# Cross-cancer genome-wide association analysis of lung, ovary, breast, prostate and colon cancer identifies a novel cancer locus at 1q22

Gordon Fehringer[1], Peter Kraft[2], Christopher A. Haiman [3], Paul Pharoah[4], Rosalind Eeles[5], Nilanjan Chatterjee[6], Fred Schumacher[3], Joellen Schildkraut[7], Paul Brennan[8], Heike Bickeböller[9], Richard Houlston[5], Maria Teresa Landi[6], Neil Caporaso[6], Angela Risch[10], Ali Amin Al Olama[11], Sonja I Berndt[6], Edward Giovannucci[2], Henrik Grönberg[12], Zsofia Kote-Jarai[5], Jing Ma[13], Kenneth Muir[14], Meir Stampfer[2], Victoria L. Stevens[15], Fredrik Wiklund[12], Walter Willett[2], Ellen L. Goode[16], Jenny Permuth-Wey[17], Harvey Risch[18], Brett M. Reid[17], Stephane Bezieau[19], Hermann Brenner[20], Andrew T. Chan[13], Thomas J. Hudson[21], Jonathan K Kocarnik[22], Polly A. Newcomb[22], Robert E. Schoen[23], Martha L. Slattery[24], Emily White[22], Muriel Adank[25] on behalf of Hereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON), Habibul Ahsan [26], Kristiina Aittomäki[27], Laura Baglietto[28], Sonja Berndt[6], Carl Blomquist[27], Federico Canzian[29], Kamila Czene [12], Isabel dos-Santos-Silva [30], A. Heather Eliassen[13], Jonine Figueroa[6], Dieter Flesch-Janys[31], Olivia Fletcher[32], Montserrat Garcia-Closas[6], Mia M. Gaudet[15], Nichola Johnson[32], Per Hall [12], Aditi Hazra[13], Rebecca Hein [33], Albert Hofman[34], John L. Hopper[35], Astrid Irwanto[36], Mattias Johansson[37], Rudolf Kaaks[10], Muhammad G. Kibriya[26], Peter Lichtner[38], Sara Lindström[2], Jianjun Liu[36], Eiliv Lund[39], Enes Makalic[35], Alfons Meindl[40], Bertram Müller-Myhsok[41], Taru A. Muranen[27], Heli Nevanlinna[27], Petra H. Peeters[42], Julian Peto[30], Ross L. Prentice[22], Nazneen Rahman[43], Maria Jose Sanchez[44], Daniel F. Schmidt[35], Rita K. Schmutzler[45], Melissa C. Southey[46], Rulla Tamimi[13], Ruth C. Travis[47], Clare Turnbull[43], Andre G. Uitterlinden[34], Zhaoming Wang[6], Alice S. Whittemore[48], Rose Yang[6], Wei Zheng [49], Thorunn Rafnar[50], Julius Gudmundsson[50], Simon N Stacey[50], Kari Stefansson[50], Patrick Sulem[50], The PRACTICAL Consortium, Ovarian Cancer Association Consortium (OCAC), Y. Ann Chen[17], Jonathan P. Tyrer[4], David C. Christiani[2], Yongyue Wei[2], African American Breast Cancer Consortium (AABC), African Ancestry Prostate Cancer Consortium (AAPC), Japanese American Prostate Cancer Consortium (JAPC), Latino American Breast Cancer Consortium (LABC), Latino American Prostate Cancer Consortium (LAPC), Hongbing Shen[51], Dr Zhibin Hu [51], Xiao-Ou Shu[49], Kouya Shiraishi[52], Atsushi Takahashi[53], Yohan Bossé[54], Ma'en Obeidat[55], David Nickle[56], Wim Timens[57], Matthew L. Freedman[58], Qiyuan Li[58], Daniela Seminara[6], Stephen J. Chanock[6], Jian Gong[22], Ulrike Peters[22], Stephen Gruber on behalf of Colorectal Transdisciplinary (CORECT) Study[3], Christopher I. Amos[59], Thomas A. Sellers[17], Douglas Easton[4], David J Hunter[2], Brian E. Henderson[3], Rayjean J Hung[1].

1. Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Canada. 2. Harvard T.H. Chan School of Public Health, Boston MA. 3. University of South California, Los Angeles, USA. 4. University of Cambridge, Cambridge, UK. 5. Institute of Cancer Research, London, UK. 6. National Cancer Instititute, Bethesda, USA. 7. Duke University, Durham, USA. 8. International Agency for Research on Cancer, Lyon, France. 9. University of Göttingen, Medical School, Göttingen, Germany. 10. National Center for Tumor Diseases and German Cancer Research Center (DKFZ), Heidelberg, Germany. 11. Cambridge University, Cambridge, United Kingdom. 12. Karolinska Institutet, Stockholm, Sweden. 13. Harvard Medical School, Boston MA; Brigham and Women's Hospital, Boston MA. 14. University of Manchester, Manchester, UK. The University of Warwick, Coventry, UK. 15. Epidemiology Research Program, American Cancer Society, Atlanta GA. 16. Mayo Clinic, Rochester Minnesota. 17. Moffitt Cancer Center, Tampa, USA. 18. Yale University, New Haven, Connecticut, USA. 19. Service de Génétique Médicale, Nantes, France. 20. Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany. 21. Ontario Institute for Cancer Research, Toronto Ontario. 22. Fred Hutchinson Cancer Research Center, Seattle, USA. 23. University of Pittsburgh Medical Center, Pittsburgh, USA. 24. University of Utah Health Sciences Center, Salt Lake City, Utah, USA. 25. VU University Medical Center, Amsterdam, The Netherlands. 26. University of Chicago, Chicago, IL, USA.

27. University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland. 28. Inserm U1018, Paris-South University, Villejuif, France. 29. Genomic Epidemiology Group, German Cancer Research Center (DKFZ) Heidelberg, Germany. 30. Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, U.K. 31. University Medical Center Hamburg-Eppendorf, Hamburg Germany. 32. Breakthrough Research Centre, The Institute of Cancer Research, London, UK. 33. Division of Cancer Epidemiology, German Cancer Research Center (DKFZ Heidelberg), Germany. 34. Erasmus Medical Center, Rotterdam, The Netherlands. 35. Melbourne School of Population Health, University of Melbourne, Melbourne, Victoria, Australia. 36. Genome Institute of Singapore, Singapore. 37. Department of Biobank Research, Umea University, Sweden; International Agency for Research on Cancer (IARC/WHO). 38. German Research Center for Environmental Health, Neuherberg, Germany . 39. Institute of Community Medicine, UiT The Artic University of Norway, Tromso, Norway. 40. Technische Universität München, Munich, Germany. 41. Max Planck Institute of Psychiatry, Munich, Germany. 42. Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht. 43. Institute of Cancer Research, Sutton, U.K. 44. Escuela Andaluza de Salud Publica, Instituto de Investigacion Biosanitaria ibs.GRANADA. Hospitales Universitarios de Granada/Universidad de Granada. CIBER de Epidemiología y Salud Pública CIBERESP, Spain. 45. University of Cologne, Cologne, Germany. 46. The University of Melbourne, Melbourne, Victoria, Australia. 47. Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, U.K. 48. Stanford University, Stanford, CA, USA. 49. Vanderbilt University, Nashville, USA. 50. deCODE genetics, Amgen, Reykjavik, Iceland. 51. Nanjing Medical University School of Public Health, Nanjing, China. 52. Division of Genome Biology, National Cancer Center Research Institute. Tokyo, Japan. 53. Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. 54. Institut universitaire de cardiologie et de pneumologie de Québec, Department of Molecular Medicine, Laval University, Québec, Canada. 55. University of British Columbia Centre for Heart Lung, Innovation, St. Paul's Hospital. Vancouver, Canada. 56. Merck & Co, MRL, Seattle, Washington, United States. 57. University of Groningen, University Medical Center Groningen, GRIAC Research Institute, Groningen, Netherlands. 58. Dana-Farber Cancer Institute, Boston, USA. 59. Geisel School of Medicine, Dartmouth College, Lebanon.

**Abstract**

Identifying genetic variants with pleiotropic associations can uncover common pathways influencing multiple cancers and further understanding of cancer susceptibility. Our 2-staged approach used genome-wide association results for lung, ovary, breast, prostate and colon cancer from the GAME-ON/GECCO Network (61,851 cases, 61,820 controls) to identify pleiotropic associations, and independent studies (55,789 cases, 330,490 controls) for replication. We identified a novel pleiotropic association at 1q22 with a variant associated with breast and lung squamous cell carcinoma (overall (both stages) P-value for both cancers combined=$8.9 \times 10^{-8}$), with eQTL analysis showing a consistent association with *ADAM15*/*THBS3* gene expression in lung tissues. New pleiotropic associations were also found at previously known cancer loci: variants at a known *BRCA2* locus for lung and breast cancer were associated with serous ovarian cancer (overall p-value=$4.0 \times 10^{-8}$); a known breast cancer locus, *CASP8/ALS2CR12*, with a variant associated with prostate cancer (overall P-value=$1.9 \times 10^{-8}$), and a known breast cancer locus, *CDKN2B-AS1*, where one variant was associated with lung adenocarcinoma (overall P-value=$1.0 \times 10^{-5}$) and a second was associated with prostate cancer (overall P-value=$9.5 \times 10^{-7}$). Our results provide important insights into common carcinogenesis across multiple major cancers and highlight the value of pleiotropy analysis.

**Introduction**

Genome wide association studies (GWAS) have identified hundreds of genetic variants that are associated with risk of specific cancers[1]. It has been observed that some chromosomal regions demonstrate pleiotropic associations, where variants at one locus are associated with multiple cancers. One of the first identified pleiotropic loci is the 8q24 locus, where genetic variants are associated with breast, prostate, colorectal and ovarian cancer risk[2], with some of the variants at this locus only associated with one cancer, while others are associated with multiple cancers[3,4] . Similarly, genetic variants at the *TERT-CLPTM1L* region at 5p15.33 are associated with risk of lung, bladder, prostate cervical, pancreatic and other cancers[5,6].

The identification of pleiotropic loci is an important step in improving our knowledge of cancer etiology by potentially identifying pathways that influence carcinogenesis of different tumors, and in furthering understanding of susceptibility for cancer. Furthermore, analyzing genomic data across multiple cancer sites might identify novel susceptibility loci, as variants that do not meet the stringent criteria for GWAS significance for any one cancer site, might show a significant association when multiple cancers are analyzed together[7].

In this study we performed a genome wide investigation of pleiotropic associations across five common cancers - lung, ovary, breast, prostate and colorectal cancer using data from the Genetic Associations and Mechanisms in Oncology (GAME-ON) Network and the Genetic and Epidemiology of Colorectal Cancer Consortium (GECCO)[8]. The GAME-ON Network was launched by the National Cancer Institute (NCI) to capitalize on the extensive investment in GWAS, with the overarching goal to integrate post-GWAS research and to facilitate analyses that address research questions that are common across multiple cancer sites. The GAME-ON Network is focused on tumors that currently represent a major public health burden and has assembled extensive genomic data from consortia investigating the cancer sites that constitute the basis of our cross-cancer analysis. Our study is the largest investigation of pleiotropic associations to date using GWAS results for 61,851 cases and 61,820 controls and testing nearly 10 million variants.

**Results**

After applying quality control filters (see methods) we analyzed 9,916,564 variants for pleiotropic associations using 61,851 cases and 61,820 controls of European ancestry across five common cancer sites in the GAME-ON Network and GECCO (GAME-ON/GECCO). The characteristics of the contributing studies are summarized in Table 1. Figure 1 displays a Circos plot showing association test results for each of the five main cancer sites investigated. Multiple peaks can be seen for each site with breast and prostate showing the most genome-wide significant associations.

We used the association analysis based on subsets (ASSET) meta-analytic approach[7] to investigate pleiotropic associations across cancer sites using summary level data from the GAME-ON/GECCO discovery set. This method generalizes the standard fixed meta-analysis by jointly examining the association between each genetic variant within subsets of cancers and allowing for subsets with opposing directions of association and null associations. We supplemented this approach with a standard fixed effects meta-analysis and by reviewing regions where multiple sites were associated with a single variant (see Methods for details and rationale). One hundred and ninety variants in 33 regions were prioritized for follow-up in replication and generalizability based on the following criteria: (1) variant associations that were significant at the P-value threshold of $5 \times 10^{-7}$ for two-sided or one-sided (positively or negatively associated) ASSET tests, or in standard fixed effect meta-analyses (using ASSET); or (2) variant associations with P-value of $5 \times 10^{-3}$ or less for at least two cancer sites (including subtypes from different cancer sites) (Supplementary Table 1). Our replication data sets included a total of 55,789 cases and 330,490 controls of European descent from deCODE (all 5 cancers), Harvard (lung cancer), iCOGS (breast cancer), PRACTICAL/iCOGS (Prostate cancer) and OCAC/iCOGS (ovarian cancer). An additional 46,785 cases and 42,892 controls from iCOG Breast cancer were used for validation of our novel pleiotropic association at 1q22 (see below). For generalizability, we conducted *in silico* look up of GWAS results from Nanjing (lung), Japan (lung), Shanghai (breast), San Francisco (breast for Latinos), the Japanese and Latino populations in Multiethnic Cohort (MEC) study (breast and prostate) the African American Breast Cancer GWAS Consortium (AABC), and the African Ancestry Prostate Cancer GWAS Consortium (AAPC) and with a total of 18,152 cancer cases and 21,410 controls (Table 1). Out of the 33 regions selected for follow-up, we replicated associations at four regions (at $P \leq 0.05$). We describe the most significant associations in each region below (see supplementary Table 1 and 2 for summary).

**<u>Novel susceptibility region: 1q22 for lung squamous cell carcinoma and breast cancer</u>**

Standard meta-analysis of GAME-ON/GECCO discovery data identified an association between rs1057941 located at 1q22 and overall risk of cancer ($P=1.74 \times 10^{-7}$, Fig. 2a). Overall lung, lung squamous cell carcinoma (lung SqCC) and breast cancer were strongly associated with this variant in GAME-ON/GECCO data (Lung: OR=1.08, 95% CI 1.05-1.12, $P=9.2 \times 10^{-6}$; lung SqCC: OR=1.10, 95% CI 1.04-1.16, P=0.001; Breast: OR=1.07, 95% CI 1.04-1.11, $P=6.08 \times 10^{-5}$). The association for lung SqCC was replicated in deCODE and Harvard studies combined, with OR of 1.12 (95% CI 1.01-1.23, P=0.03). The association with breast cancer was replicated in deCODE and iCOGS combined (OR=1.02, 95%CI 1.00-1.04, P=0.01). The p-value for association from a meta-analysis of both discovery and validation sets (i.e., GAME-

ON/GECCO-deCODE-Harvard) for lung SqCC and breast cancer approached genome-wide significance ($P=8.9 \times 10^{-8}$) (Fig. 2a).

In addition to breast, lung and lung SqCC, the aggressive form of prostate cancer was selected by ASSET as part of the subset of cancers associated with rs1057941. We have no replication data for this cancer and we did not replicate the GAME-ON/GECCO association with prostate cancer overall (deCODE and PRACTICAL/iCOGS combined: OR=1.02, 95% CI 1.00-1.05, P=0.08). We did not see associations for lung SqCC or breast cancer in other ethnic groups (data not shown).

Regional plots constructed from GAME-ON/GECCO results show a distinct peak in P-values in a 40kb region of LD at 1q22 that includes *KRTCAP2, GBA, MTX1, MUC1, TRIM46, THBS3, ADAM15 and ASH1L* (Fig. 2b). Although the strength of the meta-analysis signal led us to identify rs1057941 for replication other variants in the region with a weaker cross-cancer meta-analysis signal had the most significant site-specific associations: *MUC1* variant rs4072037 for lung SqCC ($P=3.21 \times 10^{-4}$) and the *TRIM46* variant, rs3814316, for breast cancer ($P=3.06 \times 10^{-6}$) (Fig. 2b and supplementary Fig. 1a-b).

We conducted eQTL analysis for lung SqCC associated variants in this region in non-tumor lung tissues of 1,111 patients from three studies assembled by Laval University (n=409), The University of British Columbia (n=339) and the University of Groningen (n=363). These analyses found that rs4072037 acted as a normal lung tissue eQTL for two genes in this region, *ADAM15* and *THBS3,* with two studies showing significant associations after adjustment for multiple comparisons and the third showing a nominally significant association consistent in direction with the other two (*ADAM15*: Laval $P=2.39 \times 10^{-7}$, University of British Columbia $P=4.09 \times 10^{-5}$, University of Groningen P=0.08; *THBS3*: Laval $P=1.71 \times 10^{-5}$, University of British Columbia $P=4.15 \times 10^{-6}$, University of Groningen P=0.004) (Fig. 3a and b). The risk allele, A, was consistently associated with increased gene expression for all studies.

**Previously known cancer loci with newly identified pleiotropic associations**

***13q13.1 BRCA2 (known for breast and lung cancer) and serous ovarian cancer***

We observed genome-wide significant pleiotropic associations for three rare *BRCA2* variants rs11571815, rs11571818 and rs11571833 (ASSET two-sided P-values: rs11571815 $P=5.53 \times 10^{-10}$, rs11571818 $P=5.45 \times 10^{-10}$, rs11571833 $P=6.14 \times 10^{-10}$) in GAME-ON/GECCO. These variants are in perfect LD with each other according to 1000 genomes March 2012 release data (CEU). As results were nearly identical in our data sets we focus on rs11571833 (Fig. 4a), a potentially functional variant identified in previous GWAS as associated with breast cancer and lung cancer (primarily driven by lung SqCC), where the latter study used a subset of the lung cancer data included here [9,10]. Our analysis indicates an additional association with serous ovarian cancer (OR=1.76, 95% CI 1.30-2.38; $P=2.49 \times 10^{-4}$), that was replicated in iCOGS (OR=1.45 95% CI=1.22-1.72, $P=3.08 \times 10^{-5}$), meeting our corrected significance threshold of $P \leq 0.0003$ (see methods) (Fig. 4a). Combining GAME-ON/GECCO and iCOGS resulted in a genome-wide significant association between rs11571833 and serous ovarian cancer ($P=3.95 \times 10^{-8}$). We did not find a significant association with breast cancer for this variant (OR=1.22, 95% CI 0.96-1.54, P=0.10) in GAME-ON/GECCO. The associations for lung and serous ovarian cancers are in the same

direction, and the one sided ASSET test indicated an even stronger association for the subset that included these cancers (P=9.4 X $10^{-13}$) than the two sided p-value (Fig. 4a). Regional plots show that subset meta-analysis and serous ovarian cancer associations are strongest at rs11571833 and its two neighboring variants rs11571815 and rs11571818 (Fig. 4b and supplementary Fig. 2). Our investigation of ovarian eQTL for these variants did not produce a significant association.

***2q33.1 CASP8/ALS2CR12 (known for breast cancer and malignant melanoma) and prostate cancer***

We identified a pleiotropic association between rs13016963 located in 2q33.1 and prostate (OR=1.08, 95% CI 1.04-1.13, P=3.05 x $10^{-5}$) and breast cancer (OR=0.93, 95% CI 0.90-0.96, P=5.75 x$10^{-5}$) (Fig. 5a) in GAME-ON/GECCO. This variant was associated with melanoma in a previous GWAS[11]. The association we found with breast cancer can be explained through LD with previously identified breast cancer variants in the region[12-15]. The association with prostate cancer was replicated in deCODE and iCOGS (OR=1.05, 95% CI 1.03-1.08, P=7.6 x $10^{-5}$) meeting our corrected significance threshold of P ≤ 0.0003 (Fig. 5a). The combined GAME-ON/GECCO, deCODE and iCOGS p-value was 1.9 x $10^{-8}$. This variant is in intron 5 of *ALS2CR12*, adjacent to *CASP8*. It sits in a region of high LD that includes several variants at these two genes that show similar strength associations with prostate cancer (Fig. 5b).

We conducted an eQTL analysis for variants showing the most significant associations with prostate cancer in the region, all of which were in strong LD ($R^2 \geq 0.70$) with rs13016963. We combined 145 prostate tumour samples and 33 normal tissue samples from TCGA for the analysis. In figure 6, we show results for rs1035142, which was in perfect LD with rs13016963 in this sample and was associated with *BZW1* (P=0.001, FDR=0.04).

***9p21.3 CDKN2B1 (known for lung SqCC and breast cancer) and lung adenocarcinoma***

The variant rs62560775 at *CDKN2B-AS1* located in 9p21.3 showed evidence for pleiotropy in GAME-ON/GECCO with associations for lung adenocarcinoma (OR=1.19, 95% CI = 1.08-1.31, P=2.77 x $10^{-4}$) and breast cancer (OR=1.11, 95% CI 1.05-1.17, P=5.30 x $10^{-4}$). Variant rs62560775 is in the vicinity of a previously reported breast cancer risk allele at the same gene. This region was previously reported as a lung SqCC susceptibility locus, but this is the first time that we observed an association with lung adenocarcinoma. It was replicated based on the combined data of deCODE and Harvard (OR=1.16, 95% CI 1.03-1.30, P-value=0.01). The combined p-value for GAME-ON/GECCO, deCODE and Harvard was 1.0 x $10^{-5}$ (Fig. 7a). The association of this variant with lung adenocarcinoma was the most significant in the region (Fig. 7b). Our lung eQTL investigation of this variant showed no significant association with gene expression in this region.

A second variant in the region, rs1011970, was associated with prostate cancer (OR=1.10, 95% CI = 1.05-1.15, P=7.3 x $10^{-5}$). This variant was found to be associated with breast cancer in a previous GWAS[16]. The association with prostate cancer was replicated in deCODE and iCOGS combined (P=0.001). The combined p-value for GAME-ON/GECCO, deCODE and iCOGS was 9.5 x $10^{-7}$ (Fig 8a) This variant was in LD with rs62560775 ($R^2$ = 0.58) , but the association with lung adenocarcinoma was not as strong (rs1011970: P=0.012, rs62560775: P=2.77 x $10^{-4}$). Fig 8b, shows that this variant has the second most significant association with prostate cancer in the region. The strongest association occurs at

rs72652411, a variant which is not in LD with rs1011970 and not associated with any cancer other than prostate.

**Other evidence for pleiotropic associations**

To investigate whether pleiotropic regions for pairs of cancers occurred more often than expected by chance, we used conditional QQ plots to assess enrichment of associations for a given cancer conditioned on p-value category of the other cancer (supplementary Fig 3a-c). As demonstrated by the leftward deflection of the Q-Q plots with decreasing p-value category, there is evidence of pleiotropic association for breast and ovarian cancer (supplementary Fig. 3a), breast and prostate cancer (supplemental Figure 3B), and prostate and colorectal cancer (supplementary Fig. 3c). However, there is no evidence of pleiotropic associations for prostate and ovary cancer, or lung cancer with any of the other 4 cancer sites.

**Discussion**

Using data from the GAME-ON Network and GECCO we conducted a cross-cancer GWAS analysis investigating pleiotropic associations for five cancer sites (lung, breast, colorectal, ovary and prostate) including histology and subtypes. We identified four novel pleiotropic associations that were supported by results in GAME-ON/GECCO data and our independent replication data sets. We identified a pleiotropic association at the 1q22 region involving breast cancer and lung SqCC, neither of which was previously known to be associated with genetic variation in this region. The association with lung SqCC was further supported by the eQTL analysis. We found convincing support for an association between a known lung and breast cancer locus at *BRCA2* and serous ovarian cancer risk. Our data also provide convincing support for an association of a locus at *CASP8/ALS2CR12*, known to be associated with breast cancer and melanoma, with prostate cancer; while genetic variation at the 9p21.3 region, known to be associated with breast cancer and lung SqCC, appears to be associated with lung adenocarcinoma and prostate cancer.

The locus at 1q22 represented by rs1057941 was identified through standard meta-analysis with breast cancer and lung SqCC associated with this variant in GAME-ON/GECCO and in our validation sets. It is worthwhile to mention that the same locus at 1q22 was recently found be associated with blood lipid traits in an parallel analysis in GAME-ON (Zuber et al, submitted), which provides further support for the biological importance of this locus. Rs1057941 lies in a region of LD that includes *KRTCAP2*, *MTX1*, *TRIM46*, *MUC1*, *GBA*, *THBS3, ADAM15 and ASH1L*. While rs1057941 had the strongest association in the meta-analysis, the strongest signals by individual cancer site found in the GAME-ON/GECCO data set were represented by rs3814316 for breast cancer (at *TRIM46*) and rs4072037 for lung SqCC (at *MUC1*). Both of these sequence variants are in LD with rs1057941 (rs3814316: $R^2 = 0.57$; rs4072037, $R^2 = 0.40$) and all variants are within 15kb of one another.

Of these variants, rs4072037 at *MUC1* was previously suggested to be functional as it was shown to regulate alternative splicing of the second exon in *MUC1* and modifies the gene's transcriptional activity[17]. Aberrantly glycosylated MUC1 is overexpressed in most epithelial cancers and is known to

have an oncogenic effect. It mediates the production of growth factors such as connective tissue growth factor (CTGF), and platelet driven growth factor A and B (PDGF-A and PDGF-B) that promote activation of the MAPK and PI3k/Akt pathways potentiating proliferation and survival of tumor cells[18]. It also plays a critical role in EGFR signalling, promoting survival of NSCLC cells[19].

Our lung eQTL investigation found no association with *MUC1* expression but the risk allele, A, of rs4072037 was associated with increased expression of two other genes in the region (*ADAM15* and *THBS3*). This result suggests other potential mechanisms by which this variant could influence cancer risk. *ADAM15* is of particular interest as it is overexpressed in both lung and breast cancer[20-22], which is consistent with our finding of pleiotropic associations of these two cancer sites. Overexpression in breast cancer is associated with Her2/neu expression and evidence from breast cancer cell lines indicates that *ADAM15* catalyzes the cleavage of E-cadherin which in turn binds to and enhances ErbB receptor signalling[22]. rs4072037 may also influence risk of other cancers as an association with gastric cancer was found in a GWAS conducted in China[23]. Although this result did not reach genome wide significance (at P-value of $5 \times 10^{-8}$) a recent meta-analyses provided further support for an association in Asian populations[24,25].

Our results provide strong evidence of a pleiotropic association for rs11571833 at *BRCA2*, that includes lung (previously known[10]) and ovarian cancer, as the subset meta-analysis reached genome-wide significance (for both two and one sided tests). This variant was recently also reported to be associated with upper aerodigestive tract cancer[26]. rs11571833 is an uncommon (minor allele frequency=0.01) and potentially functional variant resulting in an amino acid change (c.9976A>T) responsible for BRCA2 p.Lys3326X. Thr9976 results in the loss of the C-terminal domain of *BRCA2*, a change hypothesized to inhibit the RAD51-BRCA2 interaction in BRCA2 mediated double strand-break repair thereby increasing the risk of cancer[10].

It is possible that the association could be explained by LD with a *BRCA2* mutation. However, previous work from our consortium indicates that co-occurrence of the rare rs11571833 T allele and risk conferring *BRCA2* mutations is unlikely as co-occurrence between this variant and highly penetrant or pathogenic BRCA2 mutations was not observed in several independent samples[10].

For 2q33.1, we found evidence for a pleiotropic effect for rs13016963 (at *ALS2CR12*) on breast and prostate cancer, with association between this variant and prostate cancer reaching genome-wide significance in the combined discover and replicate sets. This region was already known to harbor breast cancer susceptibility loci, in particular rs1045485 encoding the missense alteration D302H in *CASP8*[12,15] (adjacent to *ALS2CR12)*, was previously reported to be associated with breast cancer risk. Subsequently rs1830298 *at ALS2CR12* and rs1045494 at *CASP8* were also reported to be associated with breast cancer risk[14,13].

Of these three breast cancer susceptibility variants, rs1830298 was found to have the most significant association with breast cancer ($P=1.02 \times 10^{-7}$) in our study, and was also associated with prostate cancer risk ($P=5.2X10^{-4}$); whereas rs1045485 and rs1045494 were not. Although there is strong LD between rs1830298 and rs13016963 ($R^2 = 0.74$), rs13016963 and other variants that it is in strong LD with

($R^2 \geq 0.80$), have more significant associations with prostate cancer ($P \leq 3.05 \times 10^{-5}$) than rs1830298, suggesting there might be multiple variants contributing to cancer risk in the region. Interestingly, rs13016963 was also found to be associated with risk of melanoma in a previous GWAS in subjects of European descent[11] indicating this variant may be associated with both prostate cancer and melanoma in Caucasian populations. It was also found to be associated with esophageal squamous cell carcinoma in Han Chinese[23].

Previous research has examined associations between other genetic variants in this region and prostate cancer risk. A possible association between the *CASP8* histidine variant D302H and the more aggressive form of prostate cancer in European populations was reported by two studies[27,28]. This variant was not associated with overall prostate cancer in GAME-ON/GECCO (P=0.14) and is only in very weak LD with rs13016963 ($R^2$=0.11).

Our prostate eQTL analysis suggested that rs13016963 influences the expression of *BZW1.* Previous studies indicate a role for *BZW1* in carcinogenesis. *BZW1* can activate histone H4 gene transcription and serves as a co-regulator of other transcription factors involved in cell cycling. It has been implicated in promoting mucoepidermoid carcinoma tumor growth[29]. We also found two potential functional variants in the region (rs700636 and rs1035142). These variants are in very strong LD with rs13016963 ($R^2 \geq 0.97$), have associations with prostate cancer similar in strength to rs13016963, and are predicted to sit in miRNA binding sites[30].

The 9q21.3 region encoding *CDKN2B-AS1* has been much studied in cancer research. We observed a pleiotropic association of rs62560775 (located in the intronic region of *CDKN2B-AS1*) on lung adenocarcinoma and breast cancer. Timofeeva et al, found an association between rs1333040 at this locus and lung SqCC[31], but this variant was not associated with adenocarcinoma in our data set (P=0.62) The association with breast cancer might to be due to LD ($R^2$=0.38) with a previously identified breast cancer susceptibility variant, rs1011970. This variant is not strongly associated with adenocarcinoma of the lung (P=0.01), which suggests separate loci contribute to breast and lung adenocarcinoma associations in the region. Interestingly, we did replicate an association between rs1011970 and prostate cancer, suggesting this specific variant, or variants in LD with it contribute to risk for both of these cancers. Previous GWAS also report associations between this region and risk of glioma[32,33], melanoma[11,34], and basal cell carcinoma[35] and a recent pleiotropy study indicated an association with esophageal lung SqCC[36]. LD between these variants (4 of 5 of which are associated with only one of these cancers) and rs62560775 and rs1011970 range from $R^2 = 0.21$ to $R^2 = 0.60$, indicating that multiple variants in this region contribute to cancer risk.

Modification of *CDKN2B-AS1* activity could be the mechanism through which this locus influences cancer risk. CDKN2B-AS1, also known as ANRIL (antisense non-coding RNA in the INK4 locus) is known to recruit a polycomb repression complex (PRC2) that silences CDKN2B but not CDKN2A [37-39]. Although there is no known function for rs62660775 or rs1011970, a variant with which rs62660775 is in strong LD, rs3217986 (at $R^2$=0.69), was identified to be located in a miRNA binding site[30] and classified as likely to affect binding by Regulome[40].

Our Q-Q plots for pleiotropy (supplementary Fig. 3a-c) suggest pleiotropic associations are present between some sites: breast and ovarian cancer, breast and prostate cancer and prostate and lung cancer, with some evidence provided for pleiotropic associations involving prostate and lung and colorectal and ovarian cancer. There was little evidence for pleiotropic associations involving other site combinations. However, the plots most effectively reflect strong associations at regions with extensive LD. More subtle effects could be missed.

Our initial investigation using the GAME-ON and GECCO data set identified 33 regions and 190 variants that we further examined in replication data sets. We were able to replicate the associations for four of these regions. Since our replication datasets sample sizes were often smaller than that of our discovery set (depending on the cancer site), we may have insufficient power to replicate true associations particularly for less common variants, underlining the importance of sample size for investigations of pleiotropy.

In summary, using data from the GAME-ON initiative and GECCO, we have found four regions that show associations with multiple cancers, including a novel association between genetic variation at 1q22 and breast cancer and lung squamous cell carcinoma. This is the largest study to date examining pleiotropy across multiple cancer sites. There are likely additional loci that are associated with multiple cancers but these will require additional efforts with larger data series for detection.

## Methods

### Data and contributing consortia

This study used summary level data to perform cross-cancer GWAS analysis of lung, colorectal, prostate, breast and ovarian cancers based on a subset-based meta-analytical approach. Forty-six studies from North America and Europe organized into cancer site specific consortia, within the GAME-ON Network (http://epi.grants.cancer.gov/gameon/) or GECCO, participated in this investigation. Table 1 provides details for contributing consortia and studies. In addition to the five main cancer sites the analyses also included the following cancer subtypes: adenocarcinoma and squamous cell carcinoma of the lung; the aggressive form of prostate cancer; estrogen receptor negative breast cancer, and serous and endometrioid cancers of the ovary (Table 1). All studies frequency matched cases and controls on at least age and sex, and all subjects were of European descent.

### Genotyping and imputation

Genotyping was performed on Affymetrix or Illumina platforms (Table 1). Marker exclusion criteria were applied in each cancer consortium using standard exclusion criteria[8,9,31,41-44]. Genotype imputation was conducted for each cancer site using IMPUTE, BEAGLE, MACH and Minimac, with a threshold for imputation of $R^2 > 0.3$ used throughout.

### Statistical Analysis

Logistic regression analysis using a log additive model was carried out to test the association of variants with cancer risk for all of the forty-five studies. All effect estimates represent per allele adjusted odds

ratios (ORs) adjusted for age, principal components and gender where applicable. The study-specific results were then combined for each cancer site using a fixed effects model. The methodology and the results of the cancer-specific studies have been described previously[8,9,31,41-44].

A subset-based meta-analysis approach developed by Bhattacharjee et al. (ASSET) was then used to investigate pleiotropic effects across cancer sites[7]. The method generalizes the standard fixed effect meta-analysis by examining the association between genetic variants with subsets of cancers and allowing opposing direction of effects and null associations. Associations are summarized with an overall two-sided p-value with a penalty for the subset searches to adjust for multiple comparisons.

Accounting for subsets of studies with no effects and/or effects in opposing directions (i.e., both increased and reduced risk subsets can be defined) is an advantage of the subset-based meta-analysis approach. However, in the situation where a large majority of underlying effects are in one direction subset meta-analysis can have lower power compared to standard fixed effect analysis. For this reason, we also explored results based on standard fixed effects meta-analysis (when all cancer sites provided results), again using ASSET.

Analysis was performed when at least three of five cancer sites had available data. Subjects appearing in several studies with different cancer sub-types (e.g. overlapping controls for lung adenocarcinoma and lung SqCC) and across cancer types (e.g. UK ovary and UK breast GWAS both used controls from Welcome Trust Case Control Consortium, WTCCC) were accounted for in the covariance matrix when estimating standard errors for subset-based and standard meta-analyses.

We set a significance threshold of $5 \times 10^{-7}$ to identify variants of interest (i.e., variants with evidence of a pleiotropic effect) for the two sided subset-analysis test, for positive and negative associations that contributed to the two-sided subset-analysis test signal, and for fixed effect meta-analysis. We excluded variants where the association was obviously driven by a single cancer site. We also identified variants of interest as those where p-values for association between variants and individual cancer sites, including their subtypes, were less than $5 \times 10^{-3}$ for at least two cancers. Among the variants of interest identified, we prioritized for validation those that showed the strongest pleiotropic association in a region (based on subset or standard cross cancer meta-analysis) and also included variants that showed the statistically most significant associations in site-specific analyses in a region (with $P \leq 5 \times 10^{-3}$). We then sought to validate specific variant and cancer site associations that contributed to the pleiotropic signal in the region using independent sets of study populations of European descent based on all five cancers from deCODE [10,45], lung cancer from Harvard[46,47], breast (region 1q22 only) from iCOGS, ovarian from OCAC/iCOGS and prostate from PRACTICAL/iCOGS[9,42,48], with a replication threshold of $P \leq 0.05$. We further assessed significance for these variants by number of effective tests, which accounts for the correlation among variants of interest. This resulted in an adjusted significance threshold of P=0.0003[49] after accounting for testing in multiple phenotypes. For cross-ethnicity generalizability, we also examined results for our selected variants in different race/ethnicities using data from Japan (lung[50]), Nanjing (lung[51]), Shanghai (breast[52]), MEC, African American Breast Cancer GWAS Consortium (AABC), African Ancestry Prostate Cancer GWAS Consortium (AAPC), and San Francisco (breast for Latinas)[53-56].

**Further investigation of pleiotropic effects**

We further investigated pleiotropy between pairs of cancer sites (e.g., breast and lung, colorectal and prostate) using conditional Q-Q plots to examine enrichment of association signals (over-abundance of low p-values) in one cancer when conditioning on significance of p-values in the second cancer. Enrichment is reflected in a leftward deflection in the Q-Q plot with decreasing p-value categories of the second (conditioning) cancer[57] indicating a higher degree of pleiotropy between two cancer sites than is expected by chance.

**eQTL data**

We obtained non-tumor lung eQTL data of 1,111 patients from three studies assembled by Laval University (n=409), The University of British Columbia (n=339) and the University of Groningen (n=363). Gene expression profiles were obtained using an Affymetrix array (see GEO platform GPLL0379). Genotyping was carried out using using the Illumina Human1M-Duo genotyping BeadChip. Analyses were adjusted for age, sex, and smoking status. Further details of this study are published elsewhere[58-62]. We included validated variants from our study which showed evidence of association with lung cancer and also evaluated eQTL data for variants in LD ($R^2 > 0.7$) with these. A statistically significant result for a specific variant was declared if 2 of 3 studies showed a significant p-value after Bonferroni correction for multiple comparisons. We also obtained TCGA eQTL data for 402 high-grade serous ovarian cases and 145 prostate tumor samples and 33 normal tissue samples and again investigated variants with validated associations and those in high LD with them. Gene expression values for high-grade serous ovarian cases were assessed by p-value. Gene expression values for prostate cancer were adjusted for somatic copy number and CpG methylation as previously described [63]. Significant associations were defined as those having both p-value and false discovery rate (based on Benjamini-Hochberg method) of less than 0.05.

**Functional data**

We used FuncPred from SNPinfo to assist with variant function prediction. The software determines a variant's potential function in splicing regulation, TFBS prediction, miRNA binding site prediction and regulatory potential based on in-house algorithms and tools developed elsewhere (e.g., polyphen)[30]. In addition, we used RegulomeDB to further assess regulatory potential for variants of interest. This tool includes high-throughput, experimental data sets from ENCODE and other sources, as well as computational predictions and manual annotations to identify putative regulatory potential and identify functional variants[40]. We also further examined ENCODE data by examining region tracks using UCSC genome browser[64].

**Funding sources**

TRICL (Transdisciplinary Research for Cancer of Lung): National Institute of Health U19 CA148127-01 (PI: Amos), Canadian Cancer Society Research Institute (no. 020214, PI: Hung). Harvard Lung Study: The Harvard Lung Cancer Study is supported by the US National Institutes of Health (R01 CA092824, P50 CA090578, and R01 CA074386).

DRIVE (Discovery, Biology, and Risk of Inherited Variants in Breast Cancer): National Institute of Health U19 CA148065. WHI: The Women's Health Initiative contribution to DRIVE is supported by the U.S. National Heart, Lung and Blood Institute via contract HHSN26820110046C and other contracts. BBCS: This work was funded by Cancer Research UK (C150/A5660 and C1178/A3947); Breakthrough Breast Cancer and the Institut National de Cancer. We acknowledge National Health Service funding to the NIHR Biomedical Research Centre and the National Cancer Research Network (NCRN). Funding for the project was provided by the Wellcome Trust under award 076113 and 085475. Rotterdam Study: The GWAS datasets are supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project nr. 050-060-810. The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists.

CORECT (ColoRectal Transdisciplinary Study): National Institute of Health U19 CA148107; R01 CA81488, P30 CA014089.

ELLIPSE (ELLIPSE, Elucidating Loci in Prostate Cancer Susceptibility): This work was support by the GAME-ON U19 initiative for prostate cancer (ELLIPSE), U19 CA148537.

FOCI (Transdisciplinary Cancer Genetic Association and Interacting Studies): National Institutes of Health U19 CA148112-01 (PI: Sellers), R01-CA122443, P50-CA136393, P30-CA15083 (PI: Goode), Cancer Research UK (C490/A8339, C490/A16561, C490/A10119, C490/A10124 (PI: Pharoah)).

GECCO (Genetics and Epidemiology of Colorectal Cancer Consortium): National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services (U01 CA137088; R01 CA059045). ASTERISK: a Hospital Clinical Research Program (PHRC) and supported by the Regional Council of Pays de la Loire, the Groupement des Entreprises Françaises dans la Lutte contre le Cancer (GEFLUC), the Association Anne de Bretagne Génétique and the Ligue Régionale Contre le Cancer (LRCC). DACHS: German Research Council (Deutsche Forschungsgemeinschaft, BR 1704/6-1, BR 1704/6-3, BR 1704/6-4 and CH 117/1-1), and the German Federal Ministry of Education and Research (01KH0404 and 01ER0814). DALS: National Institutes of Health (R01 CA48998 to M.L. Slattery); HPFS is supported by the National Institutes of Health (P01 CA 055075, UM1 CA167552, R01 137178, R01 CA 151993 and P50 CA

127003), NHS by the National Institutes of Health (R01 CA137178, P01 CA 087969, R01 CA151993 and P50 CA 127003,) and PHS by the National Institutes of Health (R01 CA042182). OFCCR: National Institutes of Health, through funding allocated to the Ontario Registry for Studies of Familial Colorectal Cancer (U01 CA074783);. Additional funding toward genetic analyses of OFCCR includes the Ontario Research Fund, the Canadian Institutes of Health Research, and the Ontario Institute for Cancer Research, through generous support from the Ontario Ministry of Research and Innovation. PLCO: Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS. Additionally, a subset of control samples were genotyped as part of the Cancer Genetic Markers of Susceptibility (CGEMS) Prostate Cancer GWAS (Yeager, M et al. Nat Genet 2007 May;39(5):645-9), Colon CGEMS pancreatic cancer scan (PanScan) (Amundadottir, L et al. Nat Genet. 2009 Sep;41(9):986-90 and Petersen, GM et al Nat Genet. 2010 Mar;42(3):224-8), and the Lung Cancer and Smoking study (Landi, MT et al. Am J Hum Genet. 2009 Nov;85(5):679-91). The prostate and PanScan study datasets were accessed with appropriate approval through the dbGaP online resource (http://cgems.cancer.gov/data/) accession numbers phs000207.v1.p1 and phs000206.v3.p2, respectively, and the lung datasets were accessed from the dbGaP website (http://www.ncbi.nlm.nih.gov/gap) through accession number phs000093.v2.p2. Funding for the Lung Cancer and Smoking study was provided by National Institutes of Health (NIH), Genes, Environment and Health Initiative (GEI) Z01 CP 010200, NIH U01 HG004446, and NIH GEI U01 HG 004438. For the lung study, the GENEVA Coordinating Center provided assistance with genotype cleaning and general study coordination, and the Johns Hopkins University Center for Inherited Disease Research conducted genotyping. PMH: National Institutes of Health (R01 CA076366 to P.A. Newcomb). VITAL: National Institutes of Health (K05 CA154337). WHI: The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C.

BCAC is funded by Cancer Research UK [C1287/A10118, C1287/A12014] and by the European Community´s Seventh Framework Programme under grant agreement number 223175 (grant number HEALTH-F2-2009-223175) (COGS).

Lung eQTL study: The lung eQTL study at Laval University was supported by the Chaire de pneumologie de la Fondation JD Bégin de l'Université Laval, the Fondation de l'Institut universitaire de cardiologie et de pneumologie de Québec, the Respiratory Health Network of the FRQS, the Canadian Institutes of Health Research (MOP - 123369), and the Cancer Research Society and Read for the Cure. Y. Bossé is the recipient of a Junior 2 Research Scholar award from the Fonds de recherche Québec – Santé (FRQS).

Prostate eQTL study: supported by National Institute of General Medical Science grant R01GM107427

**Acknowledgement**

This study was supported by the grant from the National Institute of Health (NIH) (U19CA148127). Additional funding was obtained from NIH grants (5R01CA055769, 5R01CA127219, 5R01CA133996, and 5R01CA121197).

GECCO: The authors would like to thank all those at the GECCO Coordinating Center for helping bring together the data and people that made this project possible. The authors acknowledge Dave Duggan and team members at TGEN (Translational Genomics Research Institute), the Broad Institute, and the Génome Québec Innovation Center for genotyping DNA samples of cases and controls, and for scientific input for GECCO. ASTERISK: We are very grateful to Dr. Bruno Buecher without whom this project would not have existed. We also thank all those who agreed to participate in this study, including the patients and the healthy control persons, as well as all the physicians, technicians and students. DACHS: We thank all participants and cooperating clinicians, and Ute Handte-Daub, Renate Hettler-Jensen, Utz Benscheid, Muhabbet Celik and Ursula Eilber for excellent technical assistance. HPFS, NHS and PHS: We would like to acknowledge Patrice Soule and Hardeep Ranu of the Dana Farber Harvard Cancer Center High-Throughput Polymorphism Core who assisted in the genotyping for NHS, HPFS, and PHS under the supervision of Dr. Immaculata Devivo and Dr. David Hunter, Qin (Carolyn) Guo and Lixue Zhu who assisted in programming for NHS and HPFS, and Haiyan Zhang who assisted in programming for the PHS. We would like to thank the participants and staff of the Nurses' Health Study and the Health Professionals Follow-Up Study, for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data. PLCO: The authors thank Drs. Christine Berg and Philip Prorok, Division of Cancer Prevention, National Cancer Institute, the Screening Center investigators and staff or the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial, Mr. Tom Riley and staff, Information Management Services, Inc., Ms. Barbara O'Brien and staff, Westat, Inc., and Drs. Bill Kopp, Wen Shao, and staff, SAIC-Frederick. Most importantly, we acknowledge the study participants for their contributions to making this study possible. The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by NCI. PMH: The authors would like to thank the study participants and staff of the Hormones and Colon Cancer study. WHI: The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf

BCAC: We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out.

Rotterdam Study: The generation and management of GWAS genotype data for the Rotterdam Study (RS I, RS II, RS III) was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera, Marjolein Peters, MSc, Carolina Medina-Gomez, MSc, and Fernando Rivadeneira, PhD for their help in creating the GWAS database, and Karol Estrada, PhD, Yurii Aulchenko, PhD, and Carolina Medina-Gomez, MSc, for the creation and analysis of imputed data.

**Consortia**

**African American Breast Cancer Consortium (AABC)**
Christine B. Ambrosone, Esther M. John, Leslie Bernstein, Wei Zheng, Andrew F. Olshan, Jennifer J. Hu, Regina G. Ziegler, Sarah Nyante, Elisa V. Bandera, Sue A. Ingles, Michael F. Press, Sandra L. Deming, Jorge L. Rodriguez-Gil, Stephen J. Chanock.

**African Ancestry Prostate Cancer Consortium (AAPC)**
Sara S. Strom, Rick A. Kittles, Benjamin A. Rybicki, Janet L. Stanford, Phyllis J. Goodman, Sonja I. Berndt, John Carpten, Graham Casey[9], Lisa Chu, Ryan W. Diver, Anselm JM Hennis, Eric A. Klein, Suzanne Kolb, Loic Le Marchand, M. Cristina Leske, Adam B. Murphy, Christine Neslund-Dudas, Curtis Pettaway, Susan M. Gapstur, S. Lilly Zheng, Suh-Yuh Wu, John S. Witte, Jianfeng Xu, William Isaacs, Barbara Nemesure, William J. Blot.

**Colorectal Transdisciplinary (CORECT) Study**
Stephen Gruber, Fred Schumacher.

**Hereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON)**
Rookus M, Hogervorst F, van Leeuwen F, Verhoef S, Schmidt M, de Lange J, Collee J, van den Ouweland A, Hooning M, Seynaeve C, van Deurzen C, Obdeijn I, van Asperen C, Wijnen J, Tollenaar R, Devilee P, van Cronenburg TC, Kets C, Mensenkamp A, Ausems M, van der Luijt R, Aalfs C, van Os T, Gille J, Waisfisz Q, Meijers-Heijboer H, Gomez-Garcia E, Blok M, Oosterwijk J, van der Hout A, Mourits M, de Bock G, Vasen H.

**Ovarian Cancer Association Consortium (OCAC)**
Hoda Anton-Culver, Elisa Bandera, Susana Banerjee, Ros Glasspool, Iain McNeish, Jim Paul, Nadeem Siddiqui, Natalia Bogdanova, Thilo Dörk-Bousset, Ralf Butzow, Heli Nevanlinna, Jenny Chang-Claude, Georgia Chenevix-Trench, Penelope Webb, Linda Cook, Nhu Le, Daniel Cramer, Kathryn Terry, Thilo Dörk-Bousset, Matthias Duerst, Peter Fasching, Simon Gayther, Usha Menon, Graham Giles, Ellen Goode, Marc Goodman, Jacek Gronwald, Michelle Hildebrandt, Karen Lu, Estrid Høgdall, Claus Høgdall, Beth Karlan, Lambertus Kiemeney, Leon Massuger, Susanne Kruger Kjaer, Jolanta Kupryjanczyk, Diether Lambrechts, Douglas Levine, Keitaro Matsuo, Francesmary Modugno, Kirsten Moysich, Roberta Ness, Kirsten Moysich, Steven Narod, Catherine Phelan, Harvey Risch, Leigh Pearce, Anna Wu, Tanja Pejovic, Paul Pharoah, Mary Anne Rossing, Dale P. Sandler, Helga B Salvesen, Joellen Schildkraut, Weiva Sieh, Alice Whittemore, Meir Stampfer, Shelley Tworoger, Walter Willett, Soo-Hwang Teo, Yin Ling Woo, Nicolas Wentzensen, Wei Zheng.

**The PRACTICAL Consortium (**http://practical.ccge.medschl.cam.ac.uk/**)**
Rosalind Eeles, Doug Easton, Zsofia Kote-Jarai, Ali Amin Al Olama, Sara Benlloch, Kenneth Muir, Graham G. Giles, Fredrik Wiklund, Henrik Gronberg, Christopher A. Haiman, Johanna Schleutker, Maren Weischer, Ruth C. Travis, David Neal, Paul Pharoah, Kay-Tee Khaw, Janet L. Stanford, William J. Blot, Stephen Thibodeau, Christiane Maier, Adam S. Kibel, Cezary Cybulski, Lisa Cannon-Albright, Hermann Brenner, Jong Park, Radka Kaneva, Jyotsna Batra, Manuel R. Teixeira, Hardev Pandha.

**San Francisco Study and Latino American Breast Cancer Consortium (LABC)**
Elad Ziv, Esther M. John, Laura Fejerman.

**Japanese American Prostate Cancer (JAPC) Consortium and Latino Prostate Cancer Consortium (LAPC)**
Brian E. Henderson, Christopher A. Haiman, Loic Le Marchand.

**References**


1 **Hindorff, L. *et al. A Catalog of Published Genome-Wide Association Studies*, <www.genome.gov/gwastudies> (2014).**

2 **Hindorff, L. A., Gillanders, E. M. & Manolio, T. A. Genetic architecture of cancer and other complex diseases: lessons learned and future directions. *Carcinogenesis* 32, 945-954, doi:10.1093/carcin/bgr056 (2011).**

3 **Lange, E. M. *et al.* Genome-wide association scan for variants associated with early-onset prostate cancer. *PloS one* 9, e93436, doi:10.1371/journal.pone.0093436 (2014).**

4 **Peters, U. *et al.* Meta-analysis of new genome-wide association studies of colorectal cancer risk. *Human genetics* 131, 217-234, doi:10.1007/s00439-011-1055-0 (2012).**

5 **Petersen, G. M. *et al.* A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. *Nature genetics* 42, 224-228, doi:10.1038/ng.522 (2010).**

6 **Rafnar, T. *et al.* Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nature genetics* 41, 221-227, doi:10.1038/ng.296 (2009).**

7 **Bhattacharjee, S. *et al.* A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of heterogeneous traits. *American journal of human genetics* 90, 821-835, doi:10.1016/j.ajhg.2012.03.015 (2012).**

8 **Peters, U. *et al.* Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. *Gastroenterology* 144, 799-807 e724, doi:10.1053/j.gastro.2012.12.020 (2013).**

9 **Michailidou, K. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nature genetics* 45, 353-361, 361e351-352, doi:10.1038/ng.2563 (2013).**

10 **Wang, Y. *et al.* Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. *Nature genetics* 46, 736-741, doi:10.1038/ng.3002 (2014).**

11 **Barrett, J. H. *et al.* Genome-wide association study identifies three new melanoma susceptibility loci. *Nature genetics* 43, 1108-1113, doi:10.1038/ng.959 (2011).**

12 **Cox, A. *et al.* A common coding variant in CASP8 is associated with breast cancer risk. *Nature genetics* 39, 352-358, doi:10.1038/ng1981 (2007).**

13 **Khan, S. *et al.* MicroRNA related polymorphisms and breast cancer risk. *PloS one* 9, e109973, doi:10.1371/journal.pone.0109973 (2014).**

14 **Lin, W. Y. *et al.* Identification and characterization of novel associations in the CASP8/ALS2CR12 region on chromosome 2 with breast cancer risk. *Human molecular genetics*, doi:10.1093/hmg/ddu431 (2014).**

15 **MacPherson, G. *et al.* Association of a common variant of the CASP8 gene with reduced risk of breast cancer. *Journal of the National Cancer Institute* 96, 1866-1869, doi:10.1093/jnci/dji001 (2004).**

16 **Turnbull, C. *et al.* Genome-wide association study identifies five new breast cancer susceptibility loci. *Nature genetics* 42, 504-507, doi:10.1038/ng.586 (2010).**

17 **Saeki, N. *et al.* A functional single nucleotide polymorphism in mucin 1, at chromosome 1q22, determines susceptibility to diffuse-type gastric cancer. *Gastroenterology* 140, 892-902, doi:10.1053/j.gastro.2010.10.058 (2011).**

18 Nath, S. & Mukherjee, P. MUC1: a multifaceted oncoprotein with a key role in cancer progression. *Trends in molecular medicine* 20, 332-342, doi:10.1016/j.molmed.2014.02.007 (2014).
19 Kharbanda, A. *et al.* Targeting the Oncogenic MUC1-C Protein Inhibits Mutant EGFR-Mediated Signaling and Survival in Non–Small Cell Lung Cancer Cells. *Clinical Cancer Research*, doi:10.1158/1078-0432.ccr-13-3168 (2014).
20 Kuefer, R. *et al.* ADAM15 disintegrin is associated with aggressive prostate and breast cancer disease. *Neoplasia* 8, 319-329, doi:10.1593/neo.05682 (2006).
21 Schutz, A. *et al.* Expression of ADAM15 in lung carcinomas. *Virchows Archiv : an international journal of pathology* 446, 421-429, doi:10.1007/s00428-004-1193-z (2005).
22 Najy, A. J., Day, K. C. & Day, M. L. The ectodomain shedding of E-cadherin by ADAM15 supports ErbB receptor activation. *The Journal of biological chemistry* 283, 18393-18401, doi:10.1074/jbc.M801329200 (2008).
23 Abnet, C. C. *et al.* Genotypic variants at 2q33 and risk of esophageal squamous cell carcinoma in China: a meta-analysis of genome-wide association studies. *Human molecular genetics* 21, 2132-2141, doi:10.1093/hmg/dds029 (2012).
24 Liu, X. *et al.* MUC1 gene polymorphism rs4072037 and susceptibility to gastric cancer: a meta-analysis. *SpringerPlus* 3, 599, doi:10.1186/2193-1801-3-599 (2014).
25 Zheng, L. *et al.* Functional polymorphism rs4072037 in MUC1 gene contributes to the susceptibility to gastric cancer: evidence from pooled 6,580 cases and 10,324 controls. *Molecular biology reports* 40, 5791-5796, doi:10.1007/s11033-013-2682-4 (2013).
26 Delahaye-Sourdeix, M. *et al.* A rare truncating BRCA2 variant and genetic susceptibility to upper aerodigestive tract cancer. *Journal of the National Cancer Institute* 107, doi:10.1093/jnci/djv037 (2015).
27 Lubahn, J. *et al.* Association of CASP8 D302H polymorphism with reduced risk of aggressive prostate carcinoma. *The Prostate* 70, 646-653, doi:10.1002/pros.21098 (2010).
28 Meyer, A. *et al.* Apoptosis gene polymorphisms and risk of prostate cancer: a hospital-based study of German patients treated with brachytherapy. *Urologic oncology* 31, 74-81, doi:10.1016/j.urolonc.2010.09.011 (2013).
29 Li, S. *et al.* BZW1, a novel proliferation regulator that promotes growth of salivary muocepodermoid carcinoma. *Cancer letters* 284, 86-94, doi:10.1016/j.canlet.2009.04.019 (2009).
30 Xu, Z. & Taylor, J. A. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic acids research* 37, W600-605, doi:10.1093/nar/gkp290 (2009).
31 Timofeeva, M. N. *et al.* Influence of common genetic variation on lung cancer risk: meta-analysis of 14 900 cases and 29 485 controls. *Human molecular genetics* 21, 4980-4995, doi:10.1093/hmg/dds334 (2012).
32 Wrensch, M. *et al.* Variants in the CDKN2B and RTEL1 regions are associated with high-grade glioma susceptibility. *Nature genetics* 41, 905-908, doi:10.1038/ng.408 (2009).
33 Shete, S. *et al.* Genome-wide association study identifies five susceptibility loci for glioma. *Nature genetics* 41, 899-904, doi:10.1038/ng.407 (2009).
34 Bishop, D. T. *et al.* Genome-wide association study identifies three loci associated with melanoma risk. *Nature genetics* 41, 920-925, doi:10.1038/ng.411 (2009).
35 Stacey, S. N. *et al.* New common variants affecting susceptibility to basal cell carcinoma. *Nature genetics* 41, 909-914, doi:10.1038/ng.412 (2009).
36 Li, W. Q. *et al.* Genetic polymorphisms in the 9p21 region associated with risk of multiple cancers. *Carcinogenesis*, doi:10.1093/carcin/bgu203 (2014).

**37** **Jarinova, O. *et al.* Functional analysis of the chromosome 9p21.3 coronary artery disease risk locus. *Arteriosclerosis, thrombosis, and vascular biology* 29, 1671-1677, doi:10.1161/ATVBAHA.109.189522 (2009).**

**38** **Kotake, Y. *et al.* Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene* 30, 1956-1962, doi:10.1038/onc.2010.568 (2011).**

**39** **Yu, W. *et al.* Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* 451, 202-206, doi:10.1038/nature06468 (2008).**

**40** **Boyle, A. P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome research* 22, 1790-1797, doi:10.1101/gr.137323.112 (2012).**

**41** **Garcia-Closas, M. *et al.* Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nature genetics* 45, 392-398, 398e391-392, doi:10.1038/ng.2561 (2013).**

**42** **Pharoah, P. D. *et al.* GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. *Nature genetics* 45, 362-370, 370e361-362, doi:10.1038/ng.2564 (2013).**

**43** **Siddiq, A. *et al.* A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. *Human molecular genetics* 21, 5373-5384, doi:10.1093/hmg/dds381 (2012).**

**44** **Wang, H. *et al.* Trans-ethnic genome-wide association study of colorectal cancer identifies a new susceptibility locus in VTI1A. *Nature communications* 5, 4613, doi:10.1038/ncomms5613 (2014).**

**45** **Gudbjartsson, D. F. *et al.* Large-scale whole-genome sequencing of the Icelandic population. *Nature genetics* 47, 435-444, doi:10.1038/ng.3247 (2015).**

**46** **Asomaning, K. *et al.* Second hand smoke, age of exposure and lung cancer risk. *Lung Cancer* 61, 13-20, doi:10.1016/j.lungcan.2007.11.013 (2008).**

**47** **Liu, G. *et al.* Genetic polymorphisms of MDM2, cumulative cigarette smoking and nonsmall cell lung cancer risk. *International journal of cancer. Journal international du cancer* 122, 915-918, doi:10.1002/ijc.23178 (2008).**

**48** **Al Olama, A. A. *et al.* A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. *Nature genetics* 46, 1103-1109, doi:10.1038/ng.3094 (2014).**

**49** **Moskvina, V. & Schmidt, K. M. On multiple-testing correction in genome-wide association studies. *Genetic epidemiology* 32, 567-573, doi:10.1002/gepi.20331 (2008).**

**50** **Shiraishi, K. *et al.* A genome-wide association study identifies two new susceptibility loci for lung adenocarcinoma in the Japanese population. *Nature genetics* 44, 900-903, doi:10.1038/ng.2353 (2012).**

**51** **Hu, Z. *et al.* A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. *Nature genetics* 43, 792-796, doi:10.1038/ng.875 (2011).**

**52** **Cai, Q. *et al.* Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1. *Nature genetics* 46, 886-890, doi:10.1038/ng.3041 (2014).**

**53** **Cheng, I. *et al.* Evaluating genetic risk for prostate cancer among Japanese and Latinos. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 21, 2048-2058, doi:10.1158/1055-9965.EPI-12-0598 (2012).**

**54** **Fejerman, L. *et al.* Genome-wide association study of breast cancer in Latinas identifies novel protective variants on 6q25. *Nature communications* 5, 5260, doi:10.1038/ncomms6260 (2014).**

**55 Chen, F. *et al.* A genome-wide association study of breast cancer in women of African ancestry. *Human genetics* 132, 39-48, doi:10.1007/s00439-012-1214-y (2013).**

**56 Haiman, C. A. *et al.* Genome-wide association study of prostate cancer in men of African ancestry identifies a susceptibility locus at 17q21. *Nature genetics* 43, 570-573, doi:10.1038/ng.839 (2011).**

**57 Andreassen, O. A. *et al.* Shared common variants in prostate cancer and blood lipids. *International journal of epidemiology* 43, 1205-1214, doi:10.1093/ije/dyu090 (2014).**

**58 Hao, K. *et al.* Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS genetics* 8, e1003029, doi:10.1371/journal.pgen.1003029 (2012).**

**59 Lamontagne, M. *et al.* Refining susceptibility loci of chronic obstructive pulmonary disease with lung eqtls. *PloS one* 8, e70220, doi:10.1371/journal.pone.0070220 (2013).**

**60 Obeidat, M. *et al.* GSTCD and INTS12 regulation and expression in the human lung. *PloS one* 8, e74630, doi:10.1371/journal.pone.0074630 (2013).**

**61 Thun, G. A. *et al.* Causal and synthetic associations of variants in the SERPINA gene cluster with alpha1-antitrypsin serum levels. *PLoS genetics* 9, e1003585, doi:10.1371/journal.pgen.1003585 (2013).**

**62 Wain, L. V. *et al.* Whole exome re-sequencing implicates CCDC38 and cilia structure and function in resistance to smoking related airflow obstruction. *PLoS genetics* 10, e1004314, doi:10.1371/journal.pgen.1004314 (2014).**

**63 Li, Q. *et al.* Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. *Cell* 152, 633-641, doi:10.1016/j.cell.2012.12.034 (2013).**

**64 Kent, W. J. *et al.* The human genome browser at UCSC. *Genome research* 12, 996-1006, doi:10.1101/gr.229102. Article published online before print in May 2002 (2002).**

## Figure Legends

**Figure 1.** Manhattan plots ($-\log_{10}(p)$) by chromosome for individual cancer sites (innermost to outermost ring – ovary(Ov) breast (Br), prostate (PR), colorectal/GECCO (Co), Lung (Lu)).

**Figure 2. Results for rs1057941: a)** Forest plot for rs1057941 showing per allele ORs for risk allele A (of A/G). Standard fixed effects meta-analysis (also indicated by dashed line), and subset meta-analysis results (two-sided, one-sided and positive and negative subset associations) are shown. **b)** Regional plot showing p-values from overall meta-analysis at region 1q22 using GAME-ON/GECCO discovery set data. The top breast cancer hit (rs3814316) and top squamous cell carcinoma hit (rs4072037) are also highlighted.

**Figure 3. Boxplots of gene expression levels in normal lung tissue for rs4072037** for A) *ADAM15* and B) *THBS3*. Results are presented by study (Laval, UBC, Groningen).

**Fig 4. Results for rs11571833 a**) Forest plot showing per allele ORs for risk allele T of (T/A). Standard fixed effects meta-analysis (also indicated by dashed line), and subset meta-analysis results (two-sided, one-sided and positive and negative subset associations) are shown. **b**) Regional plot showing two sided p-values from ASSET subset meta-analysis at region 13q13.1 using GAME-ON/GECCO discovery set data. Peak is at *BRCA2*. rs11571815 and rs11571818 have nearly identical association signals but are partially obscured by rs11571833.

**Fig 5. Results for rs13016963**: **a**) Forest plot showing per allele ORs for risk allele G (of A/G). Standard fixed effects meta-analysis (also indicated by dashed line), and subset meta-analysis results (two-sided, one-sided and positive

and negative subset associations) are shown. **b**) Regional plot showing P-values for GAME-ON prostate cancer GWAS at region 2q33.1 using GAME-ON/GECCO discovery set data. Peak is at *ALS2CR12*.

**Figure 6. Boxplot of *BZW1* gene expression levels in prostate tumor and normal tissue for rs1035142/rs13016963**. rs1035142 is presented as a surrogate for rs13016963, with which it shows perfect LD in eQTL data.

**Figure 7. Results for rs62560775**: **a**) Forest plot showing per allele ORs for risk allele G (of A/G). Standard fixed effects meta-analysis (also indicated by dashed line), and subset meta-analysis results (two-sided, one-sided and positive and negative subset associations) are shown. **b**) Regional plot showing P-values for GAME-ON lung adenocarcinoma GWAS at region 9p21.3, using GAME-ON/GECCO discovery set data. Peak is at *CDKN2B-AS1*.

**Figure 8. Results for rs1011970**: **a**) Forest plot showing per allele ORs for risk allele T (of T/G). Standard fixed effects meta-analysis (also indicated by dashed line), and subset meta-analysis results (two-sided, one-sided and positive and negative subset associations) are shown. **b**) Regional plot showing P-values for GAME-ON prostate cancer GWAS at region 9p21.3, using GAME-ON/GECCO discovery set data. Peak is at *CDKN2B-AS1*.

Table 1. Contributing consortia and characteristics of data sets.

| GAME-ON/GECCO Analysis | | | | | | | Replication Stage (European ancestry) | | | Generalizability (Other ancestry) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Cancer Site (Consortium)** | **No. studies** | **Cases** | **Controls*** | **Variants†** | **Genotyping Platform** | **Imputation threshold‡** | **Study** | **Cases** | **Controls*** | **Study** | **Cases** | **Controls*** |
| Lung (TRICL)[31] | 6 | 12160 | 16838 | 8492272 | | | deCODE[10,45] | 3865 | 196658 | Nanjing[51] | 2331 | 3077 |
| | | | | | | | Harvard[46,47] | 984 | 970 | | | |
| Adenocarcinoma | | 3718 | 15871 | 8472145 | Illumina | $R^2$>0.3 | deCODE | 1434 | 198663 | Nanjing | 1304 | 3077 |
| | | | | | | | Harvard | 597 | 970 | Japan[50] | 1575 | 3363 |
| Squamous cell carcinoma | | 3422 | 16015 | 8478230 | | | deCODE | 784 | 171059 | Nanjing | 822 | 3077 |
| | | | | | | | Harvard | 216 | 970 | | | |
| Colorectal (CORECT)[44] | 6 | 5100 | 4831 | 7229595 | Affymetrix Axiom | None | deCODE | 3546 | 236404 | | | |
| Colorectal (GECCO)[8] | 13 | 10314 | 12857 | 9193926 | Illumina, Affymetrix | $R^2$>0.3 | | | | | | |
| Prostate (ELLIPSE) | 6 | 14160 | 12724 | 9084781 | Illumina, Affymetrix | $R^2$>0.3 | deCODE | 4858 | 83103 | LAPC/MEC[53] | 1034 | 1046 |
| | | | | | | | iCOGS[48] | 20219 | 20440 | AAPC[56] | 4853 | 4678 |
| | | | | | | | | | | JAPC/MEC[53] | 980 | 1005 |
| Prostate Aggressive | 6 | 4450 | 12724 | 9003304 | | | | | | | | |
| | | | | | | | | | | Shanghai[52] | 2867 | 2285 |
| Breast (DRIVE)[9,43] | 11 | 15748 | 18084 | 9331393 | Illumina, Affymetrix | $R^2$>0.3 | deCODE | 5318 | 280808 | LABC/MEC, SF[54] | 1497 | 3213 |
| | | | | | | | iCOGS[9] | 46785 | 42892 | | | |
| | | | | | | | | | | AABC[55] | 3015 | 2743 |
| Breast ER[41] | 8 | 4939 | 13128 | 9250406 | | | | | | | | |
| Ovary (FOCI)[42] (FOCI) | 3 | 4369 | 9123 | 9911464 | Illumina | $R^2$>0.3 | deCODE | 716 | 111373 | | | |
| | | | | | | | iCOGS[42] | 16283 | 23491 | | | |
| Ovary - Serous | 3 | 2556 | 9123 | 9911279 | | | iCOGS | 10316 | 23491 | | | |
| Ovary - Endometrial | 3 | 715 | 9123 | 9910229 | | | iCOGS | 2338 | 23491 | | | |
| **Total*** | 45 | 61851 | 61820 | 9916564 | | | | 55789 | 330490 | | 18152 | 21410 |

* The number of unique individuals after accounting for cancer subtypes and overlapping controls. Breast iCOGS included only 1q22 variants, so total for replicates without breast iCOGS is shown.
† Analyses were performed for a specific variant if at least 3 sites (i.e., three of lung, colorectal, prostate, breast or ovary) contributed data.
‡ Imputation performed using the 1000 genome reference panel. Exclusion threshold shown.

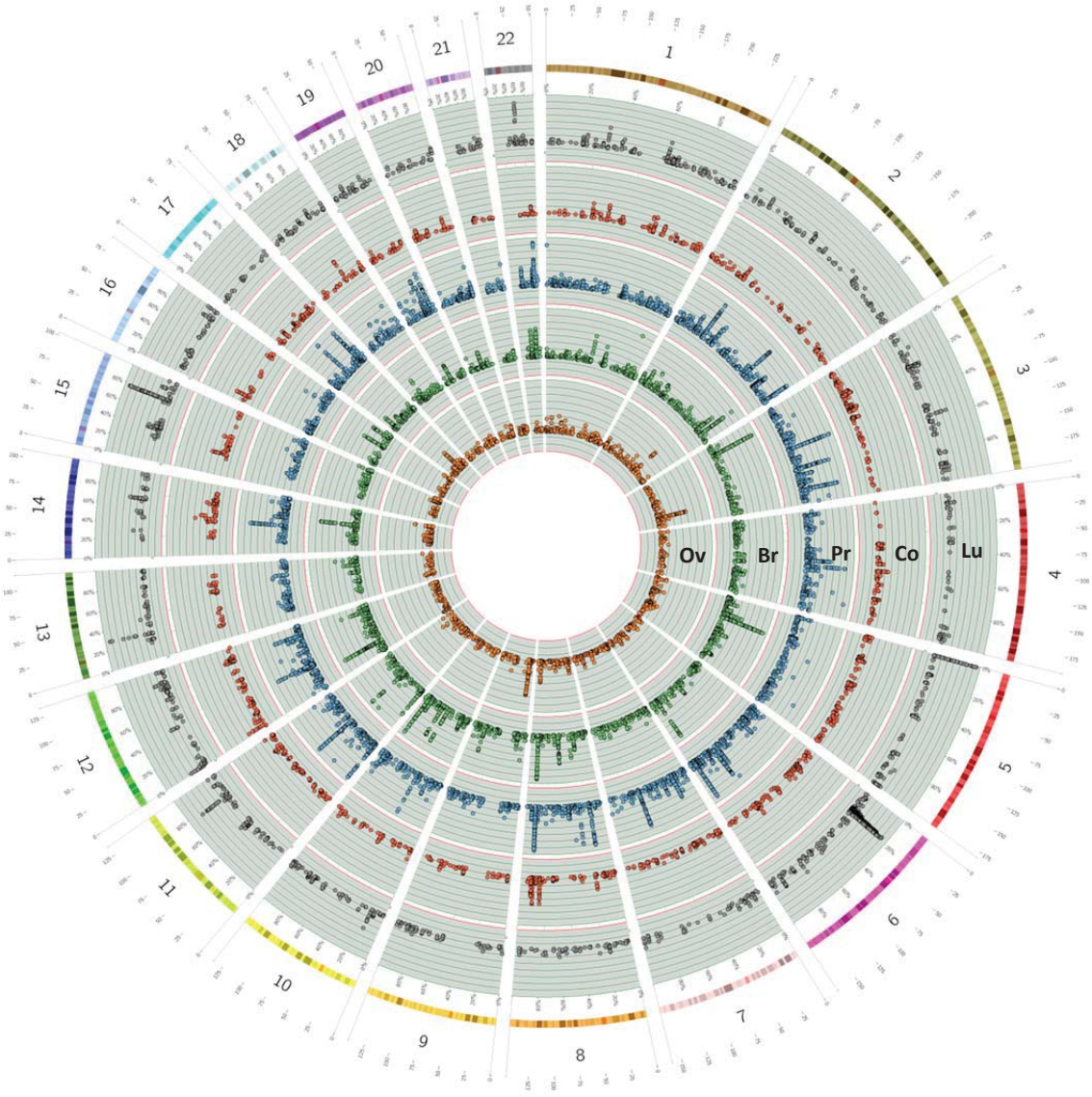
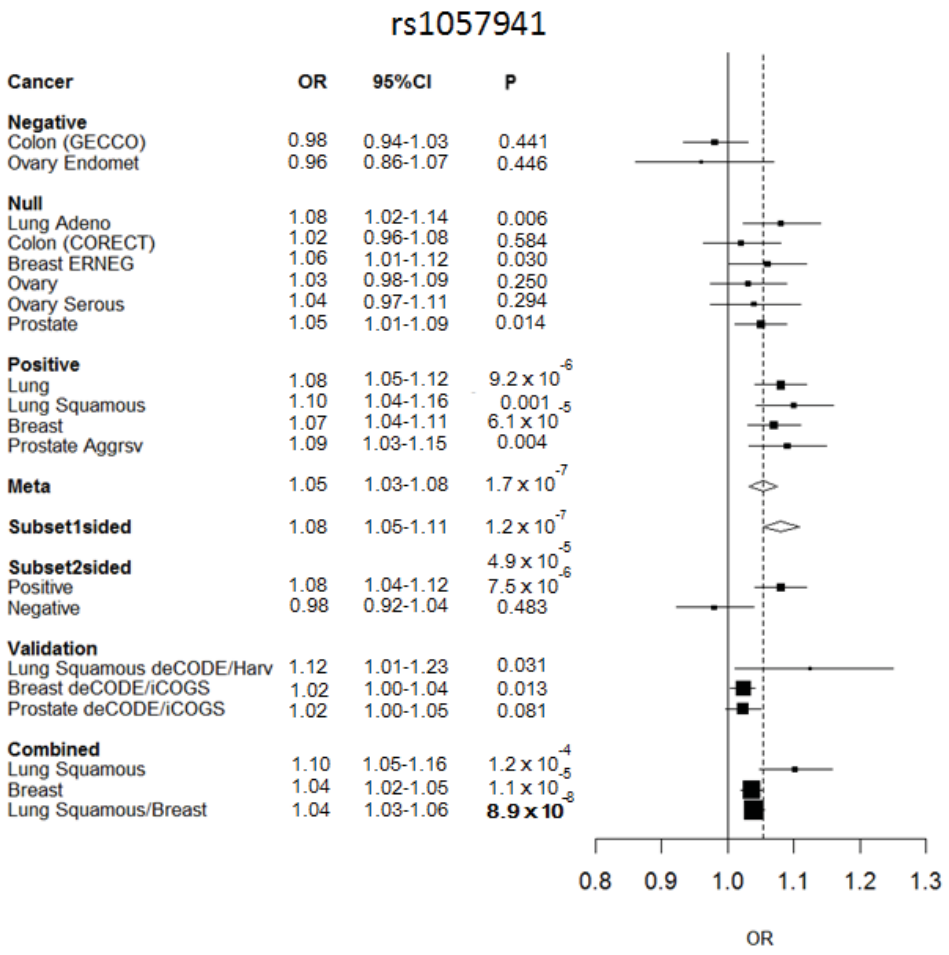



**Figure 1.** Manhattan plots ($-\log_{10}(p)$) by chromosome for individual cancer sites (innermost to outermost ring - ovary(Ov) breast (Br), prostate (PR), colorectal/GECCO (Co), Lung (Lu)).

**a)**




**b)**

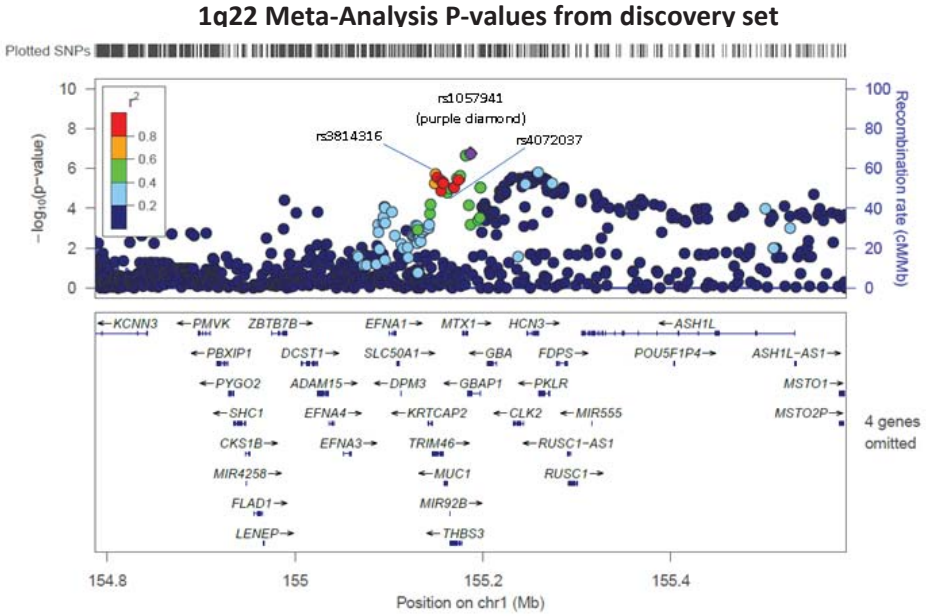
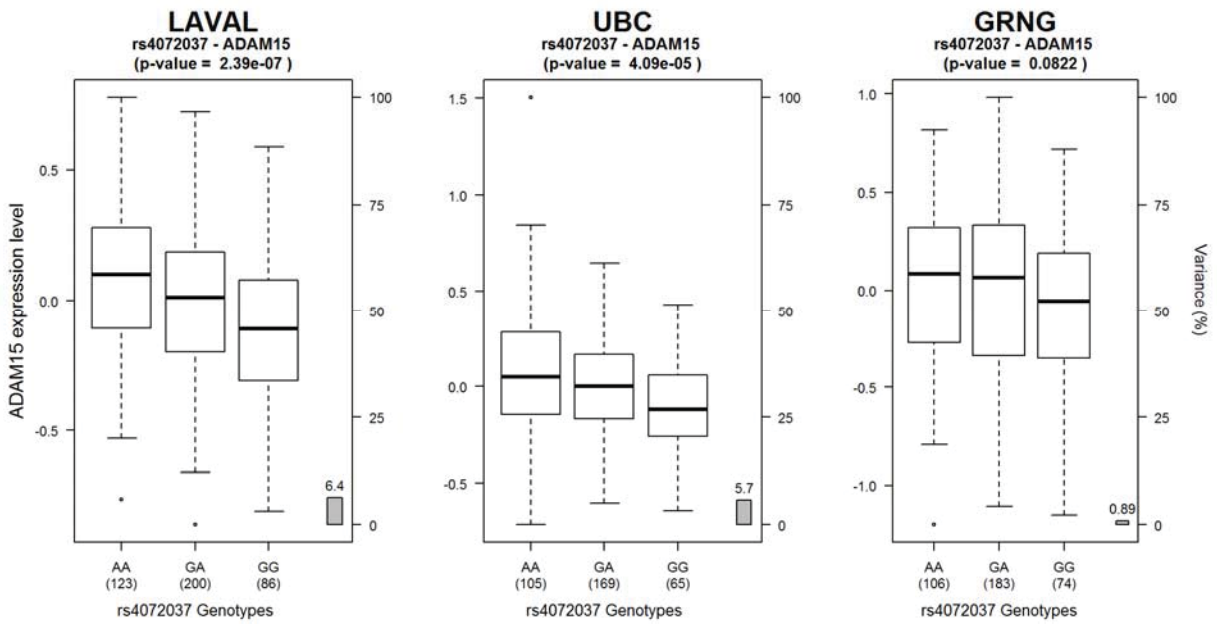



**Figure 2. Results for rs1057941:** a) Forest plot for rs1057941 showing per allele ORs for risk allele A (of A/G). Standard fixed effects meta-analysis (also indicated by dashed line), and subset meta-analysis results (two-sided, one-sided and positive and negative subset associations) are shown. b) Regional plot showing p-values from overall meta-analysis at region 1q22 using GAME-ON/GECCO discovery set data. The top breast cancer hit (rs3814316) and top squamous cell carcinoma hit (rs4072037) are also highlighted.

a)




b)

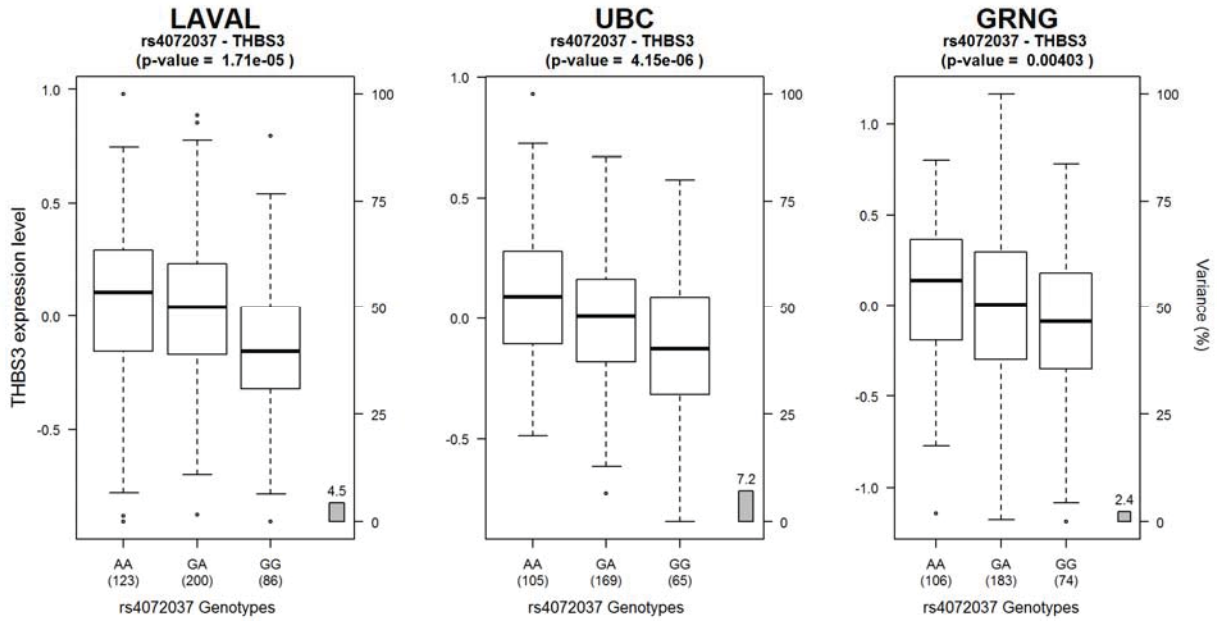
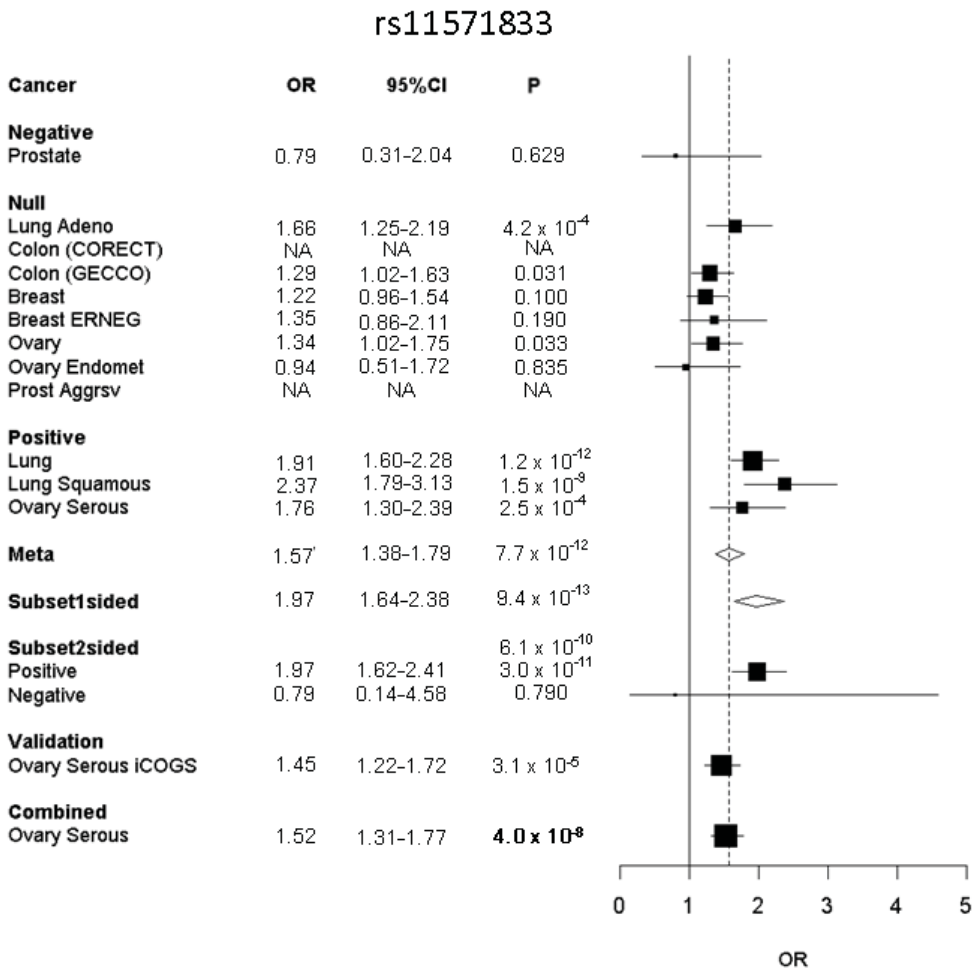



**Figure 3. Boxplots of gene expression levels in normal lung tissue for rs4072037** for A) *ADAM15* and B) *THBS3*. Results are presented by study (Laval, UBC, Groningen).

a)




b)

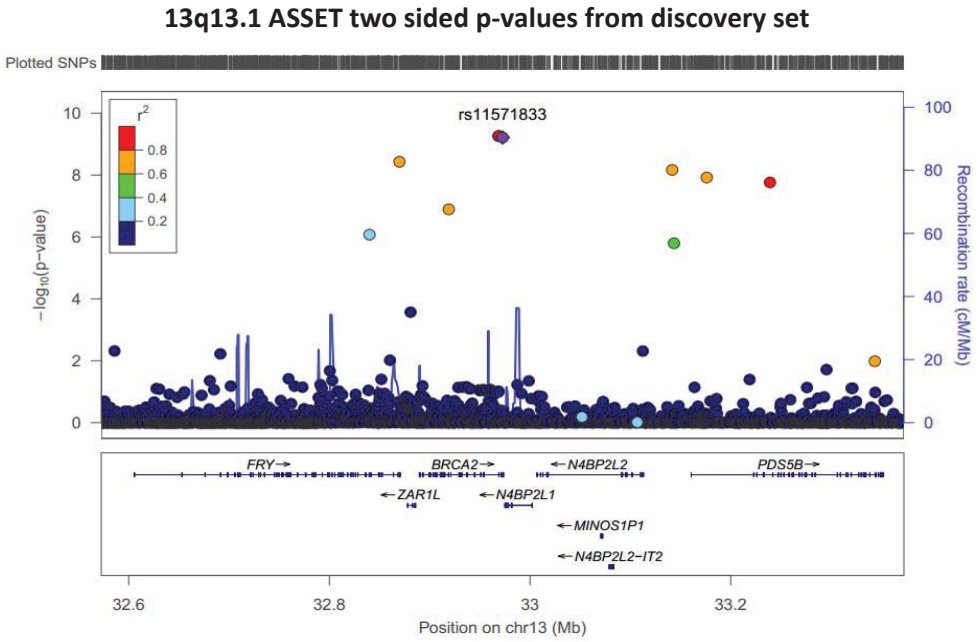
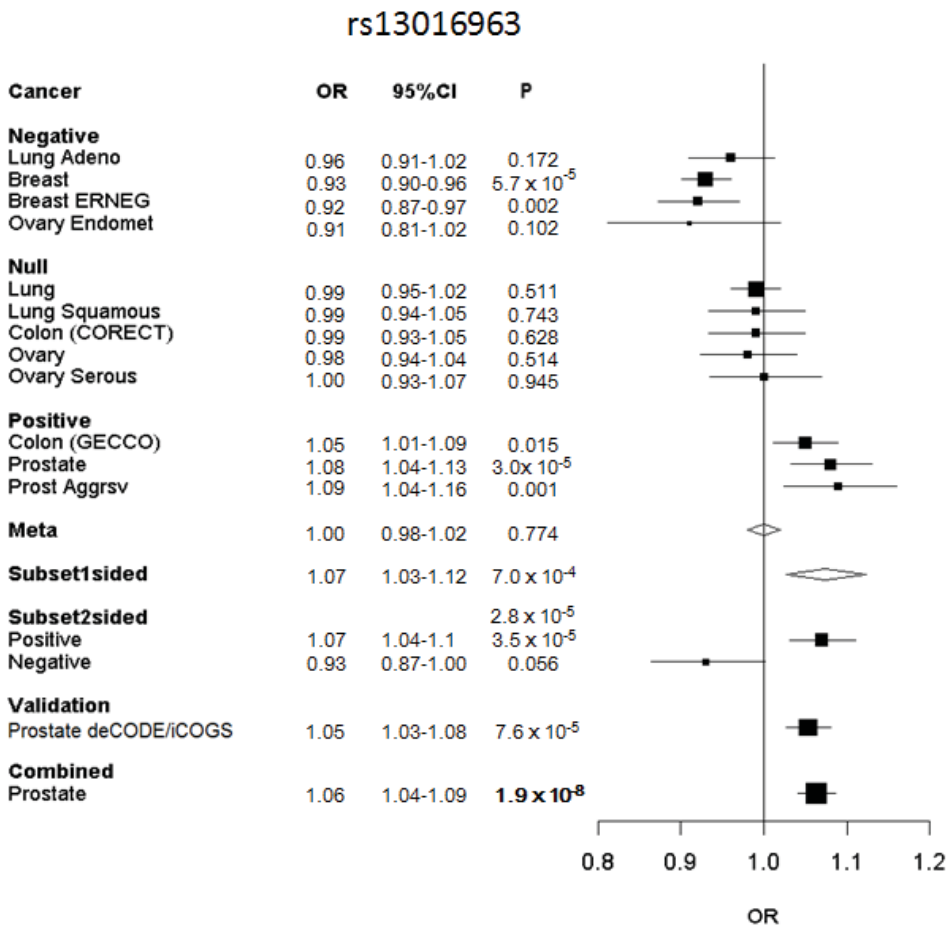



**Fig 4. Results for rs11571833 a**) Forest plot showing per allele ORs for risk allele T of (T/A). Standard fixed effects meta-analysis (also indicated by dashed line), and subset meta-analysis results (two-sided, one-sided and positive and negative subset associations) are shown. **b**) Regional plot showing two sided p-values from ASSET subset meta-analysis at region 13q13.1 using GAME-ON/GECCO discovery set data. Peak is at *BRCA2*. rs11571815 and rs11571818 have nearly identical association signals but are partially obscured by rs11571833.

a)




b)

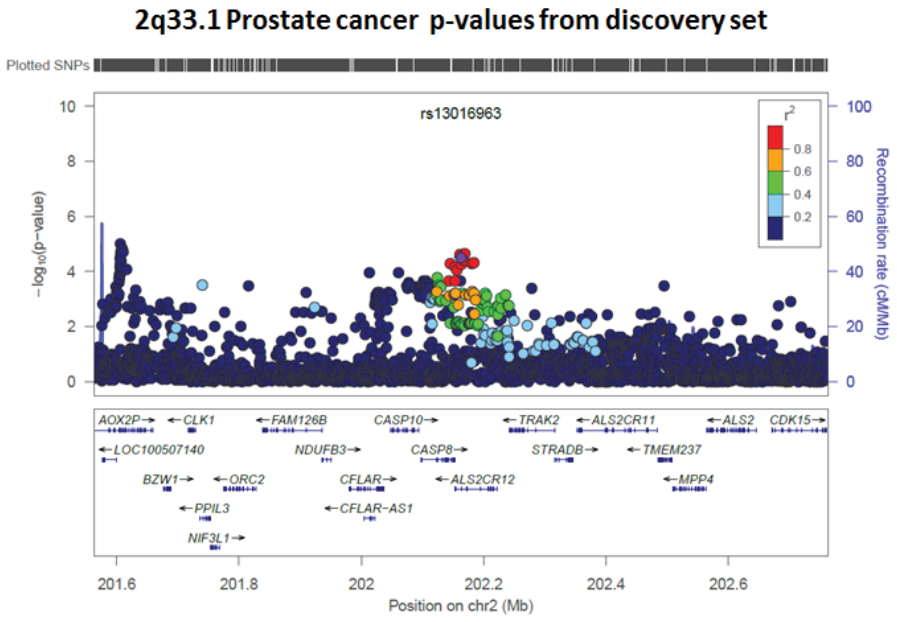



**Fig 5. Results for rs13016963**: **a**) Forest plot showing per allele ORs for risk allele G (of A/G). Standard fixed effects meta-analysis (also indicated by dashed line), and subset meta-analysis results (two-sided, one-sided and positive and negative subset associations) are shown. **b**) Regional plot showing P-values for GAME-ON prostate cancer GWAS at region 2q33.1 using GAME-ON/GECCO discovery set data. Peak is at *ALS2CR12*.

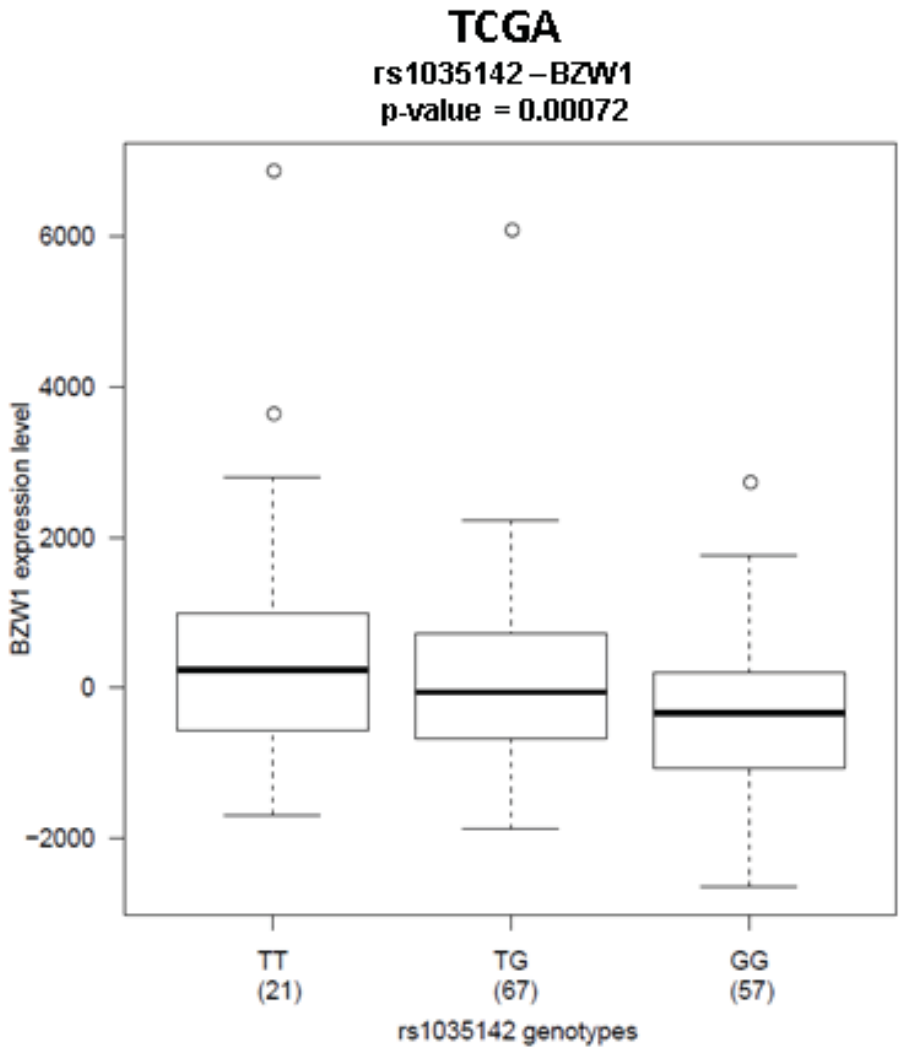
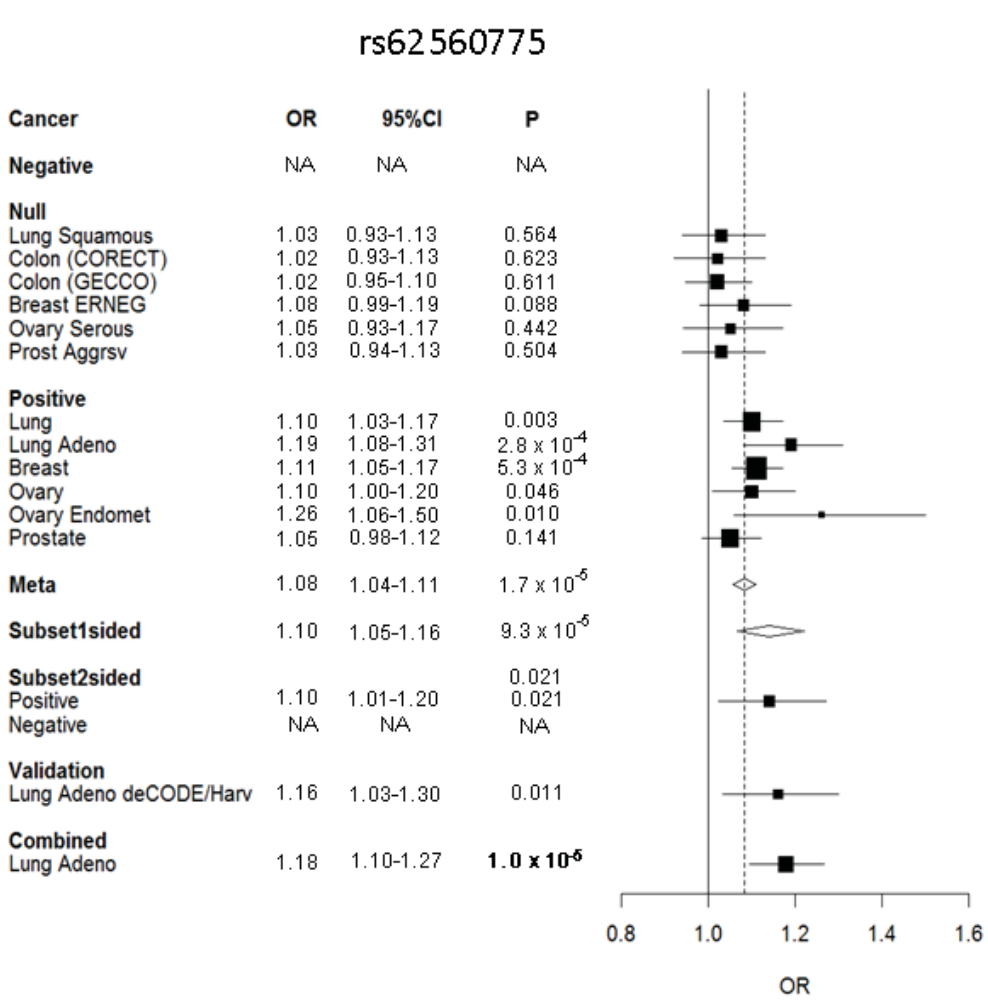



**Figure 6. Boxplot of *BZW1* gene expression levels in prostate tumor and normal tissue for rs1035142/ rs13016963**. rs1035142 is presented as a surrogate for rs13016963, with which it shows perfect LD in eQTL data.

a)




b)

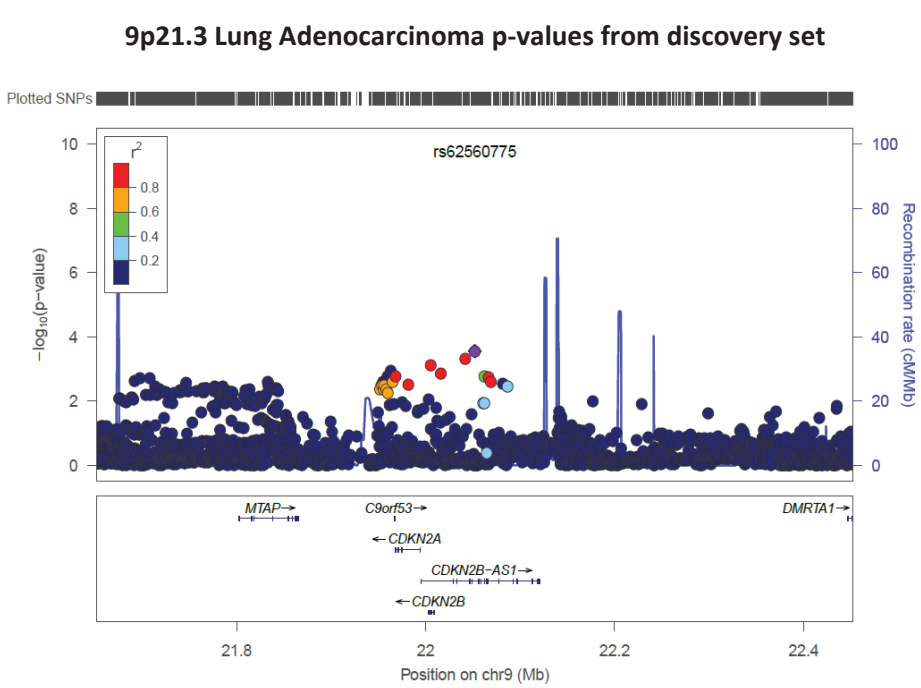
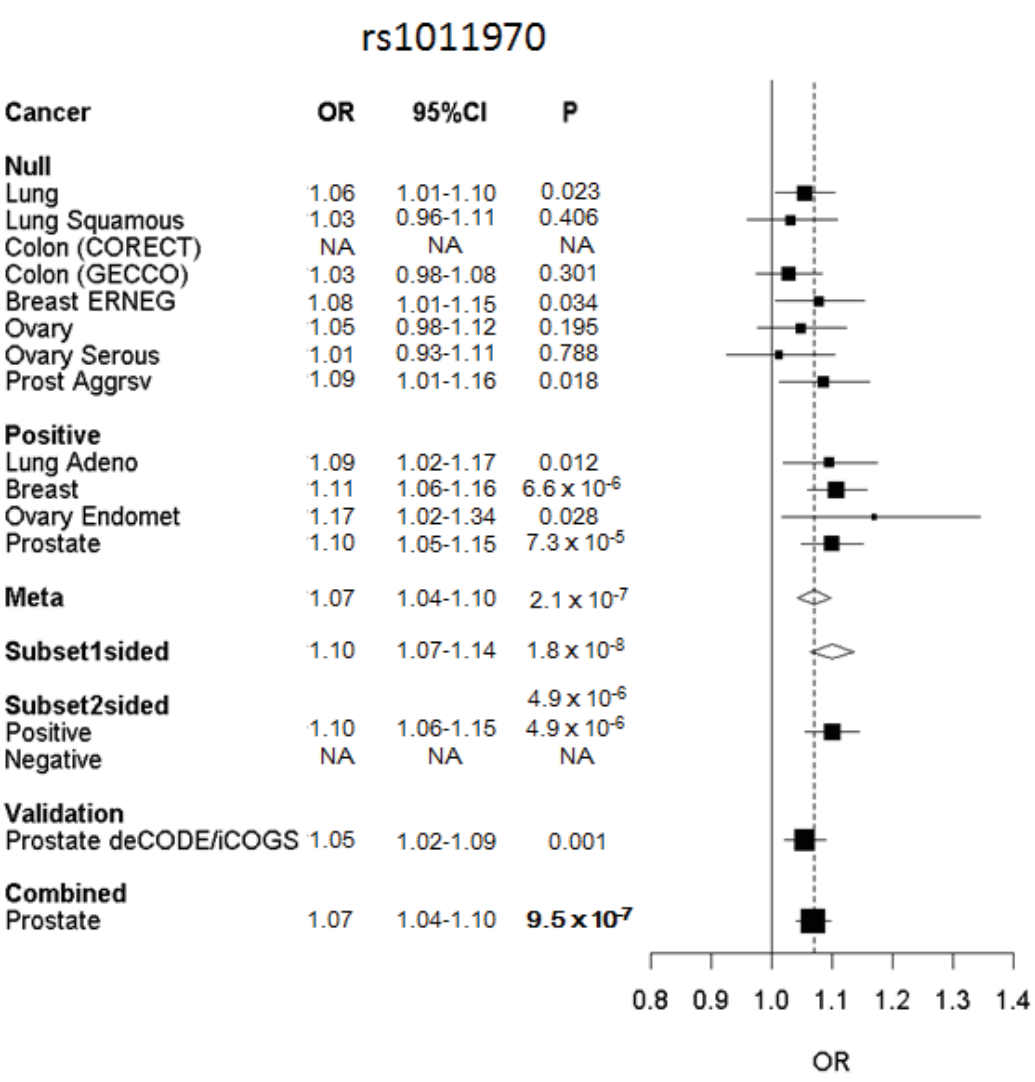



**Figure 7. Results for rs62560775**: **a**) Forest plot showing per allele ORs for risk allele G (of A/G). Standard fixed effects meta-analysis (also indicated by dashed line), and subset meta-analysis results (two-sided, one-sided and positive and negative subset associations) are shown. **b**) Regional plot showing P-values for GAME-ON lung adenocarcinoma GWAS at region 9p21.3, using GAME-ON/GECCO discovery set data. Peak is at *CDKN2B-AS1*.

a)




b)

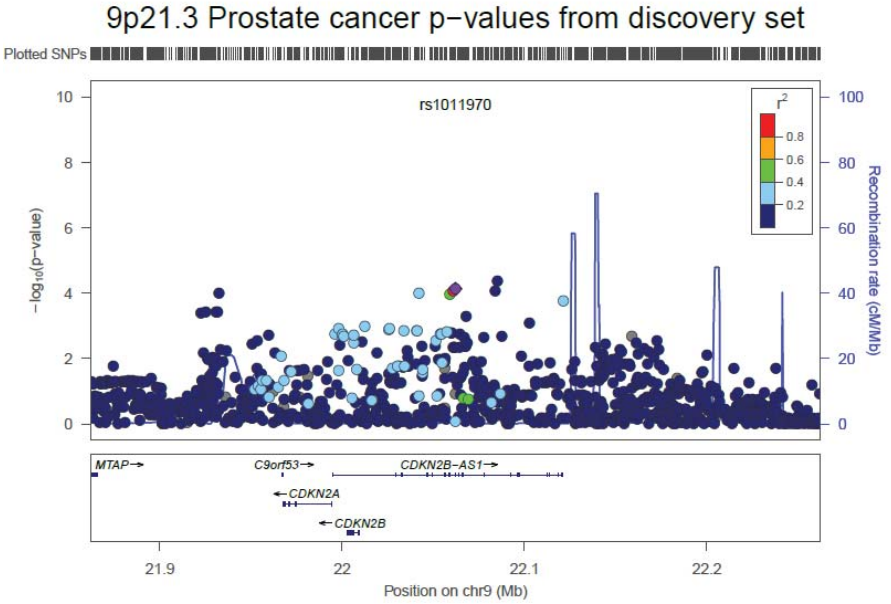



**Figure 8. Results for rs1011970**: **a**) Forest plot showing per allele ORs for risk allele T (of T/G). Standard fixed effects meta-analysis (also indicated by dashed line), and subset meta-analysis results (two-sided, one-sided and positive and negative subset associations) are shown. **b**) Regional plot showing P-values for GAME-ON prostate cancer GWAS at region 9p21.3, using GAME-ON/GECCO discovery set data. Peak is at *CDKN2B-AS1*.