Combining genome-wide studies of breast, prostate, ovarian and endometrial cancers maps cross-cancer susceptibility loci and identifies new genetic associations

Siddhartha P. Kar[1,2]†, Sara Lindström[3,4], Rayjean J. Hung[5], Kate Lawrenson[6,7], Marjanka K. Schmidt[8,9], Tracy A. O'Mara[10], Dylan M. Glubb[10], Jonathan P. Tyrer[11], Joellen M. Schildkraut[12], Jenny Chang-Claude[13,14], Ahmad G. M. Alsulimani[15], Fernando M. Antón[16], Alicia Beeghly-Fadiel[17], Line Bjørge[18,19], Clara Bodelon[20], Hiltrud Brauch[21,22,23], Stefanie Burghaus[24], Daniele Campa[25,26], Michael Carney[27], Zhihua Chen[28], Mary B. Daly[29], Andreas du Bois[30], Arif B. Ekici[31,32], Ailith Ewing[33], Peter Fasching[34,35], James M. Flanagan[36], Jan Gawelko[37], Graham G. Giles[38,39,40], Robert J. Hamilton[41], Holly R. Harris[42], Florian Heitz[43,44], Michelle Hildebrandt[45], Peter Hillemanns[46], Ruea-Yea Huang[47], Liher Imaz[48,49], Arvids Irmejs[50], Anna Jakubowska[51,52], Allan Jensen[53], Esther M. John[54,55], Päivi Kannisto[56], Beth Y. Karlan[57], Elza Khusnutdinova[58,59], Lambertus A. Kiemeney[60], Susanne K. Kjaer[53,61], Rüdiger Klapdor[46], Petra Kleiblova[62], Martin Köbel[63], Bozena Konopka[64], Camilla Krakstad[65,66], Davor Lessel[67], Artitaya Lophatananon[68], Taymaa May[69], Agnieszka D. Mieszkowska[64], Alvaro N. Monteiro[28], Kirsten Moysich[15], Kenneth Muir[68], Sune F. Nielsen[70], Kunle Odunsi[47,71], Håkan Olsson[72], Tjoung-Won Park-Simon[73], Jennifer B. Permuth[74], Paolo Peterlongo[75], Agnieszka Podgorski[64], Paolo Radice[76], Harvey A. Risch[77], Ingo B. Runnebaum[78], Iwona K. Rzepecka[64], Veronica W. Setiawan[79], Nadeem Siddiqui[80], Weiva Sieh[81], Beata Śpiewankiewicz[82], Lukasz M. Szafron[83], Cheryl L. Thompson[84], Linda J. Titus[85], Clare Turnbull[86,87,88], Nawaid Usmani[89], Anne M. van Altena[90], Ana Vega-Gliemmo[91], Ignace Vergote[92], Robert A. Vierkant[93], Joseph Vijai[94,95,96], Stacey J. Winham[93], Robert Winqvist[97,98], Herbert Yu[99], the PRACTICAL consortium, CRUK, BPC3, CAPS, PEGASUS[100], Deborah J. Thompson[11], Wei Zheng[17], Ian P. M. Tomlinson[101], Andrew Berchuck[102], Susan J. Ramus[103], Stephen J. Chanock[104], Douglas F. Easton[11,105], Georgia Chenevix-Trench[10], Simon A. Gayther[6], Amanda B. Spurdle[10], Rosalind A. Eeles[106,107], Peter Kraft[108,109], Paul D. P. Pharoah[11,105]

1 Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, UK
2 Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
3 Department of Epidemiology, University of Washington School of Public Health, Seattle, WA, USA
4 Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
5 Lunenfeld-Tanenbaum Research Institute, Sinai Health System and University of Toronto, Toronto, Canada
6 Center for Bioinformatics and Functional Genomics, Department of Biomedical Science, Cedars-Sinai Medical Center, Los Angeles, CA, USA
7 Women's Cancer Program at the Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA
8 Division of Molecular Pathology, The Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital, Amsterdam, the Netherlands
9 Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital, Amsterdam, the Netherlands
10 Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Australia
11 Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
12 Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA
13 Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany
14 Cancer Epidemiology Group, University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany
15 Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, NY, USA

16 Hospital Clínico San Carlos, Madrid, Spain
17 Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA
18 Department of Obstetrics and Gynecology, Haukeland University, Hospital, Bergen, Norway
19 Centre for Cancer Biomarkers CCBIO, Department of Clinical Science, University of Bergen, Bergen, Norway
20 Divison of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA
21 Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany
22 iFIT Cluster of Excellence, University of Tübingen, Tübingen, Germany
23 German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ), Partner Site Tübingen, Tübingen, Germany
24 Frauenklinik, Universitätsklinikum Erlangen, Erlangen, Germany
25 Division of Cancer Epidemiology, DKFZ, Heidelberg, Germany
26 Department of Biology, University of Pisa, Pisa, Italy
27 Department of Obstetrics and Gynecology, John A. Burns School of Medicine, University of Hawaii, Honolulu, HI, USA
28 Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA
29 Department of Clinical Genetics, Fox Chase Cancer Center, Philadelphia, PA, USA
30 Department of Gynecology and Gynecologic Oncology, Evang. Kliniken Essen-Mitte, Essen, Germany
31 Institute of Human Genetics, University Hospital Erlangen, Erlangen, Germany
32 Friedrich-Alexander University Erlangen-Nuremberg, Comprehensive Cancer Center Erlangen, Erlangen, Germany
33 MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, The University of Edinburgh, Western General Hospital, Edinburgh, UK
34 David Geffen School of Medicine, Department of Medicine Division of Hematology and Oncology, University of California at Los Angeles. Los Angeles, CA, USA
35 Department of Gynecology and Obstetrics, Comprehensive Cancer Center ER-EMN, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany.
36 Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London, UK
37 Medical College Rzeszow University
38 Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Australia
39 Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Australia
40 Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Melbourne, Australia
41 Division of Urology, Department of Surgery, Princess Margaret Cancer Centre, University Health Network, University of Toronto, Toronto, Canada
42 Program in Epidemiology, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
43 Department of Gynecology and Gynecologic Oncology, Kliniken Essen-Mitte/ Evang., Essen, Germany
44 Department of Gynecology, Center for Oncologic Surgery, Charité Campus Virchow-Klinikum, Berlin, Germany
45 Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA
46 Department of Gynecology and Obstetrics, Hannover Medical School, Hannover, Germany
47 Center for Immunotherapy, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA
48 Ministry of Health of the Basque Government, Public Health Division of Gipuzkoa, Donostia-San Sebastian, Spain
49 Biodonostia Health Research Institute, Donostia-San Sebastian, Spain

50 Riga Stradins University, Pauls Stradins Clinical University Hospital, Riga, Latvia
51 Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland
52 Independent Laboratory of Molecular Biology and Genetic Diagnostics, Pomeranian Medical University, Szczecin, Poland
53 Department of Virus, Lifestyle and Genes, Danish Cancer Society Research Center, Copenhagen, Denmark
54 Cancer Prevention Institute of California, Fremont, CA, USA
55 Division of Oncology and Stanford Cancer Institute, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA
56 Department of Obstetrics and Gynecology, University of Lund, Lund, Sweden
57 David Geffen School of Medicine, Department of Obstetrics and Gynecology, University of California at Los Angeles, Los Angeles, CA, USA
58 Institute of Biochemistry and Genetics, Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, Russia
59 Department of Genetics and Fundamental Medicine, Bashkir State University, Ufa, Russia
60 Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, The Netherlands
61 Department of Gynaecology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark
62 Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic
63 Department of Pathology and Laboratory Medicine, University of Calgary, Foothills Medical Center, Calgary, Canada
64 Department of Pathology and Laboratory Diagnostics, The Maria Sklodowska-Curie National Research Institute of Oncology, Warsaw, Poland
65 Department of Obstetrics and Gynecology, Haukeland University Hospital, Bergen, Norway
66 Centre for Cancer Biomarkers, Department of Clinical Science, University of Bergen, Bergen, Norway
67 Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
68 Division of Population Health, Health Services Research and Primary Care, Faculty of Biology, Medicine and Health, Centre for Epidemiology, University of Manchester, Manchester, UK
69 Division of Gynecologic Oncology, University Health Network, Princess Margaret Hospital, Toronto, Canada
70 Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark
71 Department of Gynecologic Oncology, Roswell Park Comprehensive Cancer Center, Buffalo, NY
72 Division of Oncology, Department of Clinical Sciences, Lund University, Lund, Sweden
73 Department of Gynecology, Hannover Medical School, Hannover, Germany
74 Departments of Cancer Epidemiology and Gastrointestinal Oncology, Moffitt Cancer Center and Research Institute, Tampa, FL
75 Genome Diagnostics Program, IFOM-The FIRC (Italian Foundation for Cancer Research) Institute of Molecular Oncology, Milan, Italy
76 Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Research, Fondazione IRCCS Istituto Nazionale dei Tumori (INT), Milan, Italy
77 Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT
78 Department for Gynaecology and Reproductive Medicine, Jena University Hospital-Friedrich Schiller University Jena, Jena, Germany
79 Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
80 Department of Gynaecological Oncology, Glasgow Royal Infirmary, Glasgow, UK
81 Department of Population Health Science and Policy, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

82 Department of Gynecologic Oncology, Institute of Oncology, Warsaw, Poland
83 Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland
84 Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA
85 Department of Epidemiology, Geisel School of Medicine at Dartmouth, Hanover, NH, USA
86 Division of Genetics and Epidemiology, Institute of Cancer Research, London, UK
87 Department of Clinical Genetics, Royal Marsden NHS Foundation Trust, London, UK
88 Public Health England, National Cancer Registration and Analysis Service, London, UK
89 Department of Radiation Oncology, Cross Cancer Institute, University of Alberta, Edmonton, Canada
90 Radboud University Medical Centre, Department of Obstetrics and Gynecology, Nijmegen, The Netherlands
91 Galician Public Foundation of Genomic Medicine (FPGMX), Clinical University Hospital, Santiago de Compostela, Spain
92 Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, KU University, Leuven, Belgium
93 Department of Health Sciences Research, Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, USA
94 Niehaus Center For Inherited Cancer Genomics, Memorial Sloan Kettering Cancer Center, New York, NY, USA
95 Clinical Genetics Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA
96 Department of Medicine, Weill Cornell Medical College, New York, NY, USA
97 Laboratory of Cancer Genetics and Tumor Biology, Cancer and Translational Medicine Research Unit, Biocenter Oulu, University of Oulu, Oulu, Finland
98 Laboratory of Cancer Genetics and Tumor Biology, Northern Finland Laboratory Centre Oulu, Oulu, Finland
99 Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA
100 Members from the PRACTICAL Consortium, CRUK, BPC3, CAPS and PEGASUS are provided in the Supplementmary Notes
101 Edinburgh Cancer Research Centre, University of Edinburgh, Edinburgh, UK
102 Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, Duke University Medical Center, Durham, NC, USA
103 Adult Cancer Program, Lowy Cancer Research Centre, University of New South Wales, Sydney, Australia
104 Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA
105 Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK
106 Institute of Cancer Research, London, UK
107 Royal Marsden NHS Foundation Trust, London, UK
108 Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
109 Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

† Corresponding author:

Siddhartha Kar, MBBS
MRC Integrative Epidemiology Unit
University of Bristol, Oakfield House
Oakfield Grove, Bristol BS8 2BN, UK

Email: siddhartha.kar@bristol.ac.uk

ABSTRACT

We report a meta-analysis of breast, prostate, ovarian, and endometrial cancer genome-wide association data (effective sample size: 237,483 cases/317,006 controls). This identified 465 independent lead variants ($P<5x10^{-8}$) across 192 genomic regions. Four lead variants were >1Mb from previously identified risk loci for the four cancers and an additional 23 lead variant-cancer associations were novel for one of the cancers. Bayesian models supported pleiotropic effects involving at least two cancers at 222/465 lead variants in 118/192 regions. Gene-level association analysis identified 13 shared susceptibility genes ($P<2.6x10^{-6}$) in 13 regions not previously implicated in any of the four cancers and not uncovered by our variant-level meta-analysis. Several lead variants had opposite effects across cancers, including a cluster of such variants in the TP53 pathway. Fifty-four lead variants were associated with blood cell traits and suggested genetic overlaps with clonal hematopoiesis. Our study highlights the remarkable pervasiveness of pleiotropy across hormone-related cancers, further illuminating their shared genetic and mechanistic origins at variant- and gene-level resolution.

INTRODUCTION

Cancers of the breast, prostate, ovary and endometrium together accounted for nearly 22% of all new cancer cases diagnosed and approximately 1.2 million deaths worldwide in 2018[1]. These four cancers share common etiologies. Disease risks have been linked to variations in sex steroid hormones (e.g., hormone replacement therapy increases the risks of breast, ovarian, and endometrial cancers[2–4]) and shared genetic risk factors (pleiotropy). For example, rare highly penetrant germline mutations in the *BRCA1* and *BRCA2* genes elevate risks of breast, ovarian, and prostate cancers and have yielded fundamental cross-cancer mechanistic and therapeutic insights[5]. This provides the motivation for the identification of additional shared genetic risk factors, including pleiotropic common variants that may affect the risk of multiple cancers.

In 2016, we combined summary data from individual genome-wide association studies (GWAS) of breast, prostate, and ovarian cancer susceptibility that included a total of 112,349 cancer cases and 116,421 controls to identify 12 new cancer risk loci[6]. Since then, most of these loci have been replicated at genome-wide significance ($P < 5 \times 10^{-8}$) in larger, single-cancer GWAS of breast[7], prostate[8], or ovarian cancer[9]. We have now combined the results from these larger GWAS together with the results from an endometrial cancer GWAS[10] in order to address four main aims. First, to combine single nucleotide polymorphism (SNP)-level associations across breast, prostate, ovarian, and endometrial cancers to identify susceptibility loci that have not been previously reported for any of the four cancers. Second, to use the combined data set to identify novel susceptibility loci for one of the four cancers in genomic regions that contain a known susceptibility locus for another of the four cancers. Third, to determine the combination of the four cancers that is most likely to underlie the association at the lead SNPs of all susceptibility loci identified in the combined data set in order to map shared or cross-cancer loci. Fourth, to carry out gene-level association analysis based on this powerful combined SNP-level data set to identify candidate cancer susceptibility genes in loci not previously identified by single- and multi-cancer SNP-level association analyses.

RESULTS

Cross-cancer genome-wide association meta-analysis

We used meta-analysis based on the Han and Eskin model[11,12] to combine summary results for 9,530,997 variants with minor allele frequency > 1% from the largest genome-wide association data sets for susceptibility to breast[7], prostate[8], ovarian[9], and endometrial cancers[10] published as of August 2019 (**Methods**). All cases and controls included were of European ancestry (**Supplementary Table 1**). A recent study[13] has shown that some breast cancer susceptibility SNPs have subtype-specific associations where the allele that confers risk of developing one subtype is protective for another subtype. Due to their opposite directions, such allelic associations may be masked in an analysis that

does not take into account subtype-specific associations. Subtype-specific data were available for the current study by estrogen receptor (ER)-status for breast cancer and for high-grade serous ovarian cancer (HGSOC), which is the most common histological subtype of epithelial ovarian cancer and known to be similar at the molecular level with ER-negative breast cancer[14]. Therefore, we also conducted an additional meta-analysis combining the results for ER-positive breast cancer, ER-negative breast cancer, HGSOC, prostate cancer, and endometrial cancer. Hereafter, we refer to this additional analysis as the subtype-focused meta-analysis.

The genomic control inflation statistic scaled for 1,000 cases and 1,000 controls, λ_{1000}, was 1.00 for both the main and the subtype-focused meta-analysis (**Supplementary Fig. 1**). The main meta-analysis identified 441 independent (correlated with $r^2 < 0.05$; **Methods**) lead SNPs at genome-wide significance ($P < 5 \times 10^{-8}$). These spanned 183 regions at least 1 Mb apart (**Supplementary Table 2**; **Supplementary Fig. 2a**; **Methods**). The subtype-focused meta-analysis identified 393 independent genome-wide significant lead SNPs spanning 169 regions at least 1 Mb apart (**Supplementary Table 2**; **Supplementary Fig. 2b**). Two hundred and twenty-eight lead SNPs were in common to both analyses. Of the remaining 165 lead SNPs that were unique to the subtype-focused analysis, 156 SNPs were within 1 Mb of a lead SNP identified in the same region in the main analysis. Fifteen of these 156 SNPs were independent ($r^2 < 0.05$) of the main analysis lead SNP from the same region (**Supplementary Table 2**). Nine lead SNPs were unique to the subtype-focused analysis and > 1 Mb from the main analysis lead SNPs (**Supplementary Table 2**). Thus, in total, the main and subtype-focused analyses identified 465 (441 + 15 + 9) independent lead SNPs spanning 192 (183 + 9) regions. We replicated at $P < 5 \times 10^{-8}$ in our current combined data set all 12 cancer susceptibility loci that we first identified in our previous multi-cancer study of breast, prostate, and ovarian cancer (**Supplementary Table 2**)[6].

Cancer susceptibility loci not previously identified for any of the four cancers

We identified four lead SNPs at genome-wide significance ($P < 5 \times 10^{-8}$) that were at least 1 Mb away from any previously published genome-wide significant lead risk SNPs for any of the cancers evaluated in this study (**Table 1**; **Methods**). Two lead SNPs were identified from the main meta-analysis and two were identified from the subtype-focused meta-analysis.

Two additional lead SNPs, rs66686620 in the 2p13.3 region (*ANTXR1*) and rs6065253 in the 20q12 region (*MAFB*) were also genome-wide significant and over 1 Mb away from any previously reported lead risk SNP for any of the four cancers as of August 2019 (**Supplementary Table 2**). However, since then associations ($P < 5 \times 10^{-8}$) within this 1 Mb interval have been identified in a breast cancer genome-wide association meta-analysis by Zhang *et al.*[13] (i) in the 2p13.3 region (rs4602255) through the addition of ~10% more cases and ~9% more controls to the breast cancer data set used here and (ii) in the 20q12 region (rs6065254) through the use of a statistical model that accounted for heterogeneity in associations by hormone receptor status and tumor grade in the expanded breast cancer data set. SNPs rs6065253 and rs6065254 were perfectly correlated ($r^2 = 1$; distance between SNPs = 52bp) but rs66686620 and rs4602255 were independent ($r^2 = 0.003$; distance between SNPs = 141kb) suggesting that the multi-cancer 2p13.3 signal is a distinct association.

Table 1 The cancer susceptibility loci not previously identified for any of the four cancers.[a]

Region, position[b]	Lead SNP, nearest gene	Alleles (EA/OA), EAF	Cancer	OR (95% CI)	*P*
2p24.3	rs7586503	T/A	Breast	0.98 (0.96-0.99)	2.0×10^{-3}
15995041	*RNU5E-7P*	0.28	Prostate	0.95 (0.93-0.97)	1.6×10^{-7}
			Ovarian	0.99 (0.96-1.02)	5.3×10^{-1}
			Endometrial	0.95 (0.92-0.99)	6.0×10^{-3}
			Meta-analysis[c] (main)		6.0×10^{-9}
6q16.2	rs377527229	C/CAAA	Breast	1.03 (1.02-1.05)	5.9×10^{-6}
100091541	*Y_RNA*	0.23	Prostate	1.03 (1.01-1.05)	1.1×10^{-3}
			Ovarian	1.04 (1.01-1.08)	6.4×10^{-3}
			Endometrial	1.03 (1.00-1.07)	8.5×10^{-2}
			Meta-analysis[c] (main)		2.8×10^{-9}
3q27.1	rs6784596	G/A	ER-pos breast	1.00 (0.98-1.01)	7.3×10^{-1}
184267740	*EIF2B5-AS1*	0.51	ER-neg breast	1.05 (1.03-1.07)	2.0×10^{-5}
			Prostate	0.97 (0.95-0.98)	8.5×10^{-5}
			HGSOC	1.01 (0.98-1.04)	5.3×10^{-1}
			Endometrial	1.06 (1.03-1.09)	2.3×10^{-4}
			Meta-analysis[c] (subtype)		4.8×10^{-8}
17p11.2	rs5819638	CT/C	ER-pos breast	1.03 (1.02-1.05)	1.3×10^{-4}
17857626	*TOM1L2*	0.33	ER-neg breast	1.03 (1.00-1.05)	2.8×10^{-2}
			Prostate	1.04 (1.02-1.06)	2.1×10^{-5}
			HGSOC	1.01 (0.97-1.04)	7.7×10^{-1}
			Endometrial	1.03 (0.99-1.06)	1.4×10^{-1}
			Meta-analysis[c] (subtype)		4.9×10^{-8}

[a]> 1 Mb away from any previously reported lead SNP for any of the four cancers.

[b]Build 37 position.

[c]The Han and Eskin "RE2C*" model does not report a combined effect size estimate and confidence interval.
Abbreviations: SNP, single nucleotide polymorphism; EA, effect allele; OA, other allele; EAF, effect allele frequency; OR, odds ratio; CI, confidence interval; ER-pos, estrogen receptor-positive breast cancer; ER-neg, estrogen receptor-negative breast cancer; HGSOC, high-grade serous ovarian cancer.

Cancer susceptibility loci not previously identified for one of the four cancers

In addition to the four loci described above, we identified 23 associations that were novel for at least one of the cancers evaluated in this study (**Table 2**). We did this by examining the lead SNPs (main and/or subtype-focused meta-analysis $P < 5 \times 10^{-8}$) to identify those lead SNPs that were (i) associated with one of the cancers at $P < 10^{-3}$ in the contributing single-cancer data set and (ii) > 1 Mb away from any previously identified genome-wide significant lead risk SNP for the same cancer (> 10 Mb when applying this criterion to the extended major histocompatibility complex region; chr6:25—35Mb). The 23 new associations included eight for breast cancer, five for prostate cancer, and five each for ovarian and endometrial cancer susceptibility (**Table 2**). These associations involved 21 lead SNPs since two lead SNPs, rs1806845 and rs3819772, each met these criteria for two cancer types (**Table 2**). Associations between these lead SNPs and all cancers included in this study are listed in **Supplementary Tables 3 and 4**.

Table 2 The cancer susceptibility loci not previously identified for one of the four cancers.[a]

Lead SNP	Chr	Pos	EA	OA	EAF	OR (95% CI)	Single-cancer P[b]	4-cancer P	Nearest Gene
Breast cancer (main analysis)									
rs2546628	5	177882698	G	A	0.33	1.02 (1.01-1.04)	4.4×10^{-4}	1.6×10^{-8}	*COL23A1*
rs71475909	11	47377283	G	A	0.36	0.97 (0.96-0.99)	3.2×10^{-5}	1.6×10^{-8}	*SPI1*
rs58876856	11	66825559	C	T	0.09	0.96 (0.93-0.98)	4.3×10^{-5}	1.9×10^{-8}	*RHOD*
rs11066309	12	112883476	A	G	0.43	0.97 (0.96-0.98)	3.3×10^{-6}	1.3×10^{-8}	*PTPN11*
rs368962324	15	40947788	C	CAAAA	0.84	1.04 (1.02-1.05)	2.6×10^{-5}	8.8×10^{-12}	*CASC5*
rs1806845[a]	15	50681906	T	G	0.36	0.97 (0.96-0.98)	9.6×10^{-6}	5.4×10^{-9}	*GABPB1-AS1*
rs80326387	15	66705043	A	G	0.26	1.03 (1.01-1.04)	1.2×10^{-4}	2.7×10^{-11}	*MAP2K1*
rs73179015	22	43420441	A	G	0.07	0.95 (0.93-0.98)	1.0×10^{-4}	1.5×10^{-8}	*PACSIN2*
Prostate cancer (main analysis)									
rs56168262	1	51467096	C	CT	0.47	1.04 (1.03-1.06)	4.6×10^{-7}	1.2×10^{-12}	*CDKN2C*
rs3819772[a]	3	172225624	A	C	0.44	0.97 (0.95-0.98)	2.3×10^{-5}	1.0×10^{-8}	*TNFSF10*
rs571173399	5	132441799	G	T	0.24	0.96 (0.94-0.98)	1.1×10^{-4}	2.1×10^{-10}	*HSPA4*
rs10835920	11	32489664	T	C	0.36	0.97 (0.96-0.99)	7.1×10^{-4}	2.4×10^{-8}	*WT1-AS*
rs1806845[a]	15	50681906	T	G	0.36	0.97 (0.95-0.98)	1.4×10^{-4}	5.4×10^{-9}	*GABPB1-AS1*
Ovarian cancer (main analysis)									
rs3819772[a]	3	172225624	A	C	0.44	0.96 (0.93-0.98)	6.0×10^{-4}	1.0×10^{-8}	*TNFSF10*
rs138581118	6	151844694	A	G	0.06	0.90 (0.85-0.96)	7.6×10^{-4}	4.5×10^{-9}	*CCDC170*
rs13287517	9	19064129	C	G	0.38	1.05 (1.02-1.08)	3.4×10^{-4}	2.7×10^{-14}	*HAUS6*
HGSOC (subtype-focused analysis that included HGSOC and ER-status for breast cancer)									
rs35383942	1	201437832	T	C	0.06	0.88 (0.82-0.95)	5.3×10^{-4}	5.8×10^{-10}	*PHLDA3*
rs78378222	17	7571752	G	T	0.01	0.75 (0.64-0.88)	3.7×10^{-4}	1.6×10^{-15}	*TP53*
Endometrial cancer (main analysis)									
rs11240568	1	203793793	C	G	0.27	1.07 (1.04-1.11)	3.3×10^{-5}	1.2×10^{-10}	*ZC3H11A*
rs2811476	3	127898501	C	A	0.27	0.92 (0.89-0.95)	1.7×10^{-6}	8.8×10^{-32}	*EEFSEC*
rs9265885	6	31313220	C	G	0.52	0.94 (0.91-0.97)	6.2×10^{-5}	1.0×10^{-11}	*HLA-B*
rs7463114	8	29507094	C	T	0.68	0.95 (0.92-0.98)	9.1×10^{-4}	4.9×10^{-19}	*RPL17P33*
rs796945	10	90150837	T	C	0.64	0.95 (0.92-0.98)	8.1×10^{-4}	8.2×10^{-10}	*RNLS*

[a]> 1 Mb away from any previously reported lead SNP for one of the four cancers and single-cancer $P < 10^{-3}$ for that cancer. There are 23 new associations involving 21 lead SNPs: rs1806845 is a new association for two cancers (breast and prostate) and rs3819772 is a new association for two cancers (prostate and ovarian).

[b]Associations between these lead SNPs and the other cancer types included in this study are listed in Supplementary Tables 3 and 4.

Abbreviations: SNP, single nucleotide polymorphism; Chr, chromosome; Pos, build 37 position; EA, effect allele; OA, other allele; EAF, effect allele frequency; OR, odds ratio; CI, confidence interval; HGSOC, high-grade serous ovarian cancer; ER, estrogen receptor.

We identified seven associations for breast cancer (in addition to the eight reported above) that were new in the context of the breast cancer data set used in this study but have since been reported for breast cancer at $P < 5 \times 10^{-8}$ in the meta-analysis by Zhang *et al.*[13] described above (five of the seven associations; **Supplementary Table 5**) or in a trans-ancestry genome-wide association meta-analysis by Shu *et al.*[15] (two of the seven associations; **Supplementary Table 5**). Shu *et al.* combined the breast cancer data set used in this study with a data set that included over 24,000 breast cancer cases and 24,000 controls of Asian ancestry.

We also assessed the relationship between 461 lead SNPs that were within 1 Mb of one of the 463 lead SNPs previously published for breast, prostate, ovarian, and/or endometrial cancer risk and those

lead SNPs (**Methods**). We found that 187 of the 461 lead SNPs were independent ($r^2 < 0.05$) of the previously published proximal single-cancer lead SNPs (**Supplementary Table 6**). This included 8 of the 21 SNPs in **Table 2**: rs71475909, rs58876856, rs1806845, rs80326387, rs73179015, rs56168262, rs3819772, and rs9265885.

Bayesian re-evaluation of lead variants

We used the breast, prostate, ovarian, and endometrial cancer summary genetic association data to also perform meta-analysis within a Bayesian framework[16] for each lead SNP identified in the main and subtype-focused meta-analysis (**Methods**). This was motivated by two aims. First, to identify the subset or combination of cancers that was most likely to be responsible for the association at each lead SNP. Second, to quantify the combined evidence of association for this subset of cancers at the lead SNP in terms of the posterior probability of association (PPA), assuming a prior probability of association of 1 in 100,000. The Bayesian framework evaluated the evidence for 16 combinations at each lead SNP from the main meta-analysis (four non-pleiotropic combinations involving one cancer type only, six involving pairs and four involving triplets of cancers, one with all four cancers, and one where none of the cancers are associated) and 32 combinations at each lead SNP from the subtype-focused meta-analysis.

The Bayesian meta-analysis confirmed that the combination of cancers with greatest evidence for underpinning the association at each of the lead SNPs listed in **Table 2** included the cancer for which we report the association as novel (**Supplementary Table 7**). It also confirmed that for all but one of these lead SNPs, the PPA for this combination of cancers exceeded 80% (**Supplementary Table 7**). The only exception was rs58876856 with a PPA of 67% for the top-ranked combination that included contributions from breast, prostate, and endometrial cancer. Further, the top combination included at least three cancers for each of the four lead SNPs listed in **Table 1** and these combinations all achieved PPA > 80% (**Supplementary Table 7**). The top-ranked combination involved all the cancers included in this study for 18 lead SNPs in 16 regions (PPA > 80%; **Figure 1**). The top combination involved at least two cancers and had PPA > 80% for 222 of the 465 lead SNPs (in 118 of the 192 regions; **Supplementary Table 8** and **Methods**). The associations at the remaining lead SNPs were driven by a single cancer or the combination of ER-positive and ER-negative breast cancer (**Supplementary Table 8**).

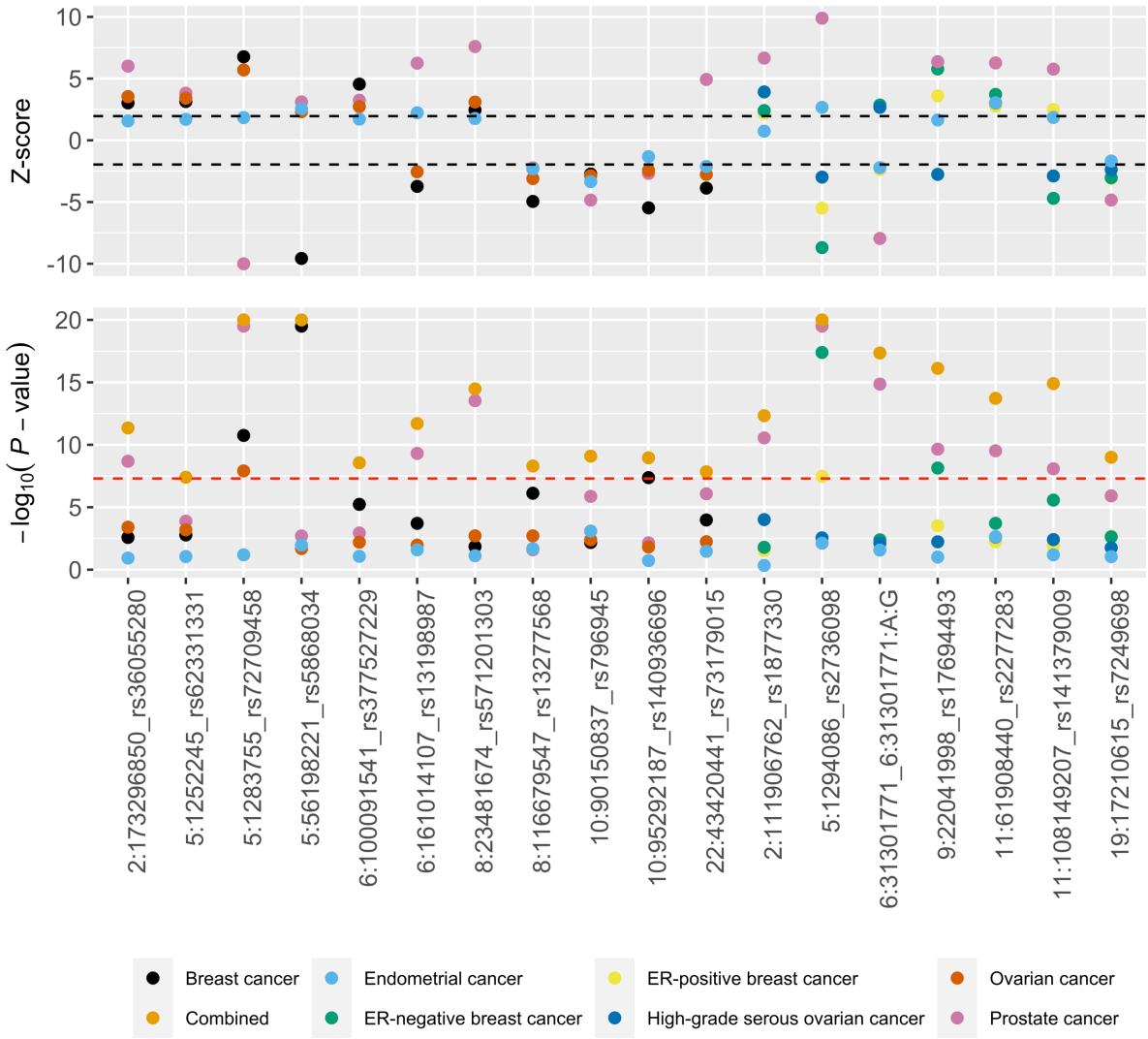

Fig. 1: Eighteen lead SNPs in 16 regions where the Bayesian framework indicated that there was evidence for association with all four cancers out of breast, prostate, ovarian, and endometrial cancer. These lead SNPs are in 16 distinct regions at least 1 Mb apart. Eleven of the 18 SNPs were identified in the main meta-analysis and seven in the subtype-focused meta-analysis. The black dashed lines indicate Z-scores of +/- 1.96 and the red dashed line indicates a *P*-value of 5 x 10^{-8}. Z-scores > 10 were set to 10 and -$\log_{10}$(*P*-values) > 20 were set to 20 to enable improved visualization of data points with less extreme values. "Combined" refers to -$\log_{10}$(*P*-values) from the Han and Eskin model-based main or subtype meta-analysis. Chromosomal positions are based on build 37.

Associations between lead variants and other diseases and traits

We investigated whether the 465 lead SNPs identified in this study or SNPs in strong linkage disequilibrium (LD; $r^2 > 0.8$) with them were associated with other diseases and traits at genome-wide significance ($P < 5 \times 10^{-8}$) in previously published studies (**Methods**)[17]. We found that 97/465 lead SNPs (or SNPs in strong LD) were associated with at least one trait or disease (other than breast, prostate, ovarian, and/or endometrial cancer; **Supplementary Table 9**). These associated traits or diseases were classified in terms of broader categories (specifically, Experimental Factor Ontologies or EFOs)[18]. The 97 SNPs were associated with 190 unique EFOs (**Supplementary Table 9**). The two EFOs associated with the largest number of lead SNPs were EFO_0004586 ("complete blood cell count"; 54/97 SNPs, including 42/222 SNPs that were associated with at least two cancers based on the Bayesian evaluation) and EFO_0004339 ("body height"; 10/97 SNPs). A recent (unpublished) study[19] has substantially increased the number of genomic loci associated with blood cell phenotypes and there are now 16,900 known genetic associations involving 29 blood cell traits. Having shown that ~12% of the lead SNPs from our study (or nearly a fifth of the lead SNPs with detectable cross-

cancer effects in the Bayesian evaluation) were associated with blood cell traits (EFO_0004586), we asked if the converse was true: were genome-wide significant SNPs from the recent blood cell trait GWAS also enriched for low *P*-values in the main four-cancer meta-analysis? We found that there was evidence for such enrichment (**Fig. 2a**), the enrichment was driven by all four cancers but particularly breast and prostate cancer (**Fig. 2a**), and the enrichment closely paralleled that seen at variants known to be associated with height (**Fig. 2b**)[20], another highly polygenic quantitative trait that ranked second in our EFO list. We found that this enrichment was not restricted to SNPs associated with a particular blood cell index class as classified by Vuckovic et al.[19] (**Fig. 2c**) or with the counts of a particular blood cell type (**Fig. 2d**).

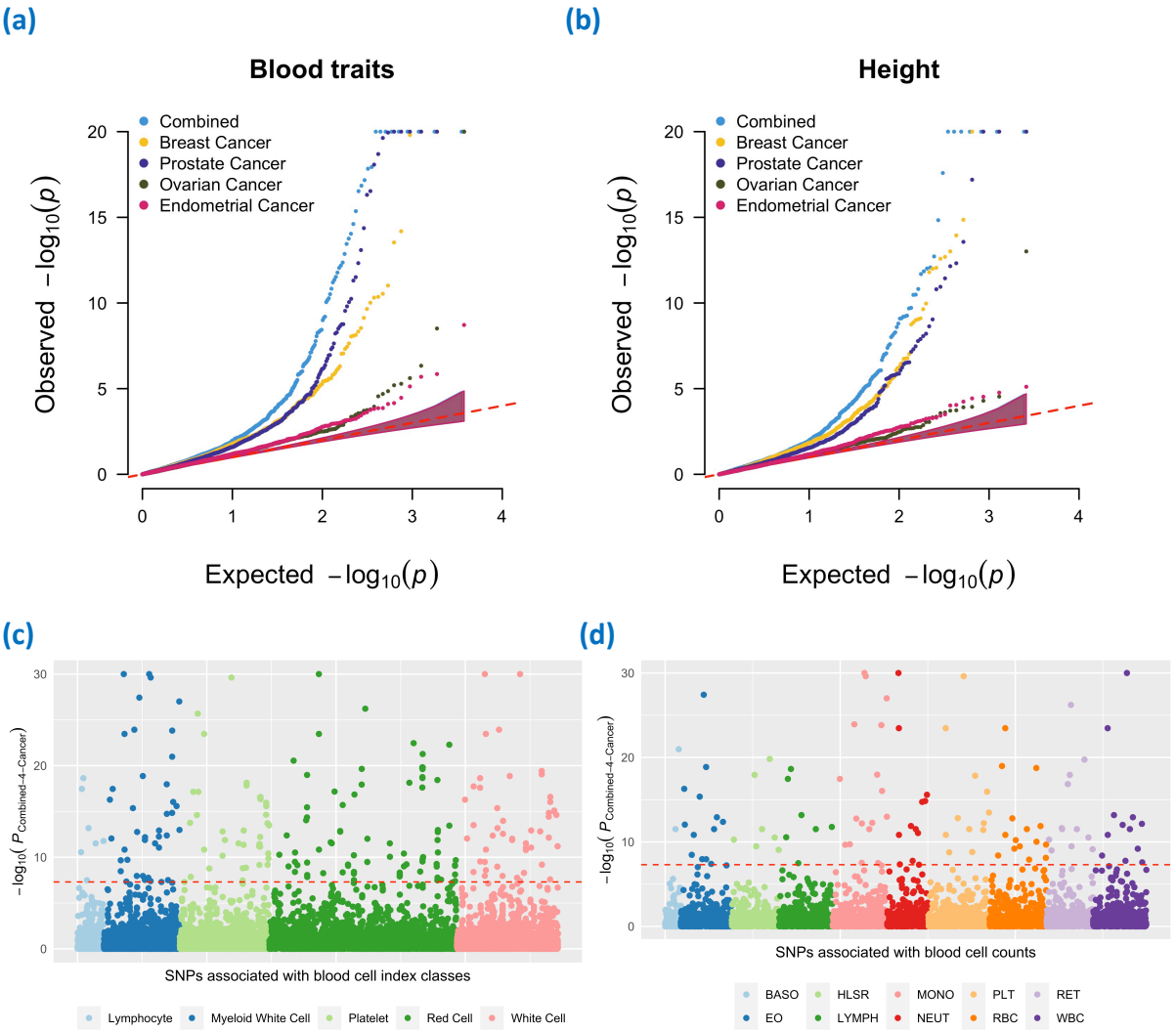

Fig. 2: Quantile-quantile (Q-Q) plots of negative logarithm (base 10) *P*-values from the main meta-analysis ("Combined") and the single-cancer data (breast, prostate, ovarian, and endometrial cancer) at independent variants ($r^2 < 0.05$) associated at genome-wide significance with (a) blood cell traits in Vuckovic et al.[19] and (b) height in Yengo et al[20]. Manhattan plot of negative logarithm (base 10) *P*-values from the main meta-analysis at variants associated at genome-wide significance in Vuckovic et al. with **(c)** blood cell index classes and **(d)** blood cell type-specific counts. Classes and counts are described in detail in Vuckovic et al. (for example, the Red Cell class includes red blood cell counts and other indices such as hemoglobin and hematocrit while the RBC count includes the red blood cell counts only). The red dashed line in the Manhattan plots indicates genome-wide significance ($P < 5 \times 10^{-8}$). Negative $\log_{10}$(*P*-values) > 20 were set to 20 in the Q-Q plots and > 30 were set to 30 in the Manhattan plots to enable improved visualization of data points with less extreme values. Abbreviations: BASO, basophil; HLSR, high light scatter reticulocyte; MONO, monocyte; PLT, platelet; RET, reticulocyte; EO, eosinophil; LYMPH, lymphocyte; NEUT, neutrophil; RBC, red blood cell; WBC, white blood cell.

Functional annotation of the new lead SNP-cancer associations

We annotated the lead SNPs (and SNPs in strong LD; $r^2 > 0.8$) marking the four loci not previously identified for any of the four cancers and the 23 new lead SNP-cancer associations using several genomics resources to identify gene targets with functional support (**Methods**)[21–28]. This included expression quantitative trait loci (eQTLs) identified at false discovery rate (FDR) < 0.05 by the Genotype-Tissue Expression (GTEx) consortium[22] in breast, prostate, ovary, and uterus (*cis*-eQTLs; **Supplementary Table 10**) and by the eQTLGen consortium[23] in blood (*cis*- and *trans*-eQTLs; **Supplementary Table 11**); and Combined Annotation Dependent Depletion (CADD) scores[24] (**Supplementary Table 12**). Among the protein coding genes nearest to the 27 lead SNPs, the GTEx database supported *COL23A1*, *HLA-B*, *HAUS6*, *RHOD*, and *GABPB1* as targets (**Supplementary Table 10**) while the eQTLGen database supported *COL23A1*, *HAUS6*, *RHOD*, *EPHB3*, *CCDC170*, *EEFSEC*, *HSPA4*, *MAP2K1*, *PACSIN2*, *RNLS*, *SPI1*, *TP53*, and *ZC3H11A* as targets (**Supplementary Table 11**). Lead SNP rs35383942 in an exon of *PHLDA3* had a CADD score of 24. This SNP is a missense variant (p.Arg28Gln) known to be associated with increased breast cancer risk[7] that we found associates with decreased ovarian cancer risk (OR = 0.88, 95% CI: 0.82—0.95; $P_{HGSOC} = 5.3 \times 10^{-4}$).

Direction of allelic association across pairs of cancers

Next, we examined the 441 genome-wide significant lead SNPs from the main analysis and the 393 genome-wide significant lead SNPs from the subtype-focused analysis to characterize the direction of allelic association across cancers with an emphasis on SNPs that were strongly associated with at least two cancers individually ($P < 10^{-3}$). We considered every possible pairwise combination of cancers in the main and subtype-focused analyses (except the pairing of ER-positive breast cancer with ER-negative breast cancer). We observed that for 94 of the 441 main lead SNPs (**Supplementary Table 3**) and 89 of the 393 subtype-focused lead SNPs (**Supplementary Table 4**), the lead SNP was associated with each cancer out of at least two cancers at $P < 10^{-3}$ in the corresponding single-cancer data set. We focused further analyses presented in this section of the results on these subsets of lead SNPs. For 29/94 main lead SNPs and 33/89 subtype lead SNPs, the allele that conferred risk of developing one of the two cancers had a protective association with the other cancer (**Supplementary Table 13**).

We explored the biological basis of these alleles that have opposite effects across cancers by analyzing over-representation[29] of pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG)[30] among the combined set of genes nearest to the lead SNPs that showed opposite effects across cancer pairs in the main and subtype-focused analysis. The strongest enrichment observed was for genes in the KEGG p53 signaling pathway (FDR = 4.4×10^{-5}; **Supplementary Table 14**). This p53 pathway enrichment was underpinned by five genes: *ATM*, *CASP8*, *CHEK2*, *MDM4*, and *TP53*. We visualized single-cancer associations at lead SNPs/alleles in or nearest to these five genes using a forest plot (**Fig. 3**). In contrast, we observed a distinct pattern of pathway over-representation (**Supplementary Table 14**) among the combined set of genes nearest to the lead SNPs that showed same direction associations (single-cancer $P < 10^{-3}$) across cancer pairs in the main analysis (65/94 lead SNPs; **Supplementary Tables 3 and 13**) and subtype-focused analysis (56/89 lead SNPs; **Supplementary Tables 4 and 13**). The KEGG apoptosis pathway was the highest-ranked pathway shared between the opposite and same direction analyses (**Supplementary Table 14**).

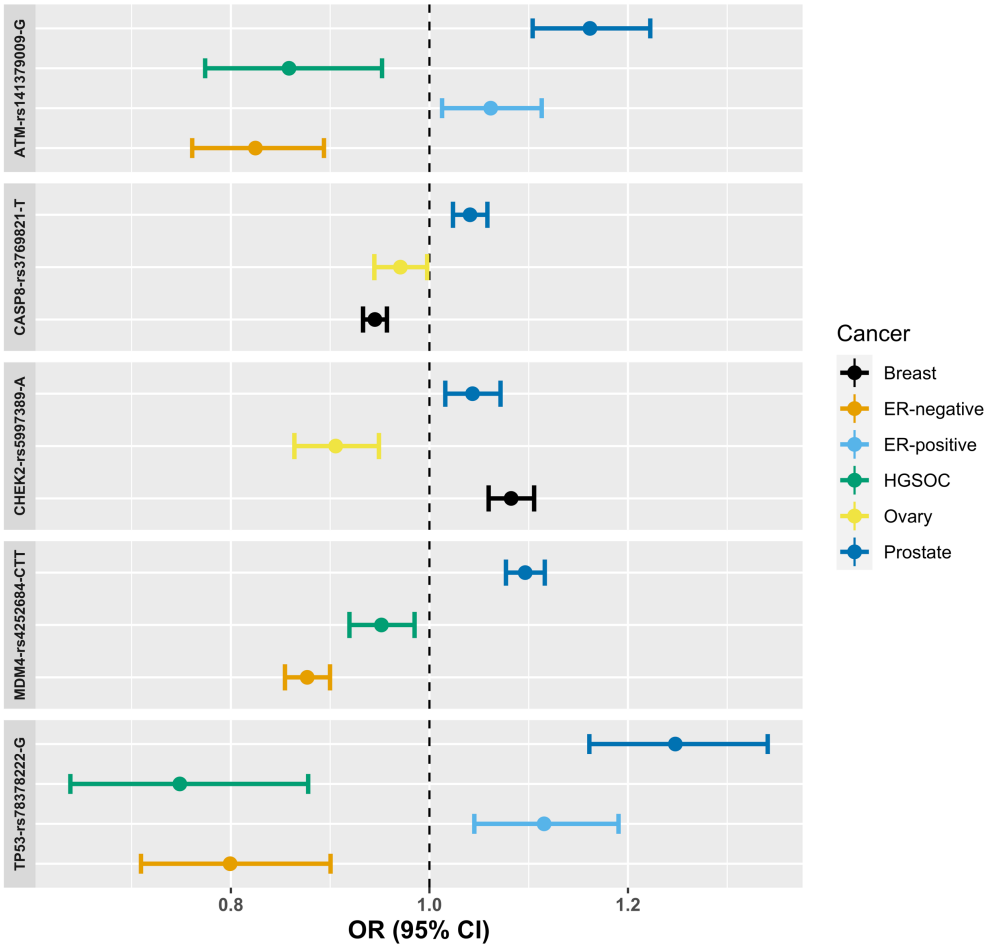

Fig. 3: Forest plots showing odds ratios (OR) and 95% confidence intervals (CI) for the lead SNPs/effect alleles in or nearest to the five genes in the TP53 signalling pathway that underlie the enrichment for this pathway. All cancer- or subtype-specific associations at $P < 0.05$ are shown in the forest plot. Abbreviations: ER, estrogen receptor; HGSOC, high-grade serous ovarian cancer.

Gene-level association analysis

We mapped all SNPs along with their association *P*-values from the main and subtype-focused genome-wide association meta-analyses to 19,100 protein-coding genes across the genome and performed gene-level association testing using Multi-marker Analysis of Genomic Annotation (MAGMA[31]). MAGMA collapses multiple SNP-level associations in a gene while taking into account LD between the SNPs and offers a gene-level association test that complements conventional single-SNP association analyses.

Before proceeding further, we evaluated the biological meaningfulness of the gene-level MAGMA associations using gene expression levels in 54 tissue types from the GTEx database[22]. Expression levels of genes in the breast, prostate, fallopian tube, and uterine tissue are predictors of the gene-level multi-cancer MAGMA association signals ($P < 1.7 \times 10^{-5}$; **Supplementary Tables 15 and 16**). This finding is consistent with these tissues being relevant to the etiology of the multi-cancer phenotype. A much weaker association was observed between gene expression in ovarian tissue and the gene-level MAGMA signal ($P = 0.03$). We also applied MAGMA to the breast, prostate, ovarian, and endometrial cancer genome-wide SNP association data sets individually. The corresponding single cancer gene-level signals were less strongly associated with tissue-specific gene expression levels than the multi-cancer gene-level signals for breast ($P_{single} = 1.7 \times 10^{-5}$ and $P_{multi} = 1.5 \times 10^{-7}$), fallopian tube ($P_{single} = 0.01$ and $P_{multi} = 1.3 \times 10^{-6}$), ovary ($P_{single} = 0.06$ and $P_{multi} = 0.04$), and uterus ($P_{single} = 0.03$ and $P_{multi} = 2.6 \times 10^{-6}$), but not in the prostate ($P_{single} = 3.9 \times 10^{-18}$ and $P_{multi} = 5.6 \times 10^{-15}$). This indicates that the gain in power obtained by combining data across cancers was, in general, improving

the association signal for genes more likely to be expressed in the relevant tissue types compared to single cancer analyses and suggests the presence of shared mechanisms underlying inherited cross-cancer risk acting across multiple anatomically separate tissue sites.

The multi-cancer gene-based association analysis identified 13 genes in as many regions (**Table 3**) at genome-wide significance ($P_{MAGMA} < 2.62 \times 10^{-6}$ after accounting for testing 19,100 genes) that were > 1 Mb away from any previously identified breast, prostate, ovarian, and/or endometrial cancer lead risk SNP and > 1 Mb away from any lead SNP identified in SNP-based analysis in this study.

Table 3 The 13 new cancer susceptibility loci identified by gene-based association (MAGMA) analysis.[a]

Gene	Chr	Start[b]	Stop[b]	Single-cancer P_{MAGMA} (if < 0.05)	4-cancer P_{MAGMA}
BSN	3	49591922	49708978	BrCa (2.5×10^{-5}), PrCa (4.3×10^{-7})	8.5×10^{-8}
SIK2	11	111473115	111601577	BrCa (1.3×10^{-6}), PrCa (8.1×10^{-6}), EnCa (7.2×10^{-3})	9.6×10^{-8}
INTS10	8	19674651	19709594	BrCa (2.1×10^{-7}), EnCa (3.3×10^{-3})	3.6×10^{-7}
CDKN1A	6	36644305	36655116	BrCa (2.0×10^{-7})	3.8×10^{-7}
UQCC1	20	33890369	33999944	PrCa (6.2×10^{-8}), EnCa (1.3×10^{-4})	5.5×10^{-7}
CLIC6	21	36041688	36090525	BrCa (2.2×10^{-4}), Prostate (1.1×10^{-3}), Ovarian (3.9×10^{-5})	5.9×10^{-7}
ATP5B	12	57031959	57039798	BrCa (3.6×10^{-3}), Prostate (1.7×10^{-3}), Ovarian (1.4×10^{-3})	7.8×10^{-7}
IKZF2	2	213864429	214017151	BrCa (3.7×10^{-2}), Prostate (1.6×10^{-8})	8.2×10^{-7}
PYGB	20	25228705	25278650	BrCa (1.4×10^{-3}), Prostate (5.9×10^{-5})	9.1×10^{-7}
TIAL1	10	121334199	121356541	BrCa (4.6×10^{-7})	1.7×10^{-6}
PRDM5	4	121606074	121844025	BrCa (6.6×10^{-3}), PrCa (1.8×10^{-4}), EnCa (2.3×10^{-2})	1.9×10^{-6}
LAMC1[c]	1	182992595	183114727	ER-pos BrCa (2.0×10^{-3}), PrCa (6.1×10^{-7}), EnCa (3.4×10^{-2})	7.8×10^{-7}
GLIS1[c]	1	53971910	54199877	ER-pos BrCa (8.3×10^{-5}), ER-neg BrCa (4.3×10^{-4}), PrCa (2.0×10^{-6})	1.1×10^{-6}

[a]> 1 Mb away from any previously reported breast, prostate, ovarian, and/or endometrial cancer lead risk SNP and > 1 Mb away from any lead risk SNP identified in this study.
[b]Build 37 position.
[c]Identified in the application of MAGMA to the subtype-focused meta-analysis results.
Abbreviations: MAGMA, Multi-marker Analysis of GenoMic Annotation; ER-pos BrCa, Estrogen Receptor-positive breast cancer; ER-neg BrCa, Estrogen Receptor-negative breast cancer; BrCa, breast cancer; PrCa, prostate cancer; OvCa, ovarian cancer; EnCa, endometrial cancer.

Pathway analysis based on gene-level associations

We performed pathway analysis on the set of genes with gene-level association $P_{MAGMA} < 2.6 \times 10^{-6}$ from the main meta-analysis or from the subtype-focussed meta-analysis and in the top 5% of the ranked gene-level associations for at least two cancers from the single-cancer analyses. Three hundred and thirteen genes met these criteria from which we excluded five HLA genes (**Supplementary Tables 17**). The final set of 308 genes was enriched for twelve pathways from KEGG at FDR < 0.05 with over five-fold enrichment in two pathways: p53 signalling and endocrine resistance (**Supplementary Tables 18**). The enrichment in the p53 signalling pathway was driven by *CCND1*, *CDKN2A*, *SIVA1*, *CCNE1*, *CASP8*, *CHEK2*, and *MDM4* and was more modest than the enrichment for the same pathway observed in the 37 genes nearest to the set of SNPs with opposite effects on at least two cancers, which was driven by *TP53*, *ATM*, *CASP8*, *CHEK2*, and *MDM4*. Performing the pathway analysis retaining the five *HLA* genes (i.e., using all 313 genes) implicated several immune-related pathways where the signal was largely driven by these genes (**Supplementary Tables 19**).

DISCUSSION

We performed the largest genome-wide association meta-analysis for shared susceptibility to four hormone-related cancers – breast, prostate, ovarian, and endometrial cancer – to date. This identified four new risk loci that were over a megabase away from previously identified risk loci for the four cancers and 21 new risk loci incorporating 23 new lead SNP-cancer type associations that were over a megabase away from previously identified risk loci for at least one of the cancers. Gene-based association analysis identified an additional 13 novel risk loci. Beyond locus discovery, we found evidence for pleiotropic cross-cancer effects at ~48% (222/465) of the genome-wide significant lead SNPs in our study. The strongest association model at these SNPs included contributions from at least two of the four cancers. We identified several risk loci where the lead SNP had strong associations with at least two of the cancers but opposite effects (for the same effect allele) on these cancers. We demonstrated enrichment of the multi-cancer association signal in breast, prostate, fallopian tube, and uterine tissues, conducted a phenome-wide association scan of cross-cancer lead SNPs revealing profound overlap with blood cell trait loci, and observed enrichment of pivotal pan-cancer pathways in our combined data set.

The 8q24.1 region, a "gene desert" with a paucity of protein coding genes other than *MYC* was one of the first regions to be recognized as one that contains SNPs associated with multiple cancer types[32]. Here we report a novel cancer susceptibility locus in a similar gene desert, 2p24.3 with lead SNP rs7586503 and nearest coding gene *MYCN*, which shows evidence for association with breast, prostate, and endometrial cancers (**Table 1**). The proto-oncogene *MYCN* is known for its ability to drive the development of multiple tumor types that arise from a common cellular lineage[33]. The lead SNP rs5819638 in the 17p11.2 region is an intronic insertion and deletion variant in *TOM1L2* (**Table 1**). This SNP is correlated (r^2 = 0.62) with rs8070624, a known genome-wide significant lead SNP for bone mineral density (BMD)[34], an estrogen-regulated trait in both men and women[35]. The T allele deletion of rs5819638 is protective for breast, prostate, and endometrial cancer. The deletion corresponds to the G allele of rs8070624 that is associated with reduced BMD, which is consistent with the observed relationship between lower BMD and decreased breast, prostate, and endometrial cancer risks[36–38]. The A allele of the new breast cancer risk lead SNP rs71475909 increases breast cancer risk but decreases prostate cancer risk (**Table 2** and **Supplementary Table 3**). The SNP lies in an intron of *SPI1*, which encodes a transcription factor that co-binds with TP53 to TP53 target genes and is crucial for the maintenance of pro-apoptotic transcriptional repression by TP53[39,40]. Expression QTL analyses in breast, prostate, ovarian, uterine, and whole blood tissue types supported several genes as candidate targets at the new risk loci. Lead SNP rs58876856 was associated with breast ($P = 4.3 \times 10^{-5}$; **Table 2**) and prostate ($P = 4.9 \times 10^{-4}$; **Supplementary Table 3**) cancer risk and with *RHOD* expression in breast ($P = 6.4 \times 10^{-25}$) and prostate ($P = 1.3 \times 10^{-16}$) tissues. *RHOD* encodes a member of the Rac subfamily of the Rho family of GTPases that has an established oncogenic role[41]. We also identified a novel association for HGSOC risk at rs35383942, a known genome-wide significant lead SNP for breast cancer risk[7] and male-pattern baldness[42]. The T allele of rs35383942 is associated with increased susceptibility to breast cancer but decreased susceptibility to HGSOC. SNP rs35383942 is a missense variant (CADD score 24; **Supplementary Table 12**) in *PHLDA3*, a tumor supressor gene that sits at the intersection of two key cancer pathways: PHLDA3 is directly regulated by TP53 and is a repressor of AKT1 signaling, contributing to TP53-dependent apoptosis[43]. The new endometrial cancer risk SNP rs2811476 (**Table 2**; at a known prostate cancer risk locus – **Supplementary Table 3**), in an intron of *EEFSEC*, was strongly correlated with rs2955117 (r^2 = 0.95) and rs2687729 (r^2 = 0.99), which are genome-wide significant lead SNPs for the hormone-related traits maternal effect on gestational age at birth[44] and age at menarche[45], respectively. The alleles associated with longer gestational duration and later menarche were protective for endometrial cancer and these findings are consistent with the "unopposed oestrogen hypothesis" for endometrial cancer[46]. The gene-based MAGMA approach identified 13 new candidate cancer susceptibility genes in 13 loci not detected by SNP-level single- and cross-cancer genome-wide association meta-analyses. Each of these genes is discussed in detail in **Supplementary Note 1**.

We identified 18 lead SNPs in 16 regions where a Bayesian evaluation of the data suggested that the combination of cancers with greatest evidence for association involved all four cancers. Some of these lead SNPs marked known multi-cancer susceptibility loci near *TERT*[47], CDKN2B-AS1[48], *INCENP*[6], and *BCL2L11*[6]. Other regions (out of the 16) have not been recognized before for such extensive cross-cancer pleiotropic effects. For example, lead SNP rs5868034 is located ~7kb from *SETD9* which encodes a lysine methyltransferase that methylates TP53, increasing its stability and promoting the expression of its targets *CDKN1A*, *BAX*, and *MDM2*[49]. While rs5868034 is ~144kb from rs62355902, a known breast cancer risk lead SNP that regulates *MAP3K1* in the same region[50], the two SNPs are independent ($r^2 = 0.03$). Further, a GTEx look-up for rs5868034 shows that it is associated with the expression of *SETD9* in breast ($P = 1.9 \times 10^{-32}$), prostate ($P = 8.5 \times 10^{-18}$), ovarian ($P = 9.6 \times 10^{-12}$), and uterine ($P = 1.6 \times 10^{-11}$) tissues[22]. Lead SNP rs140936696 is ~3kb from *CEP55*, which encodes a centrosome- and midbody-associated protein that is a major regulator of abscission or the final stage of cytokinesis[51] and of the PI3K/AKT pathway[52]. *CEP55* is a member of several common gene expression signatures (such as the "CIN70") that serve as indices of chromosomal instability, cell cycle progression, proliferation, and metastatic potential across multiple cancer types[53]. Lead SNP rs796945 near *RNLS* is associated with breast, prostate, ovarian, and endometrial cancers and has been recently identified as a lead SNP for the hormone-related gynecologic disorder endometriosis[54], all with the same direction of allelic effect.

We observed a substantial overlap between the cross-cancer genetic associations from this study and variants known to be associated with a range of blood cell phenotypes. Some loci associated with blood cell traits, in particular blood cell counts, likely represent cell-type agnostic biomarkers of cellular turnover and proliferative potential – biological processes that, in turn, may link them to solid tumor risk. Indeed, common genetic variants at *SH2B3*, *ATM*, *TERT*, *TET2*, and *CHEK2* are associated with multiple blood cell traits[55] and with myeloproliferative neoplasms[56,57], and were the genes nearest to five cross-cancer lead SNPs identified in the current study. For example, our lead SNP rs7310615 that maps to *SH2B3* ($P_{breast} = 3 \times 10^{-7}$ and $P_{endometrial} = 1.3 \times 10^{-10}$) is a genome-wide significant lead SNP associated with white blood cell, red blood cell and platelet counts and other blood cell traits[55], and with myeloproliferative neoplasms[56,57]. Rare germline missense variants in *SH2B3* and *ATM* are also associated with the blood cell (somatic) acquisition of loss-of-heterozygosity (LOH) in the chromosomal arms where these genes reside, which may result in the aberrant clonal expansion of the blood cells acquiring these LOH events or clonal hematopoeisis[58]. Genes nearest to cross-cancer risk lead SNPs in this study – *BCL2L11*, *SMC2*, *CTSK*, *USP28*, *CDKN1B*, *MAP2K1*, *TP53*, *ATM*, and *CHEK2* – also mark loci predisposing to mosaic Y chromosome loss in blood, the most common form of clonal hematopoeisis[59]. Overall, these genetic overlaps with blood cell traits add to a growing body of evidence which suggests that clonal hematopoiesis may be a proxy for genomic instability in the body that is associated with non-hematological cancer risk[59].

Pathway analysis of our 2016 breast-prostate-ovarian cancer shared susceptibility GWAS meta-analysis findings had identified a central role for induction of apoptosis through death receptor signaling[6]. In our current substantially expanded data set that also includes endometrial cancer we not only continue to detect a signal for the apoptosis pathway but also highlight roles for endocrine resistance and TP53 signaling. We specifically note a pattern involving cross-cancer susceptibility alleles at the TP53 signaling genes *CASP8*, *MDM4*, *CHEK2*, *ATM* and *TP53* wherein, in general, lead SNP alleles conferring risk of ovarian cancer/HGSOC and ER-negative breast cancer show protective associations with overall/ER-positive breast cancer and prostate cancer. This pattern likely reflects the differential prevalence of loss of tumor suppressor function versus dominant negative or gain-of-function (GOF) oncogenic *TP53* somatic mutations that is seen in these tumor types[60,61]. For example, the G allele of lead SNP rs78378222 is associated with lower expression of *TP53* ($P = 6 \times 10^{-21}$ in eQTLGen[23]; **Supplementary Table 11**). This SNP is known to alter the polyadenylation signal and impair the 3′-end processing of *TP53* mRNA[62]. The G allele is protective for HGSOC and ER-negative breast cancer (**Fig. 3**), where dominant negative or GOF *TP53* somatic mutations are more common, potentially because the lower germline-regulated expression attenuates the oncogenic effects of TP53 when such a mutation is somatically acquired. However, the same allele confers

susceptibility to ER-positive breast and prostate cancer (**Fig. 3**), where loss-of-function (LOF) *TP53* somatic mutations predominate, potentially because the lower background expression accentuates the loss of TP53 tumor suppressor function when the LOF mutation is somatically acquired. Thus, our findings hint at a potential germline-somatic interaction that strongly warrants further investigation. Antagonistic effects across cancer types at the same allele also has important implications for the use of such alleles in the development and application of polygenic risk scores that have so far been confined to a consideration of allelic associations with a single cancer. The identification of alleles with opposite effects across the cancers was helped by that fact that the Han and Eskin model[11,12] used in this study can detect effects under heterogeneity in contrast to the standard fixed-effects model used in the 2016 meta-analysis[6]. Opposite associations across related diseases at the same allele are also seen in autoimmune[63] and psychiatric[64] disease groups. While a direct comparison between cancer, autoimmune, and psychiatric studies is challenging due to differences in sample size (the cancer data set here is larger by case numbers) and design, the prevalence of such associations in cancer appears to be higher.

It is worth noting that 187 of the 465 lead SNPs highlighted in this study were within 1 Mb of known breast, prostate, ovarian, and/or endometrial cancer risk loci but not linked ($r^2 < 0.05$) to the published lead SNPs at these loci. This finding is consistent with a recent combined GWAS of multiple blood lipid traits[65] that identified several independent (in terms of LD) associations at known lipid-associated loci. These associations were not reported in the corresponding single lipid GWAS. This suggests that for related traits (groups of cancers or groups of lipids), the multi-trait genetic architecture may, at least in part, be distinct from the single-trait genetic architecture. That is, the same genomic regions may harbor distinct SNPs that have disease-specific and cross-disorder associations. This is best illustrated with an example: rs58058861 is a known genome-wide significant lead SNP for breast cancer risk[7] and our meta-analysis identifies an independent ($r^2 = 0.01$; distance between SNPs = 60kb) lead SNP rs3819772 in the same region and ~16kb from *TNFSF10* that is associated with breast cancer risk albeit less strongly ($P = 2.8 \times 10^{-4}$: **Supplementary Table 3**). However, rs3819772 from our meta-analysis is pleiotropic (**Table 2**) and also associates with prostate ($P = 2.3 \times 10^{-5}$) and ovarian ($P = 6 \times 10^{-4}$) cancer risk. *TNFSF10* encodes TRAIL, a cytokine of the tumor necrosis factor ligand superfamily that is expressed in most normal tissues and selectively induces apoptosis in transformed and tumor cells by binding to death receptors[66]. Subsequent studies should aim to validate the complex multi-trait genetic architecture in such regions by orthogonal approaches such as evaluating SNP interactions with tissue-specific and cross-tissue gene regulatory mechanisms.

In our investigation of the relationship between gene-level associations and tissue-specific gene expression we found that tissue-specific gene expression in the ovary and fallopian tube are similarly and weakly associated with gene-level associations based solely on the ovarian cancer data set used here. However, gene-level associations derived from the four-cancer meta-analysis are much more strongly associated with tissue-specific gene expression in the fallopian tube than in the ovary. This result is consistent with emerging evidence that the cell of origin of HGSOCs, the most common and aggressive histotype of ovarian cancer, is in the fallopian tube[67,68], and suggests that combining data from other cancers improves the ranking of genes associated with ovarian cancer risk that are also fallopian tube-specific. We identified nine lead SNP-breast cancer associations in our study that were new in the context of the breast cancer data set used here but have recently been identified in two larger breast cancer GWAS[13,15]. These breast and ovarian cancer findings, in particular, demonstrate that cross-cancer GWAS meta-analysis can be a powerful approach to the identification of new cancer susceptibility loci. The approach can be comparable, for pleiotropic regions, to the addition of cases and controls from the same cancer type and boost the discovery of gene-level signals that are expressed in the most relevant tissues. This is of particular relevance for elucidating the genetic architecture of rarer cancer types where larger GWAS studies may not be feasible.

In summary, this large-scale meta-analysis combining GWAS data for breast, prostate, ovarian, and endometrial cancers identified four shared susceptibility loci not previously reported for any of the four cancer types and 23 lead SNP-cancer associations in 21 shared susceptibility loci that were new

for at least one of the four cancer types. Examining these data under a gene-level lens demonstrated that multi-cancer gene-level associations were strongly enriched in the relevant tissues and yielded an additional 13 loci containing candidate breast, prostate, ovarian, and/or endometrial cancer susceptibility genes. Bayesian analysis generated the first comprehensive genome-wide catalog of lead SNPs where the association was likely to be driven by at least two of these cancers. We obtained fresh insights into the biology of the shared susceptibility loci, particularly those alleles that displayed opposite associations across cancers. Potential target genes and other phenotypic associations at the shared susceptibility loci suggest a complex interplay between inherited hormone-related cancer risk and major molecular mechanisms acting across cancer sites. These cross-cancer risk loci provide a rich substrate for the future development of laboratory-based functional studies imperative for the translation of discoveries from this GWAS into clinical and public health impact across cancer types.

METHODS

GWAS data, data harmonization, and meta-analysis

SNP genotype and sample quality control, ancestry inference, imputation, genome-wide association and meta-analysis procedures for the breast, prostate, ovarian, and endometrial cancer GWAS meta-analysis data sets have been described previously[7–10]. All analyses were based on individuals of European ancestry and used 1000 Genomes Phase 3 (Version 5)-imputed or genotyped SNPs[69]. We harmonized effect and non-effect alleles and effect size estimate (beta coefficient) signs across the GWAS meta-analysis summary statistics data sets included in this study. We focused all analyses on 9,530,997 variants with minor allele frequency > 1% and imputation quality > 0.4 in each data set.

Meta-analysis was carried out using the Han and Eskin[11] "RE2"model that has been shown to offer greater power to detect associations with heterogenous effects across traits in the context of GWAS as compared to the conventional DerSimonian and Laird[70] random-effects model which was originally developed for the meta-analysis of clinical trials. The "2" in "RE2" refers to the fact that the RE2 test statistic has two parts, a part to detect heterogeneity (random effects) and a mean effect part that is equal to the square of the fixed-effects meta-analysis test statistic[11]. The latter enables RE2 to detect associations where the effect size estimates are homogeneous across studies. We specifically used a recently developed extension of the RE2 model that was able to take into account correlation between GWAS summary statistics due to overlapping controls. This extension is referred to as "RE2C*" and was implemented using the RE2C version 1.04 software[12].

Correlation between GWAS summary statistics due to overlapping controls was estimated using the tetrachoric correlation between binary-transformed GWAS summary z-scores ($z \geq 0$ classified as 1 and $z < 0$ classified as 0)[71,72]. The tetrachoric correlation was computed using the Digby approximation (implemented in Stata 14.0) as recommended by Southam et al[71]. Southam et al. have previously demonstrated the accuracy of the tetrachoric correlation for estimating the correlation between GWAS summary statistics due to overlapping controls[71]. RE2C* uses the correlation between GWAS summary statistics due to overlapping controls to inflate the standard errors of, and "decouple"[73], the GWAS summary statistics included in a meta-analysis. Thus, the effective sample size of the RE2C* meta-analysis is the sample size obtained by simply adding up all the contributing GWAS data set sample sizes. The RE2C* model does not report a combined effect size estimate and confidence interval and only provides a *P*-value.

Independent lead SNPs, genomic regions, and known cancer susceptibility loci

The random-effects meta-analysis results were processed using the FUMA integrative platform[21]. To define independent lead SNPs, the relevant FUMA SNP2GENE tool parameters were set as (1) Minimum P-value of lead SNPs $< 5 \times 10^{-8}$, (2) r^2 threshold to define lead SNPs ≥ 0.05, (3) Reference panel population of 1000G Phase3 EUR, and (4) maximum distance between independent lead SNPs to merge into a region set as < 1000kb (i.e., lead SNPs closer than this distance were merged into a single genomic region). We also used FUMA SNP2GENE to assess the relationship between 463

lead SNPs previously reported for breast, prostate, ovarian, and/or endometrial cancer risk and the 461 lead SNPs within 1 Mb of these SNPs identified in the main and subtype-focused analyses (same parameters used as above except the maximum distance parameter that was set to 0kb). Only variants that had either been genotyped or imputed with a quality score > 0.8 were used to define lead SNPs. A comprehensive list of previously reported cancer susceptibility loci for breast, prostate, ovarian, and endometrial cancers (including susceptibility loci known to be shared by at least two of these cancers) was drawn up based on lead SNPs reported in 11 publications[6–10,13,15,74,75,48].

Bayesian meta-analysis

The exhaustive subset search function in the MetABF R package[16] was used to perform the Bayesian re-evaluation of the combined association at each lead SNP identified in the main- and subtype-focused Han and Eskin RE2C* model-based meta-analyses. The exhaustive subset search function was used to evaluate the evidence for all 16 possible combinations at each lead SNP from the main meta-analysis (four non-pleiotropic combinations involving one cancer type only, six involving pairs and four involving triplets of cancers, one with all four cancers, and one where none of the cancers are associated) and, similarly, all 32 possible combinations at each lead SNP from the subtype-focused meta-analysis. MetABF, like RE2C*, takes into account the correlation in GWAS summary statistics due to overlapping controls. This is referred to as "cryptic correlation" in the MetABF tool and the same tetrachoric correlation matrix applied to the RE2C* meta-analysis was used for the MetABF meta-analysis. Further, MetABF allows the specification of a prior expected correlation. This was set to 0 to parallel the ability of RE2C* to detect heterogeneous or random effects and set to 1 to parallel the ability of RE2C* to detect homogeneous or fixed effects and the results were averaged for the two prior correlations specified. MetABF also allows for the specification of a prior expected upper bound on the effect size estimate (odds ratio) for each study (in this case, each cancer type-specific GWAS). For example, if this prior odds ratio is set to 1.5, it reflects a belief that the prior probability that the odds ratio is larger than 1.5 is 2.5%. We used each cancer-specific summary genetic association data set to determine this prior by filtering the data set to retain only the SNPs achieving $P < 5 \times 10^{-8}$ and calculating the 97.5th percentile of the odds ratios. Thus, this reflected the odds ratios for SNPs associated with each cancer detectable given the sample size available for the cancer. The approximate Bayes factors (ABFs) calculated by MetABF were converted into posterior probabilities of association (PPAs) using the formula, $PPA = (ABF \times 10^{-5})/(ABF \times 10^{-5} + 1)$, assuming a prior probability of association at each SNP of 1 in 100,000 or 10^{-5}.

Associations between lead SNPs and other diseases and traits

The PhenoScanner version 2 bulk search tool (search date 29 May 2020) was used with query set to SNP, catalogue to diseases and traits, p-value to 5E-8, proxies to EUR, r^2 to 0.8, and build to 37. Only associations reported in publications with a PubMed identifier (PMID) were retained in the output.

Lists of genetic associations with blood cell traits and height were obtained: the Yengo et al.[20] Genetic Investigation of Anthropometric Traits (GIANT) consortium file "Meta-analysis Wood et al + UKBiobank 2018 top 3290 Height SNPs from COJO analysis GZIP" was downloaded (link in the URLs section) for height and *Supplementary Table 3* from Vuckovic et al.[19] was downloaded for blood cell traits. The FUMA pipeline (with settings as described above) was used to define independent (in terms of LD) SNPs from these genetic association files. Association *P*-values from the main four-cancer and each of the cancer-specific meta-analyses at the resultant SNP positions and CMplot R package (version 3.6.0) were used to generate the quantile-quantile plots in **Fig. 2**. Blood cell trait-associated SNP position and blood cell index class and count information from *Supplementary Table 3* from Vuckovic et al.[19] and association *P*-values from the main four-cancer meta-analysis were used to generate the Manhattan plots in **Fig. 2**.

Pathway analyses, functional and nearest gene annotation

All pathway analyses were conducted using the over-representation analysis function in WebGestalt (web-based gene set analysis toolkit)[29] and pathways from the Kyoto Encyclopedia of Genes and Genomes release 88.2, 11/01/2018[30]. Functional annotation and nearest gene identification were also undertaken using the FUMA pipeline and included expression quantitative trait loci (eQTLs) identified at false discovery rate (FDR) < 0.05 by the Genotype-Tissue Expression (GTEx) consortium (version 8 data set; *cis*-eQTLs)[22] in breast (n = 396), prostate (n = 221), ovary (n = 167), and uterus (n = 129; uterine but not specifically endometrial tissue) and by the eQTLGen consortium[23] in blood (*cis*- and *trans*-eQTLs; n = 31,684), Combined Annotation Dependent Depletion (CADD) scores[24], RegulomeDB scores[25], ANNOVAR categories[26], and ChromHMM states[27,28]. LDlink was used for all r^2 calculations presented in the Discussion section and these were based on 1000 Genomes Project Phase 3 (Version 5) European ancestry populations[76].

Gene-level association analysis and tissue-specific enrichment

Gene-level association analysis was performed using the MAGMA (multi-marker analysis of genomic annotation)[31] tool implemented in FUMA that takes into account multiple SNP-level summary genetic associations and LD between SNPs. SNPs were mapped to a gene if they were located between the start and end sites of the gene based on build 37 SNP and gene positions.

MAGMA "gene-property analysis" was performed to test for associations between gene-level signals and tissue-specific gene expression profiles. Gene-property analysis uses a multi-variable regression model that includes gene expression in a specific tissue type and the average gene expression across 54 tissue types to evaluate the relationship between tissue specificity and gene-level association. Tissue specific gene expression data were from GTEx version 8. GTEx eQTL results are only available for tissues with > 70 matched germline genotype-gene expression samples and the fallopian tube tissue samples in GTEx did not reach this sample size threshold. Therefore, fallopian tube was not included in the eQTL annotation but was available for the MAGMA gene-property analysis.

URLs

Links to download analytic tools and data sets used in the study:

RE2C: http://software.buhmhan.com/RE2C/index.php

FUMA and MAGMA: https://fuma.ctglab.nl/

MetABF: https://github.com/trochet/metabf

PhenoScanner: http://www.phenoscanner.medschl.cam.ac.uk/

GIANT consortium data:
https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files

WebGestalt: http://www.webgestalt.org/

Breast cancer GWAS data set: http://bcac.ccge.medschl.cam.ac.uk/bcacdata/oncoarray/gwas-icogs-and-oncoarray-summary-results/

Ovarian cancer GWAS data set (search for PubMed ID 28346442):
https://www.ebi.ac.uk/gwas/downloads/summary-statistics

Prostate cancer GWAS data set: http://practical.icr.ac.uk/blog/?page_id=8164

CONTRIBUTIONS

Study conception and design, statistical analysis, data interpretation, writing of first draft, guarantor: S.P.K.; Study conception, data interpretation, writing, and/or provided genetic data/samples: S.L., R.J.Hu., P.Kr., and P.D.P.P.; Data interpretation, writing, and/or provided genetic data/samples: K.L., M.K.S., T.A.O., D.M.G., J.P.T., J.M.S., J. C.-C., D.J.T., W.Z., I.P.M.T., A.B., S.J.R., S.J.C., D.F.E., G.C.-T., S.A.G., A.B.S., and R.A.E.; Provided samples and/or genetic data: A.G.M.A., F.M.A., A. B.-F., L.B., C.B., H.B., S.B., D.C., M.C., Z.C., M.B.D., A.duB., A.B.E., A.E., P.F., J.M.F., J.G., G.G.G., R.J.Ha., H.R.H., F.H., M.H., P.H., R.-Y.H., L.I., A.I., An.J. Al.J., E.M.J., P.Ka., B.Y.K., E.K., L.A.K., S.K.K., R.K., P.Kl., M.K., B.K., C.K., D.L., A.L., T.M., A.D.M., A.N.M., K.Mo., K.Mu., S.F.N., K.O., H.O., T.-W.P.-S., J.B.P., P.P., A.P., P.R., H.A.R., I.B.R., I.K.R.,V.W.S., N.S., W.S., B.Ś., L.M. S., C.L.T., L.J.T., C.T., N.U., A.M.vanA., A.V.-G., I.V., R.A.V., J.V., S.J.W., R.W., H.Y., and the PRACTICAL consortium, CRUK, BPC3, CAPS, PEGASUS; Critical review and approval of manuscript: All authors.

COMPETING INTERESTS

The authors declare no competing interests.

ACKNOWLEDGEMENTS

For the breast cancer GWAS data:

The breast cancer genome-wide association analyses were supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the 'Ministère de l'Économie, de la Science et de l'Innovation du Québec' through Genome Québec and grant PSR-SIIRI-701, The National Institutes of Health (U19 CA148065, X01HG007492), Cancer Research UK (C1287/A10118, C1287/A16563, C1287/A10710) and The European Union (HEALTH-F2-2009-223175 and H2020 633784 and 634935). All studies and funders are listed in Michailidou et al (2017).

For the prostate cancer GWAS data:

The Prostate cancer genome-wide association analyses are supported by the Canadian Institutes of Health Research, European Commission's Seventh Framework Programme grant agreement n° 223175 (HEALTH-F2-2009-223175), Cancer Research UK Grants C5047/A7357, C1287/A10118, C1287/A16563, C5047/A3354, C5047/A10692, C16913/A6135, and The National Institute of Health (NIH) Cancer Post-Cancer GWAS initiative grant: No. 1 U19 CA 148537-01 (the GAME-ON initiative).

We would also like to thank the following for funding support: The Institute of Cancer Research and The Everyman Campaign, The Prostate Cancer Research Foundation, Prostate Research Campaign UK (now PCUK), The Orchid Cancer Appeal, Rosetrees Trust, The National Cancer Research Network UK, The National Cancer Research Institute (NCRI) UK. We are grateful for support of NIHR funding to the NIHR Biomedical Research Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust.

The Prostate Cancer Program of Cancer Council Victoria also acknowledge grant support from The National Health and Medical Research Council, Australia (126402, 209057, 251533, , 396414, 450104, 504700, 504702, 504715, 623204, 940394, 614296,), VicHealth, Cancer Council Victoria, The Prostate Cancer Foundation of Australia, The Whitten Foundation, PricewaterhouseCoopers, and Tattersall's. EAO, DMK, and EMK acknowledge the Intramural Program of the National Human Genome Research Institute for their support. We also acknowledge Prostate Cancer Canada and funding by the Movember Foundation – Grant # D2013-36.

Genotyping of the OncoArray was funded by the US National Institutes of Health (NIH) [U19 CA 148537 for Elucidating Loci Involved in Prostate cancer SuscEptibility (ELLIPSE) project and X01HG007492 to the Center for Inherited Disease Research (CIDR) under contract number HHSN268201200008I] and by Cancer Research UK grant A8197/A16565. Additional analytic support was provided by NIH NCI U01 CA188392 (PI: Schumacher).

Funding for the iCOGS infrastructure came from: the European Community's Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A 10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692,

C8197/A16565), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 – the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund.

The BPC3 was supported by the U.S. National Institutes of Health, National Cancer Institute (cooperative agreements U01-CA98233 to D.J.H., U01-CA98710 to S.M.G., U01-CA98216 toE.R., and U01-CA98758 to B.E.H., and Intramural Research Program of NIH/National Cancer Institute, Division of Cancer Epidemiology and Genetics).

CAPS GWAS study was supported by the Swedish Cancer Foundation (grant no 09-0677, 11-484, 12-823), the Cancer Risk Prediction Center (CrisP; www.crispcenter.org), a Linneus Centre (Contract ID 70867902) financed by the Swedish Research Council, Swedish Research Council (grant no K2010-70X-20430-04-3, 2014-2269)

PEGASUS was supported by the Intramural Research Program, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health.

The coordination of EPIC was financially supported by the European Commission (DG-SANCO) and the International Agency for Research on Cancer. The national cohorts (that recruited male participants) are supported by Danish Cancer Society (Denmark); German Cancer Aid, German Cancer Research Center (DKFZ), Federal Ministry of Education and Research (BMBF), Deutsche Krebshilfe, Deutsches Krebsforschungszentrum and Federal Ministry of Education and Research (Germany); the Hellenic Health Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON, World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS), PI13/00061 to Granada; , PI13/01162 to EPIC-Murcia), Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, ISCIII RETIC (RD06/0020) (Spain); Swedish Cancer Society, Swedish Research Council and County Councils of Skåne and Västerbotten (Sweden); Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491 and C8221/A19170 and A29017 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk, MR/M012190/1 to EPIC-Oxford) (United Kingdom).

For the ovarian cancer GWAS data:

The ovarian cancer genome-wide association analyses were supported by the U.S. National Institutes of Health (CA1X01HG007491-01 (C.I. Amos), U19-CA148112 (T.A. Sellers), R01-CA149429 (C.M. Phelan) R01-CA058598 (M.T. Goodman) and R01-CA211707, R01-CA204954, and R01-CA211575 (S.A. Gayther); Canadian Institutes of Health Research (MOP-86727 (L.E. Kelemen)) and the Ovarian Cancer Research Fund (A. Berchuck). The COGS project was funded through a European Commission's Seventh Framework Programme grant (agreement number 223175 – HEALTH-F2–2009-223175). The Gynaecological Oncology Biobank at Westmead, a member of the Australasian Biospecimen Network-Oncology group, was funded by the National Health and Medical Research Council Enabling Grants ID 310670 & ID 628903 and the Cancer Institute NSW Grants ID 12/RIG/1-17 & 15/RIG/1-16. All studies and funders are listed in Phelan et al (2017).

For the endometrial cancer GWAS data:

The endometrial cancer genome-wide association analyses were supported by the National Health and Medical Research Council of Australia (APP552402, APP1031333, APP1109286, APP1111246 and APP1061779), the U.S. National Institutes of Health (R01-CA134958), European Research Council (EU FP7 Grant), Wellcome Trust Centre for Human Genetics (090532/Z/09Z) and Cancer Research UK. OncoArray genotyping of ECAC cases was performed with the generous assistance of the Ovarian Cancer Association Consortium (OCAC), which was funded through grants from the U.S. National Institutes of Health (CA1X01HG007491-01 (C.I. Amos), U19-CA148112 (T.A. Sellers), R01-CA149429 (C.M. Phelan) and R01-CA058598 (M.T. Goodman); Canadian Institutes of Health Research (MOP-86727 (L.E. Kelemen)) and the Ovarian Cancer Research Fund (A. Berchuck). We particularly thank the efforts of Cathy Phelan. OncoArray genotyping of the BCAC controls was funded by Genome Canada Grant GPH-129344, NIH Grant U19 CA148065, and Cancer UK Grant C1287/A16563. All studies and funders are listed in O'Mara et al (2018).

Other support:

MBCSG (Milan Breast Cancer Study Group) received funds from AIRC - Associazione Italiana per la Ricerca sul Cancro and thanks Siranoush Manoukian, Bernard Peissel, Jacopo Azzollini, Daniela Zaffaroni, Benedetta Beltrami, Bernardo Bonanni, Irene Feroce, Mariarosaria Calvello, Aliana Guerrieri Gonzaga, Monica Marabelli, Davide Bondavalli, and the personnel of the Cogentech Cancer Genetic Test Laboratory, Milan.

SL reports support from the U.S. National Institutes of Health/National Cancer Institute grant no. U01-CA194393. SPK was supported by a Homerton College Junior Research Fellowship from October 2015 to August 2019 and since September 2019 has been supported by Cancer Research UK (programme grant no. C18281/A19169) and the Integrative Epidemiology Unit, which receives funding from the UK Medical Research Council and the University of Bristol.

REFERENCES

1. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **68**, 394–424 (2018).
2. Beral, V. & Million Women Study Collaborators. Breast cancer and hormone-replacement therapy in the Million Women Study. *Lancet* **362**, 419–427 (2003).
3. Beral, V., Million Women Study Collaborators, Bull, D., Green, J. & Reeves, G. Ovarian cancer and hormone replacement therapy in the Million Women Study. *Lancet* **369**, 1703–1710 (2007).
4. Beral, V., Bull, D., Reeves, G. & Million Women Study Collaborators. Endometrial cancer and hormone-replacement therapy in the Million Women Study. *Lancet* **365**, 1543–1551 (2005).
5. Venkitaraman, A. R. Cancer suppression by the chromosome custodians, BRCA1 and BRCA2. *Science* **343**, 1470–1475 (2014).
6. Kar, S. P. *et al.* Genome-wide meta-analyses of breast, ovarian, and prostate cancer association studies identify multiple new susceptibility loci shared by at least two cancer types. *Cancer Discov.* **6**, 1052–1067 (2016).
7. Michailidou, K. *et al.* Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92–94 (2017).
8. Schumacher, F. R. *et al.* Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat. Genet.* **50**, 928–936 (2018).
9. Phelan, C. M. *et al.* Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nat. Genet.* **49**, 680–691 (2017).
10. O'Mara, T. A. *et al.* Identification of nine new susceptibility loci for endometrial cancer. *Nat. Commun.* **9**, 3166 (2018).
11. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.* **88**, 586–598 (2011).
12. Lee, C. H., Eskin, E. & Han, B. Increasing the power of meta-analysis of genome-wide association studies to detect heterogeneous effects. *Bioinformatics* **33**, i379–i388 (2017).
13. Zhang, H. *et al.* Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nat. Genet.* (2020).
14. Lawrenson, K. *et al.* Functional mechanisms underlying pleiotropic risk alleles at the 19p13.1 breast-ovarian cancer susceptibility locus. *Nat. Commun.* **7**, 12675 (2016).
15. Shu, X. *et al.* Identification of novel breast cancer susceptibility loci in meta-analyses conducted among Asian and European descendants. *Nat. Commun.* **11**, 1217 (2020).

16. Trochet, H. *et al.* Bayesian meta-analysis across genome-wide association studies of diverse phenotypes. *Genet. Epidemiol.* **43**, 532–547 (2019).

17. Kamat, M. A. *et al.* PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics* **35**, 4851–4853 (2019).

18. Malone, J. *et al.* Modeling sample variables with an Experimental Factor Ontology. *Bioinformatics* **26**, 1112–1118 (2010).

19. Vuckovic, D. *et al. The Polygenic and Monogenic Basis of Blood Traits and Diseases*. (2020) doi:10.1101/2020.02.02.20020065.

20. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).

21. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).

22. Aguet, F. *et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues*. (2019) doi:10.1101/787903.

23. Võsa, U. *et al. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis*. (2018) doi:10.1101/447367.

24. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).

25. Boyle, A. P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797 (2012).

26. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).

27. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).

28. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).

29. Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* **47**, W199–W205 (2019).

30. Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K. & Tanabe, M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* **47**, D590–D595 (2019).

31. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).

32. Cole, M. D. MYC association with cancer risk and a new model of MYC-mediated repression. *Cold Spring Harb. Perspect. Med.* **4**, a014316 (2014).

33. Rickman, D. S., Schulte, J. H. & Eilers, M. The Expanding World of N-MYC-Driven Tumors. *Cancer Discov.* **8**, 150–163 (2018).

34. Medina-Gomez, C. *et al.* Life-Course Genome-wide Association Study Meta-analysis of Total Body BMD and Assessment of Age-Specific Effects. *Am. J. Hum. Genet.* **102**, 88–102 (2018).

35. Cauley, J. A. Estrogen and bone health in men and women. *Steroids* **99**, 11–15 (2015).

36. Cauley, J. A. *et al.* Bone mineral density and risk of breast cancer in older women: the study of osteoporotic fractures. Study of Osteoporotic Fractures Research Group. *JAMA* **276**, 1404–1408 (1996).

37. Zhang, Y. *et al.* Bone mass and the risk of prostate cancer: the Framingham Study. *Am. J. Med.* **113**, 734–739 (2002).

38. Nelson, R. L., Turyk, M., Kim, J. & Persky, V. Bone mineral density and the subsequent risk of cancer in the NHANES I follow-up cohort. *BMC Cancer* **2**, 22 (2002).

39. Koutsodontis, G., Tentes, I., Papakosta, P., Moustakas, A. & Kardassis, D. Sp1 plays a critical role in the transcriptional activation of the human cyclin-dependent kinase inhibitor p21(WAF1/Cip1) gene by the p53 tumor suppressor protein. *J. Biol. Chem.* **276**, 29116–29125 (2001).

40. Li, H. *et al.* Integrated high-throughput analysis identifies Sp1 as a crucial determinant of p53-mediated apoptosis. *Cell Death Differ.* **21**, 1493–1502 (2014).

41. Vega, F. M. & Ridley, A. J. Rho GTPases in cancer cell biology. *FEBS Lett.* **582**, 2093–2101 (2008).

42. Yap, C. X. *et al.* Dissection of genetic variation and evidence for pleiotropy in male pattern baldness. *Nat. Commun.* **9**, 5407 (2018).

43. Kawase, T. *et al.* PH domain-only protein PHLDA3 is a p53-regulated repressor of Akt. *Cell* **136**, 535–550 (2009).

44. Zhang, G. *et al.* Genetic Associations with Gestational Duration and Spontaneous Preterm Birth. *N. Engl. J. Med.* **377**, 1156–1167 (2017).

45. Perry, J. R. *et al.* Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92–97 (2014).

46. Key, T. J. & Pike, M. C. The dose-effect relationship between 'unopposed' oestrogens and endometrial mitotic rate: its central role in explaining and predicting endometrial cancer risk. *Br. J. Cancer* **57**, 205–212 (1988).

47. Karami, S. *et al.* Telomere structure and maintenance gene variants and risk of five cancer types. *Int. J. Cancer* **139**, 2655–2670 (2016).

48. Fehringer, G. *et al.* Cross-Cancer Genome-Wide Analysis of Lung, Ovary, Breast, Prostate, and Colorectal Cancer Reveals Novel Pleiotropic Associations. *Cancer Res.* **76**, 5103–5114 (2016).

49. Chuikov, S. *et al.* Regulation of p53 activity through lysine methylation. *Nature* **432**, 353–360 (2004).

50. Glubb, D. M. *et al.* Fine-scale mapping of the 5q11.2 breast cancer locus reveals at least three independent risk variants regulating MAP3K1. *Am. J. Hum. Genet.* **96**, 5–20 (2015).

51. Fabbro, M. *et al.* Cdk1/Erk2- and Plk1-dependent phosphorylation of a centrosome protein, Cep55, is required for its recruitment to midbody and cytokinesis. *Dev. Cell* **9**, 477–488 (2005).

52. Sinha, D. *et al. Cep55 overexpression promotes genomic instability and tumorigenesis in mice*. (2019) doi:10.1101/780775.

53. Jeffery, J., Sinha, D., Srihari, S., Kalimutho, M. & Khanna, K. K. Beyond cytokinesis: the emerging roles of CEP55 in tumorigenesis. *Oncogene* **35**, 683–690 (2016).

54. Nilufer, R. *et al. Large-scale genome-wide association meta-analysis of endometriosis reveals 13 novel loci and genetically-associated comorbidity with other pain conditions*. (2018) doi:10.1101/406967.

55. Astle, W. J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**, 1415-1429.e19 (2016).

56. Hinds, D. A. *et al.* Germ line variants predispose to both JAK2 V617F clonal hematopoiesis and myeloproliferative neoplasms. *Blood* **128**, 1121–1128 (2016).

57. Bao, E. L. *et al. Genetic predisposition to myeloproliferative neoplasms implicates hematopoietic stem cell biology*. (2019) doi:10.1101/790626.

58. Loh, P.-R., Genovese, G. & McCarroll, S. A. *Monogenic and polygenic inheritance become instruments for clonal selection*. (2019) doi:10.1101/653691.

59. Thompson, D. J. *et al.* Genetic predisposition to mosaic Y chromosome loss in blood. *Nature* **575**, 652–657 (2019).

60. Kastenhuber, E. R. & Lowe, S. W. Putting p53 in Context. *Cell* **170**, 1062–1078 (2017).

61. Donehower, L. A. *et al.* Integrated Analysis of TP53 Gene and Pathway Alterations in The Cancer Genome Atlas. *Cell Rep.* **28**, 1370-1384.e5 (2019).

62. Stacey, S. N. *et al.* A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. *Nat. Genet.* **43**, 1098–1103 (2011).

63. Márquez, A. *et al.* Meta-analysis of Immunochip data of four autoimmune diseases reveals novel single-disease and cross-phenotype associations. *Genome Med.* **10**, 97 (2018).

64. Cross-Disorder Group of the Psychiatric Genomics Consortium & Cross-Disorder Group of the Psychiatric Genomics Consortium. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* **179**, 1469-1482.e11 (2019).

65. Turchin, M. C. & Stephens, M. Bayesian multivariate reanalysis of large genetic studies identifies many new associations. *PLoS Genet.* **15**, e1008431 (2019).

66. von Karstedt, S., Montinaro, A. & Walczak, H. Exploring the TRAILs less travelled: TRAIL in cancer biology and therapy. *Nat. Rev. Cancer* **17**, 352–366 (2017).

67. Labidi-Galy, S. I. *et al.* High grade serous ovarian carcinomas originate in the fallopian tube. *Nat. Commun.* **8**, 1093 (2017).

68. Perets, R. & Drapkin, R. It's totally tubular....riding the new wave of ovarian cancer research. *Cancer Res.* **76**, 10–17 (2016).

69. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

70. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. *Control. Clin. Trials* **7**, 177–188 (1986).

71. Southam, L. *et al.* Whole genome sequencing and imputation in isolated populations identify genetic associations with medically-relevant complex traits. *Nat. Commun.* **8**, 15606 (2017).

72. Province, M. A. & Borecki, I. B. A correlated meta-analysis strategy for data mining 'OMIC' scans. *Pac. Symp. Biocomput.* 236–246 (2013).

73. Han, B. *et al.* A general framework for meta-analyzing dependent studies with overlapping subjects in association mapping. *Hum. Mol. Genet.* **25**, 1857–1866 (2016).

74. Lilyquist, J., Ruddy, K. J., Vachon, C. M. & Couch, F. J. Common Genetic Variation and Breast Cancer Risk-Past, Present, and Future. *Cancer Epidemiol. Biomarkers Prev.* **27**, 380–394 (2018).

75. Benafif, S., Kote-Jarai, Z., Eeles, R. A. & PRACTICAL Consortium. A Review of Prostate Cancer Genome-Wide Association Studies (GWAS). *Cancer Epidemiol. Biomarkers Prev.* **27**, 845–857 (2018).

76. Machiela, M. J. & Chanock, S. J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**, 3555–3557 (2015).