OPEN ACCESS Freely available online

PLoS GENETICS

A Genome-Wide Association Study of Upper Aerodigestive Tract Cancers Conducted within the INHANCE Consortium

James D. McKay[1], Therese Truong[1], Valerie Gaborieau[1], Amelie Chabrier[1], Shu-Chun Chuang[1], Graham Byrnes[1], David Zaridze[2], Oxana Shangina[2], Neonila Szeszenia-Dabrowska[3], Jolanta Lissowska[4], Peter Rudnai[5], Eleonora Fabianova[6], Alexandru Bucur[7], Vladimir Bencko[8], Ivana Holcatova[8], Vladimir Janout[9], Lenka Foretova[10], Pagona Lagiou[11], Dimitrios Trichopoulos[11,12], Simone Benhamou[13,14], Christine Bouchardy[15], Wolfgang Ahrens[16], Franco Merletti[17], Lorenzo Richiardi[17], Renato Talamini[18], Luigi Barzan[19], Kristina Kjaerheim[20], Gary J. Macfarlane[21], Tatiana V. Macfarlane[21], Lorenzo Simonato[22], Cristina Canova[22,23], Antonio Agudo[24], Xavier Castellsagué[24,25], Ray Lowry[26], David I. Conway[27], Patricia A. McKinney[28,29], Claire M. Healy[30], Mary E. Toner[30], Ariana Znaor[31], Maria Paula Curado[1], Sergio Koifman[32], Ana Menezes[33], Victor Wünsch-Filho[34], José Eluf Neto[34], Leticia Fernández Garrote[35], Stefania Boccia[36,37], Gabriella Cadoni[36], Dario Arzani[36], Andrew F. Olshan[38], Mark C. Weissler[39], William K. Funkhouser[39], Jingchun Luo[39], Jan Lubiński[40], Joanna Trubicka[40], Marcin Lener[40], Dorota Oszutowska[40,41], Stephen M. Schwartz[42], Chu Chen[42], Sherianne Fish[42], David R. Doody[42], Joshua E. Muscat[43], Philip Lazarus[43], Carla J. Gallagher[43], Shen-Chih Chang[44], Zuo-Feng Zhang[44], Qingyi Wei[45], Erich M. Sturgis[45], Li-E Wang[45], Silvia Franceschi[1], Rolando Herrero[46], Karl T. Kelsey[47], Michael D. McClean[48], Carmen J. Marsit[47], Heather H. Nelson[49], Marjorie Romkes[50], Shama Buch[50], Tomoko Nukui[50], Shilong Zhong[50], Martin Lacko[51], Johannes J. Manni[51], Wilbert H. M. Peters[52], Rayjean J. Hung[53], John McLaughlin[54], Lars Vatten[55], Inger Njølstad[56], Gary E. Goodman[42], John K. Field[57], Triantafillos Liloglou[57], Paolo Vineis[58,59], Francoise Clavel-Chapelon[60], Domenico Palli[61], Rosario Tumino[62], Vittorio Krogh[63], Salvatore Panico[64], Carlos A. González[65], J. Ramón Quirós[66], Carmen Martínez[67], Carmen Navarro[68,25], Eva Ardanaz[25,69], Nerea Larrañaga[70], Kay-Tee Khaw[71], Timothy Key[72], H. Bas Bueno-de-Mesquita[73], Petra H. M. Peeters[74], Antonia Trichopoulou[75], Jakob Linseisen[76,77], Heiner Boeing[78], Göran Hallmans[79], Kim Overvad[80], Anne Tjønneland[81], Merethe Kumle[82], Elio Riboli[59], Kristjan Välk[83], Tõnu Voodern[83], Andres Metspalu[83], Diana Zelenika[84], Anne Boland[84], Marc Delepine[84], Mario Foglio[84], Doris Lechner[84], Hélène Blanché[85], Ivo G. Gut[84], Pilar Galan[86], Simon Heath[84], Mia Hashibe[1], Richard B. Hayes[87], Paolo Boffetta[1], Mark Lathrop[84,85], Paul Brennan[1]*

1 International Agency for Research on Cancer (IARC), Lyon, France, **2** Institute of Carcinogenesis, Cancer Research Centre, Moscow, Russia, **3** Department of Epidemiology, Institute of Occupational Medicine, Lodz, Poland, **4** The M. Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland, **5** National Institute of Environmental Health, Budapest, Hungary, **6** Regional Authority of Public Health, Banská Bystrica, Slovakia, **7** Institute of Public Health, Bucharest, Romania, **8** Institute of Hygiene and Epidemiology,1st Faculty of Medicine, Charles University, Prague, Czech Republic, **9** Palacky University, Olomouc, Czech Republic, **10** Department of Cancer Epidemiology and Genetics, Masaryk Memorial Cancer Institute, Brno, Czech Republic, **11** Department of Hygiene, Epidemiology, and Medical Statistics, University of Athens School of Medicine, Athens, Greece, **12** Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, United States of America, **13** INSERM U946, Paris, France, **14** CNRS UMR8200, Gustave Roussy Institute, Villejuif, France, **15** Geneva Cancer Registry, Institute for Social and Preventive Medicine, University of Geneva, Geneva, Switzerland, **16** Bremen Institute for Prevention Research and Social Medicine (BIPS), University of Bremen, Bremen, Germany, **17** Unit of Cancer Epidemiology, University of Turin, Turin, Italy, **18** National Cancer Institute, IRCSS, Aviano, Italy, **19** General Hospital of Pordenone, Pordenone, Italy, **20** Cancer Registry of Norway, Oslo, Norway, **21** School of Medicine and Dentistry, University of Aberdeen, Aberdeen, United Kingdom, **22** Department of Environmental Medicine and Public Health, University of Padova, Padova, Italy, **23** Respiratory Epidemiology and Public Health, National Heart and Lung Institute, Imperial College, London, United Kingdom, **24** Institut Català d'Oncologia (ICO), Barcelona, Spain, **25** CIBER Epidemiologia y Salud Publica (CIBERESP), Madrid, Spain, **26** University of Newcastle Dental School, Newcastle, United Kingdom, **27** University of Glasgow Dental School, Glasgow, Scotland, **28** University of Leeds Centre for Epidemiology and Biostatistics, Leeds, United Kingdom, **29** NHS NSS ISD, Edinburgh, Scotland, **30** Trinity College School of Dental Science, Dublin, Ireland, **31** Croatian National Cancer Registry, Croatian National Institute of Public Health, Zagreb, Croatia, **32** National School of Public Health/FIOCRUZ, Rio de Janeiro, Brazil, **33** Universidade Federal de Pelotas, Pelotas, Brazil, **34** Universidade de Sao Paulo, Sao Paulo, Brazil, **35** Institute of Oncology and Radiobiology, Havana, Cuba, **36** Institute of Hygiene, Università Cattolica del Sacro Cuore, Rome, Italy, **37** IRCCS San Raffaele Pisana, Rome, Italy, **38** Gillings School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina, United States of America, **39** School of Medicine, University of North Carolina, Chapel Hill, North Carolina, United States of America, **40** Pomeranian Medical University, Department of Genetics and Pathomorphology, International Hereditary Cancer Center, Szczecin, Poland, **41** Pomeranian Medical University, Department of Hygiene, Epidemiology, and Public Health, Szczecin, Poland, **42** Fred Hutchinson Cancer Research Centre, Seattle, Washington, United States of America, **43** Penn State College of Medicine, Hershey, Pennsylvania, United States of America, **44** University of California Los Angeles School of Public Health, Los Angeles, California, United States of America, **45** University of Texas M. D. Anderson Cancer Center, Houston, Texas, United States of America, **46** Instituto de Investigación Epidemiológica, San José, Costa Rica, **47** Brown University, Providence, Rhode Island, United States of America, **48** Boston University School of Public Health, Boston, Massachusetts, United States of America, **49** Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota, United States of America, **50** University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America, **51** Department of Otorhinolaryngology and Head and Neck Surgery, Maastricht University Medical Centre, Maastricht, The Netherlands, **52** Department of

Gastroenterology, St. Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands, **53** Samuel Lunenfeld Research Institute of the Mount Sinai Hospital, Toronto, Canada, **54** Cancer Care Ontario, Toronto, Canada, **55** Norwegian University of Science and Technology, Trondheim, Norway, **56** Department of Community Medicine, Faculty of Health Sciences, University of Tromso, Tromso, Norway, **57** Roy Castle Lung Cancer Research Programme, The University of Liverpool Cancer Research Centre, Liverpool, United Kingdom, **58** Servizio di Epidemiologia dei Tumori, Università di Torino and CPO-Piemonte, Turin, Italy, **59** Department of Epidemiology and Public Health, Imperial College, London, United Kingdom, **60** INSERM, E3N-EPIC Group Institut Gustave Roussy, Villejuif, France, **61** Molecular and Nutritional Epidemiology Unit, Cancer Research and Prevention Institute (ISPO), Florence, Italy, **62** Cancer Registry and Histopathology Unit, Azienda Ospedaliera "Civile M.P.Arezzo", Ragusa, Italy, **63** Fondazione IRCCS, Istituto Nazionale dei Tumori, Milan, Italy, **64** Dipartimento di Medicina Clinica e Sperimentale, Universita di Napoli Federico II, Naples, Italy, **65** Unit of Nutrition, Environment, and Cancer (IDIBELL, RETICC DR06-0020, Catalan Institute of Oncology (ICO), Barcelona, Spain, **66** Jefe Sección Información Sanitaria, Consejería de Servicios Sociales, Principado de Asturias, Oviedo, Spain, **67** Escuela Andaluza de Salud Pública, Granada, Spain, **68** Epidemiology Department, Murcia Health Council, Murcia, Spain, **69** Navarra Public Health Institute, Pamplona, Spain, **70** Subdirección de Salud Pública de Gipuzkoa, Gobierno Vasco, San Sebastian, Spain, **71** School of Clinical Medicine, University of Cambridge, Cambridge, United Kingdom, **72** Cancer Research UK, University of Oxford, Oxford, United Kingdom, **73** National Institute of Public Health and the Environment (RIVM), Bilthoven, The Netherlands, **74** Julius Center for Health Sciences and Primary Care, Department of Epidemiology, University Medical Center of Utrecht, Utrecht, The Netherlands, **75** WHO Collaborating Center for Nutrition, Department of Hygiene, Epidemiology, and Medical Statistics, University of Athens School of Medicine, Athens, Greece, **76** Institute of Epidemiology, Helmholtz Centre Munich, Neuherberg, Germany, **77** Division of Clinical Epidemiology, German Cancer Research Centre, Heidelberg, Germany, **78** Department of Epidemiology, Deutsches Institut für Ernährungsforschung, Potsdam-Rehbrücke, Germany, **79** Department of Public Health and Clinical Medicine, University of Umeå, Umeå, Sweden, **80** Department of Epidemiology and Social Medicine, Aarhus University, Aarhus, Denmark, **81** The Danish Cancer Society, Institute of Cancer Epidemiology, Copenhagen, Denmark, **82** University Hospital Northern Norway, Tromsø, Norway, **83** University of Tartu, Tartu, Estonia, **84** Centre National de Génotypage, Institut Génomique, Commissariat à l'énergie Atomique, Evry, France, **85** Fondation Jean Dausset-CEPH, Paris, France, **86** INSERM U557/U1125 INRA/CNAM, Université Paris 13, Bobigny, France, **87** New York University Langone Medical Center, New York, New York, United States of America

Abstract

Genome-wide association studies (GWAS) have been successful in identifying common genetic variation involved in susceptibility to etiologically complex disease. We conducted a GWAS to identify common genetic variation involved in susceptibility to upper aero-digestive tract (UADT) cancers. Genome-wide genotyping was carried out using the Illumina HumanHap300 beadchips in 2,091 UADT cancer cases and 3,513 controls from two large European multi-centre UADT cancer studies, as well as 4,821 generic controls. The 19 top-ranked variants were investigated further in an additional 6,514 UADT cancer cases and 7,892 controls of European descent from an additional 13 UADT cancer studies participating in the INHANCE consortium. Five common variants presented evidence for significant association in the combined analysis ($p \leq 5 \times 10^{-7}$). Two novel variants were identified, a 4q21 variant (rs1494961, $p = 1 \times 10^{-8}$) located near DNA repair related genes *HEL308* and *FAM175A* (or *Abraxas*) and a 12q24 variant (rs4767364, $p = 2 \times 10^{-8}$) located in an extended linkage disequilibrium region that contains multiple genes including the *aldehyde dehydrogenase 2 (ALDH2)* gene. Three remaining variants are located in the *ADH* gene cluster and were identified previously in a candidate gene study involving some of these samples. The association between these three variants and UADT cancers was independently replicated in 5,092 UADT cancer cases and 6,794 controls non-overlapping samples presented here (rs1573496-*ADH7,* $p = 5 \times 10^{-8}$; rs1229984-*ADH1B,* $p = 7 \times 10^{-9}$; and rs698-*ADH1C,* $p = 0.02$). These results implicate two variants at 4q21 and 12q24 and further highlight three *ADH* variants in UADT cancer susceptibility.

Citation: McKay JD, Truong T, Gaborieau V, Chabrier A, Chuang S-C, et al. (2011) A Genome-Wide Association Study of Upper Aerodigestive Tract Cancers Conducted within the INHANCE Consortium. PLoS Genet 7(3): e1001333. doi:10.1371/journal.pgen.1001333

Editor: Marshall S. Horwitz, University of Washington, United States of America

Received June 5, 2010; **Accepted** February 11, 2011; **Published** March 17, 2011

Funding: Support for the central Europe and ARCAGE genome-wide studies and follow-up genotyping was provided by INCa, France. Additional funding for study coordination, genotyping of replication studies, and statistical analysis was provided by the US NCI (R01 CA092039 05/05S1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: brennan@iarc.fr

Introduction

560,000 cases of upper aerodigestive tract (UADT) cancers (encompassing of the oral cavity, pharynx, larynx and esophagus) are estimated to occur each year world-wide [1]. Exposure to alcohol and tobacco [1] are the major UADT cancer risk factors in Europe and the Americas, with infection with human papillomavirus also playing an important role [2].

Elevated familial relative risks are consistently reported for UADT cancers [3-7]. While this implies that genetics contributes to UADT cancer susceptibility, the identity of the specific genes involved remains unclear. Studies of common genetic variation and UADT cancer susceptibility have mostly employed a candidate gene approach, with a particular focus on the genes that metabolize alcohol [8]. The metabolism of alcohol releases the carcinogen acetaldehyde as an intermediate [9]. As genetic variation in alcohol metabolism genes appears to influence their rate of function [10,11], variants that lead to a relative increase in exposure to acetaldehyde are expected to confer carriers to an increased risk of UADT cancers [12]. Consistent with this hypothesis, genetic variation in the *alcohol dehydrogenase* (*ADH*) *1B*, and the *aldehyde dehydrogenase 2* (*ALDH2*) genes in Asian populations have been associated with UADT cancer risk [8,12,13]. Three independent variants *ADH1B*, *ADH7* and *ADH1C* variants have also been associated with UADT cancer risk in European populations [14]. Common genetic variation in additional genetic pathways have also been considered, although with some exceptions, such as DNA repair [15,16], the results have been inconsistent [3].

Author Summary

We have used a two-phased study approach to identify common genetic variation involved in susceptibility to upper aero-digestive tract cancer. Using Illumina HumanHap300 beadchips, 2,091 UADT cancer cases and 3,513 controls from two large European multi-centre UADT cancer studies, as well as 4,821 generic controls, were genotyped for a panel 317,000 genetic variants that represent the majority of common genetic in the human genome. The 19 top-ranked variants were then studied in an additional series of 6,514 UADT cancer cases and 7,892 controls of European descent from an additional 13 UADT cancer studies. Five variants were significantly associated with UADT cancer risk after the completion of both stages, including three residing within the alcohol dehydrogenase genes (*ADH1B*, *ADH1C*, *ADH7*) that have been previously described. Two additional variants were found, one near the *ALDH2* gene and a second variant located in *HEL308*, a DNA repair gene. These results implicate two variants 4q21 and 12q24 and further highlight three *ADH* variants UADT cancer susceptibility.

The candidate gene based studies have tested only a very small proportion of common human genetic variation in relation to UADT cancer risk. To further investigate common genetic variation and susceptibility to UADT cancers, we have performed a genome-wide association study within the International Head and Neck Cancer Epidemiology (INHANCE) consortium, comprising genome wide analysis of 2,091 UADT cancer cases and 8,334 controls and replication analysis of the nineteen top ranked variants in an independent series consisting of 6,514 UADT cancer cases and 7,892 controls from thirteen additional studies.

Results

Genome-wide results

After exclusion of suboptimal DNA based on QC criteria, data from 2,091 cases and 3,513 study specific controls and 4,821 generic controls were available for statistical analyses (Table S1) with 294,620 genetic variants. The overall results did not show a large deviation from what was expected by chance ($\lambda = 1.07$) (Figure 1). One genetic variant, rs971074, was strongly associated with UADT cancers ($p < 1 \times 10^{-8}$). rs971074 is positioned in the *ADH7* locus on chromosome 4q23 and is highly correlated ($r^2 = 1.0$ CEU hapmap) with the SNP in *ADH7*, rs1573496, that we have described previously to be associated with UADT cancer risk [14]. Similarly, rs1789924, which is highly correlated ($r^2 = 0.97$ CEU hapmap) with *ADH1C* rs698, was also highly ranked ($p = 2 \times 10^{-6}$).

Variant selection for replication

For further analysis we selected the twenty top ranked genetic variants (including rs971074) for replication (Figure S1). These included those genetic variants in the discovery phase that achieved a p-value of $\leq 1 \times 10^{-5}$ (12 variants) as well as nonsynonymous variants that achieved a p-value of $\leq 1 \times 10^{-4}$ (5 additional variants). We also included variants that achieved a p-value of $\leq 5 \times 10^{-7}$ when restricting the analysis to a specific UADT cancer site (1 variant), or heavy drinkers (1 variant) (Table 1). Only one variant from each high r^2 group ($r^2 > 0.8$) was included. We additionally included the non-synonymous *ADH1B* variant, rs1229984, that has been previously associated with UADT cancers [14] but not genotyped or tagged by a proxy variant on the HumanHap300 BeadChip. The association between the top ranked genetic variants selected for replication and UADT cancer was not sensitive to adjustment for population structure using principal component analysis, or exclusion generic controls (Table S2). rs1573496 was genotyped for replication as a proxy for rs971074 ($r^2 = 1.00$) and rs698 for rs1789924 ($r^2 > 0.97$) due to availability of Taqman assays. A TaqMan assay for rs12827056 could not be designed and no highly correlated ($r^2 > 0.95$) proxy genetic variant was available, hence further investigation was not possible.

Replication and combined results

Five genetic variants at three loci, 4q21, 4q23 and 12q24, were significantly associated with UADT cancer risk in the replication series (assuming Bonferroni correction for 19 comparisons or $p \leq 0.003$, or $p = 0.05$ for previously described variants) or in the combined analysis (p-value of $\leq 5 \times 10^{-7}$) (Table 1) (Figure S2). Using imputed genotypes across the 4q21, 4q23 and 12q24 regions based on Caucasian individuals from the HapMap consortium, we did not identify any variants more strongly associated with UADT cancer risk than the SNPs genotyped on the beadchips directly (Figure 2).

Two novel variant loci were identified. rs4767364 located at 12q24 ($p_{replication} = 4 \times 10^{-4}$; $p_{combined} = 2 \times 10^{-8}$) was one of

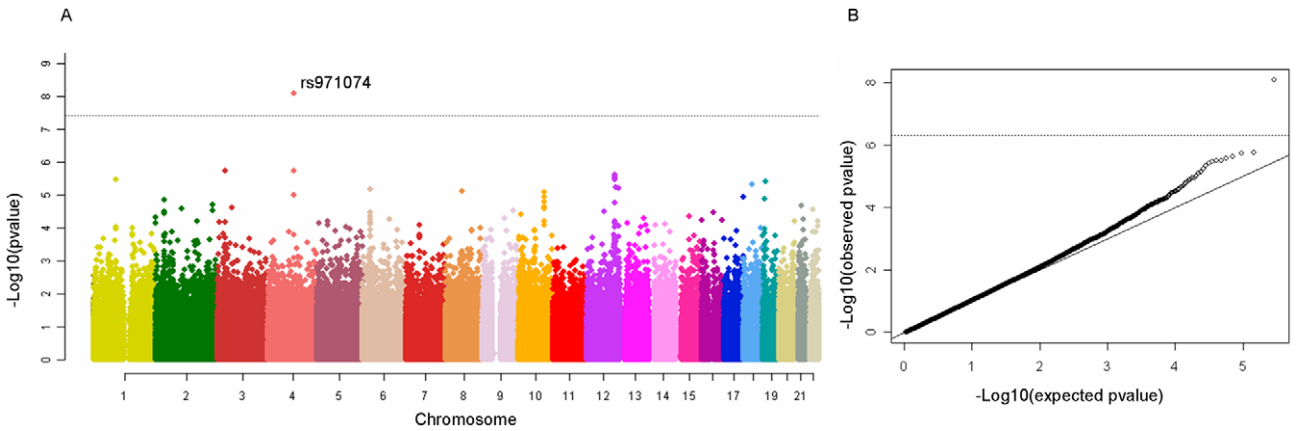

Figure 1. Manhattan plot of the ARCAGE and CE UADT cancer GWAS discovery phase. One clearly outlying marker, rs971074 is highly correlated with rs1573496, a SNP previously associated with UADT cancer risk. Right panel QQ plot for the UADT cancer GWAS.
doi:10.1371/journal.pgen.1001333.g001

Table 1. Results from the UADT cancer genome-wide and replication analysis.

		Alleles			Discovery phase[a]			Replication phase[b;f]			Combined[a]		
Marker	**Chromosome region**	**ref**	**rare**	**Reason for replication attempt**	**OR**	**95% CI**	**P_{ts}**	**OR**	**95% CI**	**P_{ts}**	**OR**	**95% CI**	**P_{ts}**
rs1229984	4q23	C	T	*ADH1B*, candidate gene	0.52	0.43–0.64	7×10^{-11}	0.68[f]	0.60–0.78	7×10^{-9}	0.64	0.59–0.71	1×10^{-20}
rs971074[c]	4q23	G	C	p_all≤1×10^{-5}	0.70	0.62–0.79	8×10^{-9}	0.78[f]	0.72–0.86	5×10^{-8}	0.75	0.70–0.80	9×10^{-17}
rs1494961	4q21	T	C	non-synonymous and p≤1×10^{-4}	1.15	1.07–1.24	1×10^{-4}	1.11	1.06–1.17	2×10^{-5}	1.12	1.08–1.17	1×10^{-8}
rs4767364	12q24	G	A	p_all≤1×10^{-5}	1.21	1.12–1.32	2×10^{-6}	1.10	1.04–1.15	4×10^{-4}	1.13	1.08–1.18	2×10^{-8}
rs1789924[c]	4q23	T	C	p_all≤1×10^{-5}	1.20	1.11–1.29	2×10^{-6}	1.07[f]	1.01–1.14	0.02	1.12	1.07–1.17	3×10^{-7}
rs1431918	8q12	G	A	p_all≤1×10^{-5}	1.19	1.10–1.28	7×10^{-6}	1.05	1.00–1.11	0.05	1.09	1.04–1.14	7×10^{-5}
rs7431530	3p24	C	T	p_all≤1×10^{-5}	0.81	0.74–0.88	2×10^{-6}	0.95	0.90–1.00	0.06	0.91	0.87–0.95	5×10^{-5}
rs3810481	20q13	G	A	non-synonymous and p≤1×10^{-4}	1.22	1.11–1.34	6×10^{-5}	1.07	0.99–1.15	0.09	1.12	1.06–1.19	2×10^{-4}
rs10801805	1p22	G	A	p_all≤1×10^{-5}	1.20	1.11–1.29	3×10^{-6}	1.04	0.98–1.10	0.15	1.09	1.04–1.14	2×10^{-4}
rs1041973	2q12	C	A	non-synonymous and p≤1×10^{-4}	0.83	0.76–0.90	3×10^{-5}	0.94	0.89–1.00	0.05	0.91	0.87–0.95	9×10^{-5}
rs4799863	18q12	A	G	p_all≤1×10^{-5}	0.84	0.78–0.91	5×10^{-6}	0.96	0.92–1.01	0.12	0.92	0.89–0.96	1×10^{-4}
rs2517452[d]	6p21	C	T	p_oral<5.10^{-7}	0.69	0.59–0.80	4×10^{-7}	1.00	0.87–1.15	0.97	0.84	0.76–0.92	5×10^{-4}
rs2012199	1q23	T	C	non-synonymous and p≤1×10^{-4}	1.24	1.11–1.38	1×10^{-4}	1.06	0.98–1.14	0.16	1.12	1.05–1.19	6×10^{-4}
rs2287802	19p13	A	G	p_all≤1×10^{-5}	1.19	1.11–1.28	4×10^{-6}	1.02	0.97–1.07	0.54	1.07	1.02–1.11	2×10^{-3}
rs16837730[e]	1p35	C	T	p heavy drinkers <5.10^{-7}	2.06	1.57–2.71	2×10^{-7}	1.02	0.86–1.22	0.79	1.25	1.08–1.45	3×10^{-3}
rs11067362	12q24	T	C	p_all≤1×10^{-5}	1.35	1.19–1.54	6×10^{-6}	1.01	0.92–1.10	0.88	1.11	1.03–1.19	8×10^{-3}
rs7924284	10q24	C	G	p_all≤1×10^{-5}	1.38	1.20–1.59	8×10^{-6}	1.00	0.90–1.11	0.97	1.12	1.03–1.21	0.01
rs484870	19p13	A	G	non-synonymous and p≤1×10^{-4}	1.16	1.08–1.26	1×10^{-4}	0.99	0.94–1.05	0.72	1.05	1.00–1.10	0.04
rs2299851	6p21	G	A	p_all≤1×10^{-5}	0.72	0.62–0.83	6×10^{-6}	1.11	1.00–1.23	0.05	0.96	0.88–1.04	0.27

a Including "generic" controls (methods) with the exception of rs1229984. Adjusted by sex, study.
b Adjusted by age, sex, study.
c rs970174 and rs1789924 were genotyped in the replication phase by highly correlated variants (r^2>0.97) rs1573496 and rs698.
d Analysis considered oral cancers only.
e Analysis considered heavy drinkers only.
f For 4q23 variants rs1229984, rs1573496, rs698, the replication phase excluded the SA Latin American study (Table 4) that had been published previously.
P_{ts}: two-sided p-value.
doi:10.1371/journal.pgen.1001333.t001

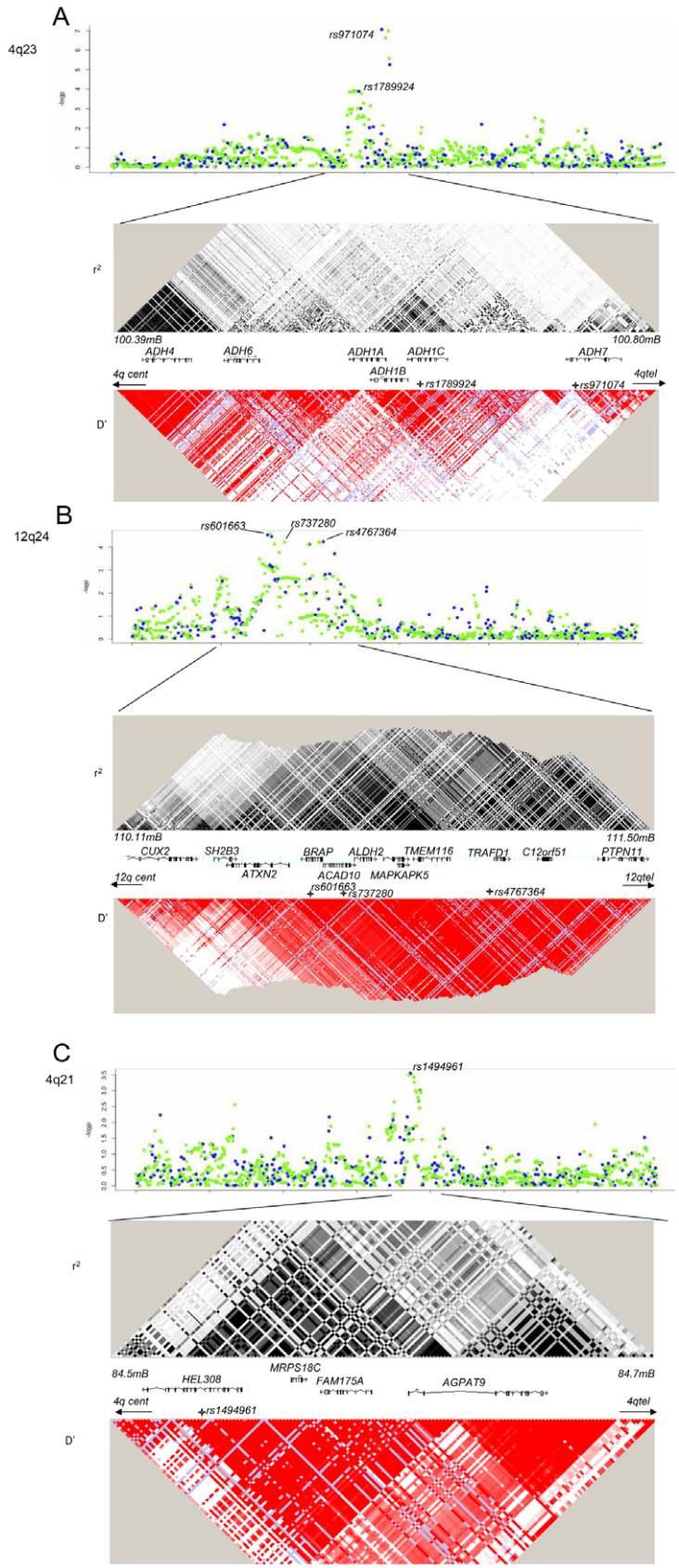

Figure 2. Imputation and LD patterns. Imputation and LD patterns across the (a) 4q23 (*ADH* loci), (b) 12q24 (*ALDH2*), and (c) 4q21 (*HEL308*). Upper panel: Single marker association results for imputed (green) and directly genotyped variants (blue). Imputation performed on 2,091 cases and 3,513 study specific controls (excluded generic controls). After adjustment for the five variants that presented with replication, no variant had a

$p<0.0005$ at any loci. Lower panel, pairwise LD estimates increasing intensities of black and red indicate higher r^2 or D′ statistics, respectively. Blue colour indicates that the pairwise comparison is not informative.
doi:10.1371/journal.pgen.1001333.g002

multiple highly correlated SNPs ($r^2 \geq 0.8$) that presented evidence for association in the GWAS stage. It is located in a LD region including multiple genes including the *aldehyde dehydrogenase 2 (ALDH2)* (Figure 2), another key gene in alcohol metabolism (Figure 2). In stratified analysis in the combined 8,744 UADT cancer cases and 11,982 controls (Table S3), the association was more pronounced in esophageal cancers compared to other UADT cancer subsites (p heterogeneity = 0.01) and exhibited borderline heterogeneity when stratifying by alcohol use (Figure 3). Some heterogeneity was noted by when stratifying by country (p = 0.004), although there was no discernable geographic distribution that could explain this heterogeneity (data not shown). We noted little evidence for association between alcohol consumption and rs4767364 (Table 2), nor was there evidence for any gene-gene interactions between associated variants in *ADH* gene cluster and rs4767364 (data not shown).

The second additional locus identified was at 4q21, with the nonsynonmous variant rs1494961 located in the *HEL308* gene ($p_{replication} = 2\times10^{-5}$, $p_{combined} = 1\times10^{-8}$) (Table 1). In combined analysis, the association tended to be more pronounced in younger ages and smokers (Figure 4). Given the possible role of the *HEL308* in DNA repair, we also investigated the possibility that rs1494961 may play a role in lung cancer susceptibility by genotyping rs1494961 in a series of 5,652 lung cancer cases and 9,338 controls. We noted a similar association between rs1494961 and lung cancer (OR = 1.09, $p = 3\times10^{-4}$) from nine lung cancer studies, even when we excluded 1,844 cases and 2,735 controls where controls overlapped with the central European UADT study (OR = 1.09, p = 0.005).

Replication of *ADH* genes associations

The association between the *ADH* variants, rs1573496, rs1229984 and rs698 at 4q23 and UADT cancer was described previously [14] in the CE, ARCAGE (excluding Bremen) and SA studies. When excluding these studies, the association with these variants was independently replicated in the additional 5,092 UADT cancer cases and 6,794 controls presented here ($p = 5\times10^{-8}$, 7×10^{-9} and 0.02 for rs1573496, rs1229984 and rs698, respectively) (Table 1). We combined all studies totaling 8,744 UADT cancer cases and 11,982 study specific controls to investigate effects of the *ADH* variants among different strata (Figure 3). For both the *ADH1B* and *ADH7* variants heterogeneity was noted by UADT cancer subsite (p heterogeneity = 0.002, and 0.06 respectively). The rs1229984 *ADH1B* variant showed strong heterogeneity when stratifying by alcohol, with little evidence for association in never drinkers. By contrast, there was little evidence for heterogeneity noted with rs1573496 and rs698, but a statistically significant association with the *ADH7* variant rs1573496 was observed never drinkers (p = 0.03).

Among ever drinkers in this pooled analysis, the minor allele carriers of rs1229984 reported consuming less alcohol than non-carriers ($p = 3\times10^{-20}$). rs1573496 minor allele carriers similarly were noted to consume somewhat less alcohol (p = 0.002), while rs698 minor allele carriers consumed slightly more (p = 0.05) (Table 2). Adjustment for alcohol consumption made little difference to the risk estimates for UADT cancer with all three variants (Table S4).

Association in African Americans

We additionally genotyped the five variants significantly associated with UADT cancer in 537 African American UADT cancer cases and 539 controls noting a significant association for the 12q24 variant rs4767364 (p = 0.004) (Table 3). Nevertheless, the smaller sample size and potential differences in genetic architecture between European and African American populations (both in terms of allele frequencies and LD structure) limits our ability to assess these five alleles in African-Americans.

Discussion

Five genetic variants at three loci, 4q23, 12q24 and 4q21, were significantly associated with UADT cancers in the independent replication series or after correction for multiple testing at a genome wide level in combined analysis ($p \leq 5\times10^{-7}$). The risk effects noted with all five variants were less prominent in the replication series when compared with the initial finding in the discovery series, consistent with the notion of "winner's curse" [17]. In combination we estimate these 5 variants are likely to explain only a small proportion (approximately 4%) of the UADT cancer familial risk.

12q24

The 12q24 variant, rs4767364, is positioned in an extended region of LD that contains multiple genes. Candidate genes include the *aldehyde dehydrogenase 2 (ALDH2)* (Figure 2), another key gene in alcohol metabolism. The minor allele carriers of *ALDH2* variants rs737280 and rs4648328, in LD with rs4767364 ($r^2 = 0.86$ and 0.67, respectively), have been associated with differences in alcohol metabolism in Europeans, leading some authors to hypothesise [18] that these alleles have a similar, albeit more modest, effect in European populations to that of the *ALDH2* rs671 variant linked to alcohol metabolism differences [10] in Asian populations. The increased UADT cancer risk we observed with the minor allele of rs4767364 (and rs737280 by imputation, Figure 2) is similar to the UADT cancer risk effect observed for heterozygote rs671 carriers [12,13] and is consistent with this hypothesis. Nevertheless, this region contains many additional plausable candidate genes. Other GWAS have implicated multiple variants in this region in many phenotypes (type 1 diabetes, arthritis, renal function, hemoglobin concentration/hematocrit, coronary artery disease and waist-to-hip ratio) [19-26] and therefore the nature of the actual causative allele and gene remains to be determined. The rs4767364 variant was also associated with UADT cancer risk in a smaller series of African Americans implying that this effect may be relevant to other populations.

4q21

The 4q21 variant significantly associated with UADT cancers was rs1494961 located (Table 1) 20 Mb proximal to the *ADH* gene cluster. There is no LD between rs1494961 and either rs1229984, rs1573496 or rs698 ($r^2 < 0.003$). rs1494961 is a non-synonymous variant positioned in the *HEL308* gene, a single stranded DNA-dependent ATPase and DNA helicase involved in DNA intra-strand cross-linking repair [27], although the residue involved, I306V, is not an evolutionary conserved site [28] suggesting that this alteration may not have a functional consequence. rs1494961 is in a LD region spanning approximately 90 kb, and is highly correlated ($r^2 > 0.95$) with more than 20 common genetic variants. This region contains additional genes (Figure 2), notably a second DNA repair-related gene, *FAM175A* (or *Abraxas* and *CCDC98*), that interacts directly with the BRCT repeat region of *BRCA1* [29]. That a comparable association was noted between this

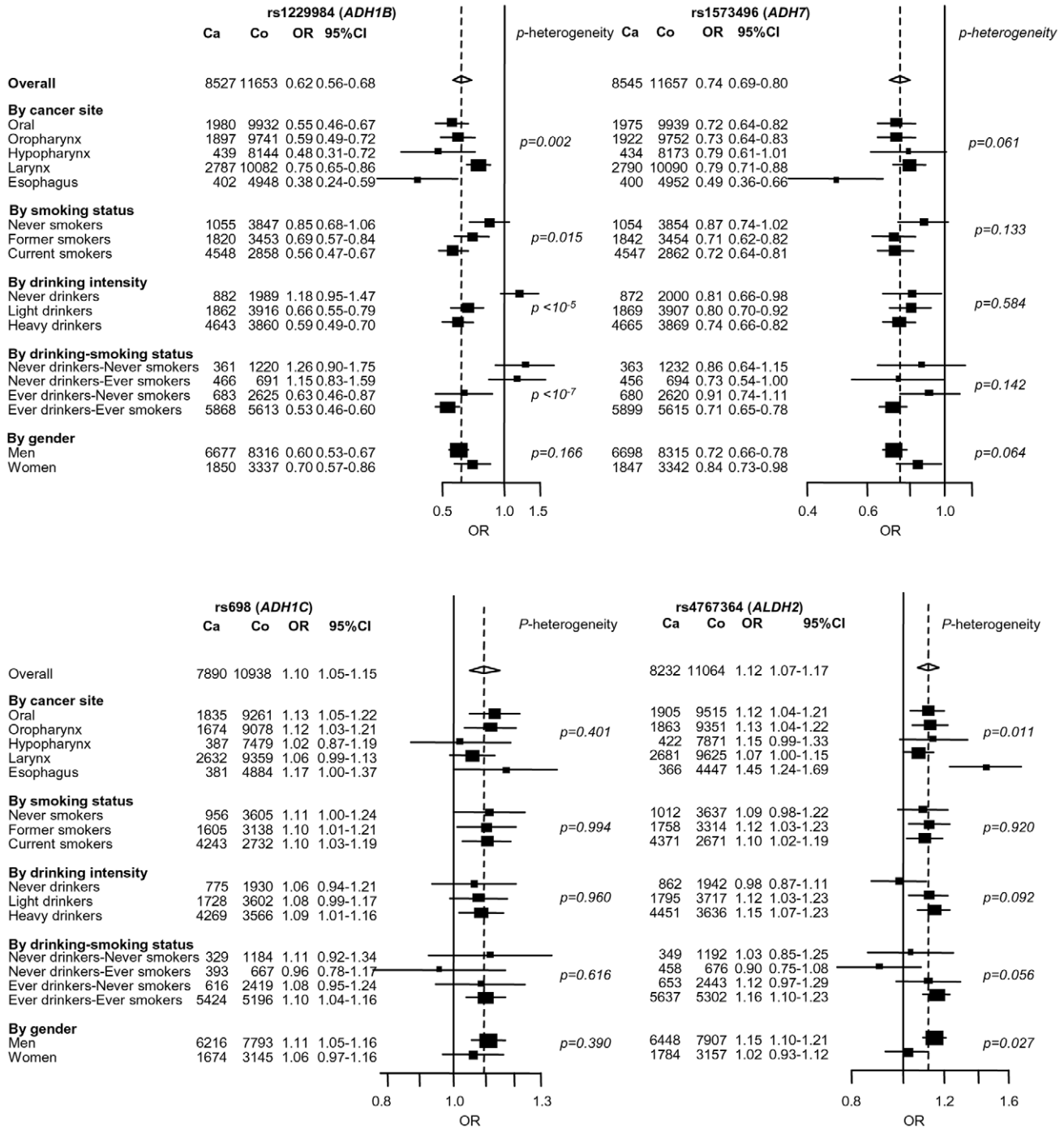

Figure 3. Stratified analysis of 4 replicated SNPs located near alcohol metabolism genes. Estimates for rs1229984 (*ADH1B*), rs1573496 (*ADH7*), rs1042758 (*ADH1C*) and rs4767364 (*ALDH2*) were derived from a log-additive genetic model. ORs were adjusted by age, sex, study and were derived from fixed effects models. "Generic" controls were not included in this analysis.
doi:10.1371/journal.pgen.1001333.g003

variant and lung cancer ($p = 3 \times 10^{-4}$) (Figure 4) suggests that the causal variant maybe relevant for cancers influenced by tobacco consumption in general.

4q23

The top two ranked variants (rs1573496 and rs698 and correlated variants) from the GWAS stage we have previously associated with UADT cancer risk [14]. The association between these variants, and a third variant, rs1229984, not included in the Humanhap300 beadchip but genotyped here based on our previous findings [14], and UADT cancer was independently replicated in the additional UADT cases and controls presented here ($p = 1 \times 10^{-7}$, 1×10^{-8} and 0.01 for rs1573496, rs1229984 and rs698, respectively).

The combined sample series presented here, totaling 8,774 UADT cancer cases and 11,982 controls, allowed further

Table 2. Association between rs1229984, rs1573496, rs698, rs4767364, and drinking intensity in ever drinkers expressed as mean of ml of ethanol consumed per day.

		All			Controls			UADT Cases		
		n	mean	CI 95%	n	mean	CI 95%	n	mean	CI 95%
rs1229984 (*ADH1B*)	CC	14,518	35.06	33.32–36.80	7,742	22.43	20.91–23.95	6,776	46.76	43.31–50.21
	CT,TT	1,323	22.85	18.73–26.98	907	15.60	12.35–18.85	416	23.97	14.68–33.25
p-trend				3×10^{-20}			5×10^{-10}			3×10^{-12}
rs1573496 (*ADH7*)	GG	12,936	35.02	33.21–36.82	6,823	22.36	20.79–23.93	6,113	46.84	43.28–50.40
	GC, CC	2,926	30.29	27.36–33.22	1,822	19.68	17.25–22.11	1,104	38.73	32.68–44.78
p-trend				*0.002*			*0.03*			*0.008*
rs698 (*ADH1C*)	TT	5,574	32.05	29.69–34.41	3,054	20.65	18.65–22.66	2,520	42.17	37.48–46.85
	TC	6,748	32.51	30.29–34.74	3,685	21.40	19.51–23.29	3,063	42.32	37.91–46.74
	CC	2,377	36.15	32.92–39.39	1,285	20.42	17.63–23.22	1,092	51.03	44.74–57.31
p-trend				*0.05*			*0.32*			*0.06*
rs4767364 (*ALDH2*)	GG	7,232	33.78	31.66–35.90	4,114	21.40	19.57–23.22	3,118	45.02	40.81–49.23
	AG	6,297	33.91	31.68–36.14	3,302	21.44	19.48–23.40	2,995	45.00	40.65–49.34
	AA	1,527	35.84	31.91–39.77	718	25.09	21.43–28.74	809	45.98	38.88–53.09
p-trend				*0.60*			*0.14*			*0.96*

Adjusted mean of ml per day were derived from ANOVA.
P-trend were derived from a linear regression with log(ml of ethanol per day) as an outcome using a log-additive genetic model.
All estimates were adjusted by sex, age, study, pack-years (and case/control status when appropriate).
doi:10.1371/journal.pgen.1001333.t002

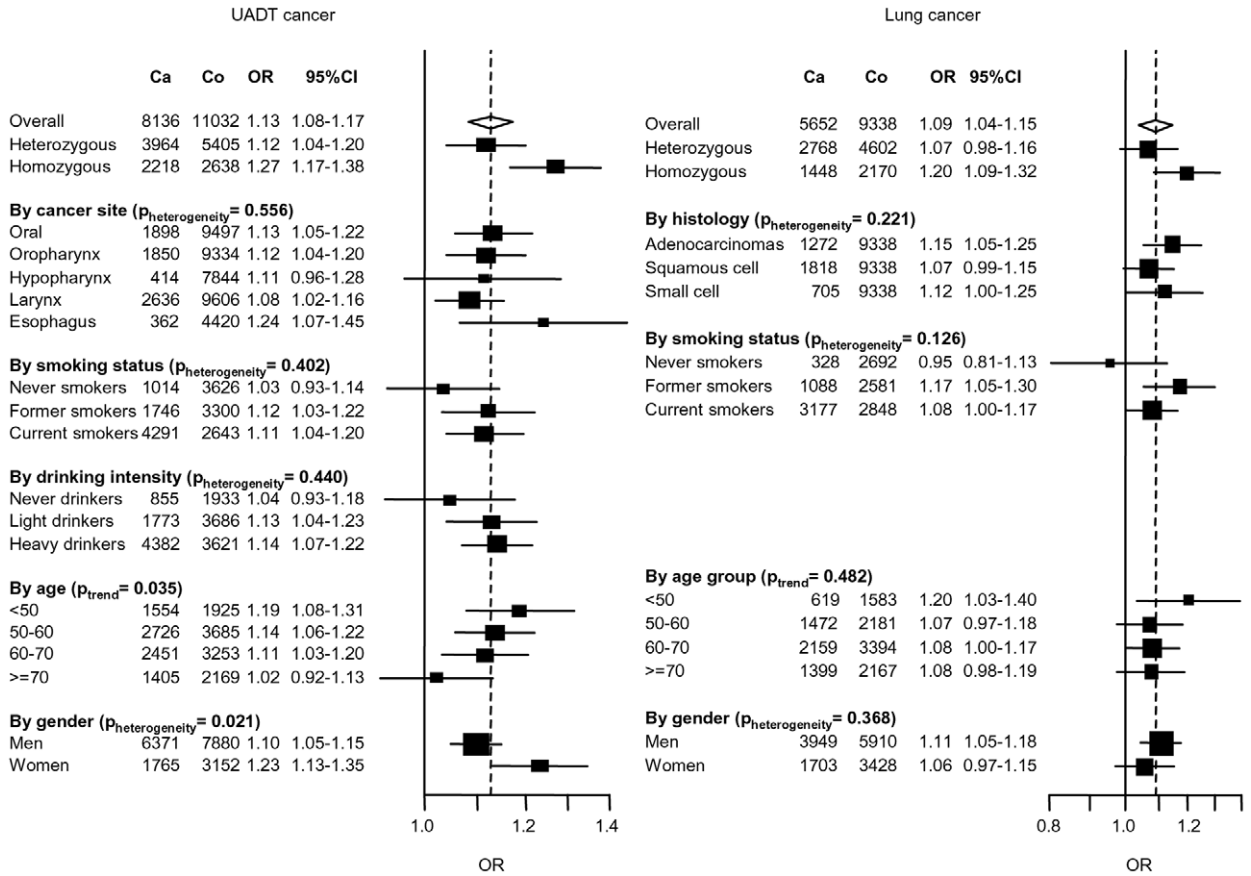

Figure 4. Association between 4q21 variant (rs1494961) and UADT and lung cancers. Estimates were derived from a log-additive model. ORs were adjusted by age, sex, study and were derived from fixed effects models. ''Generic'' controls were not included in this analysis.
doi:10.1371/journal.pgen.1001333.g004

Table 3. Comparison of results from the genome-wide analysis with analysis in a UADT case-control series of African-American origin.

Marker		Combined GWA and replication analysis (European descent)				African American				
		MAF	OR	95% CI	p-value	MAF	Ca/Co	OR	95% CI	p-value
rs1229984	(*ADH1B*)	0.06	0.64	0.59–0.71	1×10^{-20}	0.03	532/546	0.99	0.46–2.13	0.98
rs1573496	(*ADH7*)	0.11	0.75	0.70–0.80	9×10^{-17}	0.02	536/547	0.83	0.37–1.89	0.66
rs698	(*ADH1C*)	0.38	1.12	1.07–1.17	3×10^{-7}	0.18	508/527	1.10	0.87–1.39	0.44
rs4767364	(*ALDH2*)	0.29	1.13	1.08–1.18	2×10^{-8}	0.45	537/539	1.35	1.10–1.65	4×10^{-3}
rs1494961	(*HEL308*)	0.49	1.12	1.08–1.17	1×10^{-8}	0.26	544/540	1.06	0.88–1.29	0.53

Ca: number of cases; Co: number of controls.
MAF: minor allele frequency in controls.
doi:10.1371/journal.pgen.1001333.t003

exploration of these genetic effects among UADT cancer subsites and strata defined by gender, drinking and smoking. The effects of these three variants were generally present for each UADT subsites but more pronounced in esophageal cancers and males (Figure 3). Strong heterogeneity was found with rs1229984 when stratifying by alcohol consumption. Notably, an association was observed in ''Ever drinkers-Never smokers'', but not in ''Never drinkers-Ever smokers'', suggesting the effect with the rs1229984 variant is mediated through alcohol drinking rather than tobacco smoking. In contrast, the lack of heterogeneity for rs1573496 when stratifying by alcohol use may imply differences in the mechanism of carcinogenesis among these *ADH* variants.

Several studies have suggested rs1229984 may influence alcohol consumption behaviour [30-33]. We have strongly replicated this association ($p=3\times10^{-20}$). Similarly, minor allele carriers of rs1573496 and rs698 also consumed different amounts of alcohol compared with non-carriers (Table 2). Comparable to the observations made between 15q25 variants, propensity to smoke and lung cancer [34-36], adjustment for alcohol consumption did not fully explain the UADT cancer association with these variants (Table S4) suggesting, at least within the limits of this measurement of alcohol consumption, that these risks are unlikely to be explained by alcohol consumption behaviour patterns.

In conclusion, this study has identified two novel variants robustly associated with UADT cancers, and independently replicated three variants previously identified. All five variants variants are positioned near genes that appear relevant to etiology of UADT cancers, although further work is needed to identify the causative allele and gene at these loci.

Materials and Methods

Discovery phase study samples

Genome-wide genotyping was performed in two European based multi-centre UADT cancer case-control studies (Table 4), the International Agency for Research on Cancer (IARC) central europe study [14,37,34] conducted from 2000 to 2002, in 6 centers from 5 countries; and the ARCAGE [14,34,38] (**A**lcohol-**R**elated **C**ancers **a**nd **G**enetic susceptibility in **E**urope) multicentre case control study conducted by IARC from 2002 to 2005 in 12 centers from 9 European countries. DNA of sufficient quality and quantity for genome-wide genotyping was available for 2,230 UADT cancer cases (squamous cell carcinomas) and 4,090 controls from these two studies. We additionally included 4,983 generic controls to further increase statistical power. These generic controls included: 1,385 individuals from the 1958 birth cohort, (Wellcome Trust case control consortium[39]) as well as 1,823 French and 433 Norwegian controls genotyped by the Centre National Genotypage (CNG Evry France). We also included in our control series a separate group of 1,342 kidney cancer cases from the same centres as the central Europe study, inclusion or exclusion of these ''controls'' had no material effect on the results presented (Table S2). Both studies have been approved by local ethics committees as well as IARC IRB.

Genome-wide genotyping and quality control

The central Europe study and the ARCAGE study were genotyped using the Illumina Sentrix HumanHap300 BeadChip at the Centre d'Etude du Polymorphisme Humain (CEPH) and the CNG as described previously [34,40].

We conducted systematic quality control steps on the raw Illumina HumanHap300 genotyping data. Variants with a genotype call rate of less than 95% and also individuals where the overall genotype completion rate was less than 95% were excluded. We also conducted further exclusions where the genotype distribution clearly deviated from that expected by Hardy-Weinberg Equilibrium (HWE) among controls (p-value of less than 10^{-7}) and where there were discrepancies between sex based genotype and reported sex, as well as individuals with unlikely heterozygosity rates across genetic variants on the X chromosome (Table S1). Those genotyped were restricted to individuals of self – reported European ethnicity. To further increase the ethnic homogeneity of the series, we used the program STRUCTURE [41] to identify individuals of mixed ethnicity. Using a subseries of 12,898 genetic variants from the HumanHap 300 BeadChip panel evenly distributed across the genome and in low linkage disequilibrium (LD) ($r^2<0.004$) [42], we estimated the genetic profile of the study participants compared with individuals of known ethnic origins (the Caucasian, African and east-Asian individuals genotyped by the HapMap project). We excluded 34 individuals because of some evidence of ethnic admixture (Figure S3), indicating that the extent of admixture within the central Europe and ARCAGE study centers is limited.

Genome-wide statistical analysis

The association between each genetic variant and the disease risk was estimated by the odds ratio (OR) per allele and ninety-five percent confidence intervals (CI) using multivariate unconditional logistic regression assuming a log-additive genetic model with sex and country of recruitment included in the regression model as covariates. Results that obtained a level of significance of a two sided $p<5\times10^{-7}$ were considered significant at a genome wide

Table 4. The 15 UADT cancer studies participating in the genome-wide and replication analysis.

Study Name	Study setting	Coordinating centre	Genotyping centre	Principal Investigators	UADT Subsites[e]	Control source	Cases[a]	Controls[a]	Cases	Controls
GWAS									Post GWAS Qc	
ARCAGE [b]	Europe - Multicentre	IARC	CNG	Boffetta/ Brennan	UADT	Hospital-based	1,422	1,503	1,368	1,313
Central Europe [c]	Europe - Multicentre	IARC	CNG	Boffetta/ Brennan	UADT	Hospital-based	808	2,587	723	2,200
Generic controls										4,821
Replication										
SA[d]	Latin America - Multicentre	IARC	IARC	Boffetta/ Brennan	UADT	Hospital-based	1,422	1,098		
ARCAGE - Bremen	Bremen - Germany	Bremen Uni.	IARC	Ahrens	UADT	Hospital-based	164	190		
Rome	Roma - Italy	Uni. Rome	IARC	Boccia	HN	Hospital-based	251	237		
Poland	Szczecin - Poland	Szczecin Uni	IARC	Lubinski	Larynx	Hospital-based	409	1,039		
Seattle (Oral Gen study)	Washington-US	Fred Hutchinson Cancer Research Centre	FHCRC	Schwartz /Chen	HN	Population-based	193	388		
University of North Carolina (CHANCE study)	North Carolina - US	University of North Carolina	University of North Carolina	Olshan	HN	Population-based	940	1,087		
Penn State	Tampa - US	Penn State University	Penn State University	Muscat/ Lazarus		Hospital-based	310	534		
	Philadelphia, New York City - US			Lazarus	HN					
UCLA	Los Angeles - US	University of California, LA	University of California, LA	Zhang	UADT	Population-based	206	577		
MD Anderson	Houston - US	MD Anderson Cancer Centre	MD Anderson Cancer Centre	Wei/Sturgis	HN	Hospital-based	431	431		
IARC - oral cancer (ORC)	Europe - Multicentre	IARC	IARC	Franceschi	Oral	Hospital-based	611	643		
Boston (HNSCC)	Boston - US	Brown Uni.	Brown Uni.	Kelsey	HN	Population-based	513	593		
University of Pittsburgh (SCCHN-SPORE)	Pittsburgh - US	University of Pittsburgh	IARC	Romkes	HN	Hospital-based	610	771		
The Netherlands	Maastricht Hospital - Netherlands	University St Radboud		Lacko/Peters	HN	Hospital-based	454	304		
Total							**8,744**	**11,982**		

a Including only individuals of self-reported European ancestry.
b Includes countries: Czech Republic, Greece, Italy, Norway, UK, Spain, Croatia, Germany, France.
c Includes countries: Romania, Poland, Russia, Slovakia, Czech Republic.
d For the three variants at 4q23, results have been published previously, in ''replication'' analysis for these variants, the SA study was excluded.
e UADT –Oral, pharynx, laryngeal, esophageal cancers, HN – Head and neck cancers Oral, pharynx, laryngeal cancers.
doi:10.1371/journal.pgen.1001333.t004

level [39]. All analyses were conducted using PLINK [43]. We also conducted analyses restricting to UADT cancer subtypes (oral/pharyngeal cancer, laryngeal cancer, esophageal cancer) and restricting to heavy (>median) drinkers and heavy (>median) smokers.

The potential for population stratification not accounted for by adjustment by country was also investigated by principal components analysis (PCA) undertaken with the EIGENSTRAT package [44] using 12,898 markers in low LD [42]. Adjustment for population stratification using the PCA was performed by including significant eigenvectors that were associated with case control status ($p<0.05$) as covariates in the logistic regression.

Genotypes for genetic variants across 4q21, 4q23 and 12q21 not genotyped on the Illumina HumanHap300 BeadChip, but genotyped by the HAPMAP consortium, were imputed using the program MACH with phased genotypes from the CEU Hapmap genotyping as a scaffold. Unconditional logistic regression using posterior haplotype probabilities (haplotype dosages) from MACH were carried out using ProbABEL [45] including age, sex, and country of origin in the regression as covariates. Linkage Disequilibirum (LD) statistics (D' and r^2) were calculated using Haploview [46].

Replication study samples

The replication series consisted of 6,514 UADT cancer cases (squamous cell carcinomas) and 7,892 controls from 13 UADT cancer case-control studies (Table 4). With the exception of the Szczecin case-control study [16], all studies were part of the

INHANCE consortium. As previously described [1,3,47], all INHANCE studies have extensive information on tumor site and histology, as well as lifestyle characteristics. The Szczecin, Seattle, UCLA and MD Anderson studies were only able to genotype a proportion of the variants (Table S5). Results for the three *ADH* variants, rs1229984, rs1573496 and rs698 have been published previously for the Latin American study (LA). For these variants, in "replication" analysis the Latin American study was excluded. All studies have been approved by local ethics committees as well as IARC IRB.

Replication genotyping

Replication genotyping was performed using the TaqMan genotyping platform in 8 participating genotyping laboratories (Table 4). The robustness of the Taqman assays (primers and probes are available upon request) were confirmed at IARC by re-genotyping the CEPH HapMap (CEU) trios and confirming concordance with HapMap genotypes. Any discordance between Hapmap and Taqman generated genotypes was resolved by direct DNA sequencing. All Taqman assays were found to be performing robustly. IARC supplied Taqman assays and a standardized Taqman genotyping protocol to each of the 8 participating genotyping laboratories. A common series of 90 standard DNAs were genotyped at each laboratory to ensure the quality and comparability of the genotyping results across the different studies. Concordance with the consensus genotype and the results produced at the eight genotyping laboratories for the standardized DNAs was 99.75%, and no individual centre had a overall concordance of less than 99.5%. If the assay produced 2 or more discordant genotypes relative to the consensus, the study genotypes for this genetic variant were not included in the statistical analysis. Assays that had a per-centre success rate of $<90\%$ or for which genotype distributions deviated from HWE ($p<0.001$) were also excluded (Table S5).

Replication statistical analysis

The association between the nineteen variants and UADT cancer risk was estimated by per allele ORs and their 95% CI derived from multivariate unconditional logistic regression, with age, sex, and study (and country of origin where appropriate) included in the regression model as covariates. Measures of alcohol consumption have been previously harmonized across INHANCE studies [48]. The association between *ADH/ALDH2* variants and alcohol consumption was carried out in ever drinkers using multivariate linear regression using a log transformed milliliter of ethanol consumed per day as an outcome, adjusting for age, sex, study, packyears (and case-control status when appropriate). Milliliters of ethanol consumed per day was not available for 3 studies (Szczecin, Philadelphia/New York and The Netherlands study). Heterogeneity of ORs across the studies and across the stratification groups was assessed using the Cochran's Q-test. All replication and combined analyses were conducted using SAS 9.1 software. P values were two sided.

Investigation of the effects of 4q21 variant rs1494961 and lung cancer risk

The series of lung cancer cases and controls used to investigate 4q21 variant, rs1494961, and lung cancer risk included studies from central Europe (IARC), Toronto (McGill), HUNT2/Tromso, the CARET cohort, EPIC-lung, the Szczecin case-control study, Liverpool Lung Project (LLP), Paris France and Estonia as described previously [34,40,49]. All studies have been approved by local ethics committees as well as IARC IRB.

Genotyping protocol for 4q21 variant, rs1494961

Genotyping for rs1494691 was performed using the Illumina beadchips (Central Europe (IARC), Toronto (McGill), HUNT2/Tromso, the CARET cohort, France and Estonia) or the Applied Biosystems Taqman assays (EPIC-lung, the Szczecin case-control study, Liverpool Lung Project (LLP)) at IARC.

For the central European lung cancer study, the controls overlapped with the central European UADT cancer study for Bucharest (Romania), Lodz (Poland), Moscow (Russia), Banska Bystrika (Slovakia), and Olomouc and Prague (Czech Republic). We therefore performed analyses both including and excluding centres where controls overlapped.

Web resources

http://inhance.iarc.fr/ (December 2010)

http://www.hapmap.org (December 2010)

http://www.sph.umich.edu/csg/abecasis/mach/index.html (December 2010)

Supporting Information

Figure S1 Strategy for discovery and replication in the genome-wide association study.
Found at: doi:10.1371/journal.pgen.1001333.s001 (0.17 MB DOC)

Figure S2 Analysis of selected variants by study and by UADT cancer site in the replication series. For replication estimates of rs1229984, rs1573496, rs698, the SA study was excluded.
Found at: doi:10.1371/journal.pgen.1001333.s002 (0.26 MB DOC)

Figure S3 STRUCTURE Admixture plots. Individuals plotted against individuals of known Caucasian (CEU), African (YRI) and East Asian (JPT-CHB) origin. Individuals with greater than 30% admixture (dashed line) were excluded.
Found at: doi:10.1371/journal.pgen.1001333.s003 (0.30 MB DOC)

Table S1 Exclusion criteria of subjects for GWAS.
Found at: doi:10.1371/journal.pgen.1001333.s004 (0.18 MB DOC)

Table S2 Sensitivity analysis on the top variants identified by the genome-wide analysis.
Found at: doi:10.1371/journal.pgen.1001333.s005 (0.24 MB DOC)

Table S3 Selected demographic characteristics of cases and controls (GWAS and replication data combined).
Found at: doi:10.1371/journal.pgen.1001333.s006 (0.20 MB DOC)

Table S4 Comparison between analysis adjusted and unadjusted on tobacco and alcohol consumption.
Found at: doi:10.1371/journal.pgen.1001333.s007 (0.16 MB DOC)

Table S5 Minor allele frequency of each variant per study.
Found at: doi:10.1371/journal.pgen.1001333.s008 (0.22 MB DOC)

Acknowledgments

The authors thank all of the participants who took part in this research and the funders and support and technical staff who made this study possible. We thank Paul Pharoah and SEARCH for contribution of biological samples.

Author Contributions

Contributed reagents/materials/analysis tools: D Zaridze, O Shangina, N Szeszenia-Dabrowska, J Lissowska, P Rudnai, E Fabianova, A Bucur, V Bencko, I Holcatova, V Janout, L Foretova, P Lagiou, D Trichopoulos, S Benhamou, C Bouchardy, W Ahrens, F Merletti, L Richiardi, R Talamini, L Barzan, K Kjaerheim, GJ Macfarlane, TV Macfarlane, L Simonato, C Canova, A Agudo, X Castellsagué, R Lowry, DI Conway, PA McKinney, CM Healy, ME Toner, A Znaor, MP Curado, S Koifman, A Menezes, V Wünsch-Filho, J Eluf Neto, L Fernández Garrote, S Boccia, G Cadoni, D Arzani, AF Olshan, MC Weissler, WK Funkhouser, J Luo, J Lubiński, J Trubicka, M Lener, D Oszutowska, SM Schwartz, C Chen, S Fish, DR Doody, JE Muscat, P Lazarus, CJ Gallagher, S-C Chang, Z-F Zhang, Q Wei, EM Sturgis, L-E Wang, S Franceschi, R Herrero, KT Kelsey, MD McClean, CJ Marsit, HH Nelson, M Romkes, S Buch, T Nukui, S Zhong, M Lacko, JJ Manni, WHM Peters, RJ Hung, J McLaughlin, L Vatten, I Njølstad, GE Goodman, JK Field, T Liloglou, P Vineis, F Clavel-Chapelon, D Palli, R Tumino, V Krogh, S Panico, CA González, JR Quirós, C Martínez, C Navarro, E Ardanaz, N Larrañaga, K-T Khaw, T Key, HB Bueno-de-Mesquita, PHM Peeters, A Trichopoulou, J Linseisen, H Boeing, G Hallmans, K Overvad, A Tjønneland, M Kumle, E Riboli, K Välk, T Voodern, A Metspalu, P Galan, M Hashibe, RB Hayes, P Boffetta, P Brennan. Performed the experiments: JD McKay, A Chabrier, D Zelenika, A Boland, M Delepine, M Foglio, D Lechner, H Blanché, IG Gut. Analyzed the data: JD McKay, T Truong, V Gaborieau, S-C Chuang, G Byrnes, S Heath, M Hashibe. Conceived and designed the experiments: RB Hayes, P Boffetta, M Lathrop, P Brennan. Wrote the paper: JD McKay, T Truong, P Brennan.

References

1. Ferlay J, Bray F, Pisani P, Parkin DM (2004) GLOBOCAN 2002 Cancer Incidence Mortality and Prevalence Worldwide. IARC Cancer Base No 5 version 20: IARC Press Lyon.
2. Hashibe M, Brennan P, Benhamou S, Castellsague X, Chen C, et al. (2007) Alcohol drinking in never users of tobacco cigarette smoking in never drinkers and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. J Natl Cancer Inst 99: 777–789.
3. Herrero R, Castellsagué X, Pawlita M, Lissowska J, Kee F, et al. (2003) Human papillomavirus and oral cancer: the International Agency for Research on Cancer multicenter study. J Natl Cancer Inst 95: 1772–1783.
4. Negri E, Boffetta P, Berthiller J, Castellsague X, Curado MP, et al. (2009) Family history of cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. Int J Cancer 124: 394–401.
5. Goldstein AM, Blot WJ, Greenberg RS, Schoenberg JB, Austin DF, et al. (1994) Familial risk in oral and pharyngeal cancer. Eur J Cancer B Oral Oncol 30: 319–22.
6. Goldgar DE, Easton DF, Cannon-Albright LA, Skolnick MH (1994) Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. J Natl Cancer Inst 86: 1600–1608.
7. Garavello W, Foschi R, Talamini R, La Vecchia C, Rossi M, et al. (2008) Family history and the risk of oral and pharyngeal cancer. Int J Cancer 122: 1827–31.
8. Brennan P, Lewis S, Hashibe M, Bell DA, Boffetta P, et al. (2004) Pooled analysis of alcohol dehydrogenase genotypes and head and neck cancer: a HuGE review. Am J Epidemiol 159(1): 1–16.
9. IARC (1988) Alcohol drinking IARC monographs on the evaluation of carcinogenic risks to humans. Vol 44: Lyon IARC.
10. Yoshida A, Huang IY, Ikawa M (1984) Molecular abnormality of an inactive aldehyde dehydrogenase variant commonly found in Orientals. Proc Natl Acad Sci USA 81: 258–261.
11. Enomoto N, Takase S, Yasuhara M, Takada A (1991) Acetaldehyde metabolism in different aldehyde dehydrogenase-2 genotypes. Alcohol Clin Exp Res 15: 141–144.
12. Lewis SJ, Smith GD (2005) Alcohol ALDH2 and esophageal cancer: a meta-analysis which illustrates the potentials and limitations of a Mendelian randomization approach. Cancer Epidemiol Biomarkers Prev 14: 1967–1971.
13. Brooks PJ, Enoch MA, Goldman D, Li TK, Yokoyama A (2009) The alcohol flushing response: an unrecognized risk factor for esophageal cancer from alcohol consumption. PLoS Med 6: e50. doi:10.1371/journal.pmed.1000050.
14. Hashibe M, McKay JD, Curado Oliveira J, Koifman S, Koifman R, et al. (2008) Multiple ADH genes are associated with upper aero-digestive cancers in three large independent studies. Nature Genetics 40: 707–709.
15. 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: 1794–801.
16. Cybulski C, Masojc B, Oszutowska D, Jaworowska E, Grodzki T, et al. (2008) Constitutional CHEK2 mutations are associated with a decreased risk of lung and laryngeal cancers. Carcinogenesis 29: 762–765.
17. Xiao R, Boehnke M (2009) Quantifying and correcting for the winner's curse in genetic association studies. Genet Epidemiol. 33: 453–62.
18. Dickson PA, James MR, Heath AC, Montgomery GW, Martin NG, et al. (2006) Effects of variation at the ALDH2 locus on alcohol metabolism sensitivity consumption and dependence in Europeans. Alcohol Clin Exp Res 30: 1093–1100.
19. Cooper JD, Smyth DJ, Smiles AM, Plagnol V, Walker NM, et al. (2008) Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. Nat Genet. 40: 1399–401.
20. Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, et al. (2009) Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet. 41: 703–707.
21. Prahalad S, Hansen S, Whiting A, Guthery SL, Clifford B, et al. (2009) Variants in TNFAIP3, STAT4, and C12orf30 loci associated with multiple autoimmune diseases are also associated with juvenile idiopathic arthritis. Arthritis Rheum 60: 2124–30.
22. Köttgen A, Pattaro C, Böger CA, Fuchsberger C, Olden M, et al. (2010) New loci associated with kidney function and chronic kidney disease. Nat Genet. 42: 376–84.
23. Ganesh SK, Zakai NA, van Rooij FJ, Soranzo N, Smith AV, et al. (2009) Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. Nat Genet. 41: 1191–8.
24. Soranzo N, Spector TD, Mangino M, Kühnel B, Rendon A, et al. (2009) A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. Nat Genet. 41: 1182–90.
25. Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, et al. (2010) Genome-wide association study of hematological and biochemical traits in a Japanese population. Nat Genet. 42: 210–5.
26. Cho YS, Go MJ, Kim YJ, Heo JY, Oh JH, et al. (2009) A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. Nat Genet. 41: 527–34.
27. Marini F, Wood RD (2002) A human DNA helicase homologous to the DNA cross-link sensitivity protein Mus308. J Biol Chem 277: 8716–8723.
28. Marini F, Kim N, Schuffert A, Wood RD (2003) POLN a nuclear PolA family DNA polymerase homologous to the DNA cross-link sensitivity protein Mus308. J Biol Chem 278: 32014–32019.
29. Wang B, Matsuoka S, Ballif BA, Zhang D, Smogorzewska A, et al. (2007) Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response. Science 316: 1194–1198.
30. Macgregor S, Lind PA, Bucholz KK, Hansell NK, Madden PA, et al. (2009) Associations of ADH and ALDH2 gene variation with self report alcohol reactions consumption and dependence: an integrated analysis. Hum Mol Genet 18: 580–593.
31. Tolstrup JS, Nordestgaard BG, Rasmussen S, Tybjaerg-Hansen A, Grønbaek M (2008) Alcoholism and alcohol drinking habits predicted from alcohol dehydrogenase genes. Pharmacogenomics J 8(3): 220–7.
32. Zuccolo L, Fitz-Simon N, Gray R, Ring SM, Sayal K, et al. (2009) A non-synonymous variant in ADH1B is strongly associated with prenatal alcohol use in a European sample of pregnant women. Hum Mol Genet 15: 4457–66.
33. Luo X, Kranzler HR, Zuo L, Wang S, Schork NJ, et al. (2006) Diplotype trend regression analysis of the ADH gene cluster and the ALDH2 gene: multiple significant associations with alcohol dependence. Am J Hum Genet 78: 973–87.
34. Hung RJ, McKay JD, Gaborieau 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.
35. 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: 616–22.
36. 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: 638–42.
37. Scelo G, Constantinescu V, Csiki I, Zaridze D, Szeszenia-Dabrowska N, et al. Occupational exposure to vinyl chloride acrylonitrile and styrene and lung cancer risk (Europe). Cancer Causes Control 2004 15: 445–52.
38. Lagiou P, Georgila C, Minaki P, Ahrens W, Pohlabeln H, et al. (2009) Alcohol-related cancers and genetic susceptibility in Europe: the ARCAGE project: study samples and data collection. Eur J Cancer Prev 18: 76–84.
39. The Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14000 cases of seven common diseases and 3000 shared controls. Nature 447: 661–678.
40. McKay JD, Hung RJ, Gaborieau V, Boffetta P, Chabrier A, et al. (2008) Lung cancer susceptibility locus at 5p15.33. Nat Genet. 40: 1404–6.
41. Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164: 1567–1587.
42. Yu K, Wang Z, Li Q, Wacholder S, Hunter DJ, et al. (2008) Population substructure and control selection in genome-wide association studies. PLoS ONE 3: e2551. doi:10.1371/journal.pone.0002551.

43. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559–575.
44. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38: 904–909.
45. Aulchenko YS, Struchalin MV, van Duijn CM (2010) ProbABEL package for genome-wide association analysis of imputed data. BMC Bioinformatics 11: 134.
46. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21: 263–5.
47. Conway DI, Hashibe M, Boffetta P, INHANCE consortium, Wunsch-Filho V, et al. (2009) Enhancing epidemiologic research on head and neck cancer: INHANCE - The international head and neck cancer epidemiology consortium. Oral Oncol 45: 743–746.
48. Purdue MP, Hashibe M, Berthiller J, La Vecchia C, Dal Maso L, et al. (2009) Type of alcoholic beverage and risk of head and neck cancer–a pooled analysis within the INHANCE Consortium. Am J Epidemiol. 169: 132–42.
49. Lips EH, Gaborieau V, McKay JD, Chabrier A, Hung RJ, et al. (2010) Association between a 15q25 gene variant smoking quantity and tobacco-related cancers among 17 000 individuals. Int J Epidemiol. 39(2): 563–77.