and rs7501939 respectively. The risk alleles for T2D are G and T for rs4430796 and rs7501939 respectively. The two SNPs have modestly high LD in Caucasians, but low LD in Africans ($r^2 = 0.77$ and 0.22 in CEU and YRI, respectively). Investigators were requested to provide all GWA data that they had obtained for evaluation of any cancer phenotype, including any additional unpublished datasets. Additional genotyping for the two specific variants was encouraged, when a GWA platform had been used that did not directly genotype these polymorphisms (e.g. Affymetrix or Perlegen rather than Illumina). When a study had data on more than one cancer type, data were requested to be provided separately for each cancer type. Investigators were asked to provide also information and clarifications about the design of their studies, and to ensure that population stratification and cryptic relatedness had been appropriately addressed and appropriate quality controls were available for the genotyping. All GWAS investigations that contributed data on these SNPs used stringent QC standards (as described in detail in their original publications) and the two SNPs fulfilled these standards. Approval from local institutional review boards and steering committees was obtained, as deemed necessary for each study by its investigators. The contributed data were checked for completeness and with logical queries and any missing or unclear information was clarified through communication with the contributing investigators.

Meta-analysis

For each SNP, we performed meta-analyses including the data from all eligible cancer studies (regardless of the specific cancer phenotype addressed) and also subgroup meta-analyses, with each subgroup limited to studies addressing a specific cancer phenotype. A separate analysis compared the results of the association for prostate cancer versus the association for all other cancers combined.

All analyses followed the per allele (log-additive) model of inheritance with effect sizes expressed in the odds ratio (OR) scale using both fixed and random effects models[45]. Heterogeneity testing used the Q statistic (considered statistically significant at $p < 0.10$), and the I^2 metric[46] and its 95% CIs [47]. Analyses excluding data from studies with pooled genotyping gave similar results (not shown).

Based on the accumulated total sample size and given the minor allele frequencies of these two polymorphisms in HapMap CEU (47% for rs4430796 A allele and 47% for rs7501939 T allele), the meta-analysis had 95% or higher power to detect an association of $OR = 1.10$ at $\alpha = 0.05$ with each of the two polymorphisms for overall cancer risk, prostate cancer risk, or other cancer risk. Reported p-values are two-tailed. Analyses were performed in STATA 10.0 (College Station, Texas).

Acknowledgments

Australian Melanoma Family Study investigators: Graham J Mann¹, John L Hopper², Joanne F Aitken³, Richard F Kefford¹, Graham G Giles¹, Bruce K Armstrong³.

¹Westmead Institute of Cancer Research, University of Sydney at Westmead Millennium Institute and Melanoma Institute Australia, Westmead, New South Wales, Australia. ²Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, School of Population

Health, University of Melbourne, Melbourne, Victoria, Australia. ³ Viertel Centre for Research in Cancer Control, The Cancer Council Queensland, Brisbane, Queensland, Australia. ⁴Cancer Epidemiology Centre, The Cancer Council Victoria, Carlton, Victoria, Australia. ⁵School of Public Health, University of Sydney, New South Wales, Australia.

PanScan Consortium: Gloria M. Petersen^{*1}, Laufey Amundadottir^{*2,3}, Charles S. Fuchs^{*4,5}, Peter Kraft^{*6,7}, Rachael Z. Stolzenberg-Solomon^{*8}, Kevin B. Jacobs^{3,8,9}, Alan A. Arslan^{10,11,12}, H. Bas Bueno-de-Mesquita¹³, Steven Gallinger¹⁴, Myron Gross¹⁵, Kathy Helzlsouer¹⁶, Elizabeth A. Holly¹⁷, Eric J. Jacobs¹⁸, Alison P. Klein^{19,20}, Andrea LaCroix²¹, Donghui Li²², Margaret T. Mandelson^{21,23}, Sara H. Olson²⁴, Harvey A. Risch²⁵, Wei Zheng²⁶, Demetrius Albanes³, William R. Bamlet¹, Christine D. Berg²⁷, Marie-Christine Boutron-Ruault²⁸, Julie E. Buring^{29,30}, Paige M. Bracci¹⁷, Federico Canzian³¹, Sandra Clipp³², Michelle Cotterchio³³, Mariza de Andrade¹, Eric J. Duell³⁴, J. Michael Gaziano^{35,36}, Edward L. Giovannucci^{5,6,37}, Michael Goggins³⁸, Göran Hallmans³⁹, Susan E. Hankinson^{5,6}, Manal Hassan²², Barbara Howard⁴⁰, David J. Hunter^{5,6}, Amy Hutchinson^{3,8}, Mazda Jenab³⁴, Rudolf Kaaks³¹, Charles Kooperberg²¹, Vittorio Krogh¹¹, Robert C. Kurtz⁴², Shannon M. Lynch³, Robert R. McWilliams¹, Julie B. Mendelsohn³, Dominique S. Michaud^{6,43}, Hemang Parikh^{2,3}, Alpa V. Patel¹⁸, Petra H.M. Peeters^{43,44}, Aleksandar Rajkovic⁴⁵, Elio Riboli⁴³, Laudina Rodriguez⁴⁶, Daniela Seminara⁴⁷, Xiao-Ou Shu²⁶, Gilles Thomas^{3,48}, Anne Tjonneland⁴⁹, Geoffrey S. Tobias³, Dimitrios Trichopoulos^{6,50}, Stephen K. Van Den Eeden⁵¹, Jarmo Virtamo⁵², Jean Wactawski-Wende⁵³, Zhaoming Wang^{3,8}, Brian M. Wolpin^{4,5}, Herbert Yu²⁵, Kai Yu³, Anne Zeleniuch-Jacquotte^{11,12}, Joseph F. Fraumeni, Jr.³, Robert N. Hoover^{*3}, Patricia Hartge^{*3}, Stephen J. Chanock^{*2,3}.

¹Department of Health Sciences Research, College of Medicine, Mayo Clinic, Rochester, MN, USA. ²Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA. ³Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA. ⁴Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. ⁵Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. ⁶Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA. ⁷Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA. ⁸Core Genotyping Facility, SAIC-Frederick Inc., NCI-Frederick, Frederick, MD, USA. ⁹Bioinformed Consulting Services, Gaithersburg, MD, USA. ¹⁰Department of Obstetrics and Gynecology, New York University School of Medicine, New York, NY, USA. ¹¹Department of Environmental Medicine, New York University School of Medicine, New York, NY, USA. ¹²New York University Cancer Institute, New York, NY, USA. ¹³National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands and Department of Gastroenterology and Hepatology, University Medical Centre Utrecht, Utrecht, The Netherlands. ¹⁴Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada. ¹⁵Department of Laboratory Medicine/Pathology, School of Medicine, University of Minnesota, Minneapolis, MN, USA. ¹⁶Prevention and Research Center, Mercy Medical Center, Baltimore, MD, USA. ¹⁷Department of Epidemiology & Biostatistics, University of California San Francisco, San Francisco, CA, USA. ¹⁸Department of Epidemiology, American Cancer Society, Atlanta, GA, USA. ¹⁹Department of Oncology, the Johns Hopkins University School of Medicine, Baltimore, MD, USA. ²⁰Department of Epidemiology, the Bloomberg School of Public Health, The Sol Goldman Pancreatic Research Center, The Johns Hopkins Medical Institutions, Baltimore, MD, USA. ²¹Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. ²²Department of Gastrointestinal Medical Oncology, University of Texas M.D. Anderson Cancer Center, Houston, Texas, USA. ²³Group Health Center for Health Studies, Seattle, WA, USA. ²⁴Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY, USA. ²⁵Yale University School of Public Health, New Haven, CT, USA. ²⁶Department of Medicine and Vanderbilt-Ingram Cancer Center, Vanderbilt University, Nashville, TN, USA. ²⁷Division of Cancer Prevention, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. ²⁸Inserm, Paris-Sud University, Institut Gustave-Roussy, Villejuif, France. ²⁹Divisions of Preventive Medicine and Aging, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston,

MA, USA. ³⁰Department of Ambulatory Care and Prevention, Harvard Medical School, Boston, MA, USA. ³¹German Cancer Research Center (DKFZ), Heidelberg, Germany. ³²Johns Hopkins Bloomberg School of Public Health, George W. Comstock Center for Public Health Research and Prevention, Hagerstown, MD, USA. ³³Cancer Care Ontario and Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada. ³⁴Catalan Institute of Oncology (ICO), Barcelona, Spain. ³⁵Physicians' Health Study, Divisions of Aging, Cardiovascular Disease, and Preventive Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. ³⁶Massachusetts Veterans Epidemiology Research and Information Center, Veterans Affairs Boston Healthcare System, Boston, MA, USA. ³⁷Department of Nutrition, Harvard School of Public Health, Boston, MA, USA. ³⁸Departments of Oncology, Pathology and Medicine, The Johns Hopkins University School of Medicine, Baltimore, MD, USA. ³⁹Department of Public Health and Clinical Medicine, Nutritional Research, Umeå University, Umeå, Sweden. ⁴⁰MedStar Research Institute, Georgetown University, Hyattsville, MD. ⁴¹Nutritional Epidemiology Unit, Fondazione IRCCS Istituto Nazionale dei Tumori Milano, Milan, Italy. ⁴²Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY, USA. ⁴³Division of Epidemiology, Public Health and Primary Care, Imperial College London, London, UK. ⁴⁴Julius Center for Health

Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands. ⁴⁵Departments of Obstetrics and Gynecology and Reproductive Sciences, University of Pittsburgh, Pittsburgh, PA, USA. ⁴⁶Public Health and Participation Directorate, Health and Health Care Services Council, Asturias, Spain. ⁴⁷Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD, USA. ⁴⁸Synergie-Lyon-Cancer, INSERM U590, Centre Leon Berard, 69373 Lyon, Cedex ⁰⁸, France. ⁴⁹Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, Denmark. ⁵⁰Bureau of Epidemiologic Research, Academy of Athens, Greece. ⁵¹Division of Research, Kaiser Permanente, Northern California Region, Oakland, CA, USA. ⁵²Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland. ⁵³Department of Social and Preventive Medicine, University at Buffalo, SUNY Buffalo, NY, USA.

Author Contributions

Conceived and designed the experiments: KSE EZ MIM JPAI. Analyzed the data: KSE JPAI. Contributed reagents/materials/analysis tools: KSE JG PS SNS ST LTA HG JX VG RAE DEN JLD FCH KM SJH MRS BZ LCC KMB AMFSI NKH SM IPT ML CIA JMM WBI DFE PB PC RB DFG TR DJH SJC KS. Wrote the paper: KSE EZ MIM JG JPAI.

References

- Inoue M, Iwasaki M, Otani T, Sasazuki S, Noda M, et al. (2006) Diabetes mellitus and the risk of cancer: results from a large-scale population-based cohort study in Japan. *Arch Intern Med* 166: 1871–1877.
- Stattin P, Bjor O, Ferrari P, Lukanova A, Lenner P, et al. (2007) Prospective study of hyperglycemia and cancer risk. *Diabetes Care* 30: 561–567.
- Rousseau MC, Parent ME, Pollak MN, Siemiatycki J (2006) Diabetes mellitus and cancer risk in a population-based case-control study among men from Montreal, Canada. *Int J Cancer* 118: 2105–2109.
- Kasper JS, Giovannucci E (2006) A meta-analysis of diabetes mellitus and the risk of prostate cancer. *Cancer Epidemiol Biomarkers Prev* 15: 2056–2062.
- Gong Z, Neuhauser ML, Goodman PJ, Albanes D, Chi C, et al. (2006) Obesity, diabetes, and risk of prostate cancer: results from the prostate cancer prevention trial. *Cancer Epidemiol Biomarkers Prev* 15: 1977–1983.
- Calton BA, Chang SC, Wright ME, Kipnis V, Lawson K, et al. (2007) History of diabetes mellitus and subsequent prostate cancer risk in the NIH-AARP Diet and Health Study. *Cancer Causes Control* 18: 493–503.
- Will JC, Vinicor F, Calle EE (1999) Is diabetes mellitus associated with prostate cancer incidence and survival? *Epidemiology* 10: 313–318.
- Manolio TA, Brooks LD, Collins FS (2008) A HapMap harvest of insights into the genetics of common disease. *J Clin Invest* 118: 1590–1605.
- Gudmundsson J, Sulem P, Steinthorsdottir V, Bergthorsson JT, Thorleifsson G, et al. (2007) Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet* 39: 977–983.
- Sun J, Zheng SL, Wiklund F, Isaacs SD, Purcell LD, et al. (2008) Evidence for two independent prostate cancer risk-associated loci in the HNF1B gene at 17q12. *Nat Genet* 40: 1153–1155.
- Thomas G, Jacobs KB, Yeager M, Kraft P, Wacholder S, et al. (2008) Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet*.
- Winkler W, Weedon MN, Graham RR, McCarroll SA, Purcell S, et al. (2007) Evaluation of common variants in the six known maturity-onset diabetes of the young (MODY) genes for association with type 2 diabetes. *Diabetes* 56: 685–693.
- Horikawa Y, Iwasaki N, Hara M, Furuta H, Hinokio Y, et al. (1997) Mutation in hepatocyte nuclear factor-1 beta gene (TCF2) associated with MODY. *Nat Genet* 17: 384–385.
- Tenesa A, Farrington SM, Prendergast JG, Porteous ME, Walker M, et al. (2008) Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet* 40: 631–637.
- Gold B, Kirchhoff T, Stefanov S, Lautenberger J, Viale A, et al. (2008) Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. *Proc Natl Acad Sci U S A* 105: 4340–4345.
- Maris JM, Mosse YP, Bradfield JP, Hou C, Monni S, et al. (2008) Chromosome 6p22 locus associated with clinically aggressive neuroblastoma. *N Engl J Med* 358: 2585–2593.
- Cust AE, Schmid H, Maskiell JA, Jetann J, Ferguson M, et al. (2009) Population-based, case-control-family design to investigate genetic and environmental influences on melanoma risk: Australian Melanoma Family Study. *Am J Epidemiol* 170: 1541–1554.
- Eeles RA, Kote-Jarai Z, Giles GG, Olama AA, Guy M, et al. (2008) Multiple newly identified loci associated with prostate cancer susceptibility. *Nat Genet* 40: 316–321.
- Petersen GM, Amundadottir L, Fuchs CS, Kraft P, Stolzenberg-Solomon RZ, et al. (2010) A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. *Nat Genet* 42: 224–228.
- Brennan P, McKay J, Moore L, Zaridze D, Mukeria A, et al. (2007) Uncommon CHEK2 mis-sense variant and reduced risk of tobacco-related cancers: case control study. *Hum Mol Genet* 16: 1794–1801.
- Hung RJ, McKay JD, Gaboricau V, Boffetta P, Hashibe M, et al. (2008) A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* 452: 633–637.
- Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, et al. (2008) Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet* 40: 616–622.
- Zanke BW, Greenwood CM, Rangrej J, Kustra R, Tenesa A, et al. (2007) Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet* 39: 989–994.
- Brown KM, Macgregor S, Montgomery GW, Craig DW, Zhao ZZ, et al. (2008) Common sequence variants on 20q11.22 confer melanoma susceptibility. *Nat Genet* 40: 838–840.
- Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, et al. (2007) Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet* 39: 645–649.
- Duggan D, Zheng SL, Knowlton M, Benitez D, Dimitrov L, et al. (2007) Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP. *J Natl Cancer Inst* 99: 1836–1844.
- Tomlinson IP, Webb E, Carvajal-Carmona L, Broderick P, Howarth K, et al. (2008) A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat Genet* 40: 623–630.
- Stacey SN, Manolescu A, Sulem P, Thorlacius S, Gudjonsson SA, et al. (2008) Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* 40: 703–706.
- Rafnar T, Sulem P, Stacey SN, Geller F, Gudmundsson J, et al. (2009) Sequence variants at the TERT-CLPTMIL locus associate with many cancer types. *Nat Genet* 41: 221–227.
- Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, et al. (2008) A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* 452: 638–642.
- Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, et al. (2007) Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet* 39: 631–637.
- Gudmundsson J, Sulem P, Rafnar T, Bergthorsson JT, Manolescu A, et al. (2008) Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. *Nat Genet*.
- Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, et al. (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 447: 1087–1093.
- http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000007.v9.p4.
- Edghill EL, Bingham C, Ellard S, Hattersley AT (2006) Mutations in hepatocyte nuclear factor-1beta and their related phenotypes. *J Med Genet* 43: 84–90.
- Rebouissou S, Vasiliu V, Thomas C, Bellanne-Chantelot C, Bui H, et al. (2005) Germ-line hepatocyte nuclear factor 1alpha and 1beta mutations in renal cell carcinomas. *Hum Mol Genet* 14: 603–614.
- Reber M, Cereghini S (2001) Variant hepatocyte nuclear factor 1 expression in the mouse genital tract. *Mech Dev* 100: 75–78.

38. Coffinier C, Barra J, Babinet C, Yaniv M (1999) Expression of the vHNF1/HNF1beta homeoprotein gene during mouse organogenesis. *Mech Dev* 89: 211–213.
39. Ioannidis JP, Thomas G, Daly MJ (2009) Validating, augmenting and refining genome-wide association signals. *Nat Rev Genet* 10: 318–329.
40. Yamada H, Penney KL, Takahashi H, Katoh T, Yamano Y, et al. (2009) Replication of prostate cancer risk loci in a Japanese case-control association study. *J Natl Cancer Inst* 101: 1330–1336.
41. Levin AM, Machiela MJ, Zuhlke KA, Ray AM, Cooney KA, et al. (2008) Chromosome 17q12 variants contribute to risk of early-onset prostate cancer. *Cancer Res* 68: 6492–6495.
42. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, et al. (2008) Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 40: 638–645.
43. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, et al. (2007) Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 316: 1336–1341.
44. Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, et al. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* 106: 9362–9367.
45. Kavvoura FK, Ioannidis JP (2008) Methods for meta-analysis in genetic association studies: a review of their potential and pitfalls. *Hum Genet* 123: 1–14.
46. Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. *Bmj* 327: 557–560.
47. Ioannidis JP, Patsopoulos NA, Evangelou E (2007) Uncertainty in heterogeneity estimates in meta-analyses. *Bmj* 335: 914–916.
48. Petersen GM, Amundadottir L, Fuchs CS, Kraft P, Stolzenberg-Solomon RZ, et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. *Nat Genet*.